SU-SEL-69-058

On the Computational Complexity of Finite Functions

by

Philip M. Spiro

May 1968

Technical Report No. 1

Prepared under
Office of Naval Research Contract
Nonr 225(83) and
National Aeronautics and Space Administration
Grant 05-020-014

DIGITAL SYSTEMS LABORATORY
STANFORD ELECTRONICS LABORATORIES
STANFORD UNIVERSITY • STANFORD, CALIFORNIA

SU-SEL-69-058

ON THE COMPUTATIONAL COMPLEXITY OF FINITE FUNCTIONS

by

Philip M. Spira

May 1968

Distribution of this document is unlimited.

Technical Report No. 1

Prepared under

Office of Naval Research
Contract Nonr 225(83)
and
National Aeronautics and Space Administration
Grant 05-020-014

Digital Systems Laboratory

Stanford Electronics Laboratories

Stanford University Stanford, California

ABSTRACT

One of the most rapidly expanding fields of applied mathematics and engineering is automata theory, Although the term "automaton" is derived from "self-moving thing," the prime concern of automata theory is the study of information-processing devices, A specific example of information processing is computation, and thus the mathematical properties of devices which perform computations are of interest to automata theorists. In this thesis we investigate the computation by logic circuits of a certain class of functions having finite domain. To a given function f a number of so-called complexity criteria can be assigned relative to that class, e.g., the minimum computation time of or the minimum number of elements contained in any circuit of the class which is capable of computing f. Our prime criterion of interest will be computation time.

The type of circuits investigated in this thesis are called (d,r) circuits. A (d,r) circuit is composed of logical elements each having at most r inputs and one output. Each input value and output value is an element from the set $Z_d = \{0,1,\dots,d-1\}$, and each element has unit delay in computing its output. Thus a given element computes a function from Z_d^k to z_d, for some $k < r$, in unit time. The output of one element can be connected to inputs of any number of elements (including itself) and can also comprise one of the outputs of the circuit and an element receives a given one of its inputs either from the output of some element or from the inputs to the circuit. When individual elements are interconnected to form a (d,r) circuit we can associate 8 computation time with the entire circuit.

Specifically, let $f : X_1 \times \dots \times X_n \rightarrow Y$ be any function on finite

sets $X_1,\ldots,X_n$. Let C be a (d,r) circuit whose input lines are partitioned into n sets. Let $I_{C,j}$ be the set of configurations of values from Z_d on the j^{th} $(J = 1,2,\ldots,n)$, and let OC be the set of output configurations of the circuit. Then C is said to compute f in time τ if there are maps $g_j : X_j \to I_{C,j}$ $(j = 1,2,\ldots,n)$ and a 1 - 1 function $h : Y \to 0_C$ such that, if the input from time 0 through time $\tau - 1$ is $[g_1(x_1),\ldots,g_n(x_n)]$, then the output of C at time τ will be $h(f(x_1,\ldots,x_n))$.

Winograd has done pioneering work on the time of computation of finite functions by (d,r) circuits. He has derived lower bounds on computation time and has constructed near optimal circuits for many classes of finite functions.

A principal contribution of this thesis is a complete determination of the time necessary to compute multiplication in a finite group with a (d,r) circuit. A new group theoretic quantity 6(G) is defined whose reciprocal is the proper generalization of Winograd's a(G) to nonabelian groups. Then a novel method of circuit synthesis for group multiplication is given. In contrast to previous procedures, it is valid for any finite group--abelian or not. It is completely algebraic in character and is based upon our result that any finite group has a family of subgroups having a trivial intersection and minimum order $\delta(G)$. The computation time achieved is, in all cases, at most one unit greater than our lower bound. In particular, if G is abelian our computation time is never greater--and often considerably less--than Winograd's.

We then generalize the group'multiplication procedure to a method to compute any finite function, For given sets X_1, X_2 and Y and any family

of subsets of Y having a certain property called completeness, a corresponding heirarchy of functions having domain Xl x X_2 and range Y is established-- the position of a function depending upon its compatation time with our method, For reasons which we explain in the text this appears to be a very natural classification criterion. At the bottom of the heirarchy are invertible functions such as numerical addition and multiplication, and the position of a function in the heirarchy depends essentially upon how far it is from being invertible, For large $|X_1|$ and $|X_2|$ almost all functions are near the top, corresponding to the fact that nearly all $f : X_1 \times X_2 \rightarrow Y$ require computation time equal to the maximum required for any such function, The new method is then applied to the case of finite semigroup multiplication.

CONTENTS

ACKNOWLEDGMENT

I would like to express my deep gratitude to my thesis advisor, Professor Michael A. Arbib, for suggesting this research and for his excellent guidance throughout the course of its development. I am equally grateful to Professor Allen M Peterson for his friendship and support during my stay at Stanford,

Many of the ideas developed here were greatly aided by enlightening discussions I had with Dr. Shmuel Winograd, I am most appreciative of his help.

The research reported herein was supported in part by the National Aeronautics and Space Administration under Grant No. NsG-377(3208) and in part by the U.S. Air Force Office of Scientific Research, Information Sciences Directorate, under Grant No. AF-AFOSR-1198-67.

I. INTRODUCTION TO THE COMPUTATIONAL COMPLEXITY OF FINITE FUNCTIONS

1. INTRODUCTION

This thesis is concerned with the computational complexity of finite functions. There are various ways to measure complexity. Two of the most natural complexity measures of a circuit which computes a finite function are its time of operation and the amount of hardware it contains. We shall first review and make a few comments upon a number of important papers adopting one or both of these definitions of complexity.

2. CONTACT CIRCUITS

The first measure of the complexity of a switching function f in which we are interested is the minimum number of contacts in a relay contact network which realizes f as one of its transmission (or hindrance) functions.

Let

$$\mathfrak{F}_{n,m} = (f : f : \{0,1\}^n \to \{0,1\}^m\}$$

$$\mathfrak{F}_n = \mathfrak{F}_{n,1} = \{f : f : \{0,1\}^n \to \{0,1\}\}$$

$$X(f) = \text{the least number of elements in a circuit realizing } f$$

$$\lambda(n,m) = \max\{\lambda(f) : f \in \mathfrak{F}_{n,m}\}$$

$$\lambda(n) = \max\{\lambda(f) : f \in \mathfrak{F}_n\}$$

The quantity A(n) is usually called Shannon's function. It was Shannon [Ref. 1] who first introduced and studied it obtaining a lower and an upper bound on it.

Let $f \in \mathfrak{F}_n$ and let $m \leq n$. Then [Ref. 2, p. 78] f can be factored into conjunctive normal form as

$$f(x_1,\ldots,x_n) = \left[x_1 + \ldots + x_{n-m} + v_1(x_{n-m+1},\ldots x_n)\right]$$
$$\ldots\left[\bar{x}_1 + \ldots + x_{n-m} + v_{2^{n-m}}(x_{n-m+1},\ldots x_n)\right]$$

where each $v_i \in \mathfrak{F}_m$. Shannon's approach in deriving a lower bound on $\lambda(n)$ is to give a general synthesis procedure in which the terms of the form $x_1^* + \ldots + x_{n\ m}^*$, where each $x_1^* = x_1$ or $\bar{x}_i$, are realized by one type of network which is then cascaded in series with another network which realizes every function in $\mathfrak{F}_m$ as one of its hindrance functions. (Note that the set of all terms of this form in $x_1,\ldots,x_{n-m}$ are called the minterms in these variables.) This yields a circuit capable of realizing any $f \in \mathfrak{F}_n$. The number of relays in the entire circuit depends upon m; thus Shannon minimizes it for $1 < m < n$ to obtain his lower bound.

The network which Shannon uses to obtain the minterms is known as a complete tree, which requires $2^{n-m+1} - 2$ contacts. Shannon then shows by induction that a network realizing each function of m variables can be realized with 2^{2^m+1} contacts. Thus

$$\lambda(n) \leq 2^{n-m+1} - 2 + 2^{2^m+1} < 2^{n-m+1} + 2^{2^m+1}; \quad 1 < m < n$$

for contact circuits. A study and minimization of this bound yields the results:

a) $\lambda(n) < \dfrac{2^{n+3}}{n}$

b) $\lambda(n) < \dfrac{2^{n+2}}{n}$ for almost all n

c) Given $\epsilon > 0$, there is an infinite sequence

(ni) for which $\lambda(n_i) < \dfrac{2^{n_i+1}}{n_i}(1 + \epsilon)$

As we shall see later an improved synthesis procedure derived by Lupanov yields a considerably sharper bound.

Shannon uses a different approach to obtain his lower bound on x(n). When the number of functions in $\mathfrak{F}_n$ is compared with the number of different possible networks containing a given number of relay contacts it is demonstrable that:

a) For any $\epsilon > 0$ there is a finite N such that

$$n > N \Longrightarrow \lambda(f) > \frac{2^n}{n}(1 - \epsilon) \quad \text{for almost all } f \in \mathfrak{F}_n$$

b) There is a positive number A such that

$$A(n) > A\,\frac{2^n}{n} \quad \text{for all } n$$

Lupanov [Ref. 3,4] is able to sharpen Shannon's results by a different synthesis method. He finds a network to realize the minterms in n - m variables requiring only

$$\frac{2^{n-m+1}}{n-m} + \frac{2^{n-m+2}}{n-m-\log_2(n-m}$$

contacts-asymptotically half as many relay contacts as a complete tree. Such a circuit is called a Lupanov tree. This enables him to prove that

$$\lambda(n) = \frac{2^n}{n}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right)$$

for contact networks.

Lupanov's procedure has a certain weakness. Consider a network which realizes each minterm in n variables as one of its transmission functions. The network is said to have a sneak path if there is a nonzero transmission function between two output nodes. This property makes it unsuitable for certain application, and a Lupanov tree has many sneak paths. Moore [Ref. 5] shows that a network realizing all minterms in n variables and having no sneak paths requires at least $2^{n+1} - 2$ contacts. Thus the growth rate of $\lambda(n)$ exceeds $2^n/n$ for disjunctive (no sneak path) networks.

3. CIRCUITS OF FUNCTIONAL ELEMENTS

A class of switching elements is said to be functionally complete for $\mathfrak{F}_{p,q}$ if, given any $f \in \mathfrak{F}_{p,q}$, there is a network containing only elements in the class which has inputs $x_1, \ldots X_n$ and output $f(x_1 \ldots, x_n)$. (Clearly such a set will also be complete for $\mathfrak{F}_{p',q'}$ for any p' and any q'.) Let S_1 be such a class and let $f \in \mathfrak{F}_{p,q}$. With each type of element in S_1 associate a fixed positive weight. Then let $\lambda_1(f)$ be the minimal weighted sum of the number of different elements in a minimal sum circuit which computes f and let

$$\lambda_1(p, q) = \max\left\{\lambda_1(f) : f \in \mathfrak{F}_{p,q}\right\}$$

Muller [Ref. 6]shows that if there is another set S_2 with corresponding complexity measure λ_2, then there are constants K_1 and K_2 independent of p and q such that

$$K_1\lambda_1(p,q) \leq \lambda_2(p,q) \leq K_2\lambda_1(p,q)$$

This is easily seen, e.g., K_2 is the maximal complexity of an element in S_2 realized by elements in S_1. Thus to investigate growth rate of $\lambda(p,q)$ associated with any complete set S one need only do so for a particular set. Muller also looks at the rate of increase of $\lambda(p,q)$ with $n = p + \log_2 q$ for the case in which $1 < q < 2^{2^p}$, which is clearly general since 2^{2^p} is the cardinality of $\mathfrak{F}_p$. He concludes that there exist constants C_1 and C_2 for, independent of p and q, such that

$$c_1 \frac{2^n}{n} \leq \lambda(p,q) \leq c_2 \frac{2^n}{n}$$

For the case $q = 1$, i.e., for functions in $\mathfrak{F}_{p,1} = \mathfrak{F}_n$, this parallels the result for relay circuits. Unfortunately it appears that Muller's proof of the upper bound is incorrect,[f] although Lupanov showed the result is true at least for $q = 1$. His lower bound is proved by an argument comparing the number of functions in $\mathfrak{F}_{p,q}$ with the number of networks of a given complexity, analogous to Shannon's proof for contact networks.

Lupanov [Ref. 4,7] studies the same problem for the case $q = 1$ $(n = p)$. He chooses a complete set consisting of two input and gates, two imput or gates, and inverters and shows that the associated complexity measure with all weights unity, $\lambda(n)$, satisfies

$$\lambda(n) \leq \frac{2^n}{n} \left(1 + O\left(\frac{\log_2 n}{n}\right)\right)$$

4. TIME AND NUMBER OF ELEMENTS

Another measure of the complexity of a finite function is the time necessary to compute it with a circuit of a given class. In contrast to the work described in preceding sections most results in this area are for specific classes of functions more limited than, e.g., all $f \in \mathfrak{F}_n$.

