JOURNAL OF COMPUTER AND SYSTEM SCIENCES **26**, 295–310 (1983)

Groups, the Theory of Ends, and Context-Free Languages

DAVID E. MULLER

Department of Mathematics, Computer Science and Coordinated Science Laboratory, University of Illinois, Urbana, Illinois 61801–3082

AND

PAUL E. SCHUPP*

Department of Mathematics, University of Illinois, Urbana, Illinois 61801–3082

Received January 30, 1982; revised May 8, 1982

1. INTRODUCTION

This is the first of a series of three papers describing some surprising connections between group theory, the theory of ends, the theory of automata and formal languages, and second-order logic. This first paper discusses the interaction of group theory and formal language theory. Using the theory of ends and Rabin's theorem that the monadic theory of the infinite binary tree is decidable, the second paper will establish the decidability of the monadic second-order theory of a very large class of graphs. The third paper will give give a new proof of Rabin's theorem. The sketch [9] is a detailed statement of the reults of the first and second papers.

We begin our investigation with a question raised by Anisimov [1]. A finitely generated group can be described by a presentation $G = \langle X; R \rangle$ in terms of generators and defining relations. The *word problem* $W(G)$ is the set of all words on the generators and their inverses which represent the identity element of G. Anisimov asked, "If $W(G)$ is a context-free language in the usual sense of formal language theory, what can one say about the algebraic structure of the group G?" Although the set $W(G)$ depends on the presentation, an easy lemma shows that if $W(G)$ is a context-free language for one presentation of G, then $W(G)$ is a context-free laguage for every finitely generated presentation of G. Thus we shall simply say that a finitely generated group is *context-free* if the word problem is a context-free language for finitely generated presentations of G. A group is *virtually free* if it has a free subgroup of finite index. We were led to conjecture that a finitely generated group is

* This research was supported in part by National Science Foundation Grants MCS-78-13642 and MCS-80-03251, and by the Joint Services Electronics Program under Contract N00014-79-C-0424.

0022-0000/83 $3.00

context-free if and only if it is virtually free and we have essentially proven the conjecture. Precisely, we have the following result:

THEOREM. *Let G be a finitely generated group. G is free if and only if G is context-free and torsion-free. G is virtually free if and only if G is context-free and accessible.*

The condition of accessibility mentioned in the theorem is a technical condition which will be explained in detail later. While it is conjectured that all finitely generated groups are accessible, it is difficult to prove that specific groups are accessible. There is really no doubt that context-free groups are accessible but we have not been able to prove this directly.

2. Word Problems and Automata

We remark at the outset that there are considerable similarities between the idea of describing groups by presentations in terms of generators and defining relations and the idea of generating formal languages by grammars. For a good introduction to presentations see the first chapter of Magnus *et al.* [8]. For example, the group $C = \langle x; x^3 = 1 \rangle$ has only one generator whose third power is the identity and is thus the cyclic group of order three. The group $Z = \langle x \rangle$ with one generator and no defining relations is the infinite cyclic group. The group $A = \langle x, y; xy = yx \rangle$ which has two generators subject only to the relation that they commute is the free Abelian group on two generators, while the group $F = \langle x, y \rangle$ is the free group on two generators.

In order to fix our notation and terminology we give a brief formal description of the group G defined by the presentation $\langle X; R \rangle$. We choose a set X^{-1}, disjoint from X, with a one-to-one correspondence $x \mapsto x^{-1}$ between X and X^{-1}. We define $(x^{-1})^{-1}$ to be x. We write $X^{\pm 1}$ for $X \cup X^{-1}$ and call the elements of $X^{\pm 1}$ *letters*. A *word* is a finite sequence of letters, and we denote the empty word by 1 when discussing groups. The *length* $|w|$ of a word w is the number of letters in w. If $w = y_1, \ldots, y_n$ is a word, then $w^{-1} = y_n^{-1} \ldots y_1^{-1}$. Since an equation $u = v$ can always be written in the equivalent form $uv^{-1} = 1$, we shall assume that all the equations of R have the form $r = 1$ and call the word r a *defining relator*. Since we want G to be a group, we need the *trivial relators* yy^{-1}, $y \in X^{\pm 1}$. (The trivial relators are understood and are not written as part of the presentation.) Two words u and v on $X^{\pm 1}$ are *equivalent* if it is possible to tansform u into v by a finite sequence of insertions or deletions of trivial relators and defining relators. The set of equivalence classes forms a group G, where multiplication is concatenation of representatives, $[u][v] = [uv]$, the identity is the class of the empty word and $[y]^{-1} = [y^{-1}]$. The word problem $W(G)$ is simply the set of words equivalent to 1.

Considering Σ^* as the free semigroup generated by $\Sigma = X^{\pm 1}$, there is a natural homomorphism $\eta: \Sigma^* \to G$ defined by $\eta(u) = [u]$. In other words, u and v are equivalent if and only if $\eta(u) = \eta(v)$. To avoid using equivalence class notation, we write "$u = v$ in G" if $\eta(u) = \eta(v)$ and say that u and v represent the same element of G.

In discussing formal language theory and automata, we refer to Harrison [5] and to Hopcroft and Ullman [6]. An important feature of formal language theory is that such languages can be described in either of two ways—acceptance by a machine of the appropriate type or generation by a grammar of the appropriate type. A language is regular if and only if it is the set of strings accepted by a finite state automaton and we assume that the reader is familiar with this concept. The characterization of groups with regular word problem was given by Anisimov [1] but we give here a simpler proof more suited to our purposes.

