This is a preprint of a paper presented at the NATO Advanced Study Institute on New Directions in Signal Processing in Communications and Control, Darlington, England, August 5-17, 1974, that will appear in the Institute Proceedings to be published by the Noordhoff International Publishing Company of Holland.

MARKOV INFORMATION SOURCES*

James L. Massey

Freimann Professor of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana, U.S.A.

ABSTRACT

A regular Markov source is defined as the output of a deterministic, but noisy, channel driven by the state sequence of a regular finite-state Markov chain. The rate of such a source is the per letter uncertainty of its digits. The well-known result that the rate of a unifilar regular Markov source is easily calculable is demonstrated, where unifilarity means that the present state of the Markov chain and the next output of the deterministic channel uniquely determine the next state. At present, there is no known method to calculate the rate of a nonunifilar source. Two tentative approaches to this unsolved problem are given, namely source identical twins and the master-slave source, which appear to shed some light on the question of rate calculation for a nonunifilar source.

LIST OF SYMBOLS

f	input-output function for a deterministic channel
$H(X)$	uncertainty of the random variable X (bits/symbol)
X_i	i-th state in a state sequence for a Markov chain
Y_i	i-th output digit from an information source

* This research was supported by the U.S. National Aeronautics and Space Administration under NASA Grant NGL 15-004-026 at the University of Notre Dame in liaison with the Communications and Navigation Division of the Goddard Space Flight Center

This is a preprint of a paper presented at the NATO Advanced study Institute on New Directions in Signal Processing in Communications and Control, Darlington, England, August 5-17, 1974, that will appear in the Institute Proceedings to be published by the Noordhoff International Publishing Company of Holland.

1. INTRODUCTION

Our aims in this lecture are to show how interesting models of information sources can be built from Markov chains, to review what is known about the information rate of such "Markov sources", and finally to indicate two approaches that appear to us to be promising lines of enquiry for extending the present theory to cover more practical situations. In this section we give the definitions necessary for stating in the following section the known results on the rate of "Markov sources." In the third and final section, we indicate two tentative approaches to rate calculation for Markov sources and we state a conjecture whose truth would solve the rate calculation problem.

An A-ary discrete information source is a device which emits a sequence $Y_1, Y_2, Y_3, \ldots$ of random variables each taking values in the finite set A. Such a source is specified by stating, for every positive integer n, the joint frequency distribution $P(Y_1, Y_2, \ldots Y_n)$. Such a definition of an "information source" is a direct reflection of the basic information-theoretic viewpoint that a device produces "information" only when there is "uncertainty" as to what its output will be. Thus, information sources must be described probabilistically.

We shall find it convenient to write $Y_{[u,v]}$ for the subsequence $Y_u, Y_{u+1}, \ldots Y_v$ of the output sequence of a source and to write $Y_{[u,v)}$ for $Y_{[u,v-1]}$. In keeping with this notation we shall, for instance, write the joint frequency distribution $P(Y_1, Y_2, \ldots Y_n)$ as $P(Y_{[1,n]})$ and write the conditional frequency distribution $P(Y_n | Y_1, Y_2, \ldots Y_{n-1})$ as $P(Y_n | Y_{[1,n)})$.

According to information theory (the reader is referred to Gallager[1] for further details of all information-theoretic concepts and results used in this lecture), the uncertainty, or entropy, of the sequence $Y_{[1,n]}$ is the expectation of the negative logarithm of its frequency distribution $P(Y_{[1,n]})$; that is,

$$H(Y_{[1,n]}) = E\{- \log P(Y_{[1,n]})\} \tag{1}$$

and has the units of bits when, as we shall assume hereafter, the logarithm is taken to the base 2. Similarly, the conditional uncertainty of Y_n given $Y_{[1,n)}$ is defined to be the quantity

$$H(Y_n | Y_{[1,n)}) = E\{- \log P(Y_n | Y_{[1,n)})\}. \tag{2}$$

The per letter uncertainty of the sequence $Y_{[1,n]}$ is the quantity

$$H_n(Y) = \frac{1}{n} H(Y_{[1,n]}). \tag{3}$$

When the righthand side of (3) has a limit as n tends toward infinity, we write $H_\infty(Y)$ for this limit.