Ofman [Ref. 8] introduces a certain class of logical circuits which we now define in the more lucid terminology of Winograd [Ref. 1]. Call a circuit C a (d,r) circuit (or (d,r) automation) if it is composed of functional elements each having at most r input lines, one splittable

[f] In his proof Muller seems to assume that all functions of the first $k+1$ or k out of the p Boolean variables are available at no cost in elements, which indeed they are not.

output line, and carrying on all lines elements from the set $Z_d = \{0,1,\ldots,d - 1)$. In addition each element will have unit deley in computing its output. Thus we can associate a computation time with the entire circuit.

Ofman [Ref. 8]is interested in specific classes of functions which can be parameterized by a number n, e.g., addition of two n bit numbers. He is primarily concerned with rate of growth with n of the number of elements and rate of growth with n of the computation time for (d,r) circuits computing these functions.

The classe of functions of the most interest which he considers is addition of r binary n bit numbers. He notes that conventional bit-by-bit addition of two n bit numbers requires $\lambda = O(n)$ elements and $\tau = O(n)$ computation time. By a method similar to the look ahead carry method used on some modern digital computers he achieves $\tau = O(\log_2 n)$. For addition of r numbers his growth rates are $A = O(n)$ and $\tau = O(\log r \cdot n)$. His method of proof makes his results not applicable for finite n, but only in the limit as $n \to \infty$, since it uses $(3,2)$ circuits and the observation that this does not change the growth rate. By the methods of Ch. II it can be shown that the growth rate of τ in his circuits is as small as possible.

A similarly slanted paper by Karatsuba and Ofman [Ref. 9]demonstrates that for any s; $1 < s < n$ there is a $(2,2)$ circuit to multiply two n bit numbers with growth rates $\lambda = O(n^2/s)$ and $\tau = O(s \log_2 n)$. This result is derived from the previous paper. To multiply the two numbers first the product of one by each bit of the other is formed. Then the addition of the first $\lceil n/s \rceil$ products is performed by means of the circuit of [Ref. 8] where $\lceil x \rceil$ in the smallest integer > x. Then that number is added to the next $\lceil n/s \rceil$ products, and the process is continued until the final result is obtained. The interesting trade-off indicated above is the result.

Multiplication of two n bit numbers is also the subject of a paper by Toom [Ref. 10]. He shows that for sufficiently large c, e.g.; c = 32, there is a network to perform this multiplication for any n having $\lambda \leq C_1 nc^{\sqrt{\log_2 n}}$ and $\tau \leq C_2 nc^{\sqrt{\log_2 n}}$ for constants C_1 and C_2. There is a vast literature on computational complexity of Turing machines (for the definition and basic concepts of Turing machines see [Ref. 11]) which we shall not attempt to cover. However, we mention in passing that Cook [Ref. 12] has shown that there is a multitape Turing machine to multiply any two n-digit numbers within $n2^{5\sqrt{\log n}}$ steps, for all n, in the manner of Toom's algorithm.

5. WINOGRAD'S WORK ON TIME OF COMPUTATION

Winograd [Ref. 13,14] has done pioneering work on the time of computation of finite functions by (d,r) circuits. Let $f : X_1 \times \ldots \times X_n \rightarrow Y$ be a function on finite sets. Let C be a (d,r) circuit, and let the input lines of C be partitioned into n sets with $I_{C,j}$ the set of possible configurations from Z_d on the j^{th} $(j = 1,2,\ldots,n)$ and let O_C be the set of possible output configurations. Then C is said to compute f in time τ if there are maps $g_j : X_j \rightarrow I_{C,j}$ $(j = 1,2,\ldots,n)$ and a 1 - 1 function $h : Y \rightarrow O_C$ such that, if the input to C from time 0 to time $\tau - 1$ is $[g_1(x_1),\ldots,g_n(x_n)]$, then the output at time τ will be $h(f(x_1,\ldots,x_n))$. He derives lower bounds for many classes of functions and also constructs circuits which operate in near the lower bound times. We summarize these in a chart. The details appear in [Ref. 13, 14] or in his thesis [Ref. 15].

We must first give some definitions

Definition 4.1. Let H be a group. Say $P(H) = 1$ if there is an $a \in H$ with $a \neq e$ such that every nontrivial subgroup of H contains a. Let G be a group. Then

$$\alpha(G) = \max\{|H| : H < G \text{ and } P(H) = 1\}$$

Definition 4.2. Let $a(N) = \alpha(Z_N)$ where Z_N is $\{0,1,...,N - 1\}$ under addition modulo N.

Definition 4.3. Let AN be the group of positive integers less than and relatively prime to N under multiplication modulo N. Let $\beta(N) = \alpha(A_N)$.

Definition 4.4. Let $Q_m = \mathrm{lcm}\{1,2,...,m\}$ and let $\gamma(N) = \min\{m : Q_m \geq N\}$

Definition 4.5. Let $\lceil x \rceil$ be the smallest integer $\geq x$ and $\lfloor x \rfloor$ be the greatest integer $\leq x$.

We can now give Winograd's lower bounds and the realization times of his circuits. Note all lower bounds are valid for $d > 2$ and $r > 2$. All realizations are valid for $d \geq 2$ and $r \geq 3$ except that for ϕ_1, which is also valid for $r = 2$.

Function	Lower Bound Time	Realization Time
$\phi_1 : Z_N \times Z_N \rightarrow \{0,1\}$ $\phi_1(x,y) = 1$ if $x = y$ $= 0$ if $x \neq y$	$\left\lceil \log_r 2 \left\lceil \log_d N \right\rceil \right\rceil$	$1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d N \right\rceil \right\rceil \right\rceil$
$\phi_2 : Z_N \times Z_N \rightarrow \{0,1\}$ $\phi_2(x,y) = 1$ if $x \geq y$ $= 0$ if $x < y$	$\left\lceil \log_r 2 \left\lceil \log_d N \right\rceil \right\rceil$	Not Given

G a finite group
$\phi_G : G \times G \rightarrow G$
group multiplication
	$\left\lceil \log_r 2 \left\lceil \log_d \alpha(G) \right\rceil \right\rceil$	$2 + \left\lceil \log_{\lfloor r+1/2 \rfloor} \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \alpha(G) \right\rceil \right\rceil \right\rceil$
		if G is abelian. Not given
		for G not abelian.

$\phi_4 : Z_N \times Z_N \rightarrow Z_N$ $\phi_4(x,y) = x + y \mod N$	$\left\lceil \log_r 2 \left\lceil \log_d \alpha(N) \right\rceil \right\rceil$	$2 + \left\lceil \log_{\lfloor r+1/2 \rfloor} \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \alpha(N) \right\rceil \right\rceil \right\rceil$
$\phi_5 : Z_N \times Z_N \rightarrow \{0,1\}$ $\phi_5(x,y) = \left\lfloor \frac{x + y}{N} \right\rfloor$	$\left\lceil \log_r 2 \left\lceil \log_d N \right\rceil \right\rceil$	$1 + \left\lceil \log_{\lfloor r+1/2 \rfloor} \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d N \right\rceil \right\rceil \right\rceil$
$\phi_6 : Z_N \times Z_N \rightarrow Z_N$ $\phi_6(x,y) = xy \mod N$	$\left\lceil \log_r 2 \left\lceil \log_d \beta(N) \right\rceil \right\rceil$	$3 + \left\lceil \log_{\lfloor r+1/2 \rfloor} \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \beta(N) \right\rceil \right\rceil \right\rceil$

Function	Lower Bound Time	Realization Time
$\phi_7 : Z_N \times Z_N \rightarrow Z_N$ $\phi_7(x,y) = \left\lfloor \frac{x \cdot y}{N} \right\rfloor$	$\left\lceil \log_r 2 \left\lceil \log_d \left\lfloor N^{1/2} \right\rfloor \right\rceil \right\rceil$	Not Given
$\phi_8 : Z_N \times Z_N \rightarrow Z_{2N-1}$ $\phi_8(x,y) = x + y$	$\left\lceil \log_r 2 \left\lceil \log_d \gamma \left(\left\lceil \frac{N}{2} \right\rceil \right) \right\rceil \right\rceil$	$2 + \left\lceil \log_{\lfloor r+1/2 \rfloor} \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \gamma (2N - 1) \right\rceil \right\rceil \right\rceil$
$\phi_9 : \{1,\ldots,N\}^2 \rightarrow \{1,\ldots,N^2\}$ $\phi_9(x,y) = xy$	$\left\lceil \log_r 2 \left\lceil \log_d \gamma \left(\left\lceil \frac{\lfloor \log_2 2N \rfloor}{2} \right\rceil \right) \right\rceil \right\rceil$	$2 + \left\lceil \log_{\lfloor r+1/2 \rfloor} \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \gamma \left(2 \left\lfloor \log_2 N \right\rfloor - 1 \right) \right\rceil \right\rceil \right\rceil$

6. CONTRIBUTIONS OF THIS RESEARCH

A principal contribution of this thesis is a complete determination of the time necessary to compute multiplication in a finite group with a (d,r) circuit. A new group theoretic quantity 6(G) is defined whose reciprocal is the proper generalization of Winograd's $\alpha(G)$ to nonabelian groups. Then a novel method of circuit synthesis for group multiplication is given. In contrast to previous procedures, it is valid for any finite group--abelian or not, It is completely algebraic in character and is based upon our result that any finite group has a family of subgroups having a. trivial intersection and minimum order F(G). The computation time achieved is, in all cases, at most one unit greater than our lower bound. In particular, if G is abelian our computation time is never greater--and often considerably less--than Winograd's,

We then generalize the group multiplication procedure to a method to compute any finite function. For given sets X_1, X_2 and Y and any family of subsets of Y having a certain property called completness, a corresponding heirarchy of functions having domain $X_1 \times X_2$ and range Y is established--the position of a function depending upon its computation time with our method. For reasons which we explain in the text this appears to be a very natural classification criterion. At the bottom of the heirarchy are invertible functions such as numerical addition and multiplication, and the position of a function in the heirarchy depends essentially upon how far it is from being invertihlc. For large $|X_1|$ and $|X_2|$ almost all functions are near the top, corresponding to the fact that nearily all $f : X_1 \times X_2 \to Y$ require computation time equal to the maximum required for any such function. The new method Is then applied to the case of finite semigroup multiplication.

REFERENCES

[1] Shannon, C. E., "The Synthesis of Two-Terminal Switching Circuits," Bell Syst. Tech. J., vol. 28, no. 1, 1949, pp. 59-98.

[2] McCluskey, E. J., Introduction to the Theory of Switching Circuits McGraw-Hill Book Company, New York, 1965.

[3] Lupanov, 0. B., "The Synthesis of Contact Networks," Dokl. Akad. Nauk SSR, vol. 119, no. 1, 1958, pp. 23-26.

[4] Lupanov, 0. B., "On the Synthesis of Certain Classes of Control Systems," Prob. Kib., vol. 10, no. 3, 1963, PP. 63-97.

[5] Moore, E. F., "Minimal Complete Relay Decoding Networks," IBM J. Res. and Dev., vol. 4, no. 5, 1960, pp. 525-531.

[6] Muller, D. E., "Complexity in Electronic Switching Circuits," IRE, vol. EC-5, no. 1, 1956, pp. 15-19.

[7] Lupanov, 0. B., "A Method of Circuit Synthesis," Izvestia VUZ, Radiophysics Series, vol. 1, no. 1, 1959, pp. 120-140 (unavailable).

[8] Ofman, Yu, "On the Algorithmic Complexity of Discrete Functions," Dokl. Akad. Nauk SSR, **vol.** 145, no. 1, 1962, pp. 48-51.

[9] Karatsuba, A. and Yu Ofman, 'Multiplication of Multi-Digit Numbers With Computers," Dokl. Akad. Nauk SSR, vol. 145, no. 2, 1962, p. 293.

[10] Toom, A. L., 'The Complexity of a Scheme of Functional Elements Realizing the Multiplication of Integers," Dokl. Akad. Nauk, SSR, vol. 150, no. 3, 1963, pp. 496-498.

[11] Minsky, M., Computation: Finite and Infinite Machines, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1967.

[12] Cook, S. A., "On the Minimum Computation Time of Function," Ph.D. Thesis, Department of Mathematics, Harvard University, May 1966,

(Report No. BL-41 of the Harvard Computation Laboratory).

[13] Winograd, S., "On the Time Required to Perform Addition," J. ACM, vol. 12, no. 2, 1965, pp. 277-285.

[14] Winograd, S., "On the Time Required to Perform Multiplication," J. ACM, vol. 14, no. 4, 1967, pp. 793-802.