LEMMA 1. *A group has a regular word problem if and only if it is finite.*

Proof. If G is a finite group, let $\{g_1, ..., g_n\}$ be a complete list of the elements of G, where g_1 is the identity element. The *multiplication table presentation* of G is the finite presentation $\langle g_1, ..., g_n; ..., g_i g_j = g_k, ...\rangle$ with generators $g_1, ..., g_n$, and, for each choice of $1 \leqslant i, j \leqslant n$, there is a defining relation $g_i g_j = g_k$, where g_k is the product of g_i and g_j. It is now easy to construct a finite state automation M which accepts the word problem of any presentation of G. Let M have state set $\{g_1, ..., g_k\}$, and let g_1 be the initial state and the only final state of M. Any letter y of the alphabet of the given presentation represents an element g_j of G. If M is in state g_i and reads an input letter representing g_j, then M changes state to g_k, where $g_i g_j = g_k$.

Now let $G = \langle X; R\rangle$ be any finitely generated infinite group. There must be arbitrarily long words w on $X^{\pm 1}$ such that no nonempty subword of w represents the identity of G. For, if there were an upper bound b on the length of such words, every element of G could be represented by a word of length not exceeding b which, since X is finite, would imply the finiteness of G. Let M be any finite state automaton with input alphabet $X^{\pm 1}$. Let w be a word whose length is greater than the number n of states of M and such that no nonempty subword of w represents the identity of G. If M begins reading w, then M must be in the same state after reading two distinct initial segments, say u and uv, of w. Now $uu^{-1} = 1$ in G but $uvu^{-1} \neq 1$ in G since the latter is a conjugate of the nonidentity element v. Since M must either accept or reject both uu^{-1} and uvu^{-1}, M does not accept exactly the word problem of G. ∎

For basic details concerning the concept of a pushdown automaton, usually written pda, we refer to Harrison [5] and to Hopcroft and Ullman [6]. Our notation follows Harrison [5] but we differ from the usual convention in that we allow a machine to continue running when it empties the stack since this convention is by far the most natural when dealing with group word problems. We also allow a machine to start with an empty stack. If ζ denotes the contents of the stack at a given time, we consider the rightmost symbol of ζ to be the top of the stack. When discussing words on an arbitrary alphabet we use Λ to denote the empty string.

In order to accept arbitrary context-free languages it is necessary to use nondeterministic pushdown automata. To fix our notation, we consider a pda to be a seventuple $M = (Q, \Sigma, Z, \delta, q_0, z_0, \hat{Q})$, where Q is a finite set of *states*, Σ is a finite *input alphabet*, Z is a finite *stack alphabet* with $Z \supseteq \Sigma$, $q_0 \in Q$ is the *initial state*, $z_0 \in Z \cup \{\Lambda\}$ is the *start symbol*, $\hat{Q} \subseteq Q$ is the set of *final states*, and δ is the *transition function*. Since we are considering machines which may continue running on an empty stack, δ is a function from $Q \times (\Sigma \cup \{\Lambda\}) \times (Z \cup \{\Lambda\})$ to finite subsets of $Q \times Z^*$.

The interpretation of $(q_i, \zeta_i) \in \delta(q, \sigma, z)$ is that when the pda is in state q reading the input symbol σ and z is the top symbol on the stack then M can change to state q_i, replace z by ζ_i and move the reading head one square to the right. We consider ζ_i as being placed on the stack from left to right so that the rightmost symbol of ζ_i (if $\zeta_i \neq \Lambda$) becomes the top symbol of the stack. If $\delta(q, \sigma, z)$ is empty, the machine halts.

The interpretation of $(q_i, \zeta_i) \in \delta(q, \Lambda, z)$ is that when the pda is in state q with z the top symbol of the stack, then independently of the input symbol being scanned, the machine changes state to q_i and replaces z by ζ_i. In this case the input head is not moved. Such a transition is called a *Λ-move*. In the obvious way we allow Λ-moves when the stack is empty.

A pda is *deterministic* if $|\delta(q, \sigma, z)| \leqslant 1$ for all triples (q, σ, z) and the machine cannot make both a Λ-move and a non-Λ-move for a given pair (q, z). As usual, we consider deterministic machines to be a special case of nondeterministic ones. All the pda's which we construct to accept word problems will be deterministic. (In the sequel to this paper we shall prove that all context-free group word problems are indeed deterministic without assuming the accessibility condition.)

We write $M \vdash_{w}^{*} (q, \zeta)$ if it is possible for M to be in state q with ζ written on the stack after reading the input w. The language *accepted* by M by final state and empty stack is $L = \{w \mid M \vdash_{w}^{*} (q, \Lambda) \text{ for some } q \in \hat{Q}\}$. It is well known that a language L is context-free if and only if L is accepted by some pda by final state and empty stack.

We now discuss a particular pda M which accepts the word problem of the free group $F = \langle x_1, \ldots, x_n \rangle$ on n generators. It is a well-known fact that a word w on $X^{\pm 1} = \{x_1, \ldots, x_m\}^{\pm 1}$ represents the identity of F if and only if w can be reduced to the empty word by successive deletions of subwords consisting of an inverse pair $x_i^{\varepsilon} x_i^{-\varepsilon}$, $\varepsilon = \pm 1$. The machine M works as follows: M has only one state q which is also a final state; M begins with empty stack. On reading a letter x_i^{ε}, if that letter is not the inverse of the letter on the top of the stack, then the letter is added to the stack. If the letter being read is the inverse of the letter on the top of the stack, then the top of the stack is deleted. At any point, the word read so far is equal to the identity in F if and only if the stack is empty. If M reads the word $xyy^{-1}x^{-1}y^{-1}x$, then the successive stack contents are $\Lambda, x, xy, x, \Lambda, y^{-1}, y^{-1}x$.

