ERRATA SHEET

Technical Report SEL-65-1

<u>Errors in Finite Automata</u>

Philip S. Dauber

page 24 Definition 4.9, line 2:

$M_i = (M_i, \Sigma, \delta_i)$

page 40 Definition 6.1:

and $m_o \in M$ (initial state).

page 51 Corollary 6.1:

An error E is finite under a strongly connected M_1-transduction if and only if $\pi_E \leq \bigcap_{\{s_i\}} \pi_{s_i}$ where $\{s_i\}$ is the set of minimum idempotents of $S_{M_2}(W)$.

page 55

Hence $h_{M_2}(W) \supseteq \{s_{oo}\}$,

THE UNIVERSITY OF MICHIGAN

SYSTEMS ENGINEERING LABORATORY

Department of Electrical Engineering
College of Engineering

Technical Report SEL-65-1

ERRORS IN FINITE AUTOMATA

by

Philip S. Dauber

October, 1965

This research was supported in part by AF contract AF 30(602)-3546 and in part by the General Electric Research Laboratory, Information Study Section, Schenectady, New York.

Acknowledgments

I would like to express my appreciation to Professors Richard F. Arnold, Harvey L. Garner, John H. Holland, and Eugene L. Lawler for serving on my doctoral committee. I would particularly like to thank Professor Arnold for serving as chairman. I would also like to thank Miss Sharon Bauerle for her fine job of typing this report.

I would, in particular, like to thank my wife, Elayne, for her constant encouragement throughout this work.

This work was supported in part by the Information Systems Laboratory of The University of Michigan through Air Force contract AF 30(602)-3546. Parts of sections four and five were done during the summer of 1964 while the author was associated with the Information Studies Section of the General Electric Research Laboratory, Schenectady, New York. To these two I give my sincere thanks for their financial and intellectual support.

Abstract

The purpose of this research is to study the correctability properties of errors in a finite automaton driven by a random source. An error is defined to be a pair of states and is corrected by a tape if the tape takes both coordinates of the pair into the same state. Errors are then classified as one of four types: correctable, finite, definite, and non-correctable. A correctable error is an error for which there is a correcting tape. The error is finite if the probability of the set of correcting tapes approaches one as the length of the tapes gets longer. A definite error is an error for which all tapes, of length greater than some fixed length, are correcting tapes. A non-correctable error is one for which there does not exist a correcting tape.

We show that the set of finite errors induces a partition, called the finite error partition, on the set of states. Also, for a restricted class of random sources, this partition can be obtained from the set of correctable errors independent of the statistics of the source.

The notation of the semigroup of the automaton is then introduced. It is shown that many of the error properties of the automaton can be studied in terms of its semigroups. In particular, necessary and sufficient conditions are given for the automaton to have errors which are correctable but not finite and for the automaton to have only definite or non-correctable errors.

Further results are then given on analyzing the error properties of finite automata which have a large number of states but can be decomposed into a series, parallel connection of smaller automata.

Table of Contents

1. Introduction

This problem arose from an attempt to make a general study of reliability in computer like machines. The classic results of von Neumann (ref. 10), deal only with networks which do not have any feedback. Thus a malfunction only causes the network to be in the incorrect state for a fixed length of time.

However, a malfunction in the general case with feedback can cause an error which persists forever. Fortunately, not all errors are of this type. Some errors are of the type that can persist only for a bounded time. Some, although they can persist infinitely long, for certain random sources they have a probability of being corrected which approaches one as the tapes get longer. Thus "almost all" of the "long" tapes correct the error.

This is the phenomenon which we will study. We will show that errors of the latter type, which will be called finite errors, induce a partition on the set of states. In section two we will show that in the case where the set of states is a finite set, this partition can be obtained from another relation on the set of states without knowing the statistics of the same. Thus a phenomenon which will start out as probabilistic in nature will turn out to be a deterministic one. We would like to apologize now for the trivial nature of the probabilistic arguments which will be used. This is due to the nature of the problem.

In Appendix A we will show that in the case where the set of states is infinite, the phenomenon of an error being corrected by almost all the long tapes is indeed a probabilistic one. The infinite case will not be studied any further due to the fundamental difference in the nature of this problem.

The finite state problem will be studied within the framework of finite automata theory. Although we have attempted to give the definition of all the terms that we have used, we are still assuming that the reader is familiar with the main results in this area. Thus where the proof involves a construction used in the reference, we will only outline the construction.

We also found that the use of semigroups is very helpful. Thus in section four we have given the necessary definitions and fundamental results. We have not assumed that the reader has any previous knowledge of the theory of semigroups. The value of semigroups will become quite evident in the later sections.

2. Formalization of the Problem

In order to clarify the notation and to make the problem more formal, we will begin by defining a finite automaton and an error in a finite automaton.

Definition 2.1

A finite automaton M is a triple

$$M = (M', \Sigma, \delta).$$

M' is a finite set with elements m_i (set of states);

Σ is a finite set with elements σ_i (input alphabet);

δ is a function from $M' \times \Sigma \rightarrow M'$ (next state function).

Later we will use M both to denote the finite automaton and its set of states. We will also extend δ to $M' \times \Sigma^*$, the set of sequences of symbols from Σ, in the natural manner with sequences read from left to right.

Definition 2.2

a. An error, E, in a finite automaton, M, is a pair of states (m_i, m_j).

b. An error, (m_i, m_j), is corrected by a tape t ("tape" is synonymous with "sequence") if and only if

$$\delta(m_i, t) = \delta(m_j, t).$$

We can think of an error (m_i, m_j) as the situation, when due to a previous malfunction, the automaton is in state m_i and should be in state m_j, or is in state m_j and should be in state m_i. We can see from the definition of an error being corrected that these situations are equivalent.

In this work we will consider a random source, which generates sequences and drives the automaton. The next definition will make this clearer.

Definition 2.3

A random source S over an alphabet Σ is a set $\{P_n\}$, where P_n is a probability distribution on Σ^n, satisfying the requirement:

1) for all integers n and sequences x of length n

$$\sum_{\sigma \in \Sigma} P_{n+1}(x\sigma) = P_n(x)$$

If, in addition, the random source satisfies the additional requirement:

2) There is a real number $k > 0$ such that for all sequences x of length n and letter σ

$$P_{n+1}(x\sigma) \geq k\, P_n(x)$$

then we will say that the source has property P. We will call the ratio $P_{n+1}(x\sigma)/P_n(x)$ the probability that the letter σ follows the sequence x. Note that this means that if the source has property P, there is a constant, k, associated with the source such that the probability of any σ following any sequence x is always greater than k. In addition we will say that a random source $S = \{P_n\}$

drives an automaton M if the probability distribution on the set of input strings of length n for the automaton M is P_n.

Definition 2.4

Let S be a random source with property P and output symbols Σ, and let $M = (M, \Sigma, \delta)$ be a finite automaton driven by S. For an error

$E = (m_i, m_j)$ we define the following:

a. $\gamma_\ell^S(m_i,m_j)$ = probability of the set of tapes of length ℓ which correct the error (m_i,m_j).

b. $\gamma^S(m_i,m_j) = \lim_{\ell\to\infty} \gamma_\ell^S(m_i,m_j)$ if the limit exists.

The following lemma shows that for any source S and any error (m_i,m_j), $\gamma^S(m_i,m_j)$ exists.

Lemma 2.1

$\lim_{\ell\to\infty} \gamma_\ell^S(m_i,m_j)$ always exists.

PROOF: $1 \geq \gamma_{\ell+1}^S(m_i,m_j) \geq \gamma_\ell^S(m_i,m_j)$. Since the limit of a monotonic, bounded sequence always exists, the lemma is proved.

Q.E.D.

Now let us consider the following classification of errors in a finite automaton M being driven by a source S as above.

<u>Definition 2.5</u>

An error $E = (m_i, m_j)$ is

a. <u>definite</u> if and only if there is an ℓ such that $\gamma_\ell^S(E) = 1$.

b. <u>finite</u> if and only if $\gamma^S(E) = 1$.

c. <u>correctable</u> if and only if $\gamma^S(E) > 0$.

d. <u>non-correctable</u> if and only if $\gamma^S(E) = 0$.

3. Fundamental Results

In this section we will derive some fundamental properties of errors and will show the connection between the concepts of correctable and finite errors.

<u>Theorem 3.1</u>

If, for any source S, $\gamma^S(m_1,m_2) = g_1$ and $\gamma^S(m_2,m_3) = g_2$, then $1 - |g_1 - g_2| \geq \gamma^S(m_1,m_3) \geq (g_1 + g_2) - 1$.

PROOF: Let T0 be the set of tapes that do not correct (m_1,m_2) or (m_2,m_3); T1 be the set that either corrects (m_1,m_2) or (m_2,m_3) but not both; T2 be the set that corrects (m_1,m_2) and (m_2,m_3); and T3 be the set that corrects (m_1,m_3). We know that T0, T1, and T2 are mutually disjoint and that $T2 \subset T3 \subset T2 \cup T0$. We will use $Pr_\ell(T)$ to mean the probability that a tape t of length ℓ is in T. Therefore, we have

$$g_1 + g_2 = \lim_{\ell \to \infty} (Pr_\ell(T1) + 2Pr_\ell(T2))$$

$$= \lim_{\ell \to \infty} Pr_\ell(T1) + 2 \lim_{\ell \to \infty} Pr_\ell(T2).$$

But we also have for all ℓ, $Pr_\ell(T1) + Pr_\ell(T2) \leq 1$. Therefore

$$g_1 + g_2 \leq 1 + \lim_{\ell \to \infty} Pr_\ell(T2) \leq 1 + g_3$$

where $g_3 = \gamma(m_1,m_3)$. Hence $g_3 \geq g_1 + g_2 - 1$. Likewise, letting T11 be the set of tapes which correct (m_1,m_2) and not (m_2,m_3), and T12 be the set which corrects (m_2,m_3) and not (m_1,m_2), we have

$$\gamma_\ell(m_1,m_2) = Pr_\ell(T2) + Pr_\ell(T11)$$

and

$$\gamma_\ell(m_2,m_3) \;=\; Pr_\ell(T2) + Pr_\ell(T12).$$

Thus

$$|\gamma_\ell(m_1,m_2) - \gamma_\ell(m_2,m_3)| \;=\; |Pr_\ell(T11) - Pr_\ell(T12)|.$$

But

$$|Pr_\ell(T11) - Pr_\ell(T12)| \leq Pr_\ell(T1) \leq 1 - (Pr_\ell(T0) + Pr_\ell(T2)).$$

Now, taking limits as ℓ goes to infinity, we get

$$|g_1 - g_2| \leq 1 - g_3.$$

Therefore

$$g_3 \leq 1 - |g_1 - g_2|.$$

Q.E.D.

Corollary 3.1

The set of finite errors in an automaton M driven by any source S induces a partition on the set of states. That is, there is a partition π_F on the states of M so that $E = (m_i,m_j)$ is finite if and only if $m_i \equiv m_j(\pi_F)$.

PROOF: It is obvious that if $\gamma^S(m_i,m_j) = 1$, then $\gamma^S(m_j,m_i) = 1$ by the symmetry inherent in Definition 2.2b. Likewise, $\gamma^S(m_i,m_i) = 1$. Now, by Theorem 3.1, we have that if $\gamma^S(m_i,m_j) = 1$ and $\gamma^S(m_j,m_k) = 1$, then $\gamma^S(m_i,m_k) = 1$. Hence, the finiteness relation is an equivalence relation and partitions the set of states.

Q.E.D.

We will use the abbreviation lg(t) for the length of the tape t.

Theorem 3.2

Let $C \subset M \times M$ be the relation $(m_i,m_j) \in C$ if and only if (m_i,m_j) is a correctable error. Also, let S be a source with property P. Then $\gamma^S(m_i,m_j) = 1$ if and only if for all tapes t, $(\delta(m_i,t), \delta(m_j,t)) \in C$.

PROOF: If (m_i,m_j) is finite then, obviously, (m_i,m_j) is correctable. If there is a tape t such that $(\delta(m_i,t), \delta(m_j,t))$ is not correctable, then for all t' $(\delta(m_i,tt'), \delta(m_j,tt'))$ is not correctable. Hence, $\gamma(m_i,m_j) \leq 1 - (k)^{lg(t)} < 1$ where k is the constant greater than zero associated with the source. Therefore (m_i,m_j) is not a finite error. Conversely, let us assume that for all t $(\delta(m_i,t), \delta(m_j,t)) \in C$. Let $A = \{(m_k,m_\ell) \mid \text{for some t } \delta(m_i,t) = m_k \text{ and } \delta(m_j,t) = m_\ell\}$. Then, for each $(m_k,m_\ell) \in A$, pick a t' which corrects (m_k,m_ℓ). Let $p = k^r$ where $r = \max lg(t')$. Then $\gamma_\ell(m_i,m_j) \geq 1 - (1-p)^{[\ell/r]}$ where $[\ell/r]$ is the greatest integer less than ℓ/r. Hence

$$\lim_{\ell \to \infty} \gamma_\ell(m_i,m_j) \geq 1 - \lim_{\ell \to \infty} (1-p)^{[\ell/r]}.$$

Since $p > 0$, we have $\gamma(m_i,m_j) = 1$.