[15] Winograd, S., "On the Time Required to Perform Computer Operations," Ph.D. Thesis, Mathematics Department, New York University.

An annotated list of papers not reviewed above but containing similar results:

[16] Krichevskii, R. E., "Complexity of Contact Circuits Realizing a Function of Logical Algebra," Dokl. Akad, Nauk SSR, vol. 151, no. 4, 1963, pp. 803-806.

Let $f_0(x_1,\dots,x_n) = \bigvee_{1 \le i < j \le n} x_i x_j$. Then an arbitrary relay circuit realizing f_0 and not containing breaking contacts has no less than $C_2\, n \log n$ contacts for some constant C_2, Such a circuit exists as a series parallel circuit for all n.

[17] Lupanov, O. B., "Complexity of Formula Realization of Functions of Logical Algebra," Prob. Kib., vol. 3, 1960, pp. 61-80.

Let L(n) be the function analogous to $\lambda(n)$ in the case of formulae constructed from functions of a certain basis. Then $L(n) = \rho \cdot 2^n/\log_2 n$.

[18] Lupanov, O. B., "Implementing the Algebra of Logic Functions in Terms of Bounded Depth in the Basis of &, V, -," Dokl. Akad. Nauk SSR, vol. 136, no. 5, 1961, pp. 1041-1042.

In terms of asymptotic behavior of Shannon's function two input and gates, two input or gates, and inverters in formulas of depth three will achieve bounds but depth two is not sufficient.

[19] Lupanov, O. B., "On Comparing the Complexity of the Realization of Monatonic Functions by Contact Networks Containing only Closing Circuits and by Arbitrary Contact Networks," Dokl. Akad. Nauk SSR, vol. 144, no. 6, 1962, pp. 1245-1248.

Let $L(f)$ = number of contacts necessary to realize f

$L^{+}(f)$ = number of closing contacts necessary to realize f.

$A(n) = \max\limits_{f\in\mathfrak{F}_n} \frac{L^{+}(f}{L(f7)}_{4}$. Then $\lambda(n) \to \infty$.

[20] Reznik, V. I., "The Realization of Monotonic Functions by Means of Networks of Functional Elements," Dokl. Akad. Nauk SSR, vol. 139, no. 3, 1961, pp. 566-569.

A function is monotonic if $\exists$ m such that the function is 1 whenever at least m of the x_1's are. The number of elements in synthesis of such functions is here asymptotically upper bounded by $b(2^n/n^{3/2})\log^2 n$.

II. THE TIME REQUIRED FOR GROUP MULTIPLICATION[†]

1. THE MODEL

The model we will adopt is basically that of Winograd [Refs. 1,2]. We consider logical circuits composed of elements each having at most r inputs lines, one splittable output line, and unit delay in computing their outputs. Each line carries values from the set $Z_d = \{0,1,\ldots,d - 1)$. The input lines are partitioned into n sets with $I_{C,j}$ the set of possible configurations on the j^{th} $(j = 1,2,\ldots,n)$. O_C is the set of possible configurations. Such a circuit is called a (d,r) circuit.

<u>Definition 1.1</u>. Let $\phi : X_1 \times X2 \times \ldots X_n \rightarrow Y$ be a function on finite sets. A circuit C is said to compute ϕ in time τ if there are maps $g_j : X_j \rightarrow I_{C,j}$ $(j = 1,2,\ldots,n)$ and a 1 - 1 function $h : Y \rightarrow 3_C$ such that if C receives constant input $[g_1(x_1),\ldots,g_n(x_n)]$ from time 0 through time $\tau - 1$, then the output at time τ will be $h(\phi(x_1,\ldots,x_n))$.

2. THE BASIC LEMMA

We now derive a general lower bound on the time for a (d,r) circuit to compute a given finite function ϕ. It makes explicit the method underlying the results of Winograd. It is dependent upon the output code h introduced in the last section, and makes use of a new concept we shall

[†] Some of this material was presented at the Eighth IEEE Annual Symposium on Switching and Automata Theory, (See Spira, P. M. and M. A. Arbib, "Computation Time for Finite Groups, Semigroups and Automata," <u>IEEE Conference Record of the Eighth Annual Symposium on Switching and Automata Theory</u>, October 1967.)

introduce--that of separable sets. First, some preliminary definitions are necessary.

Definition 2.1. Let $\lceil x \rceil$ be the smallest integer $\geq x$; let $\lfloor x \rfloor$ be the largest integer $\leq x$; $|S|$ be the cardinality of the set S.

Definition 2.2. For a (d,r) circuit let $h_j(y)$ be the value on the j^{th} output line when the overall output configuration is $h(y)$.

Definition 2.3. Let $\phi : X_1 \times \ldots \times X_n \to Y$ and let C compute ϕ. Then $S \subseteq X_m$ is called an h_j-separable set for C in the m^{th} argument of ϕ if whenever s_1 and s_2 are distinct elements of S we can find $x_1, x_2, \ldots, x_{m-1}, x_{m+1}, \ldots, x_n$ with $x_i \in X_i$ such that

$$h_j(\phi(x_1,\ldots,x_{m-1},s_1,x_{m+1},\ldots,x_n)) \neq h_j(\phi(x_1,\ldots,x_{m-1},s_2,x_{m+1},\ldots,x_n))$$

Lemma 2.4. In a (d,r) circuit the output of an element at time τ can depend upon at most r^τ input lines.

Proof.

Just consider the fan-in with modules having r input lines to the height of τ. ▌

This observation, first made by Winograd, plus the concept of separable sets suffices to prove.

Lemma 2.5. (Tile Basic Lemma . Let C be a (d,r) circuit which computes ϕ in time τ. Then

$$\tau \geq \max_j \left\{ \left\lceil \log_r \left(\lceil \log_d |S_1(j)| \rceil + \ldots + \lceil \log_d |S_n(j)| \rceil \right) \right\rceil \right\}$$

where $S_i(j)$ is an h_j-separable set for C in the j^{th} argument of ϕ.

Proof.

The j^{th} output line at time τ must depend upon at least $\lceil \log_d |S_i(j)| \rceil$ input lines from $I_{C,i}$ or else there would be two elements of $S_i(j)$ which were not h_j-separable. Thus the j^{th} output depends upon at least $\lceil \log_d |S_1(j)| \rceil + \ldots + \lceil \log_d |S_n(j)| \rceil$ input lines and this number is at most r^{τ}. ▌

With lemma 2.5 we have exposed the methodology implicit in Winograd's treatment of the times required for addition and multiplication. By making it explicit we not only quickly obtain some of Winograd's results in the rest of this section but also shall give a deeper analysis of other concepts and shall treat a much wider class of functions in the sequel.

Corollary 2.6. Let $\phi : Z_N \times Z_N \rightarrow \{0,1\}$ be

$$\phi(x,y) = \begin{cases} 1 & \text{if } x \leq y \\ 0 & \text{if } x > y \end{cases}$$

Then if C is a (d,r) circuit to compute ϕ in time τ, we have

$$\tau \geq \lceil \log_r 2 \lceil \log_d N \rceil \rceil$$

Proof.

Pick j such that $h_j(0) \neq h_j(1)$. Then Z_N is an h_j-separable set for C in both the first and the second arguments of ϕ since, if $x > y$, $\phi(x,y) \neq \phi(y,y)$ and $\phi(x,y) \neq \phi(x,x)$. ▌

Corollary 2.7. Let $\phi : Z_N \times Z_N \rightarrow Z_N$ be

$$\phi(x_1,x_2) = \left\lfloor \frac{x_1 \cdot x_2}{N} \right\rfloor$$

Then, if C computes ϕ in time τ

$$\tau \geq \lceil \log_r 2 \lceil \log_d \lfloor N^{1/2} \rfloor \rceil \rceil$$

Proof.

Pick j such that $h_j(0) \neq h_j(1)$. Let $m = \lfloor N^{1/2} \rfloor$. Then $\{1,2,\ldots,m\}$ is an h_j-separable set for C in both arguments of ϕ, since for each $x \neq y$ with $x,y \in \{1,2,\ldots,m\}$ we may chose $w \in Z_N$ such that $x \cdot w < N < y \cdot w < Z_N$ to yield $\phi(x,w) = 0$, $\phi(y,w) = 1$. By symmetry this holds for the second argument as well and lemma 2.2 yields the result. |

We close this section with an example which shows that the size of separable sets can be strongly dependent upon the output code of the circuit which computes a given ϕ.

Example 2.8.

Let $\phi : Z_N \times Z_N \to Z_{N^2}$ be numerical multiplication with $N = 2^8$. Consider an output code in which, if the output value is M then the i^{th} line carries the i^{th} bit in the binary expansion for M. Then there are 16 output lines. Pick any $x \neq y$ with $x,y \in Z_N$. Then their binary expansions differ in at least one place, say the k^{th}. Choose $z = 2^{8-k}$. Then

$$h_8(\phi(y,z)) \neq h_8(\phi(x,z))$$

and

$$h_8(\phi(z,y)) \neq h_8(\phi(z,x))$$

So there is an h_8-separable set of size 2^8 in both arguments of ϕ.

Now consider the same ϕ but let the output code for z be the binary representation of the exponents in its prime decomposition. Let the first six output lines code the exponent of two in the result. Pick $x,y \in Z_N$ such that x and y do not have the same power of two in their prime decomposition, the powers differing in, say, the k^{th} place of their binary expansion. Then, letting $z = 2^{3-k}$,

$$h_3(\phi(x,z)) \neq h_3(\phi(y,z))$$

and

$$h_3(\phi(z,x)) \neq h_3(\phi(z,y))$$

Thus, since an element of Z_N can have eight different exponents of two in its prime decomposition, there is an h_3-separable set of size 8 in both arguments of ϕ. One easily sees that this is the maximal size of any separable set, since two is the smallest prime. Note, however, that this output code requires thirty-nine output lines.

3. REVIEW OF PREVIOUS RESULTS

Several authors have investigated the computation time necessary for a (d,r) circuit to add modulo N. Ofman [Ref. 3] gave a circuit for the special case $N = 2^n$. Significant results were obtained by Winograd [Refs. 1,2]. He derived a lower bound which we will review, and a (d,r) circuit with computation time near the lower bound. Since any finite abelian group is the direct product of cyclic groups [Ref. 4, p. 40] his results are applicable to abelian group multiplication as well.

Definition 3.1. Let H be a group. Say H has property P and write $P(H) = 1$, in case there is an element $a \in H$ with $a \neq e$ such that every nontrivial subgroup of H contains a. This will be denoted by P(a,H) = 1. Let $\alpha(G)$ be the maximal order of $H \leq G$ such that P(H) = 1.

Lemma 3.2. (Winograd). If G is abelian a(G) is the maximal order of a prime power cyclic subgroup contained in G.

Proof.

See [Ref. 1, p. 280].▮

We now give a complete characterization of a(G).

Definition 3.3. The generalized quaternion group Q_n is the group of order 2^n with two generators a and b satisfying

$$a^{2^{n-1}} = e; \; b^2 = a^{2^{n-2}}; \; ba = a^{-1}b$$

Theorem 3.4. A p-group contains a unique subgroup of order p if it is cyclic or a generalized quaternion group. (It must be cyclic if p is odd.)

Proof.

See Hall [Ref. 4, p. 189].▮

Corollary 3.5. Let G be any finite group. Then a(G) is either the order of the largest cyclic p-subgroup of G or the order of the largest generalized quaternion group contained in G, whichever is larger.

Proof.

Let H be any subgroup of G. If P(H) = 1 then $|H| = p^n$ for some prime p, for if not there would be another prime q dividing $|H|$ and consequently there would be elements u and v in H with $o(u) = p$ and $o(v) = q$. But then $\langle u\rangle \cap \langle v\rangle$ would contain only the identity. Assume $|H| = p^n$. Then every nontrivial subgroup of H contains a subgroup of order p. Thus P(H) = 1 iff H contains a unique subgroup of order p, i.e., iff H is cyclic or a generalized quaternion group. ▮

The quantity $\alpha(G)$ is critical to Winograd's lower bound time for group multiplication, which we now state. In the following section we shall give a new lower bound which is in general higher but is the same as his if the group of interest is abelian.

Theorem 3.6. (Winograd). Let G be any finite group. Let C be a (d,r) circuit which computes $\phi : G \times G \to G$ where

$$\#(a,b) = ab$$

Then C requires computation time τ where

$$\tau \geq \left\lceil \log_r 2 \left\lceil \log_d \alpha(G) \right\rceil \right\rceil$$

Proof.

