[2] P. Flandrin, "On the spectrum of fractional Brownian motions," *IEEE Trans. Inform. Theory*, vol. 35, pp. 197–199, Jan. 1989.
[3] M. S. Keshner, "$1/f$ noise," *Proc. IEEE*, vol. 70, Mar. 1982, pp. 212–218.
[4] S. Mallat, "Multiresolution representations and wavelets," Ph.D. thesis, Univ. of Pennsylvania, Phila., PA, Aug. 1988.
[5] B. Mandelbrot, "Some noises with $1/f$ spectrum, a bridge between direct current and white noise," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 289–298, Apr. 1967.
[6] B. B. Mandelbrot and H. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, pp. 422–436, Oct. 1968.
[7] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1984.

Variable-to-Fixed Length Codes are Better than Fixed-to-Variable Length Codes for Markov Sources

JACOB ZIV, FELLOW, IEEE

Abstract —**It is demonstrated that for finite-alphabet, Kth order ergodic Markov Sources (i.e., memory of K letters), a variable-to-fixed code is better than the best fixed-to-variable code (Huffman code). It is shown how to construct a variable-to-fixed length code for a Kth order ergodic Markov source, which compresses more effectively than the best fixed-to-variable code (Huffman code).**

I. Introduction

Consider the class of finite-alphabet, finite-order ergodic Markov sources, characterized by a probability distribution of the form

$$P(X)=\prod_{i=1}^{n} P\left(X_i|X_1^{i-1}\right) \tag{1}$$

where

$$X=X_1,X_2,\cdots,X_n$$

$$P\left(X_i|X_1^{i-1}\right)=P\left(X_i|X_{i-K}^{i-1}\right),\qquad \text{for any } i\geq K,$$

and where

1) $X_i^j \triangleq X_i, X_{i+1},\cdots,X_j, i<j$,
2) X_i is the output of the source at the ith instant, $X_i\in A$; $|A|=\alpha$.

A *code* is an extended alphabet C of M vectors ("words") $X_1^c, X_2^c,\cdots,X_M^c$ where $X_i^c\in A^{l(i)}$ and where $l(i)$ is the length of the vector X_i^c.

Assume also that any vector $x\in A^l$ for $l\geq \max_i l(i)$ has a prefix $X_i^c\in C$ for some $1\leq i\leq M$, and that for every i and $j(i\neq j) X_i^c\in C$ is not a prefix of $X_j^c\in C$ (i.e., the code is complete and proper [1]).

Every vector $X_i^c\in C$ is mapped into a unique binary sequence Y_i^c of length $L(Y_i^c)\triangleq L(i)$ binary letters. This sequence is called the "codeword" for the word X_i^c. A fixed-to-variable length code (FVL) is one for which

$$l(i)=l,\qquad 1\leq i\leq M. \tag{2}$$

Manuscript received September 14, 1987; revised August 25, 1988.
This work was supported by the US–Israel Binational Science Fund. This work was partially done at AT&T Bell Laboratories, Murray Hill, NJ.
The author is with the Department of Electrical Engineering, Technion, Haifa 32000, Israel.
IEEE Log Number 9034938.

A variable-to-fixed length code (VFL) is one for which

$$L(i)=L,\qquad 1\leq i\leq M;$$
$$L=\lceil \log M\rceil \tag{3}$$

where logarithms in this correspondence are taken to be of base 2.

Consider the parsing of $X=X_1^n$ into a sequence of $q_c(X)$ words of C (ignoring end-effects)

$$X=X^1,X^2,X^3,\cdots,X^j,\cdots,X^{q_c(X)}, \tag{4}$$

where $X^j\in C$, $1\leq j\leq q_c(X)$. Let

$$L_c(X)=\sum_{j=1}^{q_c(X)} L(j), \tag{5}$$

where $L(j)$ is the length of Y^j, the binary codeword that corresponds to the jth word in the parsed X. The compression-ratio for a given code C is defined by

$$\rho_c=\lim_{n\to\infty}\frac{EL_c(X)}{n\log\alpha}, \tag{6}$$

where $E(\cdot)$ denotes expectation.

It is well known [2] that

$$\rho_c\geq\frac{H}{\log\alpha} \tag{7}$$

and that there exist a sequence of FVL codes (Huffman codes) such that

$$\lim_{M\to\infty}\rho_c=\frac{H}{\log\alpha}\triangleq\rho(\infty) \tag{8}$$

where

$$H=\lim_{n\to\infty}-\frac{1}{n}E\log P(X). \tag{9}$$

Unfortunately, for finite-order Markov sources with memory $(K>1)$ and with $\rho(\infty)<1$ we have that

$$\rho(M)\triangleq\min\rho_c>\rho(\infty) \tag{10}$$

where the minimization is carried over *all* codes with M codewords.