Q.E.D.

Since the concept of an error being correctable is not dependent upon the source, the above theorem tells us that as long as we are dealing only with the class of sources that have property P, the property of an error being finite is also independent of the source. In what follows we will assume that the term "source" will refer to a source with property P unless we explicitly say otherwise. Thus we will call an error a finite error if it is finite for some source (hence all sources) with property P, and we will call π_F the finite error partition. Likewise we will drop the superscript denoting the source. We will call an error a non-trivial error if it is not of the form (m_i, m_i).

Theorem 3.2 gives us some idea of the connection between the relation C and the partition π_F. The next theorem is a stronger characterization of this connection.

We will use the canonical ordering on partitions (i.e. $\pi_1 \geq \pi_2 \Longleftrightarrow (m_1 \equiv m_2\ (\pi_2) \Longrightarrow m_1 \equiv m_2\ (\pi_1))$.

<u>Theorem 3.3</u>

π_F is the coarsest partition with the substitution property such that $m_i \equiv m_j\ (\pi_F) \Longrightarrow (m_i, m_j) \epsilon C$.

PROOF: First we must show that if $m_i \equiv m_j (\pi_F)$ then for all tapes, t, $\delta(m_i, t) \equiv \delta(m_j, t)(\pi_F)$ and hence π_F has the substitution property. Let us assume that this was not true and there was a t such that $\delta(m_i, t) \not\equiv \delta(m_j, t)(\pi_F)$. Then by Theorem 3.2 there would be a t such that the error, $(\delta(\delta(m_i, t), t'), \delta(\delta(m_j, t), t')) = (\delta(m_i, tt'), \delta(m_j, tt'))$, is not a correctable error. Hence all tapes in the set $tt'\Sigma^*$ do not correct the error. Hence, for all $\ell \geq \lg tt'$, $\gamma_\ell(m_i, m_j) \leq 1 - k^{\lg tt'}$. Thus (m_i, m_j) is not finite which is a contradiction. Now to show that π_F is the coarsest such partition, let π be a partition with the substitution property such that $m_i \equiv m_j (\pi) \Longrightarrow (m_i, m_j) \epsilon C$. Then if $m_i \equiv m_j\ (\pi)$, $(m_i, m_j) \epsilon C$. Also, since by the hypothesis π has the substitution property, for all tapes t, $\delta(m_i, t) \equiv \delta(m_j, t)(\pi)$ and hence $(\delta(m_i, t), \delta(m_j, t)) \epsilon C$. But by Theorem 3.2, this means that $m_i \equiv m_j (\pi_F)$.

Therefore $\pi \leq \pi_F$.

Q.E.D.

We will now give two definitions due to J. Hartmanis and R. Stearns (ref. 4 and 6).

Definition 3.1

A finite automaton $M_1 = (M_1, \Sigma, \delta)$ state behavior realizes $M_2 = (M_2, \Sigma, \delta')$ if and only if there is a one to one mapping, h of M_2 onto M_1 such that for all tapes t and states $m_i \in M_2$, $h(\delta'(m_i,t)) = \delta(h(m_i),t)$.

Definition 3.2

Let $M_1 = (M_1, \Sigma, \delta)$, $M_2 = (M_2, \Sigma', \delta')$ and $\lambda: M_1 \times \Sigma \to \Sigma'$. Then the series connection of M_1 with M_2 with connecting function λ is the automaton $M = (M_1 \times M_2, \Sigma, \delta'')$ where δ'' is defined as follows:

$$\delta''((m_i,m_j),\sigma) = (\delta(m_i,\sigma), \delta'(m_j, \lambda(m_i,\sigma))).$$

We will say that a finite automaton M can be state behavior realized by a series connection of finite automata M_1 and M_2 if there is a connecting function λ such that M is state behavior realized by a series connection of M_1 and M_2 with connecting function λ.

The following corollary is an immediate consequence of Theorem 3.3 and a well known result of Hartmanis.

Corollary 3.2

If M is a finite automaton with a finite partition π_F, then F can be state behavior realized by a cascade connection of two automata M/π_F and T, where all errors in T are finite, and M/π_F has no non-trivial finite errors.

PROOF: By Theorem 3.3, π_F is a partition with the substitution property. Hence, we know (Hartmanis, ref. 4) that we can decompose M into a cascade connection of two automata where the state of the front automaton distinguishes between blocks of the partition and the back automaton distinguishes the elements of a single block. Thus since an error within a single block is a finite error, all errors in the back automaton are finite. Likewise, since an error out of a block is not a finite error, the front automaton has no finite errors.

Q.E.D.

Let us now look at an example in order to demonstrate these theorems.

Example 3.1

Let M = ({a,b,c,d,e}, {0,1}, δ) where δ is the mapping shown below.

δ	0	1
a	b	d
b	a	d
c	a	b
d	b	d
e	a	d

It is easy to show that

$$C = \{(a,d),(d,a),(b,c),(c,b),(e,a),(a,e),(e,d),(d,e),(b,e),(e,b),$$
$$(c,e),(e,c),(a,a),(b,b),(c,c),(d,d),(e,e)\}.$$

There are four equivalence relations with the substitution property contained in C.

$$\pi_1 = \{\bar{a}, \bar{b}, \bar{c}, \bar{d}, \bar{e}\}$$
$$\pi_2 = \{\overline{a,d}, \bar{b}, \bar{c}, \bar{e}\}$$
$$\pi_3 = \{\bar{a}, \overline{b,c}, \bar{d}, \bar{e}\}$$
$$\pi_4 = \{\overline{a,d}, \overline{b,c}, \bar{e}\}.$$

The greatest one is π_4. Thus the only finite errors are

$$\{(a,a),(b,b),(c,c),(d,d),(e,e),(a,d),(d,a),(b,c),(c,b)\}.$$

Also using Theorem 3.3 we can get a simple proof of a special case of a theorem which was proved by Winograd (ref. 15), and also in another context by Gilbert and Moore (ref. 3).

Corollary 3.3

All errors in an automaton M are finite if and only if M has a reset tape. (A tape t is a reset tape if $\delta(m_i,t)$ is independent of m_i.)

PROOF: From Theorem 3.3 we get that all errors in an automaton M are finite if and only if all errors are correctable. Define a tape $t = t_1\ t_2\ ,,,\ t_{k-1}$ (k = number of states of M) as follows:

t_1 corrects (m_1,m_2)

t_{i+1} corrects $(\delta(m_1, t_1 \ldots t_i), \delta(m_{i+2}, t_1 \ldots t_i))$.

If it is possible to construct such a t, then t is a reset sequence. It is not possible to construct such a tape if and only if for some i, $(\delta(m_i, t_1 \ldots t_i), \delta(m_{i+2}, t_1 \ldots t_i))$ is not a correctable error. But then, this $(\delta(m_1, t_1 \ldots t_i), \delta(m_{i+2}, t_1 \ldots t_i))$ is not finite. Hence we can construct t if and only if all the errors are finite.

Q.E.D.

Let us now look at another example to show the use of this theorem.

Example 3.2

Let $M = (\{a,b,c,d\}, \{0,1\}, \delta)$ where δ is shown below.

δ	0	1
a	b	c
b	c	d
c	d	b
d	d	b

It is easy to see that all the errors are correctable. Hence $\pi_F = \{\overline{a,b,c,d}\}$ and all errors are finite. Upon examination it can be seen that the tape 000 is a reset tape since $\delta(m_i, 000) = d$ regardless of m_i.

4. Introduction to Semigroups with Applications to Definite Errors

In the course of studying errors in finite automata, it became obvious that many of the error properties (as well as other properties) of finite automata are more easily discussed in terms of the semigroups of the automata (ref. 7 and 9). In this section we will give some results about semigroups. Their usefulness will become more and more apparent in later chapters. (For a greater exposition on semigroups the reader is referred to Clifford and Preston, ref. 2).

First let us define what a semigroup is and what we mean by the semigroup of an automaton.

<u>Definition 4.1</u>

A mapping γ: $S \times S \rightarrow S$ is said to be an <u>associative mapping</u> if and only if for all elements s_1, s_2, and s_3 all in S, $((s_1 \gamma s_2) \gamma s_3) = (s_1 \gamma (s_2 \gamma s_3))$.

Thus if a mapping is associative, the order of the elements in a compound mapping determine the result independent of parenthesization. Hence we will omit the parentheses.

<u>Definition 4.2</u>

A <u>semigroup</u> is a pair, $(S, \cdot)$ where S is a set of elements, and $\cdot$ is an associative mapping of $S \times S \rightarrow S$. A semigroup will be called <u>finite</u> if the set of its elements is finite.

We will call "." multiplication and will write $s_1 s_2$ rather than $s_1 \cdot s_2$ as with multiplication of real numbers. We will also use the same letter to denote the semigroup and its set of elements. It will be clear from the context which one we mean.

Definition 4.3

Let $M = (M, \Sigma, \delta)$ be a finite automaton. S_M, the semigroup of the finite automaton M, is the semigroup whose elements are transformations, mapping the set of states M into itself, induced by the next state function δ. That is, $S_M = \{s_i \mid s_i : M \to M$ and there is a nonempty string t over Σ such that for all states m_i, $\delta(m_i, t)$ is equal to the image of the state m_i under the transformation $s_i\}$. The multiplication operation $s_1 \cdot s_2$ is the composition of s_2 and s_1.

It is clear that the multiplication operation in the above operation is associative and that the set of its elements is closed under multiplication. Hence S_M is indeed a semigroup, as desired.

Definition 4.4 (Ideals)

(a) A nonempty subset L of a semigroup S is called a left ideal if and only if $SL \subseteq L$, where $SL = \{x \mid$ for some $s \in S$ and $\ell \in L$ $s\ell = x\}$.

(b) A right ideal, R, is a nonempty subset such that $RS \subseteq R$.

(c) A two-sided ideal, T, (or just ideal for short) is a nonempty subset such that $STS \subseteq T$.

(d) A two-sided ideal is called a minimum ideal if it does not properly contain another two-sided ideal. Similarly, a right (left) ideal is a minimum right (left) ideal if it does not properly contain any right (left) ideal.

(e) The <u>kernel</u> of a semigroup is the minimum two-sided ideal, if it exists.

The following three lemmas are known results about finite semigroups (See Clifford and Preston, ref. 2.) that will be used in this study.

<u>Lemma 4.1</u>

Every finite semigroup has a kernel.

PROOF: The kernel is just the intersection of all the two-sided ideals. There is always at least one two-sided ideal since S is a two-sided ideal of itself. Likewise, since S is finite, there can be only finitely many ideals. Also, if T_1 and T_2 are two ideals of S, then $T_1T_2 \subset T_1$ and $T_1T_2 \subset T_2$. Hence $T_1T_2 \subset T_1 \cap T_2$ and thus the intersection is non-empty. Thus the theorem is proved.

Q.E.D.

<u>Lemma 4.2</u>

Let S be a finite semigroup with kernel K. Then

$$K = \cup R_i = \cup L_j$$

where $\{R_i\}$ is the set of minimum right ideals and $\{L_j\}$ is the set of minimum left ideals.

PROOF: It is clear that if R_i is a minimum right ideal $R_i \subset K$ since if $s_i \epsilon K$, $R_i s_i \subset K$ and $R_i s_i \subset R_i$. Hence $R_i \cap K$ is not empty. But $R_i \cap K$ is a right ideal, hence, $R_i \cap K = R_i$, since R_i is a minimum right ideal. Thus $\bigcup_i R_i \subset K$. We will now show that $\bigcup_i R_i$ is a left ideal. First we claim that for any $s_i \epsilon S$, s_iR_i is another minimum right ideal of S. It is clearly a right ideal. Assume $V \subset s_iR_i$ is a right ideal. Then let $U \subset R_i$, $U = \{r_i \epsilon R_i \mid s_ir_i \epsilon V\}$. Then, if V is properly contained in s_iR_i, U is properly contained in R_i. But U is a right ideal. Hence $V = s_iR_i$ and thus $s_iR_i = R_j$ a minimum right ideal. Now $\cup R_i$ is a right and left ideal and, therefore, ideal. But K is contained in every two-sided ideal. Thus $K = \cup R_i$. Likewise $K = \cup L_j$.

Q.E.D.

Lemma 4.3

Let R be a minimum right ideal and L a minimum left ideal of a semigroup S. Then $RL = R \cap L = G$ is a group.

PROOF: See Clifford and Preston, page 77 (ref. 2).

I will now give an example illustrating some of the concepts which were introduced above.

Example 4.1

Let $M = (\{a,b,c,d\}, \{0,1\}, \delta)$ where δ is given in the following table:

δ	0	1
a	b	c
b	c	d
c	d	b
d	d	b

S_M, the semigroup of the finite automaton has elements as indicated below.