The rate R of a discrete information source is defined to be the quantity $H_\infty(Y)$ provided, of course, that this limit exists. Under fairly weak conditions on the source (cf. Gallager[1]), it can be shown that $R = H_\infty(Y)$ is the least number of binary digits per source letter in a binary encoding of the source output from which the source output can be reconstructed.

A finite-state Markov chain is a sequence $X_1, X_2, X_3, \ldots$ of random variables, each of which takes values in the finite alphabet $\sum = \{\sigma_1, \sigma_2, \ldots \sigma_N\}$, such that

$$P(X_n | X_{[1,n)}) = P(X_n | X_{n-1}), \; n \geq 2 \tag{4}$$

and such that

$$P(X_n = \sigma_j | X_{n-1} = \sigma_i) = P(\sigma_j | \sigma_i) \tag{5}$$

for all i and j and all $n \geq 2$. Such a chain is often specified by a graph, as in Figure 1, in which the nodes are identified with the states σ_i, $1 \leq i \leq n$, and in which the directed edge from state σ_i to state σ_j is labelled with the transition probability $P(\sigma_j | \sigma_i)$. The chain is said to be regular if it has steady-state probabilities, i.e. when the limits

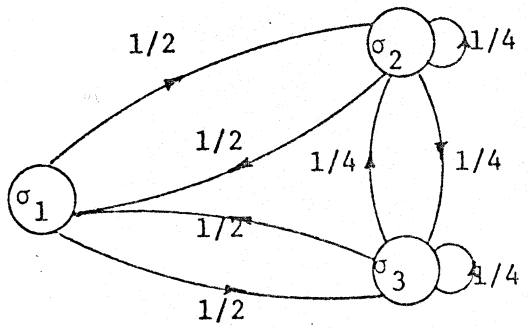

Fig. 1 Graphical Portrayal of a Finite-State Markov Chain

$$\pi_i = \lim_{n\to\infty} P(X_n = \sigma_i), \; i = 1,2,\ldots N \tag{6}$$

exist and are independent of the choice of $P(X_1)$. In this case, the steady-state probabilities π_i, $i = 1,2,\ldots N$, are also the unique stationary distribution, i.e. the unique solution of the equations

$$\pi_j = \sum_{i=1}^{N} \pi_i \, P(\sigma_j|\sigma_i) \quad j = 1,2,\ldots N \tag{7}$$

such that $\pi_1 + \pi_2 + \ldots + \pi_N = 1$. (For more details on Markov chains, the reader is referred to Ash[2] and Feller.[3]) The chain of Fig. 1 is regular and for this chain the solution of the equations (7) gives $\pi_1 = \pi_2 = \pi_3 = 1/3$ as its steady-state probabilities.

An obvious way to create an A-ary information source from a regular finite-state Markov chain is simply to take $A = \sum$, $Y_j = X_j$ and $P(X_1 = \sigma_i) = \pi_i$. In this case, the rate $R = H_\infty(X)$ is easily calculated since

$$\begin{aligned} H_n(X) &= \frac{1}{n} H(X_{[1,n]}) \\ &= \frac{1}{n} [H(X_1) + \sum_{i=2}^{n} H(X_i|X_{[1,i)})] \\ &= \frac{1}{n} [H(X_1) + \sum_{i=2}^{n} H(X_i|X_{i-1})]. \end{aligned} \tag{8}$$

Upon noting that

$$H(X_i|X_{i-1}) = -\sum_{j=1}^{N} \pi_j \sum_{k=1}^{N} P(\sigma_k|\sigma_j) \log P(\sigma_k|\sigma_j), \; i \geq 2$$

and that $0 \leq H(X_1) \leq \log N$, we have from equation (8)

$$H_\infty(X) = -\sum_{j=1}^{N} \pi_j \sum_{k=1}^{N} P(\sigma_k|\sigma_j) \log P(\sigma_k|\sigma_j) \tag{9}$$

which we shall in general refer to as the state-uncertainty of the regular finite-state Markov chain. Using the previously calculated steady-state probabilities in equation (9), we find

$$H_\infty(X) = 4/3 \text{ bits}$$

for the regular finite-state chain of Figure 1 which is the rate of this chain when its state sequence is taken to be an information

source.