LEMMA 2. *Let G be a finitely generated group. If the word problem is a context-free language in one finitely generated presentation of G, then it is a context-free language in every finitely generated presentation of G. Furthermore, if H is any finitely generated subgroup of G, then the word problem of H is also a context-free language.*

Proof. Let $\langle X; R\rangle$ be a finitely generated presentation of G with context-free word problem. Let $\langle y_1,..., y_n; S\rangle$ be a finitely generated presentation of a subgroup H of G. Then there exists an embedding $\varphi: H \to G$ which, in terms of the given presentations, is completely specified by the information $\varphi(y_i) = u_i$, $i = 1,..., n$, where each u_i is a word on $X^{\pm 1}$. Let M be the pushdown automaton which accepts the word problem of the specified presentation of G. If w is a word in the $y_i^{\pm 1}$, then since φ is an embedding, w represents the identity in the presentation $\langle Y; S\rangle$ if and only if $\varphi(w)$ represents the identity in $\langle X; R\rangle$. Construct a pushdown automaton M' for the word problem of $\langle Y; S\rangle$ as follows: On reading a letter y_i^{ε}, $\varepsilon = \pm 1$, M' simulates what M does on reading the word u_i^{ε}. (This simulation may require several Λ-moves.) We note that there is no general method for producing the embedding φ—but if one did have the finite amount of data $\varphi(y_i) = u_i$, then M' could be effectively constructed from M. ∎

In view of Lemma 2, we shall simply say that a group G is context-free if G is finitely generated and the word problem for each finitely generated presentation of G is a context-free language.

Using some results from group theory, we shall now show how to construct a pda which accepts the word problem of a finitely generated virtually free group.

LEMMA 3. *If G is a finitely generated virtually freee group, then G is context-free.*

Proof. Let H be a free subgroup which has finite index in G. A well-known general result of group theory says that any subgroup of finite index contains a subgroup N which is normal in the group G and which again has finite index. The Nielsen–Schreier theorem states that every subgroup of a free group is free, so N is a free group. A theorem of Schreier states that any subgroup having finite index in a finitely generated group is itself finitely generated. Hence, N is a finitely generated free subgroup which is normal and has finite index in G. We now need only work with N.

We shall use the subgroup N to construct a particular presentation of G which arises from considering G as an extension of N by G/N. Let N have free generators $y_1,..., y_n$. The quotient group $B = G/N$ is a finite group, say with t elements. Let $\eta: G \to B$ be the natural homomorphism from G onto B. Let $\langle b_1,..., b_t; b_i b_j = b_k\rangle$ be the multiplication table presentation of B. Let d_i, $i = 1,..., t$ be elements of G such that $\eta(d_i) = b_i$. Since N is the kernel of the homomorphism η, relations of the forms

$$d_i y_d d_i^{-1} = u_{i,d} \qquad \text{and} \qquad d_i d_j^{\varepsilon} = z_{i,\varepsilon,j} d_k$$

hold in G, where the $u_{i,d}$ and $z_{i,\varepsilon,j}$ are elements of N. Using the y's and d's as letters, we claim that

$$\langle y_1,..., y_n, d_1,..., d_t; d_i y_d d_i^{-1} = u_{i,d}, d_i d_j^{\varepsilon} = z_{i,\varepsilon,j} d_k\rangle$$

is a presentation of G. Using these relations any word can be transformed to a unique word of the form $w d_i$, where w is a freely reduced word in the y's. The latter word represents the identity of G if and only if w is empty and d_i is the symbol d_1 with $\eta(d_1)$ the identity of B.

The pda M for the word problem of this representation of G works as follows—M keeps track of the "free part" of a word in its stack and keeps track of the image in B by its states; M has states $q_1,\dots,q_t$ corresponding to $b_1,\dots,b_t$ and several other "working" states; M starts in q_1 with an empty stack. If M is in state q_i and reads y_l^ε, then, since $d_i y_l^\varepsilon = u_{i,l}^\varepsilon d_i$ in G, M uses its working states to make a sequence of Λ-moves which process the word $u_{i,l}^\varepsilon$ onto the stack (as in the description of the automaton for the free group word problem) and then returns to state q_i. If M is in state q_i and reads d_j^ε, then, since $d_i d_j^\varepsilon = z_{i,\varepsilon,j} d_k$ in G, M uses its working states to process $z_{i,\varepsilon,j}$ onto the stack and then changes state to q_k. Thus, after reading an arbitrary word v (which is equal to G to wd_i with w a word on the $y_l^{\pm 1}$) M has w on its stack and is in state q_i. Thus V represents the identity of G if and only if M is in state q_1 with empty stack. ∎

We next point out that the free Abelian group $A = \langle x, y; xy = yx\rangle$ on two generators is not context-free. This group is the "least complicated" group which is not context-free and is thus an important example. A word of the form $x^m y^n x^{-j} y^{-k}$ is equal to the identity in A if and only if $j = m$ and $k = n$. Intuitively, it is impossible for a pda to check if both of these equations hold. It is easy to give a formal proof by using the well-known "pumping lemma" for context-free languages.

3. A Geometric Characterization of Context-Free Groups

If $G = \langle X; R\rangle$ is a finitely generated group, the Cayley graph $\Gamma(G)$ of the given presentation is defined as follows. The vertex set V of $\Gamma(G)$ is the set of elements of the group G. If $g \in G$, $y \in X^{\pm 1}$, there is a labelled edge $e = (g, y, gy)$ with *initial vertex* g, *label* y, and *terminal vertex* gy. The edge $e^{-1} = (gy, y^{-1}, g)$ is the *inverse* of the edge e. Note that the graph $\Gamma(G)$ depends heavily on the given presentation. We illustrate the Cayley graphs of $Z = \langle x\rangle$, $A = \langle x, y; xy = yx\rangle$, and $F = \langle x, y\rangle$ in Fig. 1. In drawing graphs we follow the usual convention of drawing only one arc for the pair of edges e and e^{-1} and regard these two edges as being represented by traversing the arc in opposite directions.

Since we shall discuss the Cayley graph of only one group G at a time, we shall simply write Γ instead of $\Gamma(G)$. Since each edge e has a label $\varphi(e) \in X^{\pm 1}$ it is natural to define the *label* $\varphi(\alpha)$ of a path $\alpha = e_1,\dots, e_n$ as the word $\varphi(e_1),\dots, \varphi(e_n)$. From the definition of Γ, a path α in Γ is a closed path if and only if the label $\varphi(\alpha)$ represents the identity of G. The graph Γ thus certainly contains all the information given by the word problem $W(G)$. We shall prove that a group G being context-free is equivalent to a certain triangulation property of the Cayley graph $\Gamma(G)$.