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}
a	b	c	c	d	b	d	b	d	b	d	c	d	c
b	c	d	d	b	b	d	b	c	d	d	c	d	c
c	d	b	d	b	d	d	b	c	d	c	d	b	c
d	d	b	d	b	d	d	b	c	d	c	d	b	c

The kernel K of S_M is the set of elements $\{s_6, s_7, s_{13}\}$. The only minimum right ideal is $\{s_6, s_7, s_{13}\}$. The three minimum left ideals are $\{s_6\}, \{s_7\}$, and $\{s_{13}\}$. Examples of right ideals which are not minimum right ideals are $\{s_3, s_6, s_7, s_8, s_{13}\}$ and $\{s_5, s_6, s_7, s_{10}, s_{11}, s_{12}, s_{13}\}$.

Definition 4.5

Let S be a semigroup. An element $s_i \epsilon S$ is idempotent if and only if $s_i s_i = s_i$. We will call an element a minimum idempotent if it is an idempotent element and also is contained in some minimum right ideal.

Using this definition and the above lemmas we get the following well known corollary first obtained by E. H. Moore (ref. 10).

Corollary 4.1

Let s be an element of a finite semigroup S. Then for some integer k, s^k is an idempotent element of S.

PROOF: Consider the set of elements $\{s^i\}$. This is a finite semigroup and hence must contain at least one minimum right ideal and one minimum left ideal. But the intersection of a minimum right and minimum left ideal is a group and hence contains an identity. But an identity must be an idempotent. Hence the corollary is proved.

Q.E.D.

So far in this section, we have only been giving results which deal with semigroups in general. The next theorem shows the intimate connection between errors in a finite automaton and the semigroup of the automaton. For convenience we will use $h_M(x)$ to denote the element of the semigroup of the finite automaton M which corresponds to the input tape x. Note that h_M is a semigroup homomorphism of Σ^* into S_M.

Theorem 4.1

Let E be an error in a finite automaton M. Then E is

1. correctable if and only if E is corrected by some minimum idempotent of S_M.
2. finite if and only if E is corrected by every minimum idempotent of S_M.

PROOF: If E is correctable then there is a tape x such that all tapes in the set $x\ \Sigma^*$ correct E. But $h_M(x\ \Sigma^*)$ is a right ideal of S_M, which must be finite since M is finite, and hence must contain a minimum idempotent. Likewise, if s_i, a minimum idempotent of S_M corrects E, just by the definition of S_M, there must be an $x \epsilon \Sigma^*$ such that $h_M(x) = s_i$. Hence E is correctable. Thus we have proven part 1.

Let us now assume that there is a minimum idempotent which does not correct E. Then there exists a right ideal $R \subset S_M$, all of whose elements do not correct E. But the set of tapes $\{x \mid h_M(x) \epsilon R\}$ is a right ideal of Σ^* and hence contains a right ideal of form $y\ \Sigma^*$ all of whose elements do not correct E. Hence E is not finite since $\gamma(E) \leq 1 - k^{lg(y)} < 1$. Now if E is corrected by every minimum idempotent we must show that E is finite. We know for every finite semigroup there is an element s_i such that for all s_j and s_k, $s_k s_i s_j$ is in a minimum right ideal. Also, if E is corrected by every minimum idempotent, it is corrected by every element of every minimum right ideal. Now let $x \epsilon \Sigma^*$ be such that $h_M(x) = s_i$. Also, let $V = \{y \epsilon \Sigma^* \mid y \neq z \times z' \text{ for all } z \text{ and } z'\}$. Then it is obvious that $\lim_{i \to \infty} P_i V = 0$; thus $\lim_{i \to \infty} P_i(\Sigma^* - V) = 1$. But all elements of Σ^*-V correct E. Hence E is finite. We have thus proven part 2.

Q.E.D.

We will now state an immediate corollary to this theorem. We will use π_s to indicate the partition induced by the mapping associated with the semigroup element s.

Corollary 4.2

An error $E = (m_i, m_j)$ in a finite automaton M is

(1) finite if and only if $m_i \equiv m_j (\cap \pi_{s_i})$ where $\{s_i\}$ is the set of minimum idempotents of S_M.

(2) correctable if and only if $m_i \equiv m_j (\pi_{s_i})$ for some minimum idempotent of S_M, s_i.

Thus the above corollary tells us that the partitions associated with the minimum idempotents of S_M completely characterize the error properties of the automaton M. Returning again to semigroups, the next theorem will prove to be very useful in studying definite errors.

Theorem 4.2

Let S be a finite semigroup and T a two-sided ideal of S. Then there is a constant c such that for all $n > c$ and all $s_1, s_2, \ldots, s_n$, $s_1 s_2 \ldots s_n \in T$, if and only if there is no idempotent element of S contained in S-T.

PROOF: First of all, if there is an idempotent s_i in S-T, then for all j, s_i^j is in S-T. Thus there is no such constant c. We will prove the converse by contradiction. If there does not exist such a constant c, then there is an infinite set of products $s_{1_1}, s_{2_1} \cdot s_{2_2}, s_{3_1} \cdot s_{3_2} \cdot s_{3_3}, \ldots$ all in S-T. Now let n be an integer greater than $|S|$, the cardinality of the set of elements of the semigroup S, and consider the set $P = \{p \mid p_i = s_{n_1} s_{n_2} \ldots s_{n_i}\}$. Since $|P|$ must be less than $|S|$, there must be an i and j such that $p_i = p_j$; that is $s_{n_1} s_{n_2} \ldots s_{n_i} = s_{n_1} s_{n_2} \ldots s_{n_i} s_{n_{i+1}} \ldots s_{n_j}$. Now since T is a two-sided ideal and all p_i are in S-T, we get that p_i is in S-T and $x = s_{n_{i+1}} \ldots s_{n_j}$ is in S-T. Hence for any ℓ, $p_i x^\ell = p_i$ is in S-T. However from Corollary 4.1 we get that some power of x must be an idempotent, and we know it is not in T. Therefore S-T must contain an idempotent.

Q.E.D.

This theorem will get most of its use in this study in the particular instance when T is the kernel of S.

Corollary 4.3

Let S be a finite semigroup and K its kernel. Then there is a constant c such that $S^c \subset K$ if and only if all the idempotents of S are in K.

PROOF: This follows immediately from Theorem 4.2 if we let T = K.

Now let us give two more definitions.

Definition 4.6

A right ideal of a semigroup S, is a universal minimum right ideal if it is contained in every other right ideal of S. (Clifford and Miller, ref. 1)

Definition 4.7 (Perles, Rabin, and Shamir, ref. 12)

A finite automaton M = (M, Σ, δ) has a k-definite move function if and only if for all sequences $\sigma_1 \ldots \sigma_k$ of k letters from Σ; $\delta(m_i, \sigma_1 \ldots \sigma_k) = \delta(m_j, \sigma_1 \ldots \sigma_k)$ for all m_i, m_j in M. Note that this implies that

$$\delta(m_i, \sigma_1 \ldots \sigma_k) = \delta(m_j, \sigma_{k+1} \ldots \sigma_\ell \sigma_1 \ldots \sigma_k).$$

We will informally call a finite automaton k-definite if its move function is k-definite.

The tie up of definite errors and definite automata is clearer after the following two lemmas which are due to Hartmanis and Stearns (ref. 5).

Lemma 4.4

A finite automaton $M = (M, \Sigma, \delta)$ is definite if and only if all its errors are definite.

Lemma 4.5

If M is a finite automaton, then M can be decomposed into a series connection of two finite automata M_1 and M_2 as shown below with all errors in M_2 being definite and no nontrivial error in M_1 being definite. Hence M_2 is a definite automaton

This lemma follows from the fact that there is a partition with substitution property π_D on the states of M with the property that an error E is definite if and only if $\pi_E \leq \pi_D$. The following theorem gives a characterization of the semigroups associated with definite automata.

Definition 4.8

Let $S = (S, \cdot)$ be a semigroup. Then, an element z of S is a right zero if and only if for all $s \in S$, $s \cdot z = z$.

Definition 4.9

We will say that $M = (M, \Sigma, \delta)$ is a union of the finite automata $M_i = (m_i, \Sigma_i, \delta_i)$ $i=1,\ldots,k$ if the following conditions hold:

1. $i \neq j \implies M_i \cap M_j$ empty
2. $\bigcup_i M_i = M$

3. $\delta_i = \delta$ restricted to M_i.

Theorem 4.3

Let $M = (M, \Sigma, \delta)$ be a finite automaton. Then the following two conditions are equivalent:

(1) M is a union of definite automata.

(2) S_M contains a universal minimum right ideal U such that all elements of U are idempotent and all the idempotents of S_M are in U.

PROOF: If M is a union of k-definite automata then for all $\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_k}$, $\delta(m_i, \sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_k}) = \delta(m_i, t\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_k})$ for any tape t. Hence $h_M(t\ \sigma_{i_1} \ldots \sigma_{i_k})$ must equal $h_M(\sigma_1 \ldots \sigma_k)$. Thus $h_M(t)\cdot h_M(\sigma_{i_1} \ldots \sigma_{i_k})$ $= h_M(\sigma_1 \ldots \sigma_{i_k})$, since h_M is a homomorphism. But since $\sigma_i \ldots \sigma_{k_k}$ was an arbitrary type of length k we get that if t' is a tape of length k, $h_M(t')$ is a right zero and thus an idempotent. It is also clear that if x is an idempotent, then $x^k = x$, an element of $h_M(\Sigma^k)$. Thus, since every right ideal must contain an idempotent we get that $h_M(\Sigma^k)$ is a universal minimal right ideal all of whose elements are idempotents and which contains all the idempotents of S_M.

To prove the converse we must show that there is a k such that all products of k elements are in U and that every element of U is a right zero. Now since U is a universal minimum right ideal, U is the only minimum right ideal and thus by Lemma 4.2, U is the kernel of S_M. But since all the idempotents of S_M are in U, we are guaranteed of the existence of such a k by Theorem 4.2. Also, since U is the kernel of S_M, it is equal to the union of the minimum left ideals. But since the intersection of a minimum left ideal and a minimum right ideal is a group (Lemma 4.3), it must contain exactly one idempotent. Hence every element of U must be a one element minimum left ideal. Therefore, for all $u \in U$ and $s \in S_M$, $s \cdot u = u$ which means that u is a right zero.

Q.E.D.

From this theorem we get two immediate corollaries.

Corollary 4.4

M is a union of definite automata if and only if the set of idempotents of S_M is a minimum right ideal.

PROOF: If the set of idempotents is a minimum right ideal, it must be a universal minimum right ideal since every right ideal must be a finite semigroup and hence must contain an idempotent.

Corollary 4.5

M is a union of definite automata if and only if every idempotent of S_M is a right zero.

PROOF: It is clear that for any finite semigroup the set of right zeros, if it is not empty, is a universal minimum right ideal. Conversely, if all the idempotents are in a universal minimum right ideal, they are all right zeros as shown in the proof of Theorem 4.3

Q.E.D.

Let us point out here that Theorem 4.2 can be applied to give more results on definite errors. Thus, for instance, a finite automata is such that all its finite errors are definite errors if and only if there are no idempotent elements outside its kernel. As an aside, let us remark that it can be shown that linear automata only have errors which are either non-correctable or definite. Hence the above results tell us something about the structure of semigroups of linear automata.

5. Automata With Errors Which Are Correctable But Are Not Finite

In the course of studying errors, an attempt was made to classify the set of finite automata which have errors that are correctable but not finite. One conjecture of the author was that if an automaton was strongly connected [That is, for all pairs of states m_i and m_j, there is a tape t such that $\delta(m_i,t) = m_j$] and if it had correctable errors which were not finite, then the automaton could be state-behavior realized by a cascade connection of some automaton M_1 and a nontrivial permutation automaton M_2 (a permutation automaton is one where every input causes a permutation of the states).

Note that this conjecture said that correctable but not finite errors were just errors which arose in M_1 and "drifted" into M_2. That this conjecture is not true can be seen by the automaton in Example 5.1.

<u>Example 5.1</u>

$M = (\{a,b,c,d,e\}, \{0,1\}, \delta)$

δ	0	1
a	a	b
b	b	c
c	a	d
d	e	b
e	e	d

The correctable error relation is

$C = \{(a,c),(c,a),(d,e),(e,d),(c,e),(e,c),(a,d),(d,a),(a,a),(b,b),(c,c),(d,d)\}$

However, there is no nontrivial partition with the substitution property on the states of M. Hence, there are no nontrivial finite errors, and also M can not be decomposed as in the conjecture.

It turns out that there is a classification of the class of automata with errors that are correctable but not finite in terms of their semigroups. This classification is very useful in aiding our intuition as to why certain errors are correctable and not finite and also is useful if we want to construct automata in this class.

Definition 5.1

If S is a semigroup, then S^1 is equal to S if S has a two-sided identity and is equal to S ∪ {1} where $s \cdot 1 = 1 \cdot s = s$ if S does not have an identity.

Definition 5.2

Let S be a semigroup. Define a relation $\leq$ on the elements of S^1 as follows. $s_i \leq s_j$ if and only if there is an element $s_k \in S^1$ such that $s_i = s_j \cdot s_k$. Also, define a relation $\equiv$ with $s_i \equiv s_j$ if and only if $s_i \leq s_j$ and $s_j \leq s_i$.

