Information rates and power spectra of digital codes

Justesen, Jørn

[Link back to DTU Orbit](#)

[34] P. Noll and R. Zelinski, "Bounds on quantizer performance in the low bit-rate region," *IEEE Trans. Commun.*, vol. COM-26, pp. 300–304, Feb. 1978.

[35] L. Franks, *Signal Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1969, p. 7.

[36] W. A. Finamore and W. A. Pearlman, "Optimal encoding of discrete time continuous-amplitude memoryless sources with finite output alphabets," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 144–155, March 1980.

[37] J. G. Dunham, "On iterative theory of code design," submitted to *IEEE Trans. Inform. Theory*, Biomedical Computer Laboratory,

Monograph no. 389, Washington University School of Medicine, St. Louis, MO, Jan. 1981.

[38] J. S. Meditch, *Stochastic Optimal Linear Estimation and Control.* New York: McGraw-Hill, 1969.

[39] S. P. Lloyd, "Least squares quantization in PCM," unpublished memorandum, Bell Laboratories, 1957; see *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 129–137, March 1982.

[40] N. T. Gaarder and D. Slepian, "On optimal finite-state digital transmission systems," presented at the *1979 IEEE Int. Symposium on Information Theory*, Grignano, Italy, June 1979.

Information Rates and Power Spectra of Digital Codes

JØRN JUSTESEN

Abstract—The encoding of independent data symbols as a sequence of discrete amplitude, real variables with given power spectrum is considered. The maximum rate of such an encoding is determined by the achievable entropy of the discrete sequence with the given constraints. An upper bound to this entropy is expressed in terms of the rate distortion function for a memoryless finite alphabet source and mean-square error distortion measure. A class of simple dc-free power spectra is considered in detail, and a method for constructing Markov sources with such spectra is derived. It is found that these sequences have greater entropies than most codes with similar spectra that have been suggested earlier, and that they often come close to the upper bound. When the constraint on the power spectrum is replaced by a constraint on the variance of the sum of the encoded symbols, a stronger upper bound to the rate of dc-free codes is obtained. Finally, the optimality of the binary biphase code and of the ternary bipolar code is decided.

I. INTRODUCTION

IN DIGITAL communication and recording, a data sequence from a finite alphabet $B = \{b_1, b_2, \cdots, b_J\}$ is converted into a continuous real-valued function of time. It is convenient to introduce as an intermediate step a sequence of real numbers from a finite set $V = \{v_1, v_2, \cdots, v_K\}$. This sequence may be converted to the desired continuous signal by interpolation and possibly modulation. Part of the necessary shaping of the power spectrum may take place in the coding from the data sequence to the real sequence. In particular it is convenient to suppress undesirable low frequency content at this point.

We refer to [1] for an introduction to the literature on spectrum shaping codes.

A code will be defined as a one-to-one mapping from the data sequence to the real sequence. We shall assume that the data sequence may be described as a sequence of independent identically distributed random variables $\bar{A} = \cdots, A_{t-1}, A_t, A_{t+1}, \cdots$ and that the encoded sequence is a stationary ergodic random sequence $\bar{X} = \cdots, X_{t-1}, X_t, X_{t+1}, \cdots$ with given power spectrum $S_X(\omega)$. The rate of the code is the average number of data symbols per encoded symbol.

The main purpose of this investigation is to provide good upper bounds to the rate of such codes. While many different codes have been described in the literature, such bounds have not been available, and consequently it has not been clear whether significantly better results could be obtained with other codes. In Section II, it is proved that the achievable rate depends directly upon the maximum possible entropy of a sequence with alphabet V and power spectrum $S_X(\omega)$. Slepian [2] has given an excellent discussion of the problems involved in constructing maxentropic sequences over finite alphabets for given second moments. Some results were obtained for finite-order Markov processes, but the general problem of specifying the maxentropic sequence or obtaining the exact value of its entropy appears to be very difficult.

In Section III, we derive an upper bound to the entropy of a sequence with given probability distribution on V and given power spectrum. This bound involves the rate distortion function for a memoryless source with alphabet V, and the distortion in question is the minimum mean-square error of any linear predictor for $\bar{X}$. We discuss a computa-

tional procedure for evaluating this bound. In several examples in this and the following sections, the upper bound is compared to the entropies of finite order Markov sources. In many cases, the difference is only a few percent. For binary sequences, the bound is less tight. For most applications, it is desirable that the encoded sequence have a spectrum which is 0 for $\omega = 0$, remain low for ω less than a certain cut-off frequency ω_0, and be approximately constant over the remaining frequency band up to $\omega = \pi$. In order to obtain more detailed results of practical relevance, we consider the case where V is a set of integers and the spectrum is a simple rational function. In Section IV we show that the sum of the encoded symbols plays an important role in the analysis of codes whose spectrum vanishes at $\omega = 0$. In the literature [3], [4], the sum is often used in comparisons between different codes, but there has been given little justification for this approach. We prove that the sum is closely related to the best linear predictor and that an approximate value of ω_0 can be found from the variance of the sum.