Before proving the theorem we fix our notation on context-free grammars. We follow Hopcroft and Ullman [6]. A *context-free grammar* is a quadruple $C = (V, \Sigma, \mathscr{P}, S)$, where V and Σ are disjoint finite sets of *variables* and *terminals*,

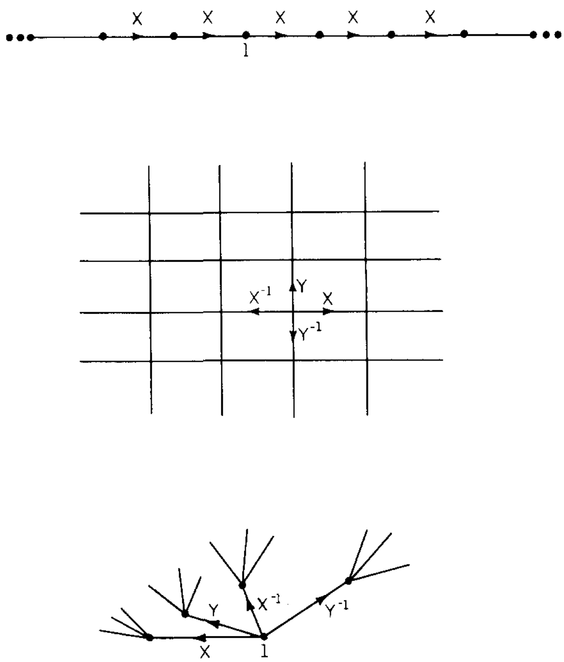

FIGURE 1

respectively, $S \in V$ is the *start symbol*, and $\mathscr{P}$ is a finite set of *productions* of the form $A \to \alpha$, where $A \in V$ and $\alpha \in (V \cup \Sigma)^*$. (If U is any set, then U^* denotes the set of all finite sequences of elements of U.) If $\alpha A \gamma$ is a string in $(V \cup \Sigma)^*$ and $A \to \beta$ is a production of $\mathscr{P}$, we write $\alpha A \gamma \Rightarrow \alpha\beta\gamma$. We write $\eta \overset{*}{\Rightarrow} \theta$ if $\eta = \theta$ or there are strings $\eta = \eta_0, \ldots, \eta_m = \theta$, $m \geqslant 1$, such that each $\eta_i \Rightarrow \eta_{i+1}$. The language generated by C is the set $L = \{w \mid w \in \Sigma^* \text{ and } S \overset{*}{\Rightarrow} w\}$ of all words on the terminal alphabet which are derivable from the start symbol S. A grammar C is in *Chomsky normal form* if all the productions of C have the form $A \to BC$, or $A \to a$, or $S \to \Lambda$, where A, B, C are variables, a is a terminal, S is the start symbol, and Λ is the empty word. Furthermore, if the production $S \to \Lambda$ occurs, then S does not appear on the right-hand side of any production.

A variable A is *useful* if there is some derivation $S \overset{*}{\Rightarrow} \alpha A \beta \overset{*}{\Rightarrow} w$ of a terminal string which uses A. Otherwise, A is *useless*. A grammar is *reduced* if all its variables are useful. A basic result on context-free languages is that every context-free language can be generated by a reduced grammar in Chomsky normal form. It is exactly such grammars which we shall need. If A is a variable of a grammar C, then $L(A) = \{w \mid w \in \Sigma^* \text{ and } A \overset{*}{\Rightarrow} w\}$ is the set of words on the terminal alphabet which are derivable from A. Note that $L(A)$ is nonempty if A is useful.

LEMMA 4. *Let C be a reduced context-free grammar which generates the word problem $W(G)$ of a group $G = \langle X; R \rangle$. If A is a variable of C and $u, v \in L(A)$, then u and v represent the same element of G.*

FIGURE 2

Proof. Since A is useful, there is a derivation $S \overset{*}{\Rightarrow} \alpha A \beta \overset{*}{\Rightarrow} w_1 w_2 w_3$, where $w_1 w_2 w_3 \in W(G)$ and w_2 is the part derived from A. If we replace the derivation of w_2 from A by derivations of u and v, respectively, then in G, $w_1 u w_2 = 1 = w_1 v w_2$. Since G is group, $u = v$ in G. ∎

In discussing a polygon P in the plane we assume that the boundary of P is a simple closed curve. A *diagonal triangulation* of P is a triangulation of P which has as vertices only the originally given vertices of P. If P is the hexagon pictured, then the triangulation of Fig. 2a is diagonal while the triagulation in Fig. 2b is not diagonal. To avoid special cases, we allow "1-gons"  and "2-gons"  and consider these figures as being triangulated.

Let α be a closed path in Γ, and let $w = y_1, \ldots, y_n$ be the label on α. Write w clockwise around the boundary of a regular n-gon in the plane. (The edges of P are thus labelled by the letters of w.) A *K-triangulation of α* is a diagonal triangulation of P with a label from the free group $F = \langle X \rangle$ assigned to each new edge such that:

(i) Reading around the boundary of each triangle gives a true relation in the group G, and

(ii) if u is the label on an edge e of the triangulation, then $|u| \leqslant K$.

The three upper triangles in Fig. 3 say, respectively, that $x_1 x_2 = u_1$, $u_1 u_2 u_3 = 1$, and $x_1^{-1} x_2^{-1} = u_2$.

The main result of this section is

THEOREM I. *A finitely generated group $G = \langle X; R \rangle$ is context-free if and only if there exists a constant K such that every closed path in the Cayley graph $\Gamma(G)$ can be K-triangulated.*

Proof. Let $G = \langle X; R \rangle$ be context-free. Let C be a reduced grammar which generates $W(G)$ and which is in Chomsky normal form.