See Winograd [Ref. 1]. ▮

Winograd also gives a procedure for constructing a circuit to multiply in an abelian group G with computation time

$$\tau = 2 + \left\lceil \log_{\lfloor (r+1)/2 \rfloor} \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \alpha(G) \right\rceil \right\rceil \right\rceil$$

which is valid for $r \geq 3$ and $d > 2$. We will give a completely different method for constructing circuits which will be valid for $r > 2$ and $d > 2$, and will work whether or not the group is abelian. Furthermore, for a given abelian group and a given d and r, our computation time will underbound Winograd's.

4. THE LOWER BOUND

In this section we shall give a new lower bound for the time required for a (d,r) circuit to perform group multiplication and shall compare it to Winograd's bound. Let G be any finite group and let $\phi : G \times G \rightarrow G$ be group multiplication. Let C be a (d,r) circuit which computes ϕ. Let $h_j(g)$ be the value on the j^{th} output line of C when the output is h(g).

Definition 4.1. Let $x,y \in G$. Then we say that x any y are R_j-equivalent if $h_j(gx) = h_j(gy)$ for all $g \in G$ and that they are L_j-equivalent if $h_j(xg) = h_j(yg)$ for all $g \in G$. Then clearly R_j and L_j are equivalence relationships and we write $R_j(g)$ for the R_j-equivalence class of g and $L_j(g)$ for the L_j-equivalence class of g.

Lemma 4.2. $R_j = R_j(e)$ and $L_j = L_j(e)$ are groups for all output lines of c. Furthermore, for any $g \in G$, $R_j(g) = R_jg$ and $L_j(g) = gL_j$.

Proof,

Say a, $b \in R_j$. Let $c \in G$. Then

$$h_j(ab^{-1}c) = h_j(bb^{-1}c) = h_j(c)$$

So $ab^{-1} \in R_j$ and it is a group. Now pick any $g \in G$. Then $d \in R_j(g)$ iff $h_j(dc) = h_j(gc)$ for all $c \in G$. But this is true iff $h_j(dg^{-1}c) = h_j(c)$, i.e., iff $dg^{-1} \in R_j$. The other half of the lemma follows dually. ▮

Maximal separable sets are determined by

Lemma 4.3. A maximal size h_j-separable set in the first argument of ϕ consists, of exactly one representative from each left coset of R_j in G. It thus has size $|G|/|R_j|$. A dual result is true for separable sets in the second argument.

Proof.

Direct from lemma 4.2 and the definition of separable sets. ∎

We now have all the pieces we need for a lower bound on group multiplication which is output code dependent.

Lemma 4.4. Let C be a (d,r) circuit to multiply in G in time τ.

Then

$$\tau \geq \max_j \left\{ \left\lceil \log_r \left(\left\lceil \log_d \frac{|G|}{|R_j|} \right\rceil + \left\lceil \log_d \frac{|G|}{|L_j|} \right\rceil \right) \right\rceil \right\}$$

Proof.

Direct from lemma 2.5 and lemma 4.3. ∎

A bound over all output codes will be derived by maximizing the minimal size of K_j and L_j for a given group.

Definition 4.5. If $G = \{e\}$ let $\delta(G) = 1$. Otherwise let S(c) be the maximal order of any subgroup of G not containing c and let

$$6(G) = \min_{c \in G-\{e\}} (S(c)).$$

Since we are only dealing with finite groups $\delta(G)$ is always well-defined and finite. Note that if $P(a,G) = 1$ then $\delta(a) = 1$ so that $\delta(G) = 1$. Note also that if G is nontrivial and $P(G) \neq 1$ then $\delta(G) > 1$ always. A simple lemma we will need in the sequel is:

Lemma 4.6. Let H and K be subgroups of a finite group G such that

$H \cap K = \{e\}$. Then $|H||K| \leq |G|$.

Proof.

Let $h_1, h_2 \in H$ and $k_1, k_2 \in K$ such that hlkl $= h_2k_2$. Then

$$h_1h_2^{-1} = k_2k_1^{-1} \in H \cap K$$

Hence $h_1 = h_2$ and $k_1 = k_2$. Thus

$$|\{hk : h \in H, k \in K\}| > |H||K|$$

But it is also a subset of G. ▮

The crucial property of 6(G) is

Lemma 4.7. For any finite group G, $\alpha(G)\delta(G) < |G|$.

Proof.

If 6(G) = 1 the lemma is true, so assume not. Pick $H < G$ and $e \neq a \in H$ with $P(a,H) = 1$ and $|H| = \alpha(G)$. Choose $K < G$ with $a \notin K$ and $|K| = \delta(a)$. Then, since $H \cap K$ is a subgroup of H not containing a, $H \cap K = \{e\}$. Hence, by lemma 4.4 and the fact that 6(G) $< \delta(a)$,

$$\alpha(G)\delta(G) < \alpha(G)\delta(a) = |H||K| \leq |G| \text{ ▮}$$

The universal lower bound for any (d,r) circuit to compute multiplication in a finite group G can now be stated.

Theorem 4.8 Let G be a finite group, $\phi : G \times G \to G$ be group multiplication, and C be a (d,r) circuit to compute ϕ for $d > 2$ and $r > 2$. Then, if C has computation time τ,

$$\tau \geq \left\lceil \log_r 2 \left\lceil \log_d \frac{|G|}{\delta(G)} \right\rceil \right\rceil$$

Proof.

Assume 6(G) > 1 and choose $a \in G$ such that 6(a) = 6(G). There must be an output line of C, say the j^{th}, such that $h_j(e) \neq h_j(a)$. But then both R_j and L_3 are subgroups of G which do not contain a. They hence have order at most 6(G). Thus, the result follows from Theorem 4.8. If 6(G) = 1 then either G = (e) or $|G| = \alpha(G)$. In the former case the theorem is true trivially. In the latter case choose $g \in G$ such that P(g,G) = 1 and pick an output line, say the i^{th}, such that $h_i(e) \neq h_i(g)$. Then $R_i = L_i = (e)$ and the result follows from Theorem 4.8.I

Lemma 4.7 implies that this lower bound is no weaker than Winograd's result given in Theorem 3.6; and, indeed, the following example shows that it is stronger.

Example 4.9.

Let p be an odd prime. Then there is a group with three generators a, b, and c and defining relations [Ref. 4, p. 52].

$$a^p = b^p = c^p = e; \quad ab = bac; \quad ca = ac; \quad cb = bc$$

which has no element of order p^2. It is easy to show that any subgroup of order p^2 must contain c. Thus

$$\delta(G) = \delta(c) = p$$

But, clearly, $\alpha(G) = p$. Thus

$$\alpha(G)\delta(G) < |G|$$

In one important case, however, the two bounds are the same.

Lemma 4.10. Let G be a finite abelian group. Then

$$\alpha(G)\delta(G) = |G|$$

Proof.

By the decomposition theorem for abelian groups [Ref. 4, p. 40]

☝ ▪ Z_1 x • ♏♏ ☒ Z_n

where each Z_i is a cyclic p-group, say $|Z_i| = Pi^{r_i}$; and, with no loss of generality

$$i < j \Rightarrow p_i^{r_i} \geq p_j^{r_j} \qquad (*)$$

If $n = 1$ the theorem is true since $P(G) = 1$ and $\delta(G) = 1$. Assume $n > 1$ and let a_i generate Z_i $(i = 1,2,\ldots,n)$. Now if we choose any $g \neq e$

$$g = \left(a_1^{k_1},\ldots,a_n^{k_n}\right)$$

where at least one exponent, say k_i is nonzero, then

$$g \notin \left(\prod_{\substack{j\neq i\\ j=1}}^{n} Z_j\right) x \{e_i\}$$

where e_i is the identity in Z_i. It follows that

$$\delta(g) \geq \prod_{\substack{j\neq i\\ j=1}}^{n} p_j^{r_j} \geq \prod_{j=2}^{n} P_j^{r_j} \text{ by } (*)$$

Thus

$$6(G) \geq \prod_{j=2}^{n} p_j^{r_j}$$

But any subgroup of order greater than $\prod_{j=2}^{n} p_j^{r_j}$ must intersect Z_1 nontrivially and thus must contain

$$\left(a_1^{p_1^{(r_1-1)}}, e_2, \dots, e_n\right) \notin \{e_1\} \times Z_2 \times \dots \times Z_n$$

Thus

$$\delta(G) \leq \delta\left(\left(a_1^{p_1^{(r_1-1)}}, e_2, \dots, e_n\right)\right)_3 = \prod_{j=2}^{n} p_j^{r_j} \qquad I$$

For the sake of completeness we give some examples of nonabelian groups G_i, each having $\alpha(G_i)\delta(G_i) = |G_i|$.

Example 4.11.

Let p be an odd prime. Let G_1 be the group generated by a and b having relations [Ref. 4, p. 52]

$$a^{p^2} = b^p = e; \quad b^{-1}ab = a^{1+p}$$

Then $\alpha(G_1) = p^2$ and any group of order p^2 must contain a^p.

Example 4.12.

Let G_2 be the direct product of two groups A and B such that

$$\alpha(A)\delta(A) = |A| \ ; \ \alpha(B)\delta(B) = |B|$$

Then it is easy to see that

$$\alpha(G_2) = \max\{\alpha(A),\alpha(B)\}; \ \delta(G_2) = \min\{|B|\ \delta(A),|A|\ \delta(B)\}$$

and thus

$$\alpha(G_2)\delta(G_2) = |G_2|$$

In particular, these properties hold if G_2 is nonabelian but all of its subgroups are normal [Ref. 4, p. 190].

5. A CIRCUIT FOR GROUP MULTIPLICATION

In this section we give a method to construct a (d,r) circuit to multiply in any finite group G which is valid for $d \geq 2$ and $r > 2$. The computation time of the circuit will be at most one unit greater than the lower bound just derived. If G is abelian and $r > 3$ our circuit can be compared to that of Winograd, It will be seen that our computation time underbounds his; and that, in fact, we can give a group for which the difference in computation time is arbitrarily large.

Lemma 5.1. Let K be any subgroup of G. Then there is a (d,r) circuit to compute $\phi : G \times G \to \{0,1\}$ in time[f]

$$\tau = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \frac{|G|}{|K|} \right\rceil \right\rceil \right\rceil$$

where

$$\phi(a,b) = 0 \text{ if } ab \in K$$

$$\phi(a,b) = 1 \text{ if } ab \notin K$$

Proof.

Let $M = |G|/|K|$. Pick a coset representative $v_i \in K\,v_i$ for each

[f] The original statement of this lemma had $\tau = 1 + \lceil \log_r \lceil \log_d (|G|/|K|) \rceil \rceil$. The refinement was pointed out to the author by Winograd.

right coset of K in G. Then $\{v_i^{-1}\}$ will be a set of left coset representatives, for $v_i^{-1}K = v_j^{-1}K$ iff $v_i v_j^{-1} \in K$. Pick a map z_1 from G to the space of $\lceil \log_d M \rceil$-ary vectors over Z_d such that

$$z_1(g_1) = z_1(g_2) \text{ iff } Kg_1 = Kg_2$$

and then define another map z_2 with same domain and range by

$$z_1(g) \oplus z_2(g^{-1}) = \bar{0}$$

where $\bar{0}$ is the all zero vector and $\oplus$ is componentwise addition modulo d. Note that z_2 maps any two elements in the same left coset to the same vector. The first level of the circuit consists of $\lceil (1/\lfloor r/2 \rfloor) \lceil \log_d M \rceil \rceil$ similar elements, If ab is being computed these modules each sum components of $z_1(a)$ and $z_2(b)$ mod d (the last adder will sum less than $\lfloor r/2 \rfloor$ if $\lfloor r/2 \rfloor$ does not divide $\lceil \log_d M \rceil$). An element has output 0 if all pairs of input components are congruent to 0 mod d. If not, its output is 1. Thus all outputs are 0 iff there is some j such that $a \in Kv_j$ and $b \in v_j^{-1}K$. The rest of the circuit is a fan-in of r input elements having output 0 iff all inputs are 0 and output 1 if at least one input is nonzero. This fan-in has depth $\lceil \log_r \lceil (1/\lfloor r/2 \rfloor) \lceil \log_d M \rceil \rceil \rceil$. Thus the circuit computes ϕ in time

$$\tau = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d M \right\rceil \right\rceil \right\rceil \quad \blacksquare$$

<u>Corollary 5.2</u>. There is a (d,r) circuit to tell if $ab \in Kv$ for any $v \in G$ with the same computation time.

<u>Definition 5.3</u>. A complete set of subgroups of a group G is a set $\{K_i\}$ of subgroups for which

$$\bigcap_i K_i = \{e\}$$

Lemma 5.4. If Ki is a complete set of subgroups of G then, for any $a \in G$, knowledge of the right cosets containing a is sufficient to determine a.