The following lemma will tell us something about the relations "$\leq$" and "$\equiv$". We will use $[s_i]$ for the equivalence class of s_i.

Lemma 5.1

(1) $\equiv$ is an equivalence relation on the elements of S^1.

(2) $\leq$ is a partial ordering on $S/\equiv$, the equivalence classes of S modulo the relation $\equiv$.

(3) If S is a finite semigroup, then $s_i s_j \geq s_j$ implies $s_i s_j \equiv s_j$.

PROOF: (1) $\equiv$ is obviously symmetric. Likewise, since S^1 has an element such that $s_j \cdot 1 = s_j$, it is reflexive. Also, $s_i \equiv s_j$ and $s_j \equiv s_k$ implies there are elements s_t, s_u, s_v, and s_w such that $s_i s_t = s_j$, $s_j s_u = s_i$, $s_j s_v = s_k$, and $s_k s_w = s_j$ by the definition of $\equiv$. Therefore, $s_k = s_i(s_t s_v)$ and $s_i = s_k(s_w s_u)$. Thus $s_i \equiv s_k$. Since $\equiv$ is symmetric, reflexive and transitive it is an equivalence relation.

(2) If $[s_i] \leq [s_j]$ and $[s_j] \leq [s_k]$, then $s_i \leq s_j$ and $s_j \leq s_k$. Thus $s_i \leq s_k$ and hence $[s_i] \leq [s_k]$. Therefore "$\leq$" is a transitive relation on $S/\equiv$. Likewise, it is antisymmetric since $x \leq y$ and $y \leq x$ implies $x \equiv y$ which means that $[x] = [y]$. A relation which is transitive and antisymmetric is a partial ordering.

(3) $s_i s_j \geq s_j$ implies there is an $s_k \epsilon S$ such that $s_i s_j s_k = s_j$. Also, since $S \cdot s_i \subseteq S$, we have $S s_i s_j s_k \subseteq S s_j s_k$. But since $s_i s_j s_k = s_j$, we get $S s_j \subseteq S s_j s_k$. However, since $(S s_j) s_k \subseteq S s_j$ we have $S s_j = S s_j s_k$. Hence, multiplication on the right by s_k causes a permutation of the finite set $S s_j$. Therefore, some power of s_k has to be the inverse transformation of s_k with respect to the set $S s_j$. Call this element s_k^{-1}.

Therefore, since $s_i s_j \epsilon S s_j$, we have $s_i s_j (s_k s_k^{-1}) = s_i s_j$. Likewise, $(s_i s_j s_k) s_k^{-1} = s_j s_k^{-1}$. Hence $s_i s_j = s_j s_k^{-1}$ and thus $s_i s_j \leq s_j$. But, since the hypothesis was that $s_i s_j \geq s_j$, we have proved that $s_i s_j \equiv s_j$.

Q.E.D.

<u>Definition 5.3</u>

Let M be a set, and S a semigroup of transformation on S. For each element s_i of S we define an equivalence relation π_{s_i} on M as follows: $m_j \equiv m_k(\pi_{s_i})$ if and only if s_i maps m_j and m_k onto the same element. Again, we will use the canonical ordering on the set of equivalence relations.

Note that we now have two partial orderings. One is on the semigroup and the other on the set of equivalence relations induced by the semigroup elements. We will now tie these two relations together.

Lemma 5.2

$s_i \leq s_j$ implies $\pi_{s_i} \geq \pi_{s_j}$.

PROOF: If $s_i \leq s_j$, then there is an s_k such that $s_j s_k = s_i$. But if $m_i \equiv m_j (\pi_{s_j})$, then $m_i \equiv m_j (\pi_{s_j s_k})$. Thus $m_i \equiv m_j (\pi_{s_i})$, and therefore $\pi_{s_i} \geq \pi_{s_j}$.

Corollary 5.1

If $s_i \equiv s_j$, then $\pi_{s_i} = \pi_{s_j}$.

PROOF: If $s_i \equiv s_j$, then both $s_i \leq s_j$ and $s_j \leq s_i$. Hence, by Lemma 5.2 $\pi_{s_i} \leq \pi_{s_j}$ and $\pi_{s_j} \leq \pi_{s_i}$. But since $\leq$ is a partial ordering on equivalence relations, we get that $\pi_{s_i} = \pi_{s_j}$.

Q.E.D.

In general the converse to Lemma 5.2 and Corollary 5.1 are not true. This can be seen, for example, in the semigroup shown in Example 4.1. In this case, $\pi_{s_1} = \pi_{s_2} = \{\bar{a}, \bar{b}, \overline{c, d}\}$. But, neither is $s_1 \leq s_2$ nor is $s_2 \leq s_1$ as would be implied by the converse. However, in an important special case the converse holds.

Theorem 5.1

Let S be a semigroup of transformations on M and let s_i be an idempotent element of S. Then, for all $s_j \in S$, if $\pi_{s_j} \geq \pi_{s_i}$ then $s_j \leq s_i$.

PROOF: Since s_i is an idempotent of S, we have $s_i s_i = s_i$. Thus, for all $m \in M$, $ms_i s_i$, the image of the point m under the transformation $(s_i s_i)$ is equal to ms_i. Hence $ms_i \equiv m(\pi_{s_i})$. Therefore, if $\pi_{s_j} \geq \pi_{s_i}$, we have $ms_i \equiv m(\pi_{s_j})$. However, this implies that $ms_i s_j = ms_j$. Hence the transformation $s_i s_j$ is equal to s_j, and therefore $s_j \leq s_i$.

Q.E.D.

<u>Corollary 5.2</u>

Let s_i and s_j be two elements of a semigroup S such that for some idempotent $s_k \in S$, $[s_i] = [s_k]$. Then $\pi_{s_i} \leq \pi_{s_j}$ if and only if $s_i \geq s_j$.

PROOF: If $s_i \geq s_j$, then by Lemma 5.2, $\pi_{s_i} \leq \pi_{s_j}$. Conversely, if $[s_i] = [s_k]$, then by Corollary 5.1, $\pi_{s_i} = \pi_{s_k}$. But, by Theorem 5.1, if $\pi_{s_k} \leq \pi_{s_j}$, then $s_k \geq s_j$. Hence, since $[s_i] = [s_k]$, we have $s_i \geq s_j$.

Q.E.D.

<u>Definition 5.4</u>

An equivalence class $[s_i]$ of a semigroup S is called a <u>minimal class</u> if and only if for all s_j, $[s_j] \leq [s_i]$ implies that $[s_j] = [s_i]$. It is a <u>minimum class</u> if for all s_j, $[s_i] \leq [s_j]$.

<u>Lemma 5.3</u>

Every minimal class of a finite semigroup contains at least one idempotent.

PROOF: We know by definition of $[s_i]$ that $[s_i]S \supseteq [s_i]$. Likewise, if $[s_i]$ is minimal, $[s_i]S \subseteq [s_i]$. Hence $[s_i]S = [s_i]$ and thus $[s_i]$ is a right ideal of S. But since a right ideal is also a finite sub-semigroup, it must contain an idempotent.

Q.E.D.

We will now apply this to characterize the finite automata which have errors which are correctable but not finite.

Theorem 5.2

A finite automaton, M, has errors which are correctable but not finite if and only if S_M has two elements s_i and s_j such that $[s_i] \neq [s_j]$ and $[s_i]$ and $[s_j]$ are both minimal classes.

PROOF: If (m,m') is an error of M which is correctable but not finite, then by Corollary 4.2, there are elements s_i, s_j in S_M such that $m \equiv m'(\pi_{s_i})$ and for all $s_k m \not\equiv m'(\pi_{s_j s_k})$. Thus for all $s_u \leq s_i$ and $s_v \leq s_j$, $\pi_{s_v} \neq \pi_{s_u}$ since $m \equiv m'(\pi_{s_u})$ and $m \not\equiv m'(\pi_{s_v})$. Therefore $[s_v] \neq [s_u]$. Conversely, assume $[s_i]$ and $[s_j]$ are two minimal classes such that $[s_i] \neq [s_j]$. Then, since there are idempotents in both $[s_i]$ and $[s_j]$ by Lemma 5.3, we get that $\pi_{s_i} \neq \pi_{s_j}$ by Theorem 5.1. Hence there are elements $m,m' \in M$ such that $m \equiv m'(\pi_{s_i})$, and $m \not\equiv m'(\pi_{s_j})$. Since for all s_k, $\pi_{s_j s_k} = \pi_{s_j}$ because $[s_j]$ is minimal, we get that (m,m') is correctable but not finite.

Q.E.D.

There are many interesting corollaries to this theorem.

Corollary 5.3

A finite automaton M has errors that are correctable but not finite if and only if its semigroup, S_M, has equivalence classes which are minimal but not minimum.

Corollary 5.4

Let M be a finite automaton with correctable, nonfinite errors. If we have two copies of M in the same unknown state, then there are tapes x and y such that if we feed one copy x and the other y, we cannot find tapes x' and y' so that by feeding x' to one and y' to the other copy they will both once again be in the same state.

Corollary 5.5

A finite automaton has no errors that are correctable but not finite if and only if for any two elements of its semigroup there is a third element less than both.

PROOF: If any two elements have a common element less than both, then there can be only one minimal class. Hence the theorem follows from Theorem 5.3.

Q.E.D.

There is one case in which the error properties come out exceptionally nice. For interest, this is given below.

Definition 5.5 (Laing and Wright, ref. 8)

A finite automaton, M, is called a commutative automaton if and only if for all s_i, s_j elements of S_M, $s_i s_j = s_j s_i$.

Corollary 5.6

A commutative automaton has no errors which are correctable but not finite.

PROOF: For any two elements s_i, s_j of its semigroup $s_i s_j = s_j s_i$, and thus $s_i s_j \leq s_i$ and $s_i s_j \leq s_j$. Hence this follows from Corollary 5.5.

Q.E.D.

Corollary 5.7

A commutative automaton, M, can be state behavior realized by an automaton of the form

where P is a permutation automaton and F is a finite automaton all of whose errors are finite.

PROOF: We know π_F, the finite error partition on M, has the substitution property. Since P must distinguish amongst blocks of π_F, no error in P can be correctable. Hence P must be a permutation automaton. Likewise, since F distinguishes amongst the elements of a single block of M, all errors in F must be finite.

Q.E.D.

Let us now give one more result concerning errors which are correctable and not finite.

Theorem 5.3

Let N be the set of errors which are correctable but are not finite in some finite automaton. Then there does not exist a tape t such that t corrects all the errors in N.

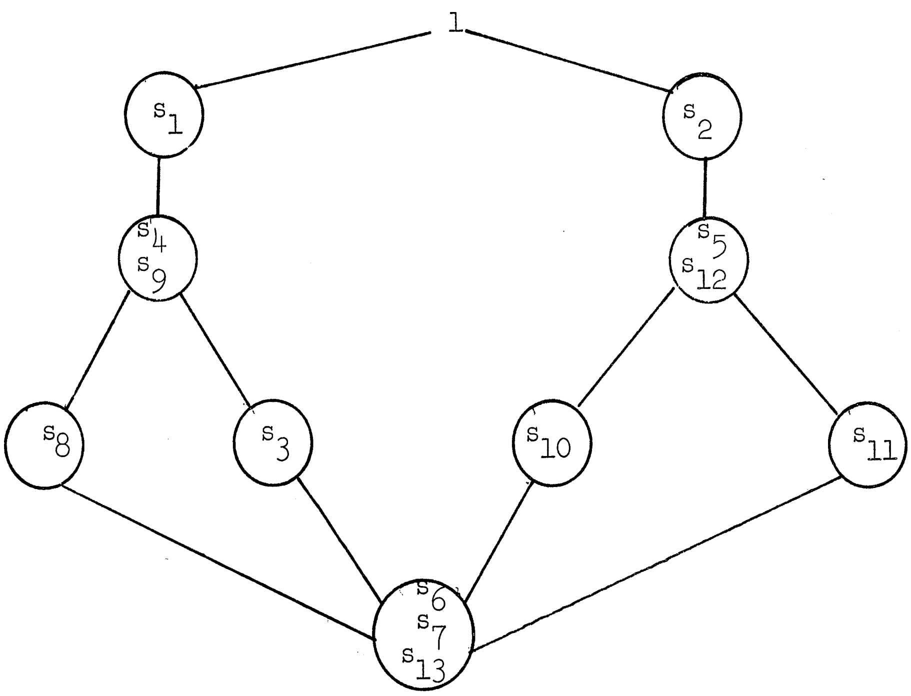
PROOF: Let $E \in N$ and let t be a tape that corrects E. We can assume without loss of generality that t corrects all finite errors. Since E is correctable but not finite, by Theorem 4.1 there is a minimum idempotent $s_i \in S_M$ such that $\pi_E \leq \pi_{s_i}$. But since $\pi_E \leq \pi_{h_M(t)}$, we get $\pi_{s_i} \not\geq \pi_{h_M(t)}$. Likewise $\pi_{s_i} \not\leq \pi_{h_M(t)}$ since if it were, then by Theorem 5.1 we would have $h_M(t) \leq s_i$. But since s_i is a minimum idempotent and $h_M(t) \neq s_i$, this cannot be true. Hence neither $\pi_{s_i} \leq \pi_{h_M(t)}$ nor $\pi_{s_i} \geq \pi_{h_M(t)}$ hold. Therefore there must be an m and m' such that $m \not\equiv m'(\pi_{h_M(t)})$ and $m \equiv m'(\pi_{s_i})$.