More generally, we shall define a regular A-ary Markov source to be an A-ary information source such that there can be found a regular finite-state Markov chain and a mapping $f: \sum \to A$ such that

$$Y_i = f(X_i) \quad j = 1,2,\ldots$$

when $P(X_1 = \sigma_j) = \pi_j$ is the initial state distribution.

As an example of a binary Markov source, consider the source defined by Equation (10) when the Markov chain is that of Figure 1 and the function f is defined by $f(\sigma_1) = f(\sigma_2) = 0$, $f(\sigma_3) = 1$. It is interesting to view the function f as a deterministic, but noisy, channel; i.e. as a channel in which all transition probabilities are either 0 or 1 but for which the channel input cannot in general be deduced unambiguously from its output. In Figure 2, we show the deterministic channel for the function f of our example. In this context, we can view any regular

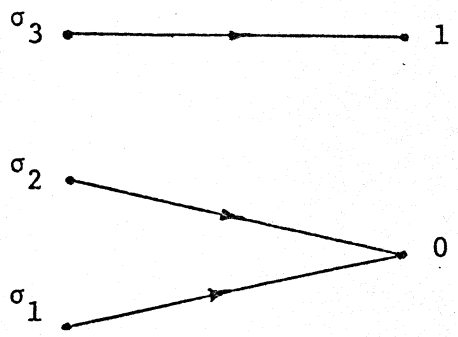

Fig. 2 The Deterministic, But Noisy, Channel Corresponding to $f(\sigma_1) = f(\sigma_2) = 0$, $f(\sigma_3) = 1$

A-ary Markov source as the cascade of a regular Markov state-source, i.e. a source whose output sequence is a regular finite-state Markov chain, and a deterministic channel f; we show this interpretation in Figure 3. It is convenient to portray a

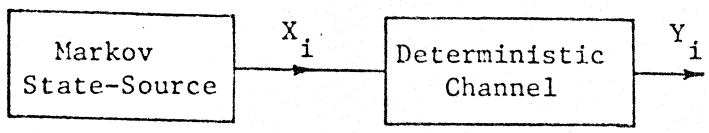

Fig. 3 General Decomposition of a Regular A-ary Markov Source

regular A-ary Markov source graphically, as we do in Figure 4 for the source obtained by cascading the Markov chain of Figure 1 with the deterministic channel of Figure 2, by the graph of its underlying Markov chain with the node labels changed from "σ_i" to "$\sigma_i/f(\sigma_i)$."

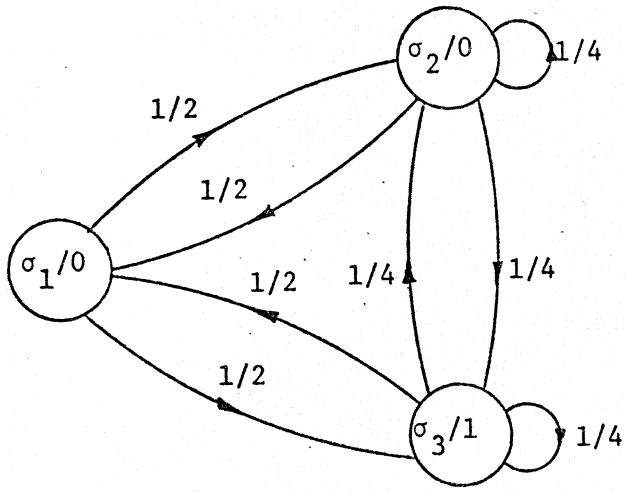

Fig. 4 Graphical Portrayal of a Regular A-ary Markov Source

In light of the ease with which $H_\infty(X)$ can be calculated, one might expect from consideration of Figure 3 that the calculation of the rate $H_\infty(Y)$ of a regular Markov source would be straightforward. As we shall see, such optimism is unfounded.