In Theorem 1, we derive lower-bounds on ρ_c, for any code such that the shortest word in C is no shorter than K. Clearly, any FVL code with more than α^K codewords is included in this family of codes.

In Theorem 2, we derive upper bounds on ρ_c for a VFL code and show that it approaches the lower bound of Theorem 1, at least for sources with large memory ($K\gg 1$).

At the same time, the rate of approach of ρ_c for the best FVL code (i.e., Huffman code) is *slower* than that of the VFL code. Thus, VFL coding takes better advantage of the source memory.

II. Derivations and Statement of Results

The coding of a sequence X was shown to be associated with parsing the sequence into $q_c(X)$ words.

Each word is encoded into one out of M codewords of the given code C. The selection of the particular codeword is independent of the past words, without taking advantage of the memory of the source. Thus, when encoding each of the $q_c(X)$ words in X, there is a certain loss in compression. We show that the accumulated average loss for X is proportional to the expected number of words $Eq_c(X)$, and demonstrate that

$Eq_c(X)$ for a variable-to-fixed code can be made *smaller* than that of the best fixed-to-variable code (namely, Huffman code), for any Kth order Markov source (provided $K \gg 1$).

Theorem 1: For any code such that

$$l(i) \geq K, \qquad i = 1,2,3,\cdots,M,$$

$$\lim_{n\to\infty} E\frac{q_c(X)}{n} = \frac{H}{H_c - I_K} \geq \frac{H}{\log M - I_K}, \tag{11}$$

where

$$I_K = E\log\frac{P(X_1^K|X_{-K}^0)}{P(X_1^K)},$$

$$H_c = -E\log P(X_i^c) \leq \log M.$$

Also

$$\rho_c \geq \lim_{n\to\infty}\frac{Eq_c(X)}{n\log\alpha}H_c = \rho(\infty)\frac{1}{1-I_K/H_c} \geq \rho(\infty)\frac{1}{1-I_K/\log M}. \tag{12}$$

Proof: Assume that $l(i) \geq K$ for all i, $1 \leq i \leq M$. By (1) and by (4)

$$P(X^i|X^1, X^2,\cdots,X^{i-1}) = P(X^i|X^{i-1}). \tag{13}$$

Thus, by (13) and since the original source is ergodic (i.e., aperiodic), we have that the probability measure of $\cdots X^1, X^2,\cdots,X^i,\cdots$ is a first-order ergodic measure [2, p. 65]. By ergodicity, by (11) and for any arbitrary small ϵ, there exist an integer $q_0 = q_0(\epsilon)$ such that for any $q > q_0$

$$\Pr\left[\left|\frac{1}{q}\sum_{j=1}^{q} L(X^j) - EL(X_i^c)\right| > \epsilon\right] < \epsilon$$

$$\Pr\left[\left|\frac{1}{q}\sum_{j=1}^{q} -\log P(X^j) - H_c\right| > \epsilon\right] < \epsilon$$

$$\Pr\left[\left|\frac{1}{q}\sum_{j=1}^{q} -\log\frac{P(X^j|X^{j-1})}{P(X^j)} - I_K\right| > \epsilon\right] < \epsilon$$

where the last result follows from the fact that since $l(i) \geq K$ for all $1 \leq i \leq M$

$$E\log\frac{P(X^j|X^{j-1})}{P(X^j)} = E\log\frac{P(X_1^K|X_{-K}^0)}{P(X_1^K)}.$$

Now, since the length of every vector X^j is bounded by $\max_{1\leq i\leq M} l(i)$, there exists an integer $n_0 = n_0(\epsilon)$ such that for any $n > n_0$, $q_c(X) > n_0/\max l(i)$ and hence

$$\Pr\left[\left|\frac{1}{q_c(X)}\sum_{j=1}^{q_c(X)} \log\frac{p(X^j|X^{j-1})}{P(X^j)} - I_K\right| > \epsilon\right] < \epsilon$$

$$\Pr\left[\left|\frac{1}{q_c(X)}\sum_{j=1}^{q_c(X)} -\log P(X^j) - H_c\right| > \epsilon\right] < \epsilon.$$