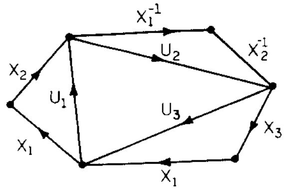

FIGURE 3

We shall show how the grammar C determines triangulations of closed paths in $\Gamma(G)$. If A is a variable of C, let u_A denote a shortest word in $L(A)$. Let α be any closed path in Γ, so its label $y_1 \dots y_n$ is in $W(G)$, and write $y_1 \dots y_n$ around the boundary of a regular n-gon P in the plane. If $n \leqslant 2$, then our convention is that P is triangulated, and if $n = 3$, then P is a triangle. In this case the maximum length of a label on a side is one. We now assume that $n \geqslant 4$. Consider any derivation $S \overset{*}{\Rightarrow} w$. The derivation must be of the form $S \Rightarrow AB \overset{*}{\Rightarrow} w_1 w_2$, where w_1 and w_2 are derivable from A and B, respectively.

Suppose that the length of both w_1 and w_2 is at least two. Construct an edge with label u_B from the vertex of P at which w_1 ends to the vertex which begins w_1 (as in Fig. 4a). By Lemma 4, $u_B = w_2$ in G, and since $w_1 w_2 = 1$ in G, $u_B = w_1^{-1}$ in G. We now have two polygons each with fewer sides. Next suppose that one of w_1 or w_2, say w_1, is a terminal symbol, so $S \overset{*}{\Rightarrow} aw_2$ and $|w_2| \geqslant 3$. Then there is a derivation $S \Rightarrow aB \Rightarrow aCD \overset{*}{\Rightarrow} aw_{21} w_{22}$, where at least one of w_{21} or w_{22}, say w_{22}, has length at least two. In this case, construct an edge with label u_D from the vertex at which w_{22} begins to the vertex at which w_{22} ends (as in Fig. 4b).

Iterating the procedure described in the smaller polygons and treating labels of the type u_A in the same way as terminal letters, we arrive at a diagonal triangulation of P in which each new edge is labelled by a label of type u_A for some variable A of the grammar C whose set V of variables is finite. Letting $K = \max_{A \in V} |u_A|$ we have a K-triangulation. Thus every closed path can be K-triangulated.

Now suppose that there exists a constant K such that every closed path in the Cayley graph of $G = \langle X; R \rangle$ can be K-triangulated. We shall describe a context-free grammar which generates $W(G)$. If u is a word on $X^{\pm 1}$ with $|u| \leqslant K$, introduce a corresponding variable A_u. (Note that since X is finite there are only finitely many words not exceeding a given length.) For each equation $u = vz$ which holds in G, where none of $|u|$, $|v|$, $|z|$ exceed K, introduce a production $A_u \to A_v A_z$. Also, if A_v is a variable and $v = y$ in G, where $y \in X^{\pm 1}$, then introduce a production $A_v \to y$. Also, introduce the production $A_1 \to 1$ and make A_1 the start symbol. It is clear that if there is a derivation $A_1 \overset{*}{\Rightarrow} w$, then $w = 1$ in G since all productions mirror true equations.

We need to show that if $w = 1$ in G, then w is derivable from A_1. We shall prove by induction on the number of triangles in T that if T is a diagonal triangulation of a polygon P such that all edges are labelled by words of length not exceeding K and reading around the boundary of every triangle gives a true equation in G, then there

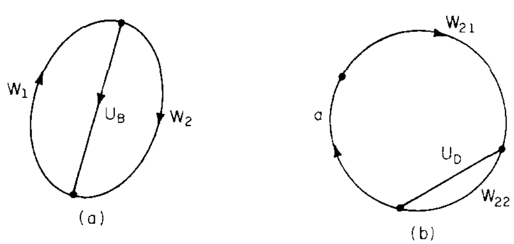

FIGURE 4

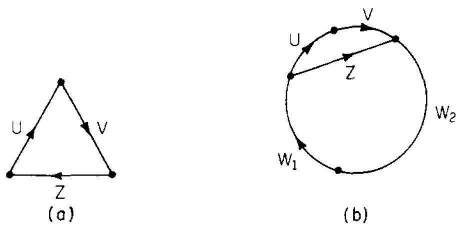

FIGURE 5

is a derivation $A_1 \overset{*}{\Rightarrow} \hat{\alpha}$, where if the edges $e_1,..., e_n$ reading around the boundary α of P are labelled by $u_1,..., u_n$, then $\hat{\alpha} = A_{u_1} \cdots A_{u_n}$. If there is only one triangle, say with boundary label uvz (as in Fig. 5a), then there is a derivation $A_1 \Rightarrow A_u A_{u^{-1}} \Rightarrow A_u A_v A_z$.

Now if T is a diagonal triangulation of a polygon P and T has more than one triangle, then there are at least two triangles which have two edges on the boundary of P. (This is intuitively clear and will be proved in Lemma 5.) Thus given P and a point at which one begins reading the label w on the boundary of P, we can write $w = w_1 uvw_2$ (as in Fig. 5b), where u and v are labels on consecutive edges of a triangle with boundary label uvz^{-1}. By the induction hypothesis, there is a derivation $A_1 \overset{*}{\Rightarrow} \hat{w}_1 A_z \hat{w}_2$ from whence $\hat{w}_1 A_u A_v \hat{w}_2$ is immediately derivable.