2. KNOWN RESULTS FOR THE RATE OF MARKOV SOURCES

A regular A-ary Markov source is said to be unifilar (Ash[2]) when the values of X_i and Y_{i+1} uniquely determine the value of X_{i+1}. (A Markov state-source is trivially unifilar since $Y_{i+1} = X_{i+1}$.) Equivalently, the source is unifilar when the values of X_1 and $Y_{[1,n]}$ uniquely determine $X_{[1,n]}$. But then, for a unifilar source,

$$H(X_{[1,n]}) = H(X_1 Y_{[1,n]})$$

$$= H(Y_{[1,n]}) + H(X_1 | Y_{[1,n]}). \qquad (10)$$

Noting that

$$0 \leq H(X_1|Y_{[1,n]}) \leq \log N, \tag{11}$$

it follows from (10), upon dividing by n and letting $n \to \infty$, that

$$H_\infty(X) = H_\infty(Y) \tag{12}$$

for a unifilar regular A-ary Markov source. Thus, the rate of such sources is just their state-uncertainty.

Unfortunately, real information sources tend to be well-modelled only by non-unifilar Markov sources. The simple Markov source of Figure 4 is non-unifilar since, for instance, if $X_1 = \sigma_3$ and $Y_2 = 0$ it is possible that $X_2 = \sigma_1$ and also possible that $X_2 = \sigma_2$. At present, there is no known way to calculate the rate $H_\infty(Y)$ of a non-unifilar source. In fact, a rather celebrated paper of Gilbert[4] is devoted to obtaining a series which converges rather quickly to the rate of the "simple" non-unifilar source of Figure 4. Using Gilbert's series, we find

$$H_\infty(Y) \simeq 0.903$$

for the non-unifilar Markov source of Figure 4.

3. TENTATIVE APPROACHES TO RATE CALCULATION FOR NON-UNIFILAR MARKOV SOURCES

We now present two approaches to non-unifilar sources which to us appear promising avenues toward the goal of calculating (not just approximating) the rate of a non-unifilar source. For simplicity, we shall consider only Markov sources which are epic in the sense that, regardless of the value of X_1, the next output letter must have a non-zero probability of taking on each value in the alphabet A. The source of Figure 4 is epic as indeed seems to be the case for realistic Markov source models of real information sources.

A. Source Identical Twins

Consider the conceptual situation of observing the output sequences of two identical, but independent, epic regular Markov sources started in the same initial state. Suppose we made many such observations but chose to ignore all observations when the output sequences did not coincide. Let $X_1, X_2, X_3, \ldots$ and $X_1', X_2', X_3', \ldots$ denote the state sequences of the two sources where we have required $X_1 = X_1'$. For all our retained observations, we would

have $(X_i, X_i') = (\sigma_j, \sigma_k)$ where $f(\sigma_j) = f(\sigma_k)$ and where the pair (σ_j, σ_k) is a possible successor of some state $(\sigma_\ell, \sigma_\ell)$ via intermediate state pairs (σ_g, σ_h) for which $f(\sigma_g) = f(\sigma_h)$. Let $\sum_T$ be the set of all such pairs (σ_j, σ_k). For example, for the source of Figure 4, $\sum_T = \{(\sigma_1,\sigma_1),(\sigma_1,\sigma_2),(\sigma_2,\sigma_1),(\sigma_2,\sigma_2),(\sigma_3,\sigma_3)\}$. Let π_{jk} be the "steady-state probability" of the state (σ_j, σ_k) in $\sum_T$ along the retained observations. Then these steady-state probabilities satisfy

$$\lambda \pi_{jk} = \sum_{(g,h)\varepsilon I_T} \pi_{gh} P(\sigma_j|\sigma_g) P(\sigma_k|\sigma_h) \tag{13}$$

where I_T is the set of all index pairs (i,j) such that $(\sigma_i, \sigma_j) \varepsilon \sum_T$ and where

$$\lambda = \lim_{n\to\infty} \Pr(Y_n = Y_n' | Y_{[1,n)} = Y'_{[1,n)}) \tag{14}$$

is the limiting conditional probability that the two independently running sources would emit the same next output letter given that their output sequences had previously coincided.