Furthermore, for any X such that $P(X) > 0$, there exists a finite number B such that

$$\left|\log\frac{P(X^j|X^{j-1})}{P(X^j)}\right| \leq B < \infty, \qquad \text{for all } j,$$

also

$$\left|\frac{q_c(X)}{n}\right| \leq 1.$$

Thus

$$\lim_{n\to\infty}\frac{1}{n}E\left\{q_c(X)\frac{1}{q_c(X)}\left[\sum_{j=1}^{q_c(X)} -\log P(X^j) + \log P(X)\right]\right\}$$

$$= \lim_{n\to\infty}\frac{1}{n}E\left\{q_c(X)\frac{1}{q_c(X)}\sum_{j=1}^{q_c(X)}\log\frac{P(X^j|X^{j-1})}{p(X^j)}\right\}$$

$$= I_K \lim_{n\to\infty} E\frac{q_c(X)}{n}. \tag{14}$$

The last step follows from the fact that $\cdots X^1, X^2,\cdots,X^i,\cdots$ is ergodic and therefore, with probability one, $\lim_{n\to\infty}[1/q_c(X)]\sum_{j=1}^{q_c(X)}\log P(X^j|X^{j-1})/P(X^j) = I_k$ and since $1/\max l(i) \leq q_c(X)/n < 1$. Similarly

$$\lim_{n\to\infty}\frac{1}{n}E\left[q_c(X)\left\{\frac{1}{q_c(X)}\sum_{i=1}^{q_c(X)} -\log P(X^i)\right\}\right] = \lim_{n\to\infty}\frac{1}{n}Eq_c(X)H_c.$$

Thus, by (9), and (14)

$$\lim_{n\to\infty}\frac{1}{n}Eq_c(X)H_c - H = \lim_{n\to\infty}\frac{1}{n}Eq_c(X)I_K. \tag{15}$$

Therefore

$$\lim_{n\to\infty}\frac{1}{n}Eq_c(X) = \frac{H}{H_c - I_K}. \tag{16}$$

Now, for any given code C

$$\rho_c = \lim_{n\to\infty}\frac{EL_c(X)}{n\log\alpha}.$$

But by (4)

$$E\frac{L_c(X)}{n} = E\left\{\frac{q_c(X)}{n}\frac{\sum_{i=1}^{q_c(X)} L(Y^i)}{q_c(X)}\right\}$$

and by the ergodicity which is implied by (13) and the arguments that led to the derivation of (14)

$$\rho_c = \lim_{n\to\infty}\frac{EL_c(X)}{n\log\alpha} = \lim_{n\to\infty}\frac{Eq_c(X)}{n\log\alpha}EL(X_i^c). \tag{17}$$

However C is a uniquely decipherable prefix code. Thus by [2]

$$EL(X_i^c) \geq H_c.$$

Therefore, for *any* code C,

$$\rho_c \geq \lim_{n\to\infty} \frac{Eq_c(X)}{n \log \alpha} H_c. \quad (18)$$

Inserting (16) into (18) yields

$$\rho_c \geq \rho(\infty)\frac{H_c}{H_c - I_K} \geq \rho(\infty)\frac{\log M}{\log M - I_K}, \quad (19)$$

which completes the proof of Theorem 1. □

Theorem 2: For every $M > \alpha / \max\{P(X_i^K)\}$ there exists a VFL code with no more than M codeword such that

$$\rho_c \leq \rho(\infty)\frac{\log M}{\log M - \log \alpha - I_K}. \quad (20)$$

Proof: Construct a VFL code as follows.

a) Start with the tree that consists of all α^K words of length K. This tree has α^K leaves.
b) Extend each leaf by all possible single-letter extensions, provided that the word that is represented by this leaf has a probability that is larger than α / M.
c) Repeat Step b) as many times as possible.

Clearly there are no more than M codewords in the code tree. At the same time, by construction, no leaf has a probability that is larger than α / M. Thus,

$$\begin{aligned} H_c &= -E \log P(X_j^c) \\ &\geq \log \frac{M}{\alpha} = \log M - \log \alpha. \end{aligned} \quad (21)$$

Therefore, by (16), (17), and (21)

$$\rho_c \leq \lim_{n\to\infty} \frac{Eq_c(X)\log M}{n \log \alpha} \leq \rho(\infty)\frac{\log M}{\log M - \log \alpha - I_K}.$$

□

Conclusion

It is clear from Theorem 1 and Theorem 2 that when $I_K \gg \log \alpha$ (i.e., $K \gg 1$), ρ_c for the VFL code of Theorem 2 approaches the lower bound in Theorem 1, namely

$$\rho_c(\text{VFL}) \approx \rho(\infty)\frac{1}{1 - I_K/\log M}.$$

At the same time, we have for the best FVL code that

$$\rho_c(\text{FVL}) \geq \rho(\infty)\frac{1}{1 - I_K/H_c}$$

and since $H_c < \log M$, its performance is inferior to that of the VFL code of Theorem 2.