Proof.

$$\cap(K_i a) = (\cap K_i)a = a \blacksquare$$

Note that a complete set of subgroups will always exist for any G, e.g., the set consisting of $\{e\}$ alone. Unless $P(G) = 1$, there will be other complete sets as well.

Lemma 5.5. Let $\{K_i\}$ be a complete set of subgroups of G. Then there is a (d,r) circuit to multiply in G in time

$$\tau = 1 + \max_i \left\{ \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \left| \frac{G}{K_i} \right| \right\rceil \right\rceil \right\rceil \right\}$$

Proof.

Follows from lemma 5.1, corollary 5.2 and lemma 5.4. ▌

Now we are able to prove

Theorem 5.6. Let G be any finite group. Then for any $d \geq 2$ and any $r > 2$ there is a (d,r) circuit to multiply in a finite group G in time

$$\tau = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \frac{|G|}{\delta(G)} \right\rceil \right\rceil \right\rceil$$

Furthermore, the circuit has computation time exceeding the lower bound by at most one time unit.

Proof.

Assume $\delta(G) > 1$. For any $g \in G$ with $g \neq e$ there is a subgroup K_g of order $\delta(g)$ not containing g. Thus $\{K_g : g \in G - \{e\}\}$ is a complete set of subgroups with

$$\min\left\{|K_g| : g \in G - \{e\}\right\} = \delta(G)$$

If 6(G) = 1 then use the complete set consisting of $\{e\}$. The second statement of the theorem follows from the fact that

$$\left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d x \right\rceil \right\rceil \right\rceil \leq \left\lceil \log_r 2 \left\lceil \log_d x \right\rceil \right\rceil$$

for $r \geq 2$. ▮

Corollary 5.7. If G is abelian or if 6(G) = 1 there is a (d,r) circuit to multiply in G in time

$$\tau = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \alpha(G) \right\rceil \right\rceil \right\rceil$$

As we have noted, Winograd's circuit for an abelian group G requires time

$$\tau = 2 + \left\lceil \log_{\lfloor (r+1)/2 \rfloor} \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \alpha(G) \right\rceil \right\rceil \right\rceil$$

since

$$\left\lfloor \frac{r+1}{2} \right\rfloor < r \quad \text{for} \quad r > 3$$

it follows that, our computation time is at least one less than his.

Example 5.8.

Say $r = 4$ and $\lceil \log_d \alpha(G) \rceil = 2^{2^k}$ for some $k \geq 1$. Then Winograd's time is $2 + 2k$ and our time is $1 + k$, i.e., his circuit requires twice as long. The reader can easily construct a myriad of similar examples.

Winograd [Ref. 2] has extended his group results to numerial addition and multiplication by noting that a circuit which can multiply in the cyclic group of order $2N - 1$ can also add two numbers between 0 and N and that numerical multiplication can be done by adding the exponents in the prime decompositions of the two factors. Since we are able to lower the time necessary to multiply in cyclic groups, we can achieve a corresponding decrease in the time for numerical addition and multiplication as well. We present this result in the framework of Winograd's definition::. The reader interested in the details of the relationship between group multiplication and these other two operations is referred to Winograd's original paper.

Definition 5.9. For an integer m let $Q_m = \text{l.c.m.}\{1,2,\ldots,m\}$ and let $y(N) = \min\{m : Q_m \geq N\}$.

Then, paralleling Winograd's application of his group multiplication time, we employ corollary 5.7 and obtain

Theorem 5.10. Let $\phi : Z_N \times Z_N \to Z_{2N-1}$ be $\phi(a,b) = a + b$. Then there is a (d,r) circuit to compute ϕ in time

$$\tau_\phi = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \gamma(2N - 1) \right\rceil \right\rceil \right\rceil \qquad r \geq 2 \quad d \geq 2$$

and

Theorem 5.11. Let $\psi : \{1,2,\ldots,N\} \times \{1,2,\ldots,N\} \to \{1,2,\ldots,N^2\}$ be $\psi(a,b) = ab$. Then for any $r \geq 2$ and any $d \geq 2$ there is a (d,r) circuit to compute ψ in time

$$\tau_{\psi} = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \gamma(2 \lfloor \log_2 N \rfloor - 1) \right\rceil \right\rceil \right\rceil$$

In closing we note for reference that Winograd has lower bounded τ_{ϕ} and τ_{ψ} as follows

Theorem 5.12. (Winograd [Ref. 2]). For any $d \geq 2$ and any $r > 2$ then any (d,r) circuit to compute ϕ requires time τ_{ϕ} where

$$\tau_{\phi} \geq \left\lceil \log_r 2 \left\lceil \log_d \gamma\left(\left\lceil \frac{N}{2} \right\rceil \right) \right\rceil \right\rceil$$

and any (d,r) circuit which computes ψ requires time

$$\tau_{\psi} \geq \left\lceil \log_r 2 \left\lceil \log_d \gamma\left(\left\lceil \frac{\lfloor \log_2 2N \rfloor}{2} \right\rceil \right) \right\rceil \right\rceil$$

The proximity of the results of theorem 5.10 and theorem 5.11 to these lower bounds is indicated by the fact that

$$\gamma(4x) \leq 2 + \gamma(x)$$

REFERENCES

[1] Winograd, S., "On the Time Required to Perform Addition," J. ACM, vol. 12, no. 2, April 1965, pp. 277-285.

[2] Winograd, S., "On the Time Required to Perform Multiplication," J. ACM, vol. 14, no. 4, October 1967,pp. 793-802.

[3] Ofman, Yu, "On the Algorithmic Complexity of Discrete Functions," Dokl. Akad. Nauk SSR, vol. 145,no. 1, 1962,pp. 48-51.

[4] Hall, M., Jr., The Theory of Groups, The Macmillan Company, New York, 1959.

III. A GENERAL METHOD AND SOME: APPLICATIONS

1. INTRODUCTION

We shall present a method for computation of any finite function $f : X_1 \times X_2 \to Y$ by means of a (d,r) circuit which generalizes results of the last chapter in a natural way. The computation of a given function is reduced to that of a set of functions of a very simple class. We show that this class of functions is basic in the sense that to compute a finite function as quickly as possible we must be able to compute some functions of this class in the least possible time,

The method is applied to various classes of functions usually yielding a near optimal circuit. Previous results are shown to be special cases within our general framework.

2. THE GENERAL METHOD

Let $f : X_1 \times X_2 \to Y$. As stated, we shall give a method to compute f which is a natural generalization of previous results.

Let W be any subset of Y. We define a function $f_W : X_1 \times X_2 \to \{0,1\}$ as follows:

$$f_W(x_1,x_2) = \begin{cases} 1 \text{ if } f(x_1,x_2) \in W \\ 0 \text{ if } f(x_1,x_2) \notin W \end{cases}$$

i.e., f_W is the characteristic function of

$$f^{-1}(W) \subseteq X_1 \times X_2$$

We define four sets associated with W, X_1, and X_2. For $x_1 \in X_1$ and $x_2 \in X_2$ let

$$A_W(x_1) = \left\{x_2 \in X_2 : f(x_1,x_2) \in W\right\}$$

$$B_W(x_2) = \left\{x_1 \in X_1 : f(x_1,x_2) \in W\right\}$$

$$C_W(x_1) = \left\{x_1' \in X_1 : A_W(x_1) = A_W(x_1')\right\}$$

$$D_W(x_2) = \left\{x_2' \in X_2 : B_W(x_2) = B_W(x_2')\right\}$$

Lemma 2.1. For any $x_1 \in X_1$, $x_2 \in X_2$ either

$$f(C_W(x_1),D_W(x_2)) \subseteq W$$

or

$$f(C_W(x_1),D_W(x_2)) \cap W = \emptyset$$

Proof.

Assume 3 $x_1' \in C_W(x_1)$ and $x_2' \in D_W(x_2)$ with $f(x_1',x_2') \in W$. Let $x_1'' \in C_W(x_1)$ and $x_2'' \in D_W(x_2)$. Then

$$f(x_1',x_2') \in W \Rightarrow f(x_1'',x_2') \in W \Rightarrow f(x_1'', x_2'') \in W$$

so that

$$f(C_W(x_1),D_W(x_2)) \subseteq W \quad ∎$$

Now let

$$\mathcal{C}_W = \left\{C_{W_1},\ldots,C_{W_m}\right\}$$

and

$$\mathcal{D}_W = \left\{D_{W_1},\ldots,D_{W_n}\right\}$$

be the collection of distinct members of $\{C_W(x_1) : xl \in Xl)$ and the collection of distinct members of $\{D_W(x_2) : x_2 \in X_2\}$ respectively. In general, given $x_1 \in X_1$ there will be more than one $D_{W_v} \in \mathfrak{D}_W$ for which

$$f(x_1, D_{W_v}) \subseteq W$$

and a similar statement holds regarding a given member of X_2. Thus define

$$M_W = \max_{x_2 \in X_2} \left| \left\{ C_{W_u} \in \mathcal{C}_W : f(C_{W_u}, x_2) \subseteq W \right\} \right|$$

and

$$N_W = \max_{x_1 \in X_1} \| D_{W_v} \in \mathfrak{D}_W : f(x_1, D_{W_v}) \subseteq W \} |$$

Then we can construct a circuit to compute f_W.

Lemma 2.2. Let $f : X_1 \times X_2 \to Y$ and let $W \subseteq Y$. Let f_W, B_W, $\mathcal{C}_W$, $\mathfrak{D}_W$, M_W, and N_W be as given above. Then, for any $d \geq 2$ and any $r > 2$, there is a (d,r) circuit to compute f_W in time τ_W, where

$$\tau_W = 1 + \min\left\{ \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d |\mathcal{C}_W| + 1 \right\rceil \right\rceil \right\rceil + \left\lceil \log_r M_W \right\rceil, \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d |\mathfrak{D}_W| + 1 \right\rceil \right\rceil \right\rceil + \left\lceil \log_r N_W \right\rceil \right\}$$

(If, for each $x_1 \in Xl$, 3 $x_2 \in X_2$ for which $f(x_1,x_2) \in W$ then $|\mathfrak{D}_W| + 1$ becomes $|\mathfrak{D}_W|$ in the above theorem statement. Similarly, if, for each $x_2' \in X_2$, 3 $x_1' \in X_1$ for which $f(x_1',x_2') \in W$, then $|\mathcal{C}_W| + 1$ becomes $|\mathcal{C}\ |$. This will be clear in the proof.)

The idea of the proof is simple: Given $x_1 \in X_1$

$$f_W(x_1,x_2) = 1 \Longleftrightarrow x_2 \in A_W(x_1)$$

$$\Longleftrightarrow x_1 \text{ is in a } C_W \text{ for which } x_2 \text{ is in the } A_W$$

Thus we shall code x_2 as the list of C_W's for which x_2 is in the A_W, code x_1 by its C_W, and check if x_1's C_W is in the List of C_W's for x_2 and thus shall then know whether or not $f_W(x_1,x_2) = 1$. Explicitly,

Proof.

We shall give a (d,r) circuit to compute f_W in time

$$1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d T \right\rceil \right\rceil \right\rceil + \left\lceil \log_r M_W \right\rceil$$

where $T = |\mathcal{C}_W| + 1$ if $3\ x_2 \in X_2$ with $B_W(x_2) = \emptyset$, and $T = |\mathcal{C}_W|$ if not. The lemma will follow from this by symmetry. If there is an $x_2 \in X_2$ for which $B_W(x_2) = \emptyset$ then let $L = \lceil \log_d |\mathcal{C}_W| + 1 \rceil$ and Let $z : X_1 \to Z_d^L - \{\bar{0}\}$, where $\bar{0}$ is the vector of all 0's, be any map satisfying

$$z(x_1) = z(x_1') \text{ iff } C_W(x_1) = C_W(x_1')$$

If there is no $x_2 \in X_2$ with $B_W(x_2) = \emptyset$ then let $L = \lceil \log_d |\mathcal{C}_W| \rceil$ and let $z : X_1 \to Z_d^L$ again be a map obeying the above condition.

Let $z_1 : X_1 \to Z_d^{LM_W}$ be given by

$$z_1(x_1) = \underbrace{z(x_1)z(x_1)\ldots z(x_1)}_{M_W \text{ times}}$$

where, if $a = (a_1,\ldots,a,)$ and $b = (b_1,\ldots,b_t)$ then the notation

$$ab = (a_1, \ldots, a_s, b_1, \ldots, b_t)$$

is used here.