Q.E.D.

We will now give two examples to demonstrate the use of these theorems.

Example 5.2

Let the automaton and semigroup be as in Example 4.1. Then the ordering on S^1 is as follows:



Thus, this automaton has no errors which are finite but not correctable since there is only one minimal class, $\{s_7, s_7, s_{13}\}$.

Example 5.3

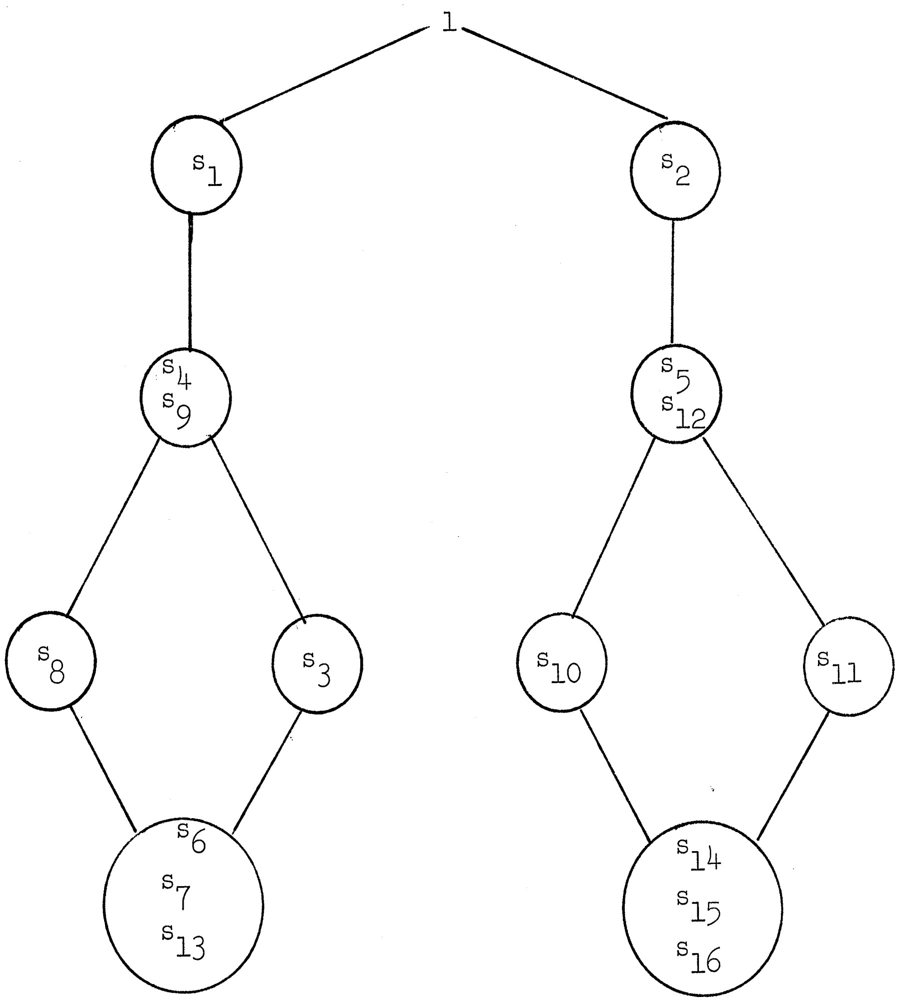
Let M = ({a,b,c,d,e,f,g}, {0,1}, δ)

δ	0	1
a	b	c
b	c	d
c	d	b
d	d	b
e	g	f
f	f	f
g	g	g

The transformation associated with the elements of the semigroup of M, S_M, are shown in the table below.

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	s_{16}
a	b	c	c	d	b	d	b	d	b	d	c	d	c	d	b	c
b	c	d	c	b	b	d	b	c	d	c	d	b	c	d	b	c
c	d	b	c	b	d	d	b	c	d	c	d	b	c	d	b	c
d	d	b	c	b	d	d	b	c	d	c	d	b	c	d	b	c
e	g	f	g	g	f	f	f	g	g	f	f	f	f	g	g	g
f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f
g	g	g	g	g	g	g	g	g	g	g	g	g	g	g	g	g

The ordering on S^1 is as follows:



We can see that there are two minimal classes $\{s_6, s_7, s_{13}\}$ and $\{s_{14}, s_{15}, s_{16}\}$. Thus the automaton must have errors which are correctable but are not finite. In fact $N = \{(e,g),(f,g)\}$. Notice that every tape which corrects (e,g) does not correct (f,g) and vice-versa as required by Theorem 5.3.

6. Errors in Series-Parallel Connections

In this section we will deal with the problem of errors in a series-parallel connection of finite-automata. This problem arises in two ways. The first is, if we are given a series-parallel connection finite automata with known error properties, what can be said about the error properties of this compound automata? The second is, if we are given a large complicated automaton, is it easier to find its error properties (or some bounds on it) by splitting the original automaton into smaller automata, find all their individual error properties, and then recombine them as above?

The parallel connection, of course, does not introduce any new problems. If we have a parallel connection of n automata, then the composite automaton is just the cartesian product of them and if the finite error partition for the i-th automaton is π_{M_i}, then the finite error partition for the composite automaton is just $\pi_{M_1} \times \pi_{M_2} \times \cdots \times \pi_{M_n}$.

Unfortunately, there is no easy solution for the series connection. We could consider the series connection problem by using the fact that the semigroup of the series connection of two automata is a subsemigroup of the wreath product of the semigroups of the two automata. In the author's attempt this led to some very complicated notation but not to any fruitful results. Hence we will go about it in another way. Now let us define a few more terms in order to be able to discuss the problem more readily.

Definition 6.1

A finite transducer M' is a 4-tuple $M' = (M, \lambda, \Sigma', m_0)$ where M is a finite automaton Σ' in a finite set (output alphabet);

λ: $M \times \Sigma \to \Sigma'$ (output function);

and $m \in M$ (initial state). We will extend λ in the natural manner to be a mapping from input strings to output strings, λ: $M \times \Sigma^* \to \Sigma'^*$. As before, we will be informal and use the same letter to denote the finite transducer, the finite automaton, and the set of states of the automaton.

We will use superscripts, viz. M^{m_i}, in order to differentiate amongst a set of transducers each of which has the same finite automaton and output function but which have different m_i's as their initial states.

A transducer M then induces a length preserving mapping from Σ^* into Σ'^*. If we drive the transducer with a random source with property P, we get a set of probability distributions $\{P_k(x)\}$ on the output strings. That is, if x is a tape of length k, over Σ',

$$P_k(x) = \sum_{v \in \Sigma^*} P_k'(v)$$

where

$$\lambda(m_0, v) = x$$

and where $P_k'(v)$ is the probability of the tape v being the output of the random source driving M, given it has generated k letters. We will call the output toegether with the induced set of probability distribution, an M-transduction of the source S, or for short, just an <u>M-transduction</u>. Keep in mind that S always has property P.

First, let us see what happens if, instead of driving the machine M_2 directly with the source, we drive it with an M-transduction of the source. Thus, the situation is as follows:

$$\text{source} \rightarrow \boxed{M_1} \rightarrow \boxed{M_2} .$$

Later, we will get to the more difficult situation where we consider specific M_1-transductions, but for now let us let M_1 range over all finite transducers and see what results we will get. We will always assume the output alphabet of M_1 is contained in the input alphabet of M_2.

<u>Theorem 6.1</u>

Let M_2 be a finite automaton. Then, all errors in M_2 are finite for any M_1-transduction of a source with property P if and only if M_2 is a definite automaton.

PROOF: If M_2 is definite then it is k-definite for some finite k. Hence all tapes of length k correct all errors in M_2 and consequently all errors in M_2 are finite. On the other hand, assume that M_2 is not definite. Then, by Lemma 4.4, there must be an error (m_i, m_j) in M_2 such that (m_i, m_j) is not definite. But this means that for all k there must be a tape t_k that does not correct (m_i, m_j). Now, since M_2 is finite, there must be tapes t and t_o such that all tapes of form $t_o t^n$ do not correct (m_i, m_j). Since the set, T, of prefixes of the set of tapes $\{t_o, t^n\}$ is a regular set, there is a finite transducer M_1 all of whose output sequences are in T. Hence (m_i, m_j) is not correctable under an M_1-transduction. Consequently, all of the errors in M_2 are not finite under an M_1-transduction.

Q.E.D.

This theorem has significant computational value. By using the techniques developed by Hartmanis and Stearn (ref. 5) we can pull off the maximal definite back machine. Then, in order to find the partition for the whole machine, we only have to find the partition for the front machine. We will also point out here the obvious result that if an error is non-correctable, it is non-correctable for any M_1-transduction. Hence, if a finite automaton has only non-correctable errors (viz. every input causes a permutation of the states), all its errors are correctable under any M_1-transduction. Thus, as a computational aid, we can also pull off automata of this type.

By using the same type of argument as in Theorem 6.1, we can get a slightly stronger result along the same line.

<u>Theorem 6.2</u>

Let E be an error in a finite automaton M_2. Then E is a correctable error for all M_1-transductions of the source if and only if E is a definite error.

PROOF: If E is a definite error, there is some k such that all tapes of length k correct E. But since every M_1-transduction is a length-preserving mapping, E remains definite for all of them. If E is not definite, then by using the same argument as was used in Theorem 6.1, to construct a transducer which does not correct (m_i, m_j), we can construct one which does not correct E. Hence E is not correctable under all M_1-transductions.

Q.E.D.

We will now give two examples to show the usefulness of these theorems.

Example 6.1

Let M = ({ge, gf, he, hf, ie, if, je, jf}, {0,1}, δ)

δ	0	1
ge	ge	gf
gf	hf	he
he	ie	if
hf	jf	je
ie	ge	gf
if	hf	he
je	ie	if
jf	jf	je

M can be decomposed into a series connection as follows.

$$M_1 = (\{e,f\}, \{0,1\}, \delta_1, \{0,1\}, \lambda, e)$$

δ_1	0	1
e	e	f
f	f	e

$$M_2 = (\{g,h,i,j\}, \{0,1\}, \delta_2)$$

δ_2	0	1
g	g	h
h	i	j
i	g	h
i	i	j

Since M_2 is a definite automaton we know that for m, π_F $\pi_F \geq \{\overline{eg,eh,ei,ej}, \overline{fg,fh,fi,fj}\}$. In fact, since (e,h) is a non-correctable error in M_1, we know that the equality holds.

Example 6.2

Let M_1 be as above and let $M_2 = (\{g,h,i\}, \{0,1\}, \delta_2)$.

δ	0	1
g	h	i
h	g	h
i	g	h

The error (h,i) is a definite error. Hence if M is the series connection for M, $\pi_F \geq \{\overline{eh, ei}, \overline{fh, fi}, \overline{eg}, \overline{fg}\}$.

Let us now attempt to apply some of the results of the earlier sections to this part. It is easy to see that if $\{P_n\}$ induced by an M-transduction is statistically indistinguishable from a set $\{P'_n\}$ induced by some source with property P over an alphabet Σ' then, for any automaton $M_2 = (M_2, \Sigma', \delta)$, an error in M_2 is correctable (or finite) under the M-transduction if and only if it is correctable (or finite). The next lemma shows a necessary and sufficient condition for this in each of two special cases. We will use $|\Sigma|$ to indicate the cardinality of the set Σ.

Lemma 6.1

Let $M = (M, \Sigma, \delta, \lambda, \Sigma', m_0)$, and let $\{P'_n\}$ be the set of probability distributions induced by it when driven by a source S' with property P.

(1) If $|\Sigma| < |\Sigma'|$ then the M-transduction is distinguishable from every source with property P over Σ'.

(2) If $|\Sigma| = |\Sigma'|$ then the M-transduction is indistinguishable from some source with property P over the alphabet Σ' if and only if for all states m_i of M such that for some tape x_i

$$\delta(m_0, x_i) = m_i,$$

for every $\sigma' \epsilon \Sigma'$ there exists an input $\sigma \epsilon \Sigma$ so that

$$\lambda(m_i, \sigma) = \sigma'.$$

PROOF: (1) If $|\Sigma| < |\Sigma'|$ then $|\{\sigma' \epsilon \Sigma |$ for some $\delta \epsilon \Sigma$ $(m_0, \sigma) = \sigma'\}| \leq |\Sigma| < |\Sigma'|$. Hence there is some output symbol which cannot occur at the beginning of an output string. Hence the M-transduction must be distinguishable from every source with property P over Σ'.