We see from equation (13) that λ is an eigenvalue of the matrix $[P(\sigma_j|\sigma_g)\ P(\sigma_k|\sigma_h)]$ with all rows and columns deleted except those for which $(j,k) \varepsilon I_T$ and $(g,h) \varepsilon I_T$. In fact, it is not too hard to verify that λ is the largest eigenvalue of this matrix. For the source of Figure 4, this matrix is

$$\begin{bmatrix} 0 & 0 & 0 & 1/4 & 1/4 \\ 0 & 0 & 1/4 & 1/8 & 1/8 \\ 0 & 1/4 & 0 & 1/8 & 1/8 \\ 1/4 & 1/8 & 1/8 & 1/16 & 1/16 \\ 1/4 & 1/8 & 1/8 & 1/16 & 1/16 \end{bmatrix}$$

and its largest eigenvalue is readily found to be $\lambda \simeq 0.555$.

It follows further from equation (14) that

$$\Pr(Y_{[1,n]} = Y'_{[1,n]}) \simeq \lambda^n, \quad n \text{ large} \tag{15}$$

in the sense that

$$\lim_{n\to\infty} [-\frac{1}{n} \log \Pr(Y_{[1,n]} = Y'_{[1,n]})] = -\log \lambda. \tag{16}$$

For the source of Figure 4, we have $-\log \lambda = -\log(.555) = .849$ which we note is slightly smaller than the rate of the source.

We now offer a plausibility argument that the rate $H_\infty(Y)$ of the source is closely connected with $-\log \lambda$. We use the typical sequence argument of Shannon.[5] For large n, there will be $2^{nH_\infty(X)}$ typical state sequences, (each with probability $2^{-nH_\infty(X)}$). Moreover $2^{nH_\infty(X|Y)}$ of these sequences will yield each of the $2^{nH_\infty(Y)}$ typical output sequences. Thus,

$$H_\infty(Y) = H_\infty(X) - H_\infty(X|Y). \tag{17}$$

But also we see that if we have two such sources running independently, then

$$\Pr(Y_{[1,n]} = Y'_{[1,n]} | Y_{[1,n]} \text{ typical}) = 2^{-n[H_\infty(X)-H_\infty(X|Y)]}$$
$$= 2^{-nH_\infty(Y)}, n \text{ large}. \tag{18}$$

Examination of equations (15) and (18) provide the basis for our claim that $H_\infty(Y)$ and $-\log \lambda$ should be closely related. In fact they will be equal for any source for which the probability is large that $Y_{[1,n]}$ is typical given that $Y_{[1,n]} = Y'_{[1,n]}$. Any difference between $H_\infty(Y)$ and $-\log \lambda$ is the result of the fact that the source is such that $Y_{[1,n]} = Y'_{[1,n]}$ tends to occur mainly for non-typical source output sequences. The unresolved question at the moment is: How can one give a general relation for $H_\infty(Y)$ in terms of $-\log \lambda$? It seems indisputable to claim, on the basis of equation (16), that $-\log \lambda$ is giving some interesting information about the Markov source.

B. The Master-Slave Source

Consider the situation illustrated in Figure 5 where an epic regular Markov source (the "master") with state sequence $X_1, X_2, X_3, \ldots$ and output sequence $Y_1, Y_2, Y_3, \ldots$ is independently running, but a second and "similar" source (the "slave") is observing only the output Y_i before making its transition to a state X'_i such that $Y'_i = f(X'_i) = Y_i$. We can model this conceptual situation by defining the <u>master-slave source</u> for the original epic regular Markov source to be the Markov source whose state set is

$$\Sigma_{MS} = \{(\sigma_i, \sigma_j) | f(\sigma_i) = f(\sigma_j)\}, \tag{18}$$

whose transition probabilities are

$$P_{MS}(\sigma_k,\sigma_\ell|\sigma_i,\sigma_j) = P(\sigma_k|\sigma_i) \frac{P(\sigma_\ell|\sigma_j)}{\sum_{\sigma:f(\sigma) = f(\sigma_k)} P(\sigma|\sigma_j)} , \tag{19}$$

and whose output function is

$$f_{MS}(\sigma_i,\sigma_j) = f(\sigma_i). \tag{20}$$

We shall write the state sequence of the master-slave source as $X_1X_1', X_2X_2', X_3X_3', \ldots$. It is not hard to see that the master-slave source will itself be an epic regular Markov source. In Figure 6, we show the master-slave source for the epic regular Markov source of Figure 4.