If $w = y_1 \ldots y_n$ is a word on $X^{\pm 1}$ such that $w = 1$ on G, then the preceding paragraph shows that there is a derivation $A_1 \overset{*}{\Rightarrow} A_{y_1} \ldots A_{y_n}$. Applying productions of the form $A_y \to y$ we have a derivation $A_1 \overset{*}{\Rightarrow} y_1 \ldots y_n$. ∎

4. The Theory of Ends and the Structure of Context-Free Groups

We now discuss the theory of ends. Our basic references here are Cohen [2] and Stallings [11]. Intuitively, an end is a way of going to infinity. Let Γ be a connected locally finite graph with a distinguished vertex v_0 selected as the *origin* of Γ. Let $\Gamma^{(n)}$ denote everything connected to v_0 by a path of length less than n. The *number of ends* of Γ is

$$e(\Gamma) = \lim_{n \to \infty} \text{(the number of infinite components of } \Gamma \backslash \Gamma^{(n)}).$$

If $G = \langle X; R \rangle$ is a finitely generated group, then the *number of ends* of G is the number of ends of the Cayley graph $\Gamma(G)$. (It is true, but not obvious from the definition which we have given, that $e(G)$ does not depend on the presentation chosen.)

Referring to Fig. 2, we see that for Γ the Cayley graph of the infinite cyclic group Z, there are always two components of $\Gamma \backslash \Gamma^{(n)}$ if $n \geqslant 1$ and thus $e(Z) = 2$. In the free group F of rank 2, $\Gamma \backslash \Gamma^{(n)}$ has 2^n components and thus $e(F) = \infty$. For the free Abelian group A of rank 2, $\Gamma \backslash \Gamma^{(n)}$ is always connected so $e(A) = 1$.

Stallings [11] has proven a remarkable theorem stating that a finitely generated group G has more than one end if and only if G can be decomposed in a particular way using certain group-theoretic constructions. We shall discuss these constructions shortly, but we shall now establish that an infinite context-free group has more than one end. We need a preliminary combinatorial lemma.

LEMMA 5. *Let T be a diagonal triangulation of a plane polygon P with at least three edges. If the boundary edges of P are divided into three consecutive nonempty arcs, then some triangle has vertices on all three arcs. (A vertex where two arcs meet is, of course, on both arcs.)*

Proof. Denote the boundary of P by ∂P and call a triangle *critical* if it has two edges on ∂P. We first observe that if T has more than one triangle, then there are at least two critical triangles. This is clear if the number k, of triangles is two and we proceed by induction on k. If all triangles with edges in ∂P are critical, the result holds. Otherwise, select a triangle t with exactly one edge e on ∂P and write $\partial P = \eta_1 e \eta_2$, $\partial t = e e_1 e_2$. The triangulation T induces triangulations T_1 of the polygon P_1 bounded by $\eta_1 e e_1$ and T_2 of the polygon P_2 bounded by $\eta_2 e_2 e$. Both T_1 and T_2 have more than one triangle and thus have at least two critical triangles, one of which may be t. The result follows. Let T be a diagonal triangulation of a plane polygon P. We next prove that if the vertices of P are colored by three colors so that all the vertices of a given color occur consecutively on the boundary of P then there is a unique triangle having vertices of all three colors. The proof is by induction on the number k of triangles, and the result is clear if $k = 1$. If $k > 1$, let t be a critical triangle. If t has vertices of all three colors, then t is the desired triangle and is unique. If t has vertices of only one or two colors, let P' be the polygon obtained by removing the two edges of t which are in ∂P. Then P' still has vertices of three colors and the result follows. The result stated in the lemma follows immediately, by taking colors a, b, c corresponding to the arcs α, β, γ and assigning colors to the vertices so that the color assigned to a vertex u corresponds to an arc containing at least one of the edges incident at u and all colors are used. ∎

LEMMA 6. *If $G = \langle X; R \rangle$ is an infinite context-free group, then G has more than one end.*

Proof. Let Γ be the Cayley graph of G. If u and v are vertices of Γ, then the distance $d(u, v)$ between u and v is the length of the shortest path connecting u and v. Choose the identity as the origin of Γ. Since G is infinite, there exist arbitrarily long words $y_1 \dots y_j$ such that the shortest word equal to $y_1 \dots y_j$ has length j. From the definition of Γ the group G acts on Γ as a transitive group of graph isomorphisms. By translating the midpoint of the path with label $y_1 \dots y_j$ to the origin (where j is even and $y_1 \dots y_j$ is as stated), we see that for any $i \geqslant 1$ there are elements u_i and v_i which are both at distance i from 1 and with $d(u_i, v_i) = 2i$.

Since G is context-free there is a constant K such that every closed path in Γ can be K-triangulated. Pick $n > \frac{3}{2}K$. As usual, let $\Gamma^{(n)}$ denote the set of all vertices and edges connected to 1 by a path of length less then n. We claim that if $i \geqslant n$, and u_i and v_i are as in the preceding paragraph, then u_i and v_i are in different components of $\Gamma \backslash \Gamma^{(n)}$. This shows that $\Gamma \backslash \Gamma^{(n)}$ has at least two infinite components when $n > \frac{3}{2}K$ and thus G has more than one end.

Suppose that u_i and v_i are in the same component of $\Gamma \backslash \Gamma^{(n)}$. Let α be a path of minimal length going from 1 to u_i, let γ be a path of minimal length going from v_i to 1, and let β be a path lying in $\Gamma \backslash \Gamma^{(n)}$ which connects u_i and v_i. (See Fig. 6). The path $\eta = \alpha\beta\gamma$ is a closed path in Γ. Let T be a K-triangulation of η. By Lemma 5, some triangle t has vertices a, b, c (not necessarily distinct) on α, β, γ, respectively. Each edge of t represents a path of length not exceeding K. Now since $b \in \Gamma \backslash \Gamma^{(n)}$ we must have $d(1, a) \geqslant n - K$, for, if $d(1, a)$ were less than $n - K$, one could go from 1 to a and then along a path of length not exceeding K to reach b, contradicting $b \in \Gamma \backslash \Gamma^{(n)}$. Thus $d(a, u_i) \leqslant i - n + K$. Similarly, $d(b, v_i) \leqslant i - n + K$. But now one can go from u_i to v_i by going along α^{-1} from u_i to a, then along the path represented by the edge of t connecting a and c, and then along α^{-1} from c to v_i. This is a distance not exceeding $2i + (3K - 2n)$, which is less than $2i$ since $2n > 3K$. But this contradicts $d(u_i, v_i) = 2i$. ∎

We are now in a position to apply Stallings' structure theorem to context-free groups. We begin with the special case of torsion-free groups. If $G = \langle X; R \rangle$ and $H = \langle Y; S \rangle$ are groups with X and Y disjoint sets, then the *free product* of G and H is the group $G * H = \langle X \cup Y; R \cup S \rangle$. The groups G and H are called the *factors* of the free product. A free product is *nontrivial* if neither of the factors is the trivial group. If G is any group, the *rank* $r(G)$ of G is the minimal number of generators of G. An important result about free products is Grushko's theorem, which states that $r(G * H) = r(G) + r(H)$.