Practical coding methods lead to encoded sequences which may be described as unifilar Markov sources. For Markov sources with $S_X(0) = 0$, there is a particular value of the sum of the encoded symbols associated with each state. In Section V we use this fact in the construction of Markov sources with simple power spectra. Comparisons of the entropies of such sources and the rates of several well-known codes show that practical codes often do not achieve the best possible spectra. We give examples of better codes derived from the Markov sources.

If the variance of the sum is given rather than the entire power spectrum, it is less complicated to find a sequence with maximum entropy. In Section VI, we carry out the maximization for binary sequences and obtain closer bounds than those derived in Section III. A similar procedure is possible in the case of ternary sequences, but the result is a great deal more involved.

In Section VII, we consider the problem of finding maxentropic sequences of spectra of the form $S_X(\omega) = A(1 - \cos \omega)$. It is proved that the binary biphase sequence has maximum entropy. On the other hand, the widely used ternary bipolar or "alternate mark inversion" code produces a sequence which is shown not to be maxentropic.

II. ENCODING OF SPECTRUM SHAPING CODES

The data sequence $\overline{A}$ is assumed to consist of independent random variables with alphabet B and identical probability distribution $p_A(b_i)$. The entropy of this sequence is $H(A)$. We wish to encode the data as a stationary ergodic sequence $\overline{X}$ with values from a finite alphabet of real numbers V in such a way that the probability distribution of each X_t is a given $p_X(v_k)$ and the power spectral density of the sequence equals a given function, $S_X(\omega)$. Equivalently, the correlation sequence

$$R_X(k) = E[X_t X_{t+k}] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_X(\omega) e^{ik\omega} \, d\omega \qquad (1)$$

may be given. We shall assume that the spectrum does not contain any discrete components, and in particular that $E[X_t] = 0$. The entropy of $\overline{X}$ may be written [5, p. 57]

$$H(X) = H(X_t \mid X_{t-1}, X_{t-2}, \cdots). \qquad (2)$$

The difference between the coding problem considered here and the more usual source coding problems is the restriction on the statistics of the encoded sequence. The basic difficulty of the problem is that there is no simple connection between the power spectrum and the entropy of $\overline{X}$.

In order to define the rate of the encoding, we collect m data symbols at a time into blocks $A_t^m = (A_{mt}, A_{mt+1}, \cdots, A_{mt+m-1})$. Similarly the encoded symbols are collected into blocks $X_t^n = (X_{nt}, X_{nt+1}, \cdots, X_{nt+n-1})$. These new data symbols are again independent identically distributed and the new encoded sequence is again stationary and ergodic. A code is a one-to-one mapping between these sequences such that the average number of encoded symbols equals the average number of data symbols. The dimensionless rate of the code is then $l = m/n$.

The most natural coding theorem in the present case is obtained by application of the Ornstein–Gray sliding block coding theorem [6]. In order to apply this result, we must restrict the sequence $\overline{X}^n$ to be a B-process, i.e., a process which can be approximated arbitrarily closely by a Markov process. It appears that this class of processes actually includes all sequences of practical interest. The following theorem is a direct consequence of the sliding block coding theorem.

Theorem 1: An encoding of rate l of a sequence $\overline{A}$ is possible if and only if there exists a B-process, $\overline{X}$, with the required alphabet, probability distribution, and power spectrum such that the entropies satisfy $l = H(X)/H(A)$.

By this result, the problem of maximizing the rate is transformed into the problem of constructing a sequence of maximum entropy.

If the restriction that $\overline{X}^n$ be a B-process is dropped, a much less satisfying coding theorem may be proved. We shall not give any details, but briefly indicate the changes involved. From the source block coding theorem [5, p. 60] it follows that if the symbols in B are equiprobable and $H(X) > lH(A)$, then blocks from $\overline{A}$ may be encoded as blocks from $\overline{X}$. The statistics of the encoded sequence are changed a little in the encoding process and not all blocks are used. However, for large block lengths, the spectrum will be close to the desired function. If the data symbols are not equiprobable, a set of small total probability will not be encoded.

Most practical encoding algorithms may be described as finite state transducers [7, p. 281]. A transducer, or a sequential machine, is defined by a set of states, $R = \{s_0, s_1, \cdots, s_N\}$, a set of input symbols, I, a set of output symbols, O, a state transition function $\delta: I \times R \to R$ and an output function $\xi: I \times R \to O$ or $\zeta: R \to O$. If the state at time t is $S(t) = s_j$ and the input symbol is i_k, then

$S(t + 1) = \delta(s_j, i_k)$. In most of our examples, the input at time t is A_t and the output X_t is associated with the state transition. Later we shall define a function Z_t which is a function of $S(t)$. Usually the alphabets of $\overline{X}$ and $\overline{Z}$ will be different.

Since the A_t are independent, they determine via δ a set of transition probabilities $p_{ij} = P[S(t + 1) = s_j | S(t) = s_i]$. If the input is suppressed and replaced by these probabilities, the output sequence is a Markov source [6, pp. 63–69]. We shall assume that, given $S(t_0)$ and the output for $t \geq t_0$, the sequence of states may be reconstructed. A Markov source with this property is often called unifilar. We shall further assume that the state sequence $S(t)$ is an ergodic Markov chain [8, pp. 35–37], which may be regular or cyclic. Thus there is a unique stationary probability distribution on the states $p_S(s_j)$ satisfying

$$p_S(s_j) = \sum_i p_S(s_i) p_{ij} \tag{3}$$

or, in terms of the transition matrix $Q^T = \{p_{ij}\}$ and the vector $\bar{p}_S^T = (p_S(s_0), p_S(s_i), \cdots, p_S(s_N))$,

$$Q\bar{p}_S = \bar{p}_S.$$

If the probability distribution of the source is assumed to be $p_S(s_j)$ at all times, the output sequence is stationary and ergodic. The entropy of the source may be expressed as

$$H(X) = -\sum_i p_S(s_i) \sum_j p_{ij} \log p_{ij}. \tag{4}$$

Even though cyclic sources can be made stationary by proper choice of the probability distribution, their sample functions may contain undesirable periodicities. Thus special attention is required in order to avoid discrete frequencies in their spectra.

Example 1: Perhaps the most widely used spectrum shaping code is the bipolar or the alternate mark inversion (AMI) code. It is used for encoding independent equiprobable binary symbols into a sequence with alphabet $V = \{0, \pm 1\}$, probability distribution $p_X(0) = 1/2$, $p_X(1) = p_X(-1) = 1/4$, and power spectrum $S_X(\omega) = (1 - \cos \omega)/2$. The encoding may be described by the transition function δ and the output function ξ:

$$\delta(s_0, 0) = s_0, \quad \xi(s_0, 0) = 0,$$
$$\delta(s_0, 1) = s_1, \quad \xi(s_0, 1) = 1,$$
$$\delta(s_1, 0) = s_1, \quad \xi(s_1, 0) = 0,$$
$$\delta(s_1, 1) = s_0, \quad \xi(s_1, 1) = -1.$$

The same sequence may be obtained by interpreting the input as the real numbers ± 1 and using the linear encoding rule

$$X_t = (A_t - A_{t-1})/2.$$

The power spectrum follows immediately from this expression. The rate of the code is one and the entropy of the output sequence 1 bit. Much of the research reported here was motivated by the question whether a code with this probability distribution and power spectrum could have a rate greater than one.

III. AN UPPER BOUND TO THE ENTROPY OF A DIGITAL SEQUENCE

In this section, we shall derive a general upper bound to the entropy of a sequence with alphabet V, probability distribution $p_X(v_k)$, and power spectrum $S_X(\omega)$. First we replace the past in (2) by a function of the past to obtain

$$H(X_t | X_{t-1}, X_{t-2}, \cdots) \leq H(X_t | \phi(X_{t-1}, X_{t-2}, \cdots)). \tag{5}$$

The type of function we have in mind is a predictor for $\overline{X}$, i.e., a function which provides an estimate of X_t. The idea is that a sequence which has a predictor with small prediction error must have a small entropy, whereas a sequence which cannot be predicted well has a large entropy. The predictor that minimizes the variance of the prediction error is $E[X_t | X_{t-1}, X_{t-2}, \cdots]$. However, this function depends on properties of the sequence about which we have no information. Knowledge of the correlation sequence, however, suffices to determine the linear minimum mean-square error (LMMSE) predictor

$$\hat{X}_t = \sum_{j=1}^{\infty} h_j X_{t-j}. \tag{6}$$

From the orthogonality condition for LMMSE estimation [9, p. 236] we get

$$E[e_t X_{t-j}] = 0, \quad j > 0, \tag{7}$$

where $e_t = X_t - \hat{X}_t$ is the prediction error. Combining (6) and (7) we get the Wiener–Hopf equation

$$\sum_{j=1}^{\infty} h_j R_X(k - j) = R_X(k), \quad k = 1, 2, \cdots. \tag{8}$$

The variance of the LMMSE prediction error may be expressed directly in terms of the power spectrum [10, p. 301] as

$$\sigma_e^2 = \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log S_X(\omega) \, d\omega \right\}. \tag{9}$$

If $S_X(\omega)$ vanishes for certain isolated real frequencies, the predictor must be expressed in a somewhat different form. We shall assume that $\overline{X}$ has no deterministic component. In that case [10, p. 292]

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log S_X(\omega) \, d\omega > -\infty \tag{10}$$

and the prediction error is still given by (9). A good approximation to the LMMSE predictor is obtained by restricting the sum in (6) to a finite number of terms. However, for rational power spectra we can get an exact expression by using recursive estimation. Let the power spectrum be given by

$$S_X(\omega) = \left| \sum_{j=0}^{N} f_j e^{-i\omega j} \bigg/ \sum_{j=0}^{N} g_j e^{-i\omega j} \right|^2. \tag{11}$$

For such spectra the analysis is much simpler [11, pp. 121–125] and the predictor has a more satisfactory interpretation. The condition (10) assures that $\overline{X}$ can be obtained by passing a sequence of uncorrelated random variables $\overline{U}$ through a linear filter [10, p. 293], and from (11) it follows that the transfer function of the filter can be taken as

$$G(\omega) = \sum_{j=0}^{N} f_j e^{-i\omega j} \Big/ \sum_{j=0}^{N} g_j e^{-i\omega j}, \qquad (12)$$

where we assume that $g_0 = 1$ and that all zeros of the denominator are outside the unit circle. The transfer function (12) corresponds to the difference equation

$$\sum_{j=0}^{N} g_j X_{t-j} = \sum_{j=0}^{N} f_j U_{t-j}. \qquad (13)$$

In the literature on recursive filtering [9, pp. 255–262], this equation is usually written as a first-order vector equation. For the present purpose, however, we prefer a scalar equation of higher order. The term $f_0 U_t$ is uncorrelated with the past and may be identified as the prediction error e_t. Rewriting (13), we find that $\hat{X}_t = X_t - e_t$ is determined recursively as

$$\sum_{j=0}^{N} g_j X_{t-j} = e_t + \sum_{j=1}^{N} f_j \left(X_{t-j} - \hat{X}_{t-j} \right)$$

or

$$\hat{X}_t = -\sum_{j=1}^{N} f_j \hat{X}_{t-j} + \sum_{j=1}^{N} (f_j - g_j) X_{t-j}. \qquad (14)$$

If $S_X(\omega)$ has no real zeros, the numerator of $G(\omega)$ does not have any zeros on the unit circle. In this case the initial conditions are unimportant if the recursion was started in the distant past, and the stochastic sequence $\overline{X}$ is determined by (14). However, if $S_X(\omega') = 0$ for some frequency ω', the initial conditions must be chosen in such a way that $\overline{X}$ does not contain any deterministic component with this frequency. In particular, if $S_X(0) = 0$, a sample function of $\overline{X}$ is determined by the corresponding *sample function of $\overline{X}$ only up to an additive constant which must be chosen to make the time average vanish.* This point will be discussed in greater detail in Section IV.

We now return to the bound on the entropy. We choose the LMMSE predictor $\hat{X}_t$ as the function ϕ in (5) so that

$$H(X) \leq H\left(X_t \mid \hat{X}_t \right). \qquad (15)$$

Now, let F be set of real-valued random variables, Y, satisfying

$$E\left[(X_t - Y)^2 \right] = \sigma_e^2.$$

Thus $\hat{X}_t \in F$. We may then obtain the desired bound as

$$H(X) \leq H\left(X_t \mid \hat{X}_t \right) \leq \max_{Y \in F} \left\{ H(X_t \mid Y) \right\}$$

$$= H(X_t) - \min_{Y \in F} \left\{ I(X_t; Y) \right\}. \qquad (16)$$

Here $H(X_t)$ depends only on the probability distribution

on V, and the term $\min\{I(X_t; Y)\}$ is a point on the rate distortion function $r^*(D)$ for a memoryless source with probability distribution $p_X(v_k)$ [12, p. 23]. We restate this result as a theorem.

Theorem 2: The entropy of a stationary ergodic sequence with alphabet V, probability distribution $p_X(v_k)$, and power spectrum $S_X(\omega)$ is upper-bounded by

$$H(X) \leq H(X_t) - r^*(\sigma_e^2),$$

where r^* is the rate distortion function for a memoryless source with probability distribution $p_X(v_k)$, for the set of real numbers as the reproducing alphabet and mean-square error as the distortion measure and the distortion σ_e^2 is the minimum mean-square error for linear prediction.

The rate distortion problem defined above is somewhat unusual since it combines a discrete source alphabet and a continuous reproducing alphabet. However, the basic properties of the $r^*(D)$ curve follow immediately from the analysis of discrete sources in [12, pp. 20–46]. $r^*(D)$ is a convex function with $r^*(0) = H(X_t)$ and $r^*(\sigma_x^2) = 0$. It has infinite slope at $D = 0$ and usually nonzero slope at $D = \sigma_x^2$. The probability distribution of Y is discrete with at most $K + 1$ active reproducing symbols for a given value of D.

It is convenient to distinguish between the $K + 1$ abstract reproducing symbols $y_0, y_1, \cdots, y_K$ and the set of real numbers consisting of the value $w_j(D)$ associated with each y_j for a particular value of D. The w_j satisfy

$$E[X_t \mid y_j] = w_j(D) \qquad (17)$$

since this choice of the real numbers minimizes D for a given set of conditional probabilities. Thus

$$\sigma_y^2 = \sigma_x^2 - \sigma_e^2 = \sigma_{\hat{x}}^2. \qquad (18)$$

$r^*(D)$ may be computed by an extended version of Blahut's algorithm [13]. For a given value of the slope s, this algorithm generates a convergent set of conditional probabilities. We have extended the algorithm by using (17) to adjust the reproducing alphabet and hence the distortions between v_k and y_j.

We conjecture that, under mild restrictions on the probability distribution of X_t, the $w_j(D)$ are continuous functions of D in an interval of the form $0 \leq D \leq D_j$. Clearly we can take $w_j(0) = v_j$ and at least two functions are defined in the entire interval from 0 to σ_x^2 and have $w_j(\sigma_x^2) = 0$. We expect the number of active reproducing symbols to be monotonically decreasing for increasing D. A symbol y_j may disappear for $D = D_j$ when its probability reaches 0, or it may merge with another symbol with the same value $w_k(D_j) = w_j(D_j)$. These assumptions facilitate the use of the iterative algorithm, but we have been unable to prove that they are always true. However, we may verify that a computed set of conditional probabilities corresponds to a point on the $r^*(D)$ curve by testing the condition [12, p. 35]

$$c(w) = \sum_k p_X(v_k)\lambda_k \exp\left[s(w - v_k)^2 \right] \leq 1, \qquad (19)$$
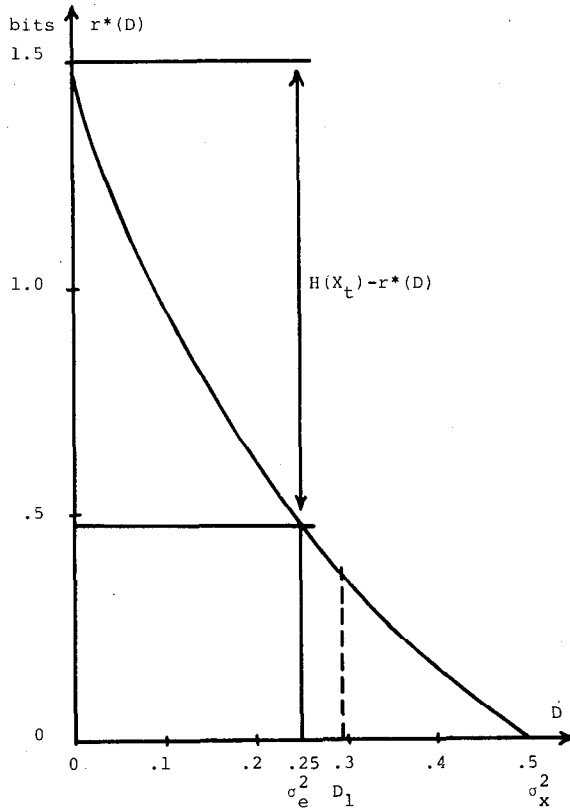
Fig. 1. Rate distortion function for ternary source and construction of the upper bound to the rate of the bipolar code.

where

$$\lambda_j^{-1} = \sum_k p_Y(y_k) \exp\left[s(w_k - v_j)^2\right].$$

Equation (19) must be satisfied for all real w, and for each active reproducing symbol y_j it is satisfied with equality when $w = w_j$. Thus it is sufficient to determine the values of the relative maxima for $c(w)$, all of which must be less than or equal to one.

Example 2: Fig. 1 shows the rate distortion function for a ternary source with $V = \{0, \pm 1\}$ and $p(0) = 1/2$, $p(1) = p(-1) = 1/4$. In Example 1, we described a code with spectrum $(1 - \cos \omega)/2$. The LMMSE predictor for a sequence with this spectrum is

$$\hat{X}_t = \hat{X}_{t-1} - X_{t-1}$$

and the prediction error has variance $\sigma_e^2 = 1/4$. For this case the upper bound of Theorem 2 becomes $H(X) \le 1.024$ bit. Thus at most a small increase in the rate of the AMI code is possible.

Fig. 2 presents the rate distortion function for a source with five symbols. For small distortions, there are five reproducing symbols, while in an interval $D_1 \le D \le D_2$ only three symbols are used, and finally, for large distortions, only two reproducing symbols are active.

In most of the cases considered in the following sections the rate distortion functions are evaluated at points where the only active reproducing symbols are $\pm \sigma_{\hat{X}}$.
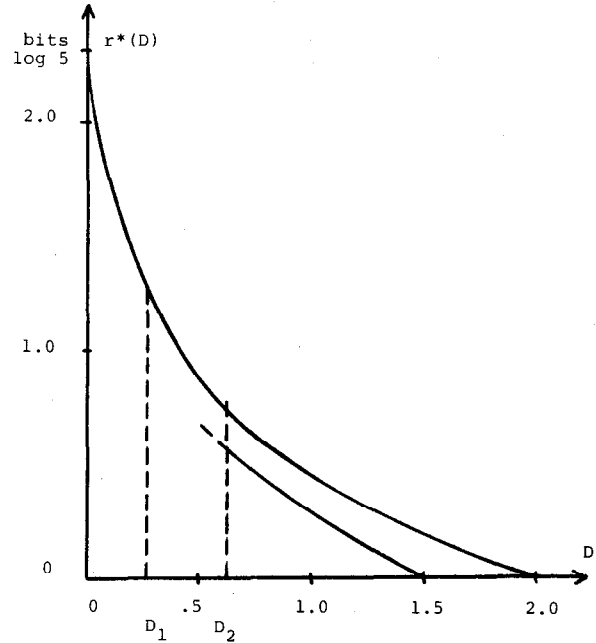

Fig. 2. Rate distortion function for a source with five equiprobable symbols and a segment of the function for a source with symbol distribution $(1/8, 1/4, 1/4, 1/4, 1/8)$.

IV. PROPERTIES OF DC-FREE SEQUENCES

In most applications of spectrum shaping codes, it is desirable that the code be dc-free, i.e., that $S_X(0) = 0$, that the spectrum remain low up to some cut-off frequency and be approximately constant over the rest of the frequency band. The exact form of the power spectrum is usually less important than the complexity of the coding algorithm.

The performance of a dc-free code is often expressed in terms of the set of values that the sum of the encoded symbols

$$Z'_{t_1 t_2} = \sum_{k=t_1}^{t_2} X_k$$

can take [3], [4], [14]. The most common performance index is the digital sum variation

$$\Delta_Z = \max_{t, t_2} \left\{ Z'_{t_1 t_2} \right\}.$$

In this section we shall make this concept more precise and give some analytical justification for using the variance of the sum as a measure of the bandwidth of the encoded sequence.

Assume, as in the previous section, that $\overline{X}$ has no deterministic component, so that (10) is satisfied, and $\overline{X}$ can be obtained by passing a sequence $\overline{U}$ of uncorrelated random variables through a causal linear filter. We shall not require that the transfer function $G(\omega)$ be rational. Further, let $S_X(0) = 0$ and let the derivative of $S_X(\omega)$ be continuous at $\omega = 0$. Then $S_X(\omega)/(1 - \cos \omega)$ has a finite limit for $\omega \to 0$, and we can write $G(\omega)$ as a product of the transfer functions $G'(\omega)$ and $G''(\omega)$ for two causal filters

$$G'(\omega) = G(\omega)/\left(1 - e^{-i\omega}\right)$$

$$G''(\omega) = 1 - e^{-i\omega}. \tag{20}$$

When $\bar{U}$ is the input to the first filter, the output is a stationary random sequence $\bar{Z}$ with power spectrum

$$S_Z(\omega) = (1/2)S_X(\omega)/(1 - \cos \omega). \qquad (21)$$

The difference equation for the second filter is then

$$Z_t = Z_{t-1} + X_t. \qquad (22)$$

As noted in Section III, the sequence $\bar{Z}$ is only uniquely defined if we specify that it contains no deterministic component. It follows from (22) that $\bar{Z}$ may be interpreted as the sum of the encoded symbols with a correction to ensure that the time average vanishes. Thus this sum is well defined for any purely random dc-free sequence whose spectrum is smooth at $\omega = 0$. The variance of the sum may be found by integration of (21) or it may be expressed directly in terms of $R_X(k)$. Consider

$$Z_k - Z_0 = X_1 + X_2 + \cdots + X_k.$$

By taking the variance of this variable we get

$$2\sigma_z^2 - 2R_Z(k) = E\left[\left(\sum_{j=1}^{k} X_j\right)^2\right]$$

$$= \sum_{j=-k+1}^{k-1} (k - |j|)R_X(j)$$

$$= k \sum_{j=-k+1}^{k-1} R_X(j) - 2\sum_{j=1}^{k-1} jR_X(j). \qquad (23)$$

When we take the limit for $k \to \infty$ in (23) and use

$$\lim_{k\to\infty} R_Z(k) = 0,$$

and

$$\lim_{k\to\infty} \sum_{j=-k+1}^{k-1} R_X(j) = S_X(0) = 0,$$

we get

$$\sigma_z^2 = -\sum_{k=1}^{\infty} kR_X(k). \qquad (24)$$

With this result $R_Z(k)$ may be found from (23) in terms of the correlation function for $\bar{X}$. In many cases σ_z^2 can be determined directly from the definition of the code even when no closed form expression for the spectrum is available.

It is not clear from the results above how the sum is related to the cut-off frequency. As discussed in [4] the obvious approximation to the power spectrum for small values of ω would be

$$S_X(\omega) \simeq -\omega^2 \sum_{k=1}^{\infty} k^2 R_X(k), \qquad (25)$$

which is not related to the sum in any simple way.

We can obtain a class of simple power spectra with the properties mentioned in the beginning of this section by
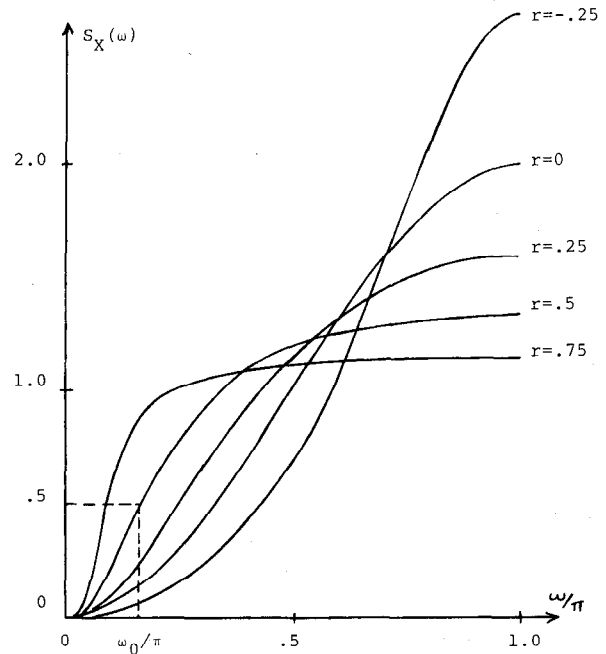

Fig. 3.   First-order power spectra. As an example the cut-off frequency is shown for $r = 1/2$.

taking the first-order rational functions

$$S_X^r(\omega) = \sigma_x^2(1 + r)(1 - \cos \omega)/(1 + r^2 - 2r\cos \omega),$$

$$|r| < 1, \quad (26)$$

or equivalently

$$R_X^r(k) = -(1/2)\sigma_x^2(1 - r)r^{|k|-1}, \qquad k \neq 0. \quad (27)$$

The peak values of these spectra are $S_X^r(\pi) = 2(1 + r)^{-1}\sigma_x^2$. In Fig. 3, $S_X^r(\omega)$ is plotted for several values of r. A sequence with spectrum $S_X^r(\omega)$ may be obtained by passing an uncorrelated sequence $\bar{U}$ through a filter with transfer function

$$G^r(\omega) = (1 - e^{-i\omega})/(1 - re^{-i\omega}). \qquad (30)$$

It is convenient to define the cut-off frequency [15], ω_0, by

$$S_X(\omega_0) = \sigma_x^2/2. \qquad (31)$$

For $S_X^r(\omega)$ we get

$$1 - \cos \omega_0 = (1 - r)^2/2. \qquad (32)$$

For small values of $1 - r$ we may use the approximation

$$\omega_0 \simeq 1 - r. \qquad (33)$$

Actually, this approximation is within one percent even for $r = 1/2$. Using (14) we find the best linear predictor to be

$$\hat{X}_t = \hat{X}_{t-1} - (1 - r)X_{t-1}. \qquad (34)$$

Taking variances on both sides of this equation, and using (7) we get

$$\sigma_{\hat{x}}^2 = (1 - r)\sigma_x^2/2, \quad \sigma_e^2 = (1 + r)\sigma_x^2/2. \qquad (35)$$

Comparing (34) and (22), we see that

$$\hat{X}_t = -(1 - r)Z_{t-1} \qquad (36)$$

and thus

$$\sigma_z^2 = (1 - r)^{-1}\sigma_x^2/2. \qquad (37)$$

Thus the cut-off frequency may be approximated by

$$\omega_0 \simeq \sigma_x^2/(2\sigma_z^2). \qquad (38)$$

This expression is very convenient because it can often be evaluated by simple calculations even though the spectrum is a complicated function. For first-order spectra the approximation (25) becomes $\omega_0 \simeq (1 + r)^{-1/2}(1 - r)$, which is not so close to the acutal value (32) as (33). Thus, for first-order spectra, the cut-off frequency is proportional to the inverse of the variance of the sum, and the sum has the additional interpretation of being proportional to the best linear predictor. For spectra that are similar to first-order spectra, we shall use (38) as an approximation to the cut-off frequency. In all cases where the spectrum can readily be obtained by calculation or from the literature, we have confirmed the value obtained in this way.

The codes of greatest practical interest have alphabets of the form

$$V_{2K} = \{\pm 1, \pm 3, \pm(2K - 1)\}$$

or

$$V_{2K+1} = \{0, \pm 1, \pm 2, \cdots, \pm K\}. \qquad (39)$$

In this case Z_t must take values from a set of the form

$$W_\alpha = \{\alpha, \alpha \pm 1, \alpha \pm 2, \cdots, \},$$

where finitely or infinitely many elements may have non-zero probabilities. It follows from (36) that $\hat{X}_t$ for first-order spectra takes values of the form $(1 - r)(\alpha \pm j)$. If V has an even number of elements, $\overline{Z}$ must alternate between two subsets of W_α since an odd integer is added in each step. Most of the codes that have been suggested in the literature have a symmetry that leads to $\alpha = 0$ or $\alpha = 1/2$. However, one of the difficult problems in constructing a sequence of maximum entropy is to determine the best value of α.

If $\overline{X}$ is a Markov source, the situation is particularly simple. We shall prove that Z_t is a function of the state $S(t)$. Let $\zeta(s_j) = z_j$ indicate the value of the sum associated with state s_j and let the symbol from V that is produced on transition from state s_j to s_k be denoted by x_{jk}.

Theorem 3: If $\overline{X}$ is an ergodic dc-free Markov source, there is a unique function of the state, $\zeta(S(t))$, such that

$$\sum_j p_S(s_j)z_j = 0 \qquad (40)$$

and

$$x_{jk} = z_k - z_j. \qquad (41)$$

Proof: There is at most one function which satisfies (40) and (41). For a given z_0, the remaining z_j can be found by repeated application of (41) since all states can be reached. If a constant is added to z_0, the same constant is added to all other z_j. Thus there is exactly one choice for z_0

that makes the mean value vanish. This set of z_j provides the correct initial conditions for the recursion (22), and for first-order spectra, the initial conditions for (14) are obtained via (36). In order to prove that the function ζ exists, we must prove that (41) cannot lead to conflicting assignments of z_j. It is convenient to think of a sequence of state transitions as a path through the graph that has a node representing each state and a directed branch for each transition [5, p. 64]. A sample function $\overline{x} = \cdots, x_{t-1}, x_t, x_{t+1}, \cdots$ may be divided into cycles which start and end in state s_0 and which do not pass through s_0. Let the sum of the x_{ij} along the jth cycle be ν_j, and let μ denote the finite average length of the cycles. The power spectrum of the source can be calculated from the sample function, and in particular for $\omega = 0$ we have

$$S_X(0) = \lim_{L \to \infty} L^{-1}\left(\sum_{j=1}^{L} x_j\right)^2. \qquad (42)$$

Rewriting this sum in terms of the cycles, we get

$$\lim_{M \to \infty} (\mu M)^{-1}\left(\sum_{j=1}^{M} \nu_j\right)^2 = 0. \qquad (43)$$

However, in (43) we could interpret the ν_j as a sample function for a sequence of independent random variables. Consequently, (43) implies that $\nu_j = 0$ for all j. Assume that, starting from a particular value z_0, we reach a state s_j with two different sums, $z_j' \neq z_j''$, by following two different paths. These paths could be extended to cycles with different sums, and at least one cycle would have a nonzero sum. Thus, the assumption leads to a contradiction.

V. Construction of Markov Sources with Given Power Spectra

If the transition probabilities and the output function for a Markov source are given, the power spectrum may be determined by a straightforward calculation. We shall restrict our presentation to the case of a state output, $Z_t = \zeta(S(t))$. It follows from Theorem 3 that, for dc-free Markov sources, the sum is a function of the state. The spectrum of $\overline{X}$ is related to $S_Z(\omega)$ by (21).

For a source with transition matrix Q, and stationary probability distribution $\overline{p}_S$, $R_Z(k)$ may be expressed as

$$R_Z(k) = \overline{\zeta}^T Q^k \overline{\zeta}_p, \qquad k \geq 0, \qquad (44)$$

where $\overline{\zeta}$ and $\overline{\zeta}_p$ indicate the column vectors

$$\overline{\zeta}^T = (\zeta(s_0), \zeta(s_1), \cdots, \zeta(s_N))$$

$$\overline{\zeta}_p^T = (\zeta(s_0)p_S(s_0), \zeta(s_1)p_S(s_1), \cdots, \zeta(s_N)p_S(s_N)).$$

From the Cayley–Hamilton theorem Q satisfies the linear recursion $C(Q) = 0$ where $C(\lambda) = |E - \lambda Q|$ and E is a unit matrix. From (44) $R_Z(k)$ satisfies the same recursion. If Q has distinct eigenvalues $1, \lambda_1, \lambda_2, \cdots, \lambda_N$ and right

eigenvectors $\bar{p}_S, \bar{u}_1, \bar{u}_2, \cdots, \bar{u}_N$, we can write (44) as

$$R_Z(k) = \bar{\zeta}^T \left(\eta_0 \bar{p}_S + \sum_{j=1}^{N} \lambda_j^k \eta_j \bar{u}_j \right), \qquad k \geq 0. \quad (45)$$

Usually, the most efficient method for computing the spectrum will be to use the recursion for $R_Z(k)$ and to find the initial conditions from (44). However, it may be possible to simplify the recursion if it is known that some of the terms in (45) are zero. Thus, if it has been assured that

$$\sum_{j=0}^{N} p_S(s_j)\zeta(s_j) = \bar{\zeta}^T \bar{p}_S = 0,$$

the factor $\lambda - 1$ in the recursion may be canceled. As a more important example, many sources of interest satisfy the symmetry conditions $\zeta(s_j) = -\zeta(s_{N-j})$ and $p_{ij} = p_{N-i, N-j}$. If N is odd, the transition matrix may be expressed as

$$Q = \begin{bmatrix} Q' & Q'' \\ \tilde{Q}'' & \tilde{Q}' \end{bmatrix},$$

where $\tilde{Q}'$ indicates that the rows and columns of Q' have been reversed. In this case half of the terms in (45) vanish because the eigenvectors satisfy $u_{jk} = u_{j, N-k}$. The recursion may be simplified to $C(\lambda) = |E - \lambda(Q' - \bar{Q}'')|$, where $\bar{Q}''$ is obtained by reversing the order of the columns in Q''. The power spectrum may be expressed in closed form as

$$S_Z(\omega) = \sum_{-\infty}^{\infty} R_Z e^{-i\omega k}$$

$$= -R_Z(0) + \bar{\zeta}^T (E - e^{i\omega}Q)^{-1} \bar{\zeta}_p$$

$$+ \bar{\zeta}^T (E - e^{-i\omega}Q)^{-1} \bar{\zeta}_p. \quad (46)$$

It follows from (46) that the denominator of $S_Z(\omega)$ and $S_X(\omega)$ divides $\|E - Qe^{-i\omega}\|^2$. However, it is a rather time-consuming task to invert the matrix in order to find the numerator of the spectrum.

The method outlined above for computing the power spectrum is also useful for constructing Markov sources with given spectra when the number of states is sufficiently small. In general there seems to be no practical procedure for constructing a source with a prescribed spectrum. We shall describe a method which can be applied in many cases of interest in the present context, and we shall use the entropies of these