We now define $z_2 : X_2 \rightarrow Z_d^{LM_W}$. If $B_W(x_2) = \emptyset$ then $z_2(x_2) = \bar{0}$. If not, let $C_1, C_2, \ldots, C_{m(x_2)}$ be the elements of $\mathcal{C}_W$ satisfying

$$f(C_j, x_2) \subsetneq W \quad i < j < m(x_2) < M_W$$

and note by lemma 2.1 that if C is any other element of $\mathcal{C}$, then

$$f(C, x_2) \cap w = \emptyset$$

Then define

$$z_2(x_2) = - z(x_{11})z(x_{12}). \; . \; .z(X_{1m(x_2)})z(x_{1m(x_2)}). \; . \; .z(X_{1m(x_2)})$$

where the equality is componentwise modulo d,

$$x_{1j}\ E\ C_j ; \quad 1 \leq j \leq m(x_2)$$

and $z(x_{1m(x_2)})$ appears $M_W - m(x_2) + 1$ times. To compute $f_W(x_1, x_2)$ suppose

$$z_1(x_1) = \left(a_1, a_2, \ldots, a_{LM_W}\right)$$

and

$$z_2(x_2) - \left(b_1, b_2, \ldots, b_{LM_W}\right)$$

Then $f_W(x_1, x_2) = 1$ iff there is an integer s, with $0 \leq s \leq M_W$, such that

$$a_j + b_j \equiv 0 \pmod{d}; \quad sL + 1 \leq j \leq (s + 1)L$$

A separate circuit will determine if this is so for each possible such value of s. The first stage of the circuit for a fixed s will contain $\lceil \lceil 1/\lfloor r/2 \rfloor \rceil L \rceil$ elements each testing $\lfloor r/2 \rfloor$ of the a_i's and $\lfloor r/2 \rfloor$ of the b_i's to see if their pairwise sums modulo d are all 0 (if $\lfloor r/2 \rfloor$ does not divide L the last element will test less than $\lfloor r/2 \rfloor$ pairs). If the sums are all 0 for a given element, its output will be 1. Otherwise its output is 0. A fan-in of elements having output 1 iff all inputs are 1 and each having at most r inputs comprises the rest of the circuit for the given s. This fan-in has depth $\lceil \log_r \lceil (1/\lfloor r/2 \rfloor) L \rceil \rceil$. These circuits yield M_W outputs at least one of which is 1 iff $f(x_1,x_2) \in W$. An additional fan-in of $\lceil \log_r M_W \rceil$ stages will determine, whether any of the values are 1 or not. Thus the time to compute f_W is as claimed. ▮

Definition 2.3. Let $\mathcal{W} = \{W_1, W_2, \ldots, W_n\}$ be a family of subsets of Y. Then $\mathcal{W}$ is said to be complete if

$$\left\{W_i : y_1 \in W_i\right\} = \left\{W_i : y_2 \in W_i\right\} \Longrightarrow y_1 = y_2$$

By repeated application of lemma 2.2 we obtain

Theorem 2.4. Let $f : X_1 \times X_2 \to Y$, $\mathcal{W}$ be a complete family of subsets of Y, and τ_{W_i} be as in lemma 2.2 for each $W_i \in \mathcal{W}$. Then there is a (d,r)circuit to compute f in time

$$\tau = \min_{W_i \in \mathcal{W}} \left\{\tau_{W_i}\right\}$$

Proof.

The theorem follows from the completeness of $\mathcal{W}$, since if we then compute f_{W_i} for each $W_i \in \mathcal{W}$ in the manner of lemma 2.2 we have effectively computed f in the time given above. ▮

We now have a general method to compute any finite function $f: X_1 \times X_2 \to Y$ given a complete set of subsets of Y. One can always--in principal at least--find the complete set of subsets yielding minimum computation time under our scheme. It is felt that the set of functions $\{f_W : W \subseteq Y\}$ related to an $f : X_1 \times X_2 \to Y$ is a basic class, since if C is a (d,r) circuit having minimum computation time over the class of (d,r) circuits computing f, then there is another (d,r) circuit C' computing f in the same time each of whose output lines is the value of f_W for some $W \subseteq Y$. This is simply because if d = 2 each output line already computes f_W for some $W \subseteq Y$; whereas otherwise we can replace the element whose output is the j^{th} output line by $\leq d$ elements--the i^{th} having output 1 iff the original element had output i - 1 and otherwise having output 0.

Thus we have made explicit that to compute a finite function as quickly as possible we must be able to compute some functions of this special class in the least possible time. The problem, of course, is to determine which functions of the special class to choose, reminding us [Ref. 1, Example 2.8] of the vital role played by coding in affecting the computation time.

Example 2.5.

Let G be a group and let $f : G \times G \to G$ be group multiplication. Then a complete set of subsets of G is e.g., the right cosets of subgroups $K_1, K_2, \ldots, K_n$ where

$$\bigcap_{i=1}^{n} K_i = \{e\}$$

For a given coset $W = K_r g$ and a given $h \in G$

$$A_W(h) = \left\{x \in G : hx \in K_r g\right\} = h^{-1}K_r g$$

$$B_W(h) = \left\{x \in G : xh \in K_r g\right\} = K_r g h^{-1}$$

$$C_W(h) = \left\{x \in G : A_W(x) = A_W(h)\right\} = h^{-1}K_r$$

$$D_W(h) = \left\{x \in G : B_W(x) = B_W(h)\right\} = hg^{-1}K_r g$$

Note that $N_W = M_W = 1$. Thus there is a (d,r) circuit to compute f_W in time

$$\tau = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \frac{|G|}{|K_r|} \right\rceil \right\rceil \right\rceil$$

since $\mathcal{C}_W = |G|/|K_r|$. The condition for decreasing the argument in the inner brackets from $|\mathcal{C}_W| + 1$ to $\mathcal{C}_W$ is satisfied.

3. SEMIGROUP MULTIPLICATION

In this section we specialize the general method to computation of multiplication in any finite semigroup. Use will be made of the concept of one-sided congruences--the natural generalization of a subgroup K of a group G. The reader not acquainted with basic semigroup notions is referred to the excellent book of Clifford and Preston [Ref. 2], or to Arbib [Ref. 3].

Let ρ be an equivalence relation on a set S. If $a \in S$ and $b \in S$ are equivalent we shall write $a\rho b$, and shall write [a] for the block $\{b : a\rho b\}$ of S containing a

Definition 3.1. Let s be a semigroup with ρ an equivalence relation on s. Then ρ is a right (left) congruence on S if aspbs (sapsb) for all $s \in S$ whenever $a\rho b$.

In what follows we comply with the notation introduced in the preceding sections where we now replace $f : X_1 \times X_2 \to Y$ by the semigroup multiplication $S \times S \to S$.

Lemma 3.2. Let ρ be a right congruence on a finite semigroup S. Then, if $u\rho v$,

$$C_{W_j}(u) = C_{W_j}(v)$$

for all blocks W_j of ρ. In addition for any $t \in S$ $B_{W_j}(t)$ is a union of blocks of ρ.

Proof.

If $u\rho v$ then $us\rho vs$, so that for any block W_j of ρ,

$$A_{W_j}(u) = \left\{s \in S : us \in W_j\right\} = \left\{s \in S : vs \in W_j\right\} = A_{W_j}(v),$$

and hence $C_{W_j}(u) = C_{W_j}(v)$.

For any $t \in S$ and any block W_k of ρ, either

$$W_k t \subseteq W_j$$

or

$$W_k t \cap W_j = \emptyset$$

Hence $B_{W_j}(t) = \{s : st \in W_j\}$ is union of blocks of ρ. ∎

Corollary 3.3. Let ρ be a right congruence, and let W_j be a congruence block of ρ. Let $f_{W_j} : S \times S \to \{0,1\}$ be

$$f_{W_3}(a,b) = 1 \text{ if } ab \in W_3$$
$$= 0 \text{ if } ab \notin W_3$$

Let

$$M_j = \max_{s \in S} \left| \left\{ W_k : W_k \text{ is a block of } \rho \text{ and } W_k s \in W_j \right\} \right|$$

and let $|\rho|$ be the number of distinct blocks of ρ. Then there is a (d,r) circuit to compute f_{W_j} in time τ_j, where

$$\tau_j = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d |\rho| + 1 \right\rceil \right\rceil \right\rceil + \left\lceil \log_r M_j \right\rceil$$

In case there is, for any $s \in S$, an element $t \in S$ with $ts \in W_j$ then the value in the inner brackets is $|\rho|$ instead of $|\rho| + 1$. Hence there is a (d,r) circuit to tell which block of ρ contains the product of two elements of S in time

$$7_\rho = \max\left\{\tau_j : W_j \text{ is a congruence class of } \rho\right\}$$

Corollary 3.4. Let S be any finite semigroup. Let

$$M = \max_{y \in S}\left\{\max_{t \in S}\left\{\left|\{s \in S : st = y\}\right|\right\}\right\}$$
$$N = \max_{y \in S}\left\{\max_{t \in S}\left\{\left|\{s \in S : ts = y)\right|\right\}\right\}$$

Then there is a (d,r) circuit to multiply in S in time

$$\tau = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor}\left(\log_d |S| + 1\right)\right\rceil\right\rceil + \min\left\lceil \log_r \{M,N\}\right\rceil$$

Proof.

Let ρ be the right (and left) congruence each class of which is a single element of S. ▮

Of course this is not, in general, the congruence to choose to minimize computation time in S, e.g., if S is a group [Ref. 1, Thm. 5.61.

Corollary 3.5. Let $\rho_1,\rho_2,\ldots,\rho_n$ be a set of right and left congruences on a finite semigroup S which have the property that

$$a\rho_i b;\ 1 < i < n \text{ iff } a = b$$

Then there is a (d,r) circuit to multiply in S in time τ where

$$\tau = \max\left\{\tau_{\rho_i}\right\}$$

and each τ_{ρ_i} is as defined in corollary 3.3.

Proof.

The set of blocks of the congruences form a complete collection of subsets of S. ▮

It is not, in general, obvious which congruences to choose to minimize computation time. Corollary 3.4 gives an upper bound in computation time necessary for a given semigroup S. One would usually expect to be able to do better.

The rest of this section gives methods for some important special classes of finite semigroups.

Definition 3.6. A semigroup S is cyclic if there is an element a such that

$$S = \left\{a^k : k > 1;\ k \text{ an integer}\right\} = \langle a\rangle$$

If S is finite there are minimal integers m and n [Ref. 2, p. 19] such that

$$a^{k+n} = a^k \quad k > m$$

We call m the index of S and n is called the period of S.

Definition 3.7. Let a(n) be the largest prime power factor of n; let $Q_m = \mathrm{lcm}\{1,2,\ldots,m\}$; and let $\gamma(N) = \min\{m : Q_m \geq N\}$. Then we have

Lemma 3.8. Let S be a cyclic semigroup of index m and period n, i.e., S = (a). Let

$$\tau_1 = \left\lceil \log_r 2 \left\lceil \log_d \alpha(n) \right\rceil \right\rceil$$

$$\tau_2 = \left\lceil \log_r 2 \left\lceil \log_d \gamma\left(\left\lfloor \frac{\lfloor (m+n-3)/2 \rfloor}{2} \right\rfloor \right) \right\rceil \right\rceil$$

$$\tau_3 = \left\lceil \log_r 2 \left\lceil \log_d \left(\left\lfloor \frac{(m-2)}{n} \right\rfloor + 1 \right) \right\rceil \right\rceil$$

Then if C is a (d,r) circuit which computes $\phi : S \times S \to S$, where ϕ is multiplication in S, in time τ,

$$\tau \geq \max\left\{\tau_1, \tau_2, \tau_3\right\}$$

Proof.

That $\tau \geq \max\{\tau_1, \tau_2\}$ follows from previous results of [Ref. 1] and the fact that C must be able to multiply in the cyclic group of order n and to add two numbers between 0 and $\lfloor (m + n - 3)/2 \rfloor/2$. This is so because both these operations constitute the function ϕ restricted to a subset of S. To show $\tau \geq \tau_3$, pick an output line, say the j^{th}, for which

$$h_j\left(a^{m-1}\right) \neq h_j\left(a^{m+n-1}\right)$$

Then the set $\{a, a^{n+1}, \ldots, a^{\lfloor (m-2)/n \rfloor n+1}\}$ is h_j-separable in 'both arguments of ϕ [see Defn. 2.3, Ref. 1] and we can apply lemma 2.5 of [Ref. 1]. ▮

By proper choice of congruences on S a circuit to multiply in S which is very often near the lower bound can be constructed.

Lemma 3.9. Let $S = (a)$ be the cyclic semigroup generated by a which has index m and period n. Then, for any $d \geq 2$ and any $r > 2$, there is a (d,r) circuit to multiply in S in time $\hat{\tau}$, where

$$\hat{\tau} = \max\{\tau_1, \tau_2\}$$

$$\tau_1 = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \alpha(n) \right\rceil \right\rceil \right\rceil$$

$$\tau_2 = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d m \right\rceil \right\rceil \right\rceil$$

Proof.