In the torsion-free case, the Stallings structure theorem says that a finitely generated torsion-free group G has more than one end if and only if G is the infinite cyclic group or G is a nontrivial free product, $G = G_1 * G_2$. We now prove our theorem in the special case of torsion-free groups.

THEOREM II. *A finitely generated torsion-free group G is free if and only if it is context-free.*

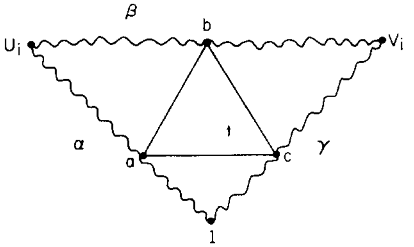

FIGURE 6

Proof. Suppose that G is context-free. We prove that G is free by induction in the rank of G. If $r(G)=0$, then G is the trivial group. If $r(G)=1$, then G is the infinite cyclic group since G is torsion-free. Now suppose that $r(G) \geqslant 2$. G is infinite since G is a nontrivial torsion-free group. By Lemma 6, G has more than one end so, by the Stallings structure theorem, G is a nontrivial free product, say $G=G_1 * G_2$. By Grushko's theorem, G_1 and G_2 have rank less than that of G. By Lemma 2, G_1 and G_2 are context-free. Hence, G_1 and G_2 are free groups by the induction hypothesis. Since the free product of free groups is free, G is free. ∎

In order to discuss the general situation we must first discuss two closely related group-theoretic constructions. Let G and H be groups and let A and B be subgroups of G and H, respectively, with an isomorphism $\psi: A \to B$ from A onto B. The group $\langle G * H; a=\psi(a), a \in A\rangle$ is called the *free product of* G *and* H *amalgamating the subgroups* A and B. We have here adopted the convention that if a presentation has been chosen for a group, then using the letter denoting that group in another presentation means that one has both the generators and relators of the given group. Thus a free product with amalgamation is obtained from the free product by identifying the elements of A with their images in B. The groups G and H are called the *factors* while A and B are called the *amalgamated subgroups*. A free product with amalgamation is *nontrivial* if the amalgamated subgroups are both proper subgroups in their respective factors.

A closely related construction is the following. Let G be a single group and let A and B be subgroups with $\psi: A \to B$ an isomorphism. The group $\langle G, t; t^{-}at=\psi(a), a \in A\rangle$ is called the HNN *extension* of G with *stable letter* t and *associated subgroups* A and B. (The letters HNN refer to G. Higman, B. H. Neumann, and H. Neumann, to whom the construction is credited). The group G is called the *base* of the HNN extension. For a detailed discussion of these constructions see, for example, [10].

We can now state the general Stallings structure theorem: A finitely generated group G has more than one end if and only if G is either a nontrivial free product with amalgamation or an HNN extension, where the amalgamated or associated subgroups are finite.

In order to use the general Stallings structure theorem in an inductive proof one must introduce the following definitions. Let G be a finitely generated group. An *accessible series* for G is a series of subgroups $G=G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_n$ such that each G_i has a decomposition as a nontrivial free product with amalgamation or as an HNN extension, where one of the factors or the base is G_{i+1} and the amalgamated or associated subgroups are finite. In short, $G=\langle G_{i+1} * G'_{i+1}; F_{i+1}=\psi(F_{i+1})\rangle$ or $G_i=\langle G_{i+1}, t_i; t_i F_{i+1} t_i^{-1}=\psi(F_{i+1})\rangle$, where G_{i+1}, G'_{i+1} are subgroups of G_i, and F_{i+1} is finite. The integer n is the *length* of the series. A group G is *accessible* if there is an upper bound on the lengths of accessible series for G, and the least upper bound s is the *accessibility length* of G. It is conjectured that all finitely generated groups are accessible but it is difficult to prove that specific groups are accessible. Since a nontrivial free product with amalgamation or an HNN extension is infinite, the accessibility length of a finite group is zero. It is an easy consequence of Grushko's theorem that torsion-free groups are accessible.

It is known that virtually free groups are accessible but we provide a fairly simple proof.

LEMMA 7. *A finitely generated virtually free group is accessible.*

Proof. As usual, if H is a subgroup of a group G, we denote the index of H in G by $[G:H]$. If G is a group define,

$$c(G) = \min\{r(H) + [G:H] : H \text{ a free subgroup of } G\}.$$

If G is finitely generated virtually free, then $c(G)$ is finite since a subgroup of finite index is again finitely generated. We prove by induction on $c(G)$ that the maximum length of an accessible series for G is less than $c(G)$. If $c(G) = 1$, then G is trivial and the result holds.

Now suppose that $G = \langle G_1 * G_1'; F = \psi(F)\rangle$, where F is finite, so that G_1 is the second group in an accessible series for G. If G_1 is finite the accessibility length of G is one, so assume that G_1 is infinite. Pick a free subgroup H so that $c(G) = r(H) + [G:H]$. Since H is free, H has trivial intersection with all conjugates of the subgroup F. The Hanna–Neumann subgroup theorem (see [10]) says that H has the structure $H = H_1 * \cdots * H_l * K$, where the H_i are distinct conjugates of subgroups of G_1 or G_1' and K is a free group having trivial intersection with the factors. Now the intersection of H and G_1 cannot be trivial since $[G_1 : G_1 \cap H] \leqslant [G:H]$ and G_1 is infinite. We may assume that $H_1 \subseteq G_1$. If there is only one factor in the displayed free product decomposition of H, then $H = H_1$ and it is clear that the index of H in G_1 is less than that of H in G. Thus $c(G_1) < c(G)$ and the result follows from the induction hypothesis. If there is more than one factor in the free product decomposition of H, we have $[G_1 : H_1] \leqslant [G:H]$ but $r(H_1) < r(H)$ by Grushko's theorem. Again, the result follows from the induction hypothesis.