(2) If $|\Sigma| = |\Sigma'|$ then for all k, $|\Sigma^k| = |\Sigma'^k|$. Hence, if the output is indistinguishable from a source with property k over Σ', M must induce a one to one mapping Σ^k onto Σ'^k since we know the mapping is length preserving. Therefore, for every output string, a unique input string is determined and thus a unique state of M is determined. But, since every output string can occur, we get that for every state m, reachable from the initial state, and for every output symbol

$\sigma'\epsilon\Sigma'$ there is an input symbol $\sigma\epsilon\Sigma$. $\lambda(m,\sigma) = \sigma'$. Conversely, if for every m and σ' there is a σ such that $\lambda(m,\sigma) = \sigma'$, the output must have property P since the symbol σ has a probability of occurring greater than k which is strictly greater than zero. Hence σ' has a probability of occurring strictly greater than zero. Thus the output is indistinguishable from a source with property P.

Q.E.D.

In sections four and five we derived various results pertaining to the connection between the error properties of the automaton and its semigroup. It seems reasonable to expect that we would be able to get analogous results for the error properties of the automaton under an M-transduction. Thus, we would hope to get a result like the following analogous to Theorem 4.1. As before, let $h_M(x)$ be the element of the semigroup S_M associated with the tape x.

Conjecture 6.1

Let $\{P_n\}$ be a set of probability distribution induced by an M-transduction over Σ and let

$$V = \{x\epsilon\Sigma^* \mid P_{lg(x)}x > 0\}.$$

Then an error E in M_2 is finite if and only if it is corrected by every minimum idempotent of $h_{M_2}(V)$.

This is not a valid conjecture as can be seen by the following counterexample.

Counterexample 6.1

Let $M = (\{a,b,c,d\}, \{0,1\}, \delta, \lambda, \{0,1\}, a)$

δ	0	1
a	b	c
b	b	b
c	c	d
d	d	c

λ	0	1
a	0	0
b	0	0
c	0	0
d	1	1

Let $M_2 = ((f,g), (0,1), \delta)$

δ	0	1
f	g	f
g	f	f

$S_{M_2} = (\{s_1,s_2,s_3,s_4\},\cdot)$.

·	s_1	s_2	s_3	s_4
s_1	s_1	s_2	s_3	s_4
s_2	s_2	s_1	s_3	s_4
s_3	s_3	s_4	s_3	s_4
s_4	s_4	s_3	s_3	s_4

The mappings associated with S_{M_2} are as follows:

	s_1	s_2	s_3	s_4
f	f	g	f	g
g	g	f	f	g

And

x	$h_{M_2}(x)$
0	s_2
1	s_3

Now

$$h_M(V) = \{s_1,s_2,s_3,s_4\},$$

but the error (f,g) is correctable but not finite under an M-transduction. Hence the conjecture cannot be true.

However, if M is strongly connected we can get somewhat similar results. We will call an M-transduction a strongly connected M-transduction whenever M is strongly connected.

Theorem 6.3

Let $\{P_n\}$ be the set of probability distributions on $\{\Sigma^n\}$ induced by a strongly connected M_1-transduction, and let

$$V = \{x \mid P_{lg(x)}x > 0\}.$$

Then there is a subsemigroup W, of Σ^* such that $W \subseteq V$ and for any finite automaton M_2 and error E in M_2, the following are true:

(1) E is correctable under an M_1-transduction if and only if E is corrected by a minimum idempotent of $h_{M_2}(W)$.

(2) E is finite under an M_1-transduction if and only if E is corrected by every minimum idempotent of $h_{M_2}(W)$.

(3) M_2 has errors which are correctable but not finite under an M_1-transduction if and only if $h_{M_2}(W)$ has two minimal classes.

PROOF: Before we begin the proof let us just point out that all the above terms are well defined since $h_{M_2}(W)$ is a semigroup, which is a subsemigroup of S_{M_2}. This follows from the fact that h_{M_2} is a homomorphism of Σ^* onto S_{M_2}. Also note that parts one and two of the theorem are generalizations of Theorem 4.1 and part three is a generalization of

Theorem 5.2.

Now, let $M_1 = (M_1, \lambda, \Sigma', m_0)$, and let $W' = \{\lambda(m_0,x) \mid \delta(m_0,x) = m_0$ and $\delta(m_0,y \mid \neq m_0$ for all prefixes y of $x\}$. Thus, W' is the set of output strings which are generated by an input string which takes the state m_0 to itself for the first time. Now let W be the free semigroup generated by W'. Thus W is the set of output strings generated by a set of input strings which takes the state m_0 to itself.

Now, let us prove claim number one. If E is correctable, under an M_1-transduction it is corrected by some tape x. But since M_1 is strongly connected there is a y such that xy is in W. Now, the set of strings in the set xyW is a right ideal of W and hence $h_{M_2}(xyW)$ is a right ideal of $h_{M_2}(W)$ all of whose members correct E. But since $h_{M_2}(xy\ W'*)$ is a right ideal it must contain a minimum idempotent of $h_{M_2}(W)$. Conversely, since every minimum idempotent of $h_{M_2}(W)$ corresponds to at least one input tape of M_1, if E is corrected by a minimum idempotent of $h_{M_2}(W)$ it is correctable. Thus we have proved claim one.

If E is not corrected by some minimum idempotent of $h_{M_2}(W)$, it is not corrected by some right ideal $h_{M_2}(x(W')*) \subseteq h_{M_2}(W)$. But since M_1 is strongly connected, this means that all tapes in the set $x\Sigma* \cap V$ do not correct E. Hence, there is a set of input tapes of form $y\Sigma*$ such that the image of any tape in the set under an M_1-transduction does not correct E. Hence E cannot be finite. Conversely, let us assume that E is corrected by every minimum idempotent. Then E is corrected by all elements of minimum right ideals, and thus corrected by the kernel of $h_{M_2}(W)$. Thus there is an $s_i \epsilon h_{M_2}(W)$ such that for all s_j, $s_j s_i$ corrects E. But corresponding to this there is an element x of W such that

$h_{M_2}(x) = s_i$. Hence all tapes of form yxz correct E. We will show that if $\{P'_n\}$ is the set of probability distribution associated with a source with property P and if A is the set of tapes which is not of form yxz and M_1 maps a set B into A, then

$$\lim_{n \to \infty} P'_n(B) = 0.$$

Thus since B', theset of tapes whose transduction does not correct E, is contained in B, E is a finite error under an M_1-transduction. If j is the number of states in M_1, we know for every input tape v in B there is a tape u of length less than j such that vu leaves M_1 in state m_0. Then there is an input tape x whose output corrects E. Also, if the constant associated with the source is k, then

$$P'_{lg(vux)}vux \geq (k)^{lg(x)+j} P'_{lg\ v}v.$$

Hence, if $P'_n(B) = a$, then

$$P'_{n+j+lg(x)}(B) \leq a(1-k^{lg(x)+j}).$$

Thus, since $k > 0$, the

$$\lim_{n \to \infty} P'_n(B) = 0.$$

Hence we have proven part two.

The proof of part three is much easier. We can just note that $h_{M_2}(W)$ is a finite semigroup (contained in S_{M_2}) and hence all the results of section five hold without modification.

Q.E.D.

Corollary 6.1

An error E is finite under a strongly connected M_1-transduction if and only if $\pi_E \leq \bigcap_{s_i} \pi_{s_i}$ where s_i is a minimum idempotent of $S_{M_2}(W)$.

PROOF: This follows immediately from part two of the above theorem.

We will call W the output semigroup associated with the strongly connected M_1-transduction. This output semigroup gives us a convenient way of talking about strongly connected transductions.

Lemma 6.2

If W_0 is the output semigroup associated with a strongly connected $M_1^{m_0}$-transduction and W_i is the output semigroup associated with the $M_1^{m_i}$-transduction, then there are tapes x and y such that $x\ W_0 y \subseteq W_i$ and $y\ W_i\ x \subseteq W_0$.

PROOF: Let x be an output string associated with an input string x' such that $\delta(m_i, x') = m_0$ and y be an output string associated with an input string y' such that $\delta(m_0, y') = m_i$. Since M_1 is strongly connected, such an x and y exist. Obviously, x and y have the required properties.

Q.E.D.

Putting together the results of Theorem 6.3 and Lemma 6.2 we get the following result.

Theorem 6.4

Let M_2 be a finite automaton all of whose errors are finite under a strongly connected $M_1^{m_0}$-transduction. Then all errors in M_2 are finite under an $M_1^{m_i}$-transduction for all states $m_i \in M_1$.

PROOF: Let W_0 be the semigroup associated with $M_1^{m_0}$-transduction and W_1 be the semigroup associated with a $M_1^{m_i}$-transduction. Now, since all errors in M_2 are finite under an $M_1^{m_0}$-transduction, by Theorem 6.3 there must be an $s_i \in h_{M_2}(W_0)$ which corrects all errors in M_2. Hence there must be a tape z in W_0 which corrects all errors. By Lemma 6.2 there is an x and y such that $x\, W_0\, y \subseteq W_1$. Hence $xzy \in W_1$ and $h_{M_2}(xzy) \in h_{M_2}(W)$. But, since z corrects all errors, xzy must correct all the errors in M_2. Also, there cannot be a minimum idempotent $s_j \in h_{M_2}(W_1)$ which does not correct all errors since if there were then $s_j \cdot h_{M_2}(xzy) \leq s_j$ would correct all errors. Hence $\pi_{s_j} h_{M_2}(xzy) \geq \pi_{s_j}$ and $\pi_{s_j} h_{M_2}(xzy) \neq \pi_{s_j}$. This contradicts the fact that s_j is a minimum idempotent.

Q.E.D.

Besides its theoretical value, this theorem has some nice computational properties. For instance, if we can factor a finite automaton into a series connection of two simple automata, and if for any state m_1 in a strongly connected component of M_1, all the errors in M_2 are finite, then all errors of form $((m_i,m_j),(m_i,m_j'))$ are finite where m_i is in the same strongly connected components of M_1 as M_2 where m_j and m_j' are any states in M_2.

We can note that the proof of Theorem 6.4 depends mainly on the fact that the set of strings which correct all errors form a two-sided ideal of Σ^* and hence the set of semigroup elements which correct all errors form a two-sided ideal of $h_{M_2}(W)$. Using the same type of argument we can get a stronger theoretical result albeit one which is not so useful.

Theorem 6.5

> Let ξ be a two-sided ideal of S_{M_2} all of whose elements correct a set of errors A in M_2. Also, let W_0 be the semigroup associated with the strongly connected $M_1^{m_0}$-transduction. Then, if $h_{M_2}(W_0) \cap \xi$ is not empty, then all errors in the set A are finite for an $M_1^{m_i}$-transduction for all $m_i \in M_1$.

PROOF: If $h_{M_2}(W) \cap \xi$ is not empty then clearly all errors are finite under an $M_1^{m_0}$-transduction. This is so since if there were a minimum idempotent s_i of S_{M_2} which did not correct an error in A, then there would be a right ideal $s_i \cdot S_{M_2}$ which did not correct A. But $s_i \cdot S_{M_2} \supseteq s_i \xi$ and $s_i \xi \subseteq \xi$ since ξ is an ideal. Hence $\xi \cap s_i S_{M_2}$ is not empty. But this is a contradiction since this intersection must correct all of A since it is contained in $s_i S_{M_2}$. Hence, by Theorem 6.3, all errors in A are finite errors under an $M_1^{m_0}$- transduction. Now let W_i be the semigroup associated with an $M_1^{m_i}$-transduction, and x and y be tapes such that $x\, W_i\, y \subseteq W_0$ and $y\, W_0\, x \subseteq W_i$. The existence of such tapes is guaranteed, by Lemma 6.1. If $h_{M_2}(W_0) \cap \xi = \xi'$ is not empty, then $h_{M_2}(y)\xi' \; h_{M_2}(x)$ is not empty. But $h_{M_2}(y)\xi' \; h_{M_2}(x) = h_{M_2}(y)(H_{M_2}(W_0) \cap \xi)h_{M_2}(x) =$ $h_{M_2}(y)\; h_{M_2}(W_0)\; h_{M_2}(x) \cap h_{M_2}(y)\; \xi\; h_{M_2}(x) \;=\; h_{M_2}(y\, W_0\, x) \cap \xi'' \neq \emptyset$. But $y\, W_0\, x \subseteq W_1$ and $\xi'' \subseteq \xi$ since ξ is a two-sided ideal. Hence $h_{M_2}(W_1) \cap \xi$

is not empty. Thus by reapplying the first part of the theorem we have shown that all errors in A are finite under an $M_1^{m_i}$-transduction.

Q.E.D.

Although this theorem may be difficult to use for computational purposes, a corollary of it is well suited for this purpose.

Corollary 6.2

Let W_0 be the semigroup associated with a strongly connected $M_1^{m_0}$-transduction and M_2 a finite automata. Then, if the intersection of W_0 with the kernel of S_{M_2} is not empty, all finite errors of M_2 are finite under any $M_1^{m_i}$-transduction.

PROOF: If we let A be the set of finite errors we know from Theorem 4.1 that all the elements of the kernel of S_{M_2} correct all the elements of A. Hence the corollary follows immediately.

Q.E.D.