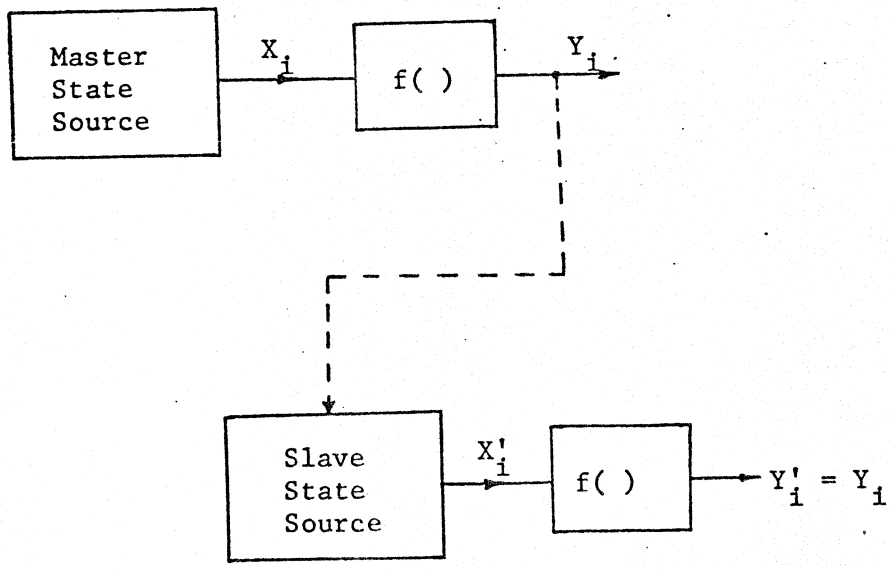

Fig. 5 Conceptual Situation for the Master-Slave Source

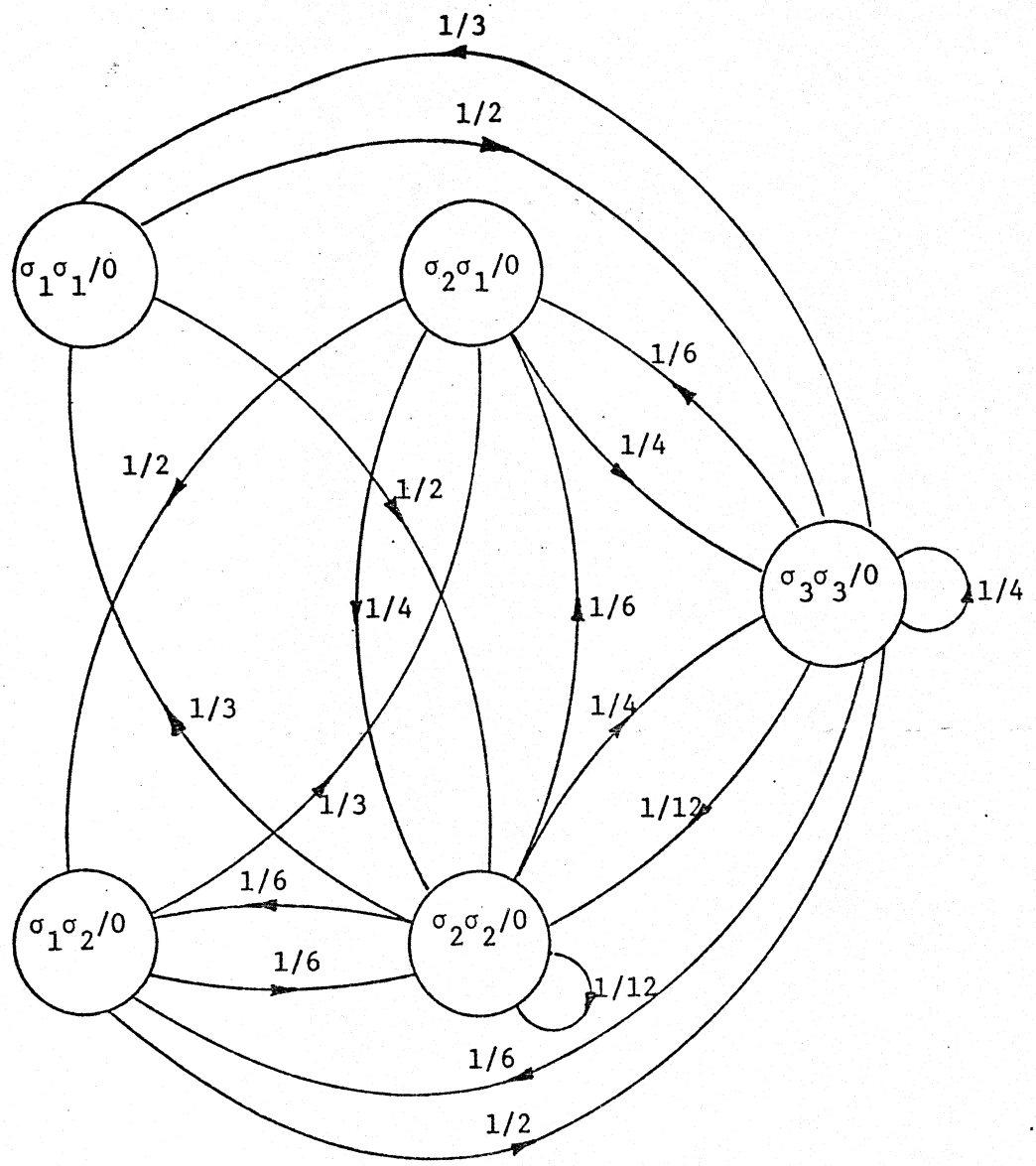

Fig. 6 The Master-Slave Source for the Source of Fig. 4

Since the slave source acts in all ways like the free-running master source except that it conditions its state transitions so as to produce the same output sequence, it seems reasonable to <u>conjecture</u> that

$$H_\infty(X'|Y) \stackrel{?}{=} H_\infty(X|Y). \tag{21}$$

In any case, we surely have

$$H_\infty(XX') = H_\infty(X) + H_\infty(X'|X)$$
$$= H_\infty(X) + H_\infty(X'|Y) \tag{22}$$

since the slave state sequence depends upon the master state sequence only through the intermediary of the output sequence. From equation (22) it follows that

$$2\,H_\infty(X) - H_\infty(XX') = H_\infty(X) - H_\infty(X'|Y). \tag{23}$$

If the conjecture (21) is true, we can use (17) and (23) to obtain

$$H_\infty(Y) \stackrel{?}{=} 2\,H_\infty(X) - H_\infty(XX'). \tag{24}$$

If equation (24) is valid, then we have a direct method to compute the rate of the original epic regular Markov source since the two terms on the righthand side are state-uncertainties of regular Markov chains and thus caculable by equation (9).

The key question remains: Is the conjecture (21) true? This conjecture is trivially true for a unifilar source since $H_\infty(X'|Y) = H_\infty(X|Y) = 0$ because the output sequence determines the state sequence except for the initial state. Thus, equation (24) holds for a unifilar source--but of course we already know how to calculate the rate of such sources. Equation (24) holds for the source of Figure 4 to within the accuracy that we have been able to calculate the Gilbert approximation to the rate and the accuracy that we have been able to calculate the steady-state probabilities for the master-slave source of Figure 6. As yet, we have been unable to produce a case where equation (24) could be shown invalid. We leave the validation or invalidation of equation (24) as an open problem.

4. REFERENCES

1. Gallager, R. G., Information Theory and Reliable Communication, John Wiley and Sons, New York, 1968.

2. Ash, R., Information Theory, Interscience Publishers, New York, 1965.

3. Feller, W., An Introduction to Probability Theory and its Applications, Vol. I, John Wiley and Sons, New York, 1950.

4. Gilbert, E. N., Capacity of a Burst Noise Channel, Bell System Technical Journal, 30, 1253, 1952.

5. Shannon, C. E., A Mathematical Theory of Communication, Bell System Technical Journal, 27, 379, 1948.