Acknowledgment

Helpful discussions with A. D. Wyner are acknowledged.

References

[1] F. Jelinek and K. S. Schneider, "On variable length-to-block coding," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 6, pp. 765–774, Nov. 1972.
[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

A Limit Theorem for n-Phase Barker Sequences

NING ZHANG AND S. W. GOLOMB

Abstract —**It is proven that for $3 \leq L \leq 19$, except for $L = 6$, the total number of normalized n-phase Barker sequences of length L increases without limit, as n goes to infinitity.**

I. Introduction

In [1], a *generalized Barker sequence of length* L is defined as an L-term sequence, $a_1, a_2, \cdots, a_L$, of complex numbers with $|a_j| = 1$ and $|C(\tau)| \leq 1$ for all j, $1 \leq j \leq L$ and all τ, $1 \leq \tau \leq L-1$, where $C(\tau) = \sum_{j=1}^{L-\tau} a_j a^*_{j+\tau}$. (Here, z^* denotes the complex conjugate of z.) We call $\{a_i\}$ an n-phase Barker sequence, if each a_i is an nth root of unity.

Under the group of n-phase Barker-preserving transformations, taking the lexicographically smallest representative of its equivalence class,[1] we may assume that $a_1 = a_2 = 1$ and a_3 is in the upper half plane. Such a sequence is called a *normalized* Barker sequence.

Definition: We let $N_L(n)$ denote the total number of normalized n-phase Barker sequences, of length L.

In [2], we have proved that for all $n \geq 1$,

$$N_6(n) = \begin{cases} 1, & \text{if } n = 6k, \\ 0, & \text{otherwise.} \end{cases}$$

Here we shall prove that for $3 \leq L \leq 19$, except for $L = 6$,

$$\lim_{n \to +\infty} N_L(n) = +\infty.$$

II. Limit of $N_L(n)$ ($3 \leq L \leq 19$ except $L = 6$)

From [1], we have the following two theorems.

Theorem 1: The sum of two unit vectors lies within the unit circle if and only if the angle between those vectors is at least 120° and at most 240°.

Theorem 2: The sum of three unit vectors lies within the unit circle if and only if there is no semicircle properly containing all three vectors.

Notation: In this section, we will use $(\alpha_1, \alpha_2, \cdots, \alpha_L)$ to denote the Barker sequence $(e^{i\alpha_1}, e^{i\alpha_2}, \cdots, e^{i\alpha_L})$, and we will use $\alpha_j \Theta \alpha_k$ to denote the sum $e^{i\alpha_j} + e^{i\alpha_k}$. The values of α_j are given in *radians*, unless the symbol for *degrees* is used.

Lemma 1: For any positive integer $n \geq 1$, we have

$$N_1(n) \equiv N_2(n) \equiv 1.$$

Lemma 2: For any positive integer $n \geq 1$, we have

1) $N_3(n) = |\{j\colon 2n \leq 6j \leq 3n\}|$;
2) $\lim_{n \to +\infty} N_3(n) = +\infty$.

Manuscript received June 1988; revised October 1989.
N. Zhang is with Broadband Network Research, Pacific Bell, 2600 Camino Ramon, Room 1S900G, San Ramon, CA 94583.
S. W. Golomb is with the University of Southern California, PHE-506, Los Angeles, CA 90089-0272.
IEEE Log Number 9034950.

[1] A group of n-phase Barker-preserving transformations: if $\{\alpha_i\}$ is an n-phase Barker sequence of length L, so too are $\{\alpha_i^*\}$ and $\{\gamma\delta^i a_i\}$ for all γ and δ with both γ and δ equal to nth roots of unity.
Each term a_i can be regarded as a positive integer power of $e^{i2\pi/n}$. Hence n-phase Barker sequences correspond to L-tuples of integers. The "lexicographically smallest representative" refers to these corresponding L-tuples of integers.

0018-9448/90/0700-0863$01.00 ©1990 IEEE