Example 6.3

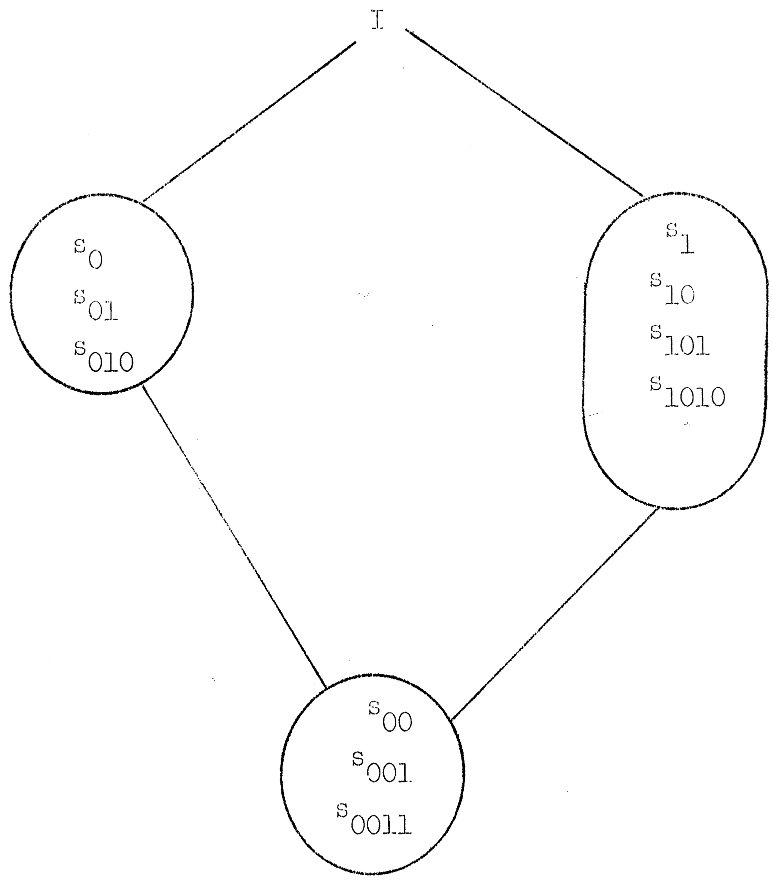
Let M_2 = ({a,b,c}, {0,1}, δ)

δ	0	1
a	b	a
b	c	b
c	c	a

Then the mappings of the elements of S_{M_2} are as shown below:

	s_0	s_1	s_{00}	s_{01}	s_{10}	s_{001}	s_{011}	s_{101}	s_{0011}	s_{0101}	s_{1010}
a	b	a	c	b	b	a	c	b	b	a	c
b	c	b	c	a	c	a	b	a	b	b	b
c	c	a	c	a	b	a	b	b	b	b	c

The ordering is as follows:



Now let $M_1 = (\{d,e,f\}, \{0,1\}, \delta, \lambda, \{0,1\}, d)$

δ	0	1
d	e	e
e	f	d
f	d	d

λ	0	1
d	1	1
e	0	0
f	0	0

W, the associated semigroup, is the free semigroup on the set of generators $V = \{100, 10\}$. Hence $h_{M_2}(W) \supseteq \{s_\infty\}$, and therefore the intersection of $h_{M_2}(W)$ and the kernel of S_{M_2} is not empty. Thus since all errors in M_2 are finite for a source with property P, by Corollary 6.2, they are all finite under any $M_1^{m_i}$-transduction.

On the other hand, assume $M_1 = (\{d,e,f\}, \{0,1\}, \delta, \lambda, \{0,1\}, d)$

δ	0	1
d	d	e
e	f	f
f	d	d

λ	0	1
d	1	1
e	1	1
f	0	0

Then W is the free-semigroup on the set of generators V = {110, 1}. Thus the only idempotent in $h_{M_2}(W)$ is s_1. Hence by Theorem 6.3 (a,c) is the only correctable and only finite error in M_2 under this M_1-transduction.

In the third case assume M_1 = ({d,e}, {0,1}, δ, λ, {0,1}, d).

δ	0	1
d	d	e
e	d	d

λ	0	1
d	1	0
e	1	1

Then W is the free semigroup on the set of generators V = {1,01}. In this case s_1 and s_{0101} are both minimum idempotents of $h_{M_2}(W)$. Hence by Theorem 6.3 we get that although both the errors (a,c) and (b,c) are correctable, the automaton M_2 has no errors which are funite under the M_1-transduction.

Thus in the case where we are dealing with strongly connected M-transductions, there are some reasonable tools. Let us now somewhat relax the strongly connected restriction and see what we get.

Definition 6.2

An M-transduction, M = (M, Σ, δ, λ, Σ', m_0), is called a connected M-transduction if and only if there is a state $m_1 \in M$ such that for all states $m_j \in M$ there is a tape x_j such that $\delta(m_j, x_j) = m_1$. States reachable from all states which they can reach, (as m_1 is) will be called stable states. All others will be called transient states.

Note that if we let M' be the set of states reachable from m_1, and let δ' and λ' be equal to δ and λ when restricted to the set of states, M', respectively, then the M'-transduction $(M', \Sigma, \delta', \lambda', \Sigma', m_1)$ is strongly connected, and hence, the M-transduction $(M, \Sigma, \delta, \lambda, \Sigma', m_1)$ is indistinguishable from a strongly connected M-transduction.

Making use of the results in the preceding part we get the following.

Theorem 6.6

Let an $M_1^{m_0}$-transduction be connected and M_2 be a finite automaton. Then an error (m_i, m_j) in M_2 is finite under an $M_1^{m_0}$-transduction if and only if for all strings x, which is the output of M_1 induced by an input string which takes M_1 from state m_0 to a stable state m_1 in M_2 for the first time, the error $(\delta(m_i,x), \delta(m_j,x))$ is finite under the $M_1^{m_1}$-transduction.

PROOF: It is clear that if for some string x as above, the error $(\delta(m_i,x), \delta(m_j,x))$ were not finite under the $M_1^{m_1}$-transduction, then (m_i, m_j) would not be finite. To show the converse all we have to show is that if $V = \{z \mid z=uv\ \delta(m_0,u) = m_1\ v\epsilon\Sigma^*\}$ and if $\{P_n\}$ is associated with a source with property P, then

$$\lim_{n \to \infty} P_n(V) = 1.$$

But this is clearly so since there is a tape u' such that for all m_i there is a prefix of u', u" such that $\delta(m_i, u") = m_1$. Thus $V \supseteq \Sigma^* u" \Sigma^*$, and

$$\lim_{n \to \infty} P_n(\Sigma^* u" \Sigma^*) = 1.$$

Thus, "almost all" the long tapes take m_0 to m_1 from where almost all the long tapes correct the error of the form $(\delta(m_i,x), \delta(m_j,x))$.

Q.E.D.

From this we get the following corollary.

Corollary 6.3

Let an $M_1^{m_0}$-transduction be connected, and M_2 a finite automaton. Then, if all errors in M_2 are finite under an $M_1^{m_1}$-transduction for some stable state $m_1 \epsilon M$, all errors in M_2 are finite under all $M_1^{m_i}$-transductions for any $m_i \epsilon M_1$.

PROOF: If all errors are finite under an $M_1^{m_1}$-transduction then obviously, all errors of the form $(\delta(m_i,x), \delta(m_j,x))$ are finite for all (m_i,m_j) errors in M_2. Hence the corollary follows immediately from Theorem 6.6.

Q.E.D.

It would appear that all of the above results are only of limited use since they only consider the case of the connected or strongly connected M-transduction. However, as the next theorem will show, they are quite general.

Theorem 6.7

Let $M_0 = (M_0, \Sigma, \delta, \lambda, \Sigma', m_0)$ be an arbitrary M-transduction. Then there exists a finite number of connected M-transductions $M_1, M_2, \ldots, M_k$ with the following property. For any finite automaton M_{k+1}, an error E in M_{k+1} is

(1) correctable with respect to an M_0-transduction if and only if it is correctable with respect to some M_i-transduction for $1 \leq i \leq k$.

(2) finite with respect to an M_0-transduction if and only if it is finite with respect to all M_i-transductions $1 \leq i \leq k$.

PROOF: First let us define $[m_i] = \{m_j \mid \text{for some } x \epsilon \Sigma \;\; \delta(m_i, x) = m_j\}$, and $|m_i| = \{m_j \mid \text{for some } x \epsilon \Sigma^* \; \delta(m_j, x) = m_i\}$. Now let $\{m_{i_1}, m_{i_2}, \ldots, m_{i_k}\}$ be a maximal set of stable states of M_0 such that the set $\{[m_{i_1}], \ldots, [m_{i_k}]\}$ are pair-wise disjoint and if for some $m_j \epsilon M_0$, $[m_j] \cap [m_{i_\ell}] \neq 0$, then $[m_j] \supseteq [m_{i_\ell}]$. Now let $M_\ell = (|m_{i_\ell}|, \Sigma, \delta_\ell, \lambda, \Sigma', m_0)$ where $\delta_\ell = \delta$ restricted to $|m_{i_\ell}|$ if $\delta(m, \sigma_j) \in |m_{i_\ell}|$ otherwise $\delta_\ell(m, \sigma_j) = \delta(m_1, \sigma_i)$ where σ_i is any element of Σ such that for m, $\delta(m, \sigma_i) \in |m_i|$, and $\lambda_\ell(m, \sigma_j) = \lambda(m, \sigma_j)$ if $\delta(m, \sigma_j) \epsilon M_\ell$ and otherwise $\lambda_\ell(m, \sigma_j) = \lambda(m, \sigma_i)$ where σ_i is as above.

It is easy to see that x is an output tape of some M_i, $1 \leq i \leq k$, if and only if x is an output tape of M_0. This follows immediately from the definition of the M_i. Hence part one of the theorem is trivially true.

Now let us assume that E is not finite with respect to some M_i-transduction $1 \leq i \leq k$. Then, by Theorem 6.6, there is an input tape x to M_i which takes m_0 to a state stable state m_i with the property that if x' is the output tape associated with input x and if y is an output tape obtainable starting in state $m_i \in M_i$ then x'y does not correct E. But if this is true then there is an input tape x" which takes M_0 to state $m_i \in M_0$ and which has x' as an output. But then all outputs which result from the set of inputs x"Σ^* do not correct E. Hence E is not finite with respect to an M_0-transduction. Conversely, let us assume E is not finite with respect to the M_0-transduction. Then there is an input x" such that all outputs of M_0 resulting from the inputs x"Σ^* do not correct E. Also, let x be the output of M_0 resulting from the input x". But from the definition of the set $\{M_i\}$ there must be an M_i-transduction with x" as an output. Hence, E is not finite with respect to that M_i. Thus the theorem is proved.

Q.E.D.

If we apply this theorem to the automaton in counterexample 6.1 we see that M decomposes into two transducers as follows.

$$M_\alpha = (\{a,b\}, \{0,1\}, \delta_\alpha, \lambda, \{0,1\}, a)$$

$$M_\beta = (\{a,c,d\}, \{0,1\}, \delta_\beta, \lambda, \{0,1\}, a)$$

δ_α	0	1
a	b	b
b	b	b

λ_α	0	1
a	0	0
b	0	0

δ_β	0	1
a	c	c
c	c	d
d	d	c

λ_β	0	1
a	0	0
c	0	0
d	1	1

The error (f,g) in M_2 is finite under the M_β-transduction but not under the M_α-transduction. Thus it is correctable but not finite under the M-transduction.

In the preceding part of this section we have only been considering half of the problem. We have been considering automata which decompose as follows:

and we have just considered errors which occur in M_2 (the "back" automaton). Let us now consider errors which occur in M_1 (the "front" automaton).

Definition 6.3

Let M be a finite automaton which can be decomposed into a series connection of a finite transducer $M_1 = (m_1, \Sigma, \delta_1, \lambda, \Sigma', m_0)$, and a finite automaton $M_2 = (M_2, \Sigma', \delta_2)$. Then an error $E = (m_i, m_j)$ in M_1 is m_i-correctable, $m_1 \in M_2$ if and only if there is an x such that $\delta_1(m_i, x) = \delta_2(m_j, x)$ and if $y = \lambda(m_i, x)$ and $z = \lambda(m_j, x)$, then $\delta_2(m_1, y) = \delta_2(m_1, z)$. Likewise, it is m_1-finite if

$$\lim_{n \to \infty} P_n(V) = 1$$

where V is the set of tapes in Σ^* which have the same properties as the tape x has above.

The next few results will relate this case to the case that was previously studied.

Lemma 6.3

Let M_1 be a finite transducer, M_2 a finite automaton, and m_1 a state of M_2. Then an error E in M_1 is m_1-finite if and only if

(1) E is finite; and

(2) if x is such that $\delta_1(m_i,x) = \delta_1(m_j,x) = m_k$ and y and z are the respective outputs of the $M_1^{m_i}$- and $M_1^{m_j}$-transducers then the error $(\delta_2(m_1,y), \delta_2(m_1,z))$ is finite under the $M_1^{m_k}$-transduction.

The proof of this lemma is obvious and follows directly from the definition. Hence, it will not be given here. Note that, in general, condition two is very difficult to check for. However, in some common cases it is very easy. Thus, for instance, if M_2 is a definite automaton, or if M_1 is a connected transduction which corrects all errors in M_2, then condition two is vacuously satisfied.