Let $p_1, p_2, \ldots, pk$ be the distinct primes dividing n, so that

$$n = p_1^{s_1} p_2^{s_2} \cdots p_k^{s_k}$$

Define congruences $\rho_1, \ldots, \rho_k$ on S such that

$$a^u \rho_i a^v \text{ iff } u \equiv v \pmod{p_i^{s_i}}$$

and a congruence λ on S such that

$$a^u \lambda a^v \text{ iff } u = v \text{ or } u > m - 1 \text{ and } v > m - 1$$

Then the blocks of these congruences form a complete set of subsets of S. The congruence class of the product of two factors being multiplied can be computed in time

$$\tau_i = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d p_i^{r_i} \right\rceil \right\rceil \right\rceil$$

for congruences $\rho_1,\ldots,\rho_k$ and in time

$$\tau_\lambda = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d m \right\rceil \right\rceil \right\rceil$$

for the congruence λ. This and the lemma follow from corollary 3.3 and corollary 3.5.▮

This circuit is close to the lower bound time if m is not too much greater than

$$\max\left\{\alpha(n), r\left\lfloor\frac{\lfloor(m+n-3)/2\rfloor}{2}\right\rfloor, \left\lfloor\frac{m-2}{n}\right\rfloor + 1\right\}$$

The reader can convince himself that this condition is satisfied quite often.

We shall now treat another very important class of semigroups.

Definition 3.10. Let S be a finite semigroup. An ideal A is a subset of S such that SAS $\subseteq$ A. A left (right) ideal is a subset A such that SA $\subseteq$ A (AS $\subseteq$ A). A semigroup is called simple if it contains no proper ideals and is left (right) simple if it contains no proper ideals.

Use will be made of Rees's elegant structure theorem [Ref. 2, p. 90] for simple semigroups.

Definition 3.11. Let G be a group and let I and Λ be arbitrary sets. Let P be a matrix of elements of G with $|\Lambda|$ rows and $|I|$ columns. Then $\mathcal{M}(G;I,\Lambda;P)$ is the set of elements of the form

$$(a)_{i\lambda} : a \in G, \quad i \in I, \quad \lambda \in \Lambda$$

With multiplication given by

$$(a)_{i\lambda} \cdot (b)_{j\mu} = (ap_{\lambda j}b)_{i\mu}$$

it is said to form a Rees matrix semigroup. P is called the sandwich matrix.

Theorem 3.12. Let S be a simple semigroup. Then S is isomorphic to the Rees matrix semigroup $\mathcal{M}(G;I,\Lambda;P)$ for some group G sets I and A, and sandwich matrix P.

Proof.

See [Ref. 2, pp. 91-99]. ∎

Left and right ideals of $\mathcal{M}(G;I,\Lambda;P)$ are easily characterized.

Lemma 3.13. Let $s = \mathcal{M}(G;I,\Lambda;P)$ be a Rees matrix semigroup. Then any left ideal of S is the form

$$L_\lambda = \left\{(a)_{i\lambda} : i \in I, \ a \in G\right\}$$

and any right ideal is of the form

$$R_i = \left\{(a)_{i\lambda} : \lambda \in \Lambda, \ g \in G\right\}$$

Proof.

Pick any $s \in L_\lambda$ and any $t \in S$. Clearly $ts \in L_\lambda$. Conversely, given any $(a)_{i\lambda},(b)_{j\lambda} \in L_\lambda$ then $(ba^{-1}(p_{\sigma i})^{-1})_{j\sigma}(a)_{i\lambda} = (b)_{j\lambda}$ for any $\sigma \in$ A, i.e.,

$$Sa = L_\lambda$$

so there are no smaller left ideals. Right ideals are dually characterized. ▌

Definition 3.15. Let $K_1, \ldots. K_n$ be subgroups of a group G. Then they are said to be a complete set of subgroups if

$$\bigcap_{i=1}^{n} K_i = \{e\}$$

Definition 3.16. Let G be a finite group. Then, if G = {e) or G is a cyclic p-group let 6(G) = 1. Otherwise for $a \in G - \{e\}$. Let

$$\delta(a) = \max\{|K| : K < G \text{ and } a \notin K)$$

$$\delta(G) = \min_{a \in G-\{e\}} \{\delta(a)\}$$

It is clear from these definitions that any finite group G possesses a complete set of subgroups whose minimal order is $\delta(G)$.

Lemma 3.17. Let $S = \mathcal{M}(G;I,\Lambda;P)$ be a finite simple semigroup. Then, for any $r \geq 2$ and any $d \geq 2$ there is a (d,r) circuit to compute $\phi : S \times S \to S$, where ϕ is semigroup multiplication, in time equal or less than τ, where

$$\tau = 1 + \Bigg\lceil \log_r \Bigg\lceil \frac{1}{\lfloor r/2 \rfloor} \Bigg\lceil \log_d \Bigg(\min\{|I|,|\Lambda|\} \frac{|G|}{\delta(G)} \Bigg\rceil \Bigg) \Bigg\rceil \Bigg\rceil$$
$$+ \Big\lceil \log_r (\min\{|I|,|\Lambda|\}) \Big\rceil$$

Proof.

We shall show that there is a (d,r) circuit to multiply in S in time

$$\tau_{\Lambda} = 1 + \lceil \log_r \lceil \frac{1}{\lfloor r/2 \rfloor} \lceil \log_d |\Lambda| \frac{|G|}{\delta(G)} \rceil \rceil \rceil + \lceil \log_r |\Lambda| \rceil \rceil$$

and the result will follow by symmetry. For an element $(a)_{i\lambda} \in S$ let

$$G\left((a)_{i\lambda}\right) = a \text{ be the G-index of } (a)_{i\lambda}$$

$$I\left((a)_{i\lambda}\right) = i \text{ be the I-index of } (a)_{i\lambda}$$

$$\Lambda\left((a)_{i\lambda}\right) = \lambda \text{ be the A-index of } (a)_{i\lambda}$$

Multiplication of left factor $(a)_{i\lambda}$ and right factor $(b)_{j\mu}$ yields an I-index i and a A-index μ which can be obtained in time 0 by feeding them straight through to the output. This amounts to considering trivial congruences, say λ and ρ, where two elements are X-equivalent iff they have the same A-index and are p-equivalent iff they have the same I-index. Thus it only remains to show how to compute the G-index of the product of two elements in S.

Let K be any subgroup of G and, for any $g \in G$, let

$$\overline{Kg} = \{s \in S : G(s) \in Kg\}$$

Following the terminology introduced in Section 2 we shall show that f_{Kg} can be computed in time at most

$$\tau_{\Lambda,K} = 1 + \lceil \log_r \lceil \frac{1}{\lfloor r/2 \rfloor} \lceil \log_d |\Lambda| \frac{|G|}{|K|} \rceil \rceil \rceil + \lceil \log_r |\Lambda| \rceil$$

The result will then follow from the fact that the same can be done for each coset of a complete set of subgroups $K_1, \ldots, K_n$ having order at least 6(G). Simple computation yields the fact that

$$A_{\overline{Kg}}\left((a)_{i\lambda}\right) = \left\{(b)_{j\mu} \in S : b \in p_{\lambda j}^{-1} a^{-1} Kg\right\}$$

Thus

$$C_{\overline{Kg}}\left((a)_{i\lambda}\right) = \left\{(c)_{s\sigma} \in S : p_{\lambda j}^{-1}a^{-1}K = p_{\sigma j}^{-1}c^{-1}K \;\forall\, j \in I\right\}$$

Hence there is at most one element of $\mathcal{C}_{\overline{Kg}}$ for each distinct pair of elements of

$$\Lambda \times \{Kg : g \in G\}$$

i.e.,

$$\left|\mathcal{C}_{\overline{Kg}}\right| \leq |\Lambda| \; \frac{|G|}{|K|}$$

Also, for any $(b)_{j\mu} \in S$ and a fixed value of $\lambda \in A$

$$\left\{(a)_{i\lambda} : G\left((a)_{i\lambda}(b)_{j\mu}\right) \in Kg\right\} = \left\{(a)_{i\lambda} : a \in Kgb^{-1}p_{\lambda j}^{-1}\right\}$$

This set is a subset (not necessarily proper) of some element of $\mathcal{C}_{\overline{Kg}}$.

Thus

$$M_{\overline{Kg}} = \max_{(b)_{j\mu} \in S} \left|\left\{C_{\overline{Kg}} \in \mathcal{C}_{\overline{Kg}} : \phi\left(C_{\overline{Kg}}, (b)j\mu\right) \leq \overline{Kg}\right\}\right| \leq |\Lambda|$$

The inequalities plus lemma 2.2 yield the result.[†] ▮

<u>Corollary 3.18</u>. If $S = \mathcal{M}(G;I,\Lambda;P)$ is left-simple, right-simple or is a group, there is a (d,r) circuit to multiply in S in time

$$\tau = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \left(\log_d \frac{|G|}{\delta(G)}\right) \right\rceil \right\rceil \right\rceil$$

[†] This is an instance where our method is not optimal. It can be shown that there is a (d,r) circuit to multiply in S in time $\tau = 2 + \left\lceil \log_r \left\lceil (1/\lfloor r/2 \rfloor)\log_d |G|/\delta(G) \right\rceil \right\rceil + \left\lceil \log_r \min\{|I|, |\Lambda|\} \right\rceil$. However, the difference is usually small.

for $r > 2$ and $d > 2$.

Proof.

By lemma 3.13 S is left-simple if $|\Lambda| = 1$, right-simple if $|I| = 1$ andagroupif $|I| = |\Lambda| = 1$. ▮

Note that by [Ref. 1, Thm. 4.8] the lower bound below is valid.

Lemma 3.19. Let $S = \mathcal{M}(G;I,\Lambda;P)$ and let C be a (d,r) circuit to multiply in S in time τ. Then

$$\tau \geq \left\lceil \log_r 2 \left\lceil \log_d \frac{|G|}{\delta(G)} \right\rceil \right\rceil$$

Proof.

C must be capable of multiplying in G by a trivial recoding of inputs. ▮

We close with mention of a special case of simple semigroup multiplication.

Example 3.20.

Let $S = \mathcal{M}(G;I,\Lambda;P)$ be simple and let P have the property that

$$p_{\lambda j} = q_\lambda r_j \quad \forall\, i \in I,\ \lambda \in \Lambda$$

Then one can multiply in G in time

$$\tau = 1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \frac{|G|}{\delta(G)} \right\rceil \right\rceil \right\rceil$$

using the group multiplication circuit of [Thm. 3.6, Ref. 1] and recoding a left factor $(a)_{i\lambda}$ as aq_λ and recoding a right factor $(b)_{j\mu}$ as $r_j b$. Note that the three cases of lemma 3.17 are special cases of this example.

REFERENCES

[1] Spira, P. M., "The Time Required for Group Multiplication," Ph.D. Thesis, Chapter II, Department of Electrical Engineering, Stanford University, Stanford, California, 1968.

[2] Clifford, A. H. and G. B Preston, The Algebraic Theory of Semigroups, vol. 1, American Mathematical Society, Providence, Phode Island, 1961.

[3] Arbib, M. A., Algebraic Theory of Machines, Languages, and Semigroups, Academic Press, New York, 1968.

IV. PRELIMINARY RELATED RESULTS AND SUGGESTIONS FOR FUTURE WORK

1. INTRODUCTION

In the preceding chapters we have given results concerned with computation time of finite functions. This is only one complexity criterion of interest. Another would be the necessary fan-out of a (d,r) circuit rapidly computing a finite function. Though we claim no significant results here, we do have several preliminary results which we present in the next section. The final section gives conclusions and some suggestions for further work.

2. THE NUMBER OF OUTPUT LINES IN A CIRCUIT COMPUTING GROUP MULTIPLICATION

Given a finite group G having a complete set of subgroups $\{K_i\}$, the methods in Chapter II can be applied to yield a circuit C to multiply in G having computation time

$$1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d \max_i \left\{ \frac{|G|}{|K_i|} \right\} \right\rceil \right\rceil \right\rceil$$

It is of some interest to ask which choices of a complete set of subgroups will minimize the number of output lines of the corresponding circuit. Note that, for any subgroup $K < G$ it is only necessary to answer the coset membership question for $|G|/|K| - 1$ of the cosets--the membership question for the remaining coset then being automatically answered. Thus, for a complete set $\mathcal{K} = \{K_i\}$ of subgroups of G a circuit with

$$s_G(\mathcal{K}) = \sum_{K_i \in \mathcal{K}} \left(\frac{|G|}{|K_i|} - 1 \right)$$

output lines can be constructed which performs multiplication in G.