A similar proof applies if $G = \langle G_1, t; tFt^{-1} = \psi(F)\rangle$, where F is finite. We may assume that G_1 is infinite and pick a free subgroup H of G so that $c(G) = r(H) + [G:H]$. The Hanna–Neumann subgroup theorem again says that $H = H_1 * \cdots * H_l * K$, where the H_i are distinct conjugates of subgroups of the base G_1 and K is a free group having trivial intersection with G_1. As before we may assume that $H_1 \subseteq G_1$ and conclude that $c(G_1) < c(G)$. ∎

We can now complete the proof of our main theorem.

THEOREM III. *Let G be a finitely generated group. Then G is virtually free if and only if G is context-free and accessible.*

Proof. The only part left to prove is that if G is context-free and accessible, then G is virtually free. The proof is by induction on the accessibility length s of G. If $s = 0$, then G has no decomposition as a nontrivial free product with amalgamation or an HNN extension with finite amalgamated or associated subgroups. But

Lemma 6 and the Stallings structure theorem then say that G is finite and thus is certainly virtually free. Suppose that $s > 0$. First suppose that $G = \langle G_1 * G_1'; F = \psi(F) \rangle$, where F is finite. Now G_1 and G_1' have accessibility length at most $(s-1)$. By Lemma 2, G_1 and G_1' are context-free. Thus G_1 and G_1' are virtually free by the induction hypothesis. We now need the results of Gregorac [3] and Karrass *et al.* [7] which say that the class of finitely generated virtually free groups is closed under taking free products with amalgamation or HNN extensions where the amalgamated or associated subgroups are finite. Thus G is virtually free. The proof is the same if $G = \langle G_1, t; tFt^{-1} = \psi(F) \rangle$ where F is finite. ∎

We wish to point out that there are interesting possibilities of correlating special classes of context-free languages with subclasses of the virtually free groups. In our second paper we shall show that the word problem of a context-free group is always a deterministic language. An interesting class of deterministic languages is the class of simple languages. For definitions and a thorough discussion of simple languages see Harrison [5]. If $G = \langle X; R \rangle$ is a group, define the *reduced word problem*, $W_0(G)$, to be the set of all words w on $X^{\pm 1}$ such that $w = 1$ in G but no nonempty proper prefix of w is equal to 1 in G. If L is a set of strings, then L^* is the set of all finite concatenations $\sigma_1 \dots \sigma_n$, where each $\sigma_i \in L$. It is easy to see that $W(G) = W_0(G)^*$. The class of simple languages seems very natural group-theoretically and we were led to conjecture the following result which Haring–Smith [4] has recently proven: A finitely generated group G is the free product of finitely many finite groups and infinite cyclic groups if and only if G has a presentation $\langle X; R \rangle$ whose reduced word problem is a simple language. Haring–Smith also geometrically characterizes such presentations: A presentation $G = \langle X; R \rangle$ has simple reduced word problem if and only if the Cayley graph $\Gamma(G)$ has the property that there are only finitely many simple closed curves passing through any vertex v of Γ.

After finishing this paper we discovered the work of Letichevskii and Smikun [12, 13] on context-free groups. Although phrased in a rather more complicated manner a characterization essentially equivalent to our Theorem I on Cayley graphs is given in [13].

REFERENCES

1. A. V. ANISIMOV, Group languages, *Kibernetika* **4** (1971), 18–24.
2. D. E. COHEN, "Groups of Cohomological Dimension One," Lecture Notes in Mathematics. No. 245, Springer-Verlag, Heidelberg, 1972.
3. R. GREGORAC, On generalized free products of finite extensions of free groups, *J. London Math. Soc.* **41**(1966), 662–666.
4. R. H. HARING–SMITH, "Groups and Simple Languages," Ph. D. Thesis, University of Illinois, Urbana, Ill., 1981.
5. M. A. HARRISON, "Introduction to Formal Language Theory," Addison–Wesley, Reading, Mass., 1978.
6. J. E. HOPCROFT AND J. D. ULLMAN, "Introduction to Automata Theory, Languages, and Computation," Addition–Wesley, Reading, Mass., 1979.

7. A. KARRASS, A. PIETROWSKI, AND SOLITAR, Finite and infinite cyclic extensions of free groups, *J. Austral. Math. Soc.* **16** (1973), 458–466.
8. W. MAGNUS, A. KARRASS, AND D. SOLITAR, "Combinatorial Group Theory," Wiley, New York, 1966.
9. D. E. MULLER AND P. E. SCHUPP, Pushdown automata, end, second-order logic, and reachability problems, *in* "Proceedings of the Thirteenth Annual ACM Symposium on the Theory of Computing," pp. 46–54, 1981.
10. P. E. SCHUPP, Reflections on HNN extensions, *in* "The Theory of Groups," Lecture Notes in Mathematics No. 372, Springer-Verlag, Heidelberg, 1974.
11. J. STALLINGS, "Group-Theory and Three-dimensional Manifolds," Yale Univ. Press, New Haven, Conn., 1971.
12. A. A. LETICHEVSKII AND L. B. SMIKUN, On a class of groups with solvable problem of automata equivalence, *Dokl. Akad. Nauk SSSR* **227** (1976).
13. L. B. SMIKUN, Relations between context-free groups and groups with a decidable problem of automata equivalence, *Kibernetika* (1976).