Let us now give a corollary which will give a necessary and sufficient condition for all errors in the series connection to be finite.

Corollary 6.4

Let M be a finite automaton which can be decomposed into a series connection of an M_1-transduction and a finite automaton M_2. Then all the errors in M are finite if and only if all the errors in M_1 are finite and there is a stable state m_1 of M_1 such that all errors in M_2 are finite under the $M_1^{m_1}$-transduction.

PROOF: First, if all the errors in M_1 are finite then M_1 must be a connected transducer. Hence, by Corollary 6.3, since all the errors in M_2 are finite under an $M_1^{m_1}$-transduction, they are finite under all $M_1^{m_i}$-transductions, $m_i \epsilon M_1$. Thus if we represent the states as pairs (m_i,m_j) where $m_i \epsilon M_1$ and $m_j \epsilon M_2$, we know that all errors of the form $((m_i,m_j)(m_i',m_j))$ are finite and likewise errors of the form $((m_i,m_j)(m_i,m_j'))$ are finite. Hence, since we know the relation of finiteness is transitive, it follows that all errors in M are finite. Conversely, if either M_1 has errors which are not finite or if for some state, m_i of M_1, there is an error of M_2 which is not finite under an $M_1^{m_i}$-transduction, then it is obvious that the automaton M has errors which are not finite.

Q.E.D.

Note that the second part of the condition can be checked for using the results of Theorems 6.3 and 6.4.

Before we close this chapter we will give one more example which will illustrate the use of many of the results in this chapter.

<u>Example 6.4</u>

Let M be the finite automaton whose move function δ is tabulated below.

δ	0	1
da	ha	ea
db	hb	eb
dc	ha	ea
ea	gb	fb
eb	gc	fc
ec	gc	fc
fa	eb	gb
fb	ec	gc
fc	ec	gc
ga	fa	ea
gb	fb	eb
gc	fa	ea
ha	jb	jb
hb	jc	jc
hc	jc	jc
ia	ja	ha
ib	jb	hb
ic	ja	ha
ja	ia	ha
jb	ib	hb
jc	ia	ha

It appears at first that the calculation of the finite error partition for M would be a very difficult task. However, as we will show below, once we notice that M can be decomposed into a series connection of M_1 and M_2 as below, the calculation will be much easier.

$$M_1 = (\{d,e,f,g,h,i,j\}, \{0,1\}, \delta_1, \lambda, \{0,1\}, x)$$

δ_1	0	1
d	h	e
e	g	f
f	e	g
g	f	e
h	j	j
i	j	h
j	i	h

λ	0	1
d	1	1
e	0	0
f	0	0
g	1	1
h	0	0
i	1	1
j	1	1

$$M_2 = (\{a,b,c\}, \{0,1\}, \delta_2)$$

δ_2	0	1
a	b	a
b	c	b
c	c	a

First let us note that we can calculate the finite error partitions for M_1 and M_2 by observation:

for M_1 $\quad \pi_F = \{\overline{a,b,c}\}$

and

for M_2 $\quad \pi_F = \{\bar{d},\bar{e},\bar{f},\bar{g},\overline{h,i,j}\}$.

Hence for M, $\pi_F \leq \{\overline{da,\ db,\ dc},\ \overline{ea,\ eb,\ ec},\ \overline{fa,\ fb,\ fc},\ \overline{ga,\ gb,\ gc},$

$\overline{ha,\ hb,\ hc,\ ia,\ ib,\ ic,\ ja,\ jb,\ jc}\}$.

Now since 001 is an element of the semigroup associated with the strongly connected transduction M_1^e, and $h_{M_2}(001)$ is an element of the kernel of S_{M_2} (See Example 6.4), by Theorem 6.3 we get

$$\pi_F \geq \pi_1 = \{\overline{ea,\ eb,\ ec},\ \overline{fa,\ fb,\ fc},\ \overline{ga,\ gb,\ gc}\}.$$

Likewise, $1 \in W$ of M_1^h and no string starting with 0 is in it. Hence

$$\pi_F \geq \pi_2 = \{\overline{ha,\ hc}\}.$$

By similar agreement we get that the errors (hb, ha), (hb, hc), and all errors of the form (jx,jy) and (ix,iy) are not finite errors.

Next, as per Theorem 6.7, instead of considering the M_1-transduction, we can consider the connected transduction M_α and M_β with moves and outputs as below.

δ_α	0	1
d	e	e
e	g	f
f	e	g
g	f	e

λ_α	0	1
d	1	1
e	0	0
f	0	0
g	1	1

δ_β	0	1
d	h	h
h	j	j
i	j	h
j	i	h

λ_β	0	1
d	1	1
h	0	0
i	1	1
j	1	1

We see that although all errors in M_2 are finite under the M_α^d-transduction, only the error (a,c) if finite under the M_β^d-transduction. Hence we get

$$\pi_F \geq \pi_3 = \{\overline{da,dc}\}$$

and that (da,db) and (db,dc) are not finite errors. Now applying the results of Lemma 6.2 we get that

$$\pi_F \geq \pi_4 = \{\overline{ia,ja},\ \overline{ib,jb},\ \overline{ic,jc}\}.$$

Likewise, since a 0 input corrects (h,i) and causes errors which are not finite regardless of the state of M_2, we get that all errors of form (hx,ix) are not finite. Similarly, 01 corrects (h,j) but causes errors which are not finite, all errors of form (hx,jx) are not finite.

Now putting this all together we get

$$\pi_F \geq \bigcup_{i=1}^{4} \pi_i = \pi_5 = \{\overline{ea,eb,ec},\ \overline{fa,fb,fc},\ \overline{ga,gb,gc},\ \overline{ha,hc},$$
$$\overline{da,dc},\ \overline{ia,ja},\ \overline{ib,jb},\ \overline{ic,jc},\ \overline{hb,db}\}.$$

Using all the errors that we know cannot be finite we find that all the finite error classes of π_5 are indeed maximal classes and that hence

$$\pi_F = \pi_5.$$

7. Conclusion

In the beginning we set out to classify and to study state errors in an automata driven by a random source. We feel that by now the reader should have a reasonable feeling of the types of state errors and their properties in the finite state case. One area for further study is that of state errors in various classes of restricted infinite automata. It is our conjecture that in most of the commonly studied classes of automata, many of the interesting questions will be recursively unvolvable. However, this may not be the situation, and in any event, the area deserves further study.

Input errors and output errors are two other types of errors that we might have looked at. Input errors can intuitively be thought of as a state error which arises due to a substitution of one input string for another. Thus they are just special types of state errors. However, they also give rise to certain structures on the set of input strings. This has been studied more extensively by Winograd (ref. 15). We can also rewrite many of the results in the last section so that they tell us about input errors.

In order to consider output errors we must consider automata with output functions. An output error is then an incorrect output due to the finite automaton being in the incorrect state. If we are considering reduced finite automata (i.e., any two states are distinguishable) (ref. 11), then for any state error, there is at least one input string which causes an output error. If the state error is not correctable then in an intuitive sense, for most of the infinite tapes, the number of output errors is unbounded. Similarly, if the state error is finite, then for most of the infinite tapes, the number of output errors is finite. If we are dealing with a finite automaton which is not reduced, we can get the same results by consdering the state errors in the derived reduced machines. Thus it can be seen that the study of state errors reported here is a prerequisite to a study of output errors. However, more work along this line can certainly be done.

Appendix

Some Remarks on the ∞ Case

One obvious way to generalize the previous results is to alter Definition 2.1 so that M, the set of states, need not be finite. We have not looked into this area extensively but will give some examples to indicate that the results do not carry over. In this appendix, "automaton" will indicate a system of the type defined in Definition 2.1 without the requirement that the set of states of M be finite.

Definition A.1

An error $E = (m_1, m_2)$ will be called a G-error if E is correctable and, for all tapes t, $(\delta(m_1,t), \delta(m_2,t))$ is correctable.

We will state without proof that the G-errors partition the set of states and that G is the largest partition with the substitution property continued in C the correctable error relation. The proofs are analogous to the proofs of the corresponding theorems in the finite case (Theorem 3.3 and Corollary 3.1). In the case of finite automatons, an error is finite if and only if it is a G-error. However, as Example A.1 will show, this is not so for arbitrary automata.

Example A.1

$M = (I \times I \times I, \{0,1,2,0',1',2'\}, \delta)$ where I is the set of integers, and δ is the move function described below.

For all $(i,j,k) \neq (0,0,0)$

$$\delta[(i,j,k),\ 0] \quad = \quad (i+1,j,k)$$

$$\delta[(i,j,k),0'] \quad = \quad (i-1,j,k)$$

$$\delta[(i,j,k),\ 1] \quad = \quad (i,j+1,k)$$

$$\delta[(i,j,k),1'] \quad = \quad (i,j-1,k)$$

$$\delta[(i,j,k),\ 2] \quad = \quad (i,j,k+1)$$

$$\delta[(i,j,k),2'] \quad = \quad (i,j,k-1)$$

and

$$\delta[(0,0,0),\ u] \quad = \quad (0,0,0)$$

where u is any input.

It is obvious that any error

$$E \quad = \quad [(i,j,k),\ (l,m,n)]$$

is correctable since there is a tape that takes it to $[(0,0,0),(0,0,0)]$. However, if S is a source which generates each symbol independently with probability 1/6, then there is a probability greater than zero that (i,j,k) never goes to $(0,0,0)$. (See Spitzer, ref. 13). Thus in the limit, there is a probability greater than zero that E is not corrected; and hence it is not finite. Thus, in the machine M, all errors are correctable; thus all errors are G-errors. However, there are no finite errors.

In this example, it is still true that the finiteness of an error does not depend upon the source, as long as it is the property P. That there are some automata where it does can be seen by the following example.

Example A.2

$$A = (I, \{0,1\}, \delta)$$

where I is the set of integers. For $i \neq 0$

$$\delta(i,0) = i+1$$

$$\delta(i,1) = i-1$$

For i=0, $\delta(0,0) = \delta(0,1) = 0.$

Let S be the source that generates x with probability p and x' with probability (1-p) independent of its past history.

Then, all errors are G-errors and if $p = 1/2$ all errors are finite. But, if $p > 1/2$ only errors of the form (i,j) $i \leq 0$, $j \leq 0$ are finite, and if $p < 1/2$ only errors of the form (k,l) $k \geq 0$, $l \geq 0$ are finite. This follows from the fact that each state is undergoing a one-dimensional random walk with an absorbing barrier at zero. The one-dimensional symmetric random walk ($p=1/2$) is recurrent whereas the other ones ($p \neq 1/2$) are not. Hence we get the above results. Thus we can see that the finiteness of an error depends upon the source.

References

1. A. H. Clifford and D. O. Miller, "Semigroups having zeroid elements," Amer. J. Math., 70, 117-125 (1948).

2. A. H. Clifford and G. B. Preston, The Algebraic Theory of Semigroups, Vol. 7, American Mathematical Society Survey (1962).

3. E. Gilbert and E. Moore, "Variable-length binary encodings," Bell Systems Tech. J., 38, 933-967 (1959).

4. J. Hartmanis, "Loop-free structure of sequential machines," Information and Control, 5, 25-43 (March 1962).

5. J. Hartmanis and R. E. Stearns, "A study of feedback and errors in sequential machines," IEEE Trans. on Electr. Computers, EC-12, 223-232 (June 1963).

6. J. Hartmanis and R. E. Stearns, "Pair algebra and its application to automata theory," Information and Control, 7, (December 1964).

7. K. B. Krohn and J. L. Rhodes, "Algebraic theory of machines," Mathematical Theory of Automata, Polytechnic Press, Brooklyn, New York (1963).

8. R. Laing and J. Wright, Commutative Machines, Technical Note 04422, 03105-24-T, Department of Philosophy, The University of Michigan, Ann Arbor (1962).

9. Yu. T. Medvedev, "On the class of events representable in a finite automaton," Sequential Machines: Selected Papers (edited by E. F. Moore), Addison-Wesley, Reading, Massachusetts (1964). Originally, in Automaty, Moscow, Izdatel'stva Inostrannoi Literaturg (1956), pp. 385-401.

10. E. H. Moore, "A definition of abstract groups, Trans. Amer. Math. Society, 3, 485-492 (1902).

11. E. F. Moore, "Gedanken-Experiments on sequential machines," Automata Studies (edited by C. Shannon and J. McCarthy) Princeton, New Jersey, 129-153 (1956).

12. M. Perles, M. D. Rabin, and E. Shamir, "The theory of definite automata," IEEE Trans. on Electr. Computers, EC-12, 233-242 (June 1963).

13. F. Spitzer, Principles of Random Walks, Van Nostrand, Princeton, New Jersey (1964).

14. J. von Newmann, "Probabilistic logics and the synthesis of reliable organisms from unreliable components," Automata Studies (edited by C. Shannon and J. McCarthy,) Princeton University Press, Princeton, New Jersey, 43-98 (1956).

15. S. Winograd, "Input error limiting automata," J. Assoc. Computing Mach., 11, 338-351 (July 1964).