<u>Definition 2.1</u>. Let G be a finite group and let $\mathcal{K}$ be a complete set

of subgroups of G. Then the quantity

$$\sum_{K_i \in \mathcal{K}} \left(\frac{|G|}{|K_i|} - 1 \right)$$

is called the index of K in G.

Definition 2.2. Let K be a subgroup of a group G. Then K is called intersection generable if there are subgroups A and B which properly contain K for which

$$K = A \cap B$$

Not all subgroups are intersection generable, e.g., for a prime p, any subgroup of Z_{p^3} of order p is not. Similarly a maximal subgroup of any group is not intersection generable. A surprising fact is

Lemma 2.3. Let G be a finite group and let K be a complete set of subgroups of G. Then, if $\mathcal{K}$ is a complete set of minimal index in G, there is no $K \in K$ which is intersection generable.

Proof.

Assume there is some $K \in K$ with $A > K$ and $B > K$ for which $A \cap B = K$. Then, since A and B each contain at least two cosets of K,

$$|A| \geq 2|K| \; ; \quad |B| \geq 2|K|$$

so that

$$\left(\frac{|G|}{|A|} - 1 \right) + \left(\frac{|G|}{|B|} - 1 \right) < \left(\frac{|G|}{|K|} - 1 \right)$$

Clearly $(K - \{K\}) \cup \{A,B\}$ is also a complete set of subgroups of G; and, by the above inequality, it has index less than the index of $\mathcal{K}$.∎

This lemma should provide a tool for the determination of a minimal index complete set of subgroups of a given finite group G. It does, in fact, if G is abelian.

<u>Lemma 2.4</u>. Let G be abelian. Then a subgroup $K < G$ is intersection generable if it has order less than 8(G).

<u>Proof</u>.

Let $|K| < 6(G)$. Since K is abelian it can be written as a direct product of prime power cyclic groups

$$K = W_1 \times \dots \times W_t$$

NOW (see lemma 4.10, Chapter II) since $|K| < \delta(G)$ there must be at least two of these groups contained in larger cyclic p-subgroups of G, say $W_r < Z_r$ and $W_s < Z_s$. Then

$$K = (W_1 \times \dots \times Z_r \times \dots \times W_t) \cap (W_1 \times \dots \times Z_s \times \dots \times W_t)$$

so K is intersection generable. ▌

<u>Lemma 2.5</u>. Let G be abelian and express it as a product of cyclic p-groups

$$G = Z_1 \times \dots \times Z_n$$

Let $\{K_i\}$ be a complete set of subgroups of G. Then, if $1 \leq i \leq r$, there is some $K_i \in K$ such that

$$K_i \cap (\{e_1\} \times \{e_2\} \times \dots \times Z_i \times \dots \times \{e_n\}) = \{e\}$$

where e is the identity in G and e_j is the identity in Z_j $(1 \leq j \leq n)$.

Proof.

Let $V_i = \{e_1\} \times \{e_2\} \times \dots \times Z_i \times \dots \times$ (en). Then, since Vi is a cyclic group, say $|V_1| = p_i^{r_i}$, it has only one group of order p_i^m for $1 < m \leq r_i$ [Ref. 1, p. 55]. Thus, if the lemma is false, the subgroup of K_i having order p_i is contained in every $K \in \mathcal{K}$, contradicting the completeness of $\mathcal{K}$. ∎

Noting the fact that $\mathcal{K}$ is complete if, for all $1 < i < n$, it contains one subgroup K_i such that $K_i \cap V_i = \{e\}$, and using lemmas 2.4 and 2.5 we have

Theorem 2.6. Let G be abelian with

$$G = Z_1 \times \dots \times Z_n$$

where each Z_i is a cyclic p-group. Then a complete set of subgroups having minimal index in G is

$$\mathcal{K} = \{\{e_1\} \times Z_2 \times \dots \times Z_r, Z_1 \times \{e_2\} \times \dots \times Z_n, \dots, Z_1 \times \dots \times Z_{n-1} \times \{e_n\}\}$$

Hence, for any abelian group G, a (d,r) circuit constructed according to lemma 5.5 of Chapter II with the minimum number of output lines also has minimum computation time for this class of network.

We thus have the, perhaps not very surprising, fact that the best such circuit--both from the standpoint of computation time and from the standpoint of number of output lines--is one which multiplies in parallel in each maximal cyclic p-group Z_i.

It is unfortunate that, to date, we have been unable to determine a method to construct minimum index complete sets of subgroups for arbitrary finite groups. Given a complete set, however, one can always look for intersection generable subgroups within it.

We close with a special case in which a method of circuit construction different from that of Chapter II can decrease the number of output lines.

Lemma 2.7. Let G be a finite group having a cyclic subgroup H of order d. Then there are maps z_1 and z_2 from G into $Z_{d'}^N$ where $N = \lceil \log_d G \rceil$, such that if $a \in G$ is fixed and $x \in Ha$ then

$$z_2(g^{-1}x) = z_1(x) - z_1(g) \text{ for all } g \in G$$

where equality is componentwise modulo d.

Proof.

Let $\{c_1, \ldots, c_M\}$ be a set of right coset representatives of H in G, $M = |G| / I H|$. Choose elements $v_1, \ldots, v_M$ of Z_d^N such that no two differ by $\bar{k}$ for any $k \in Z_d$, where $\bar{k}$ is the vector of all k's. Let b be any generator of H. Then any $g \in G$ is uniquely representable as

$$g = b^k c_j a$$

Define

$$z_1(b^k c_j a) = \bar{k} + v_j$$

To complete the proof, it is only necessary to show that z_2 can be consistently defined as required in the lemma statement; i.e., that if

$$g_1^{-1} x_1 = g_2^{-1} x_2 \quad x_1, x_2 \in Ha$$

then

$$z_1(x_1) - z_1(g_1) = z_1(x_2) - z_1(g_2)$$

But

$$g_1g_2^{-1} = x_1x_2^{-1} = b^s; \text{ some } s \in Z_d$$

Hence

$$z_1(g_1) - z_1(g_2) = z_1(x_1) - z_1(x_2) = \bar{s} \quad \blacksquare$$

It is this lemma which implies the existence of a $(d+1,r)$ circuit to multiply in G with computation time possibly somewhat higher than the least attainable, but sometimes having less output lines.

Theorem 2.8. Let G be a finite group having a cyclic group $H = (b)$ of order d. Then there is a $(d+1,r)$ circuit which computes multiplication in G in time

$$\tau = \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil \log_d |G| \right\rceil \right\rceil \right\rceil + 1$$

having $|G|/|H|$ output lines.

Proof.

Define $\psi_a : G \times G \to Z_{d+1}$ by

$$\psi_a(g_1,g_2) = \begin{cases} k \; ; & g_1g_2 = b^k a, \; k \in Z_d \\ d \; ; & g_1g_2 \notin Ha \end{cases}$$

It suffices to show computability of ψ_a in time τ, since a similar function can be computed for each left coset of H. Define z_1 and z_2 as in the above lemma where, with no loss of generality, we take $z_1(a) = \bar{0}$. Thence $z_1(b^k a) = \bar{k}$. To compute $\psi_a(g_1,g_2)$ the first stage of the circuit inspects $z_1(g_1)$ and $z_2(g_2)$ comparing them componentwise with $\lceil 1/\lfloor r/2 \rfloor \lceil \log_d |G| \rceil \rceil$ elements. An element has output k iff all pairwise sums are k modulo d. Otherwise the element has output d. All outputs are k iff $g_1 g_2 = b^k a$. This can be determined in time $\lceil \log_r \lceil 1/\lfloor r/2 \rfloor \lceil \log_d |G| \rceil\rceil\rceil$ by a fan-in of elements with at most r inputs. That the number of output lines of the circuit is as claimed is true since this is the number of cosets of H in G.▌

Example 2.9.

Let $G = Z_{p^n}$. Then there is a $(p^k + 1, r)$ circuit to compute multiplication in G, where k is any $1 \leq k \leq n$, in time

$$1 + \left\lceil \log_r \left\lceil \frac{1}{\lfloor r/2 \rfloor} \left\lceil n \log_{p^k+1} p \right\rceil\right\rceil\right\rceil$$

with p^{n-k} output lines. From previous methods we would obtain a $(p^k + 1, r)$ circuit with computation time the same but having $p^n - 1$ output lines.

3. CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

The main contributions of this thesis have been lower and upper bounds on the computation time of finite functions. The method of deriving lower bounds has been to find the largest possible separable sets for a specific function or class of functions and then to apply the basic lemma of Chapter II. Upper bounds have been derived by constructing (d,r) circuits to realize the computations of interest. An immediate consequence of a remark of Winograd's in [Ref. 2] is that, given finite sets X_1 and X_2, any function f with domain $X_1 \times X_2$ is computable by a (d,r) circuit in time $\lceil \log_d |X_1||X_2| \rceil \lceil \log_r d \rceil + \lceil \log_r \lceil \log_d |X_1||X_2| \rceil\rceil$ and that furthermore given any $\epsilon > 0$, there is an N such that, if $|X_1||X_2| > N$, then the percentage of functions on $X_1 \times X_2$ computable in less time is less than ϵ. As he further remarks his (and our) method of deriving lower bounds never yields a bound greater than $\lceil \log_r \lceil \log_d |X_1||X_2| \rceil\rceil$, which is much less than $\lceil \log_d |X_1| |X_2| \rceil \lceil \log_r d \rceil$ in general. Thus it follows that techniques used here cannot yield tight lower bounds except for a few functions.

One class of functions for which the lower bound given here is tight is group multiplication $f : G \times G \to G$ for a finite group G. One might ask what special property group multiplication posseses which most functions do not which causes it to be computable so much more quickly. A characteristic of such an $f : G \times G \to G$ shared by very few finite functions is that, given $a, y \in G$ there is one and only one $x \in G$ such that $f(a,x) = y$ and one and only one $x' \in G$ for which $f(x',a) = y$. Let us recall Chapter III and choose $g : X_1 \times X_2 \to Y$ to be any finite function. If we now select $\{\{y\} : y \in Y))$ as a complete family of subsets of Y we note that the computation time for g of the circuit of lemma 2.2 is the sum of $1 + \lceil \log_r \lceil 1/\lfloor r/2 \rfloor \lceil \log_d |X_i| \rceil \rceil \rceil$ (i = 1 or 2) and a term dependent logarithmically upon the maximum number of solutions for any $y \in Y$ either of $g(x_1, \cdot) = y$ or of $g(\cdot, x_2) = y$. Hence the two terms which determine the computation time have dependencies analogous to those at the two terms giving the maximum computation time for any finite function with domain $X_1 \times X_2$. The singly logarithmic dependent term vanishes for group multiplication but is dominant for almost all functions.

The above remarks indicate a way in which the general method of Chapter III gives a heirarchical classification of all functions with a given domain and range. At the bottom of the heirarchy are functions such as group multiplication, which are invertible. Nearer the top of the heirarchy of functions from $X_1 \times X_2 \to Y$ would be, e.g., an $f : X_1 \times X_2 \to Y$ in which for some $x_1 \in X_1$, $x_2 \in X_2$ and $y_1, y_2 \in Y$ both $f(x_1, \cdot) = y_1$ and

$f(.,x_2) = y_2$ have many solutions. Furthermore, a similar heirarchy is established by any choice of a complete family of subsets of Y. A little thought reveals that invertible functions would be at the bottom of this heirarchy also. There is clearly much more research to be done pursuing these questions further, since we do not as yet even know, e.g., what complete family of subsets of the range of a function allows its computation in minimal time except in special cases.

Some remarks regarding parallel computation in general are in order. It appears that our methods allow rapid computation of the expense of much breadth, i.e., many elements and output lines. Indeed this is often true. On the other hand it is sometimes possible to attain the lowest possible computation time while concidentally minimizing elements and output lines, e.g., there is a $(2,2)$ circuit to multiply in Z_2^n with n output lines and computation time $\tau = 1$. In any event it should be investigated for what functions it is absolutely necessary to increase breadth in order to decrease computation time and similarly it would be of interest to know for what functions one must tolerate large computation time to achieve minimal breadth.

Finally, we note that no restrictions have been made as to the input codes we have used and that the only restriction upon the output code has been that it be 1 - 1. It would be useful to further investigate properties of (d,r) circuits with specific input and output codes, e.g., we might want to multiply in a finite group G and have both input and output codes be the same or to add two n bit numbers using binary arithmetic. Much information relevant here is implicit in the basic lemma of Chapter II, but many questions remain unanswered.

REFERENCES

[1] Miller, G. A., H. F. Blichtfeld and L. E. Dickson, Theory and Applications of Finite Groups, Dover Publications, Inc., New York, 1961.

[2] Winograd, S., "On the Time Required to Perform Multiplication," J. ACM, vol. 14, no. 4, 1967, pp. 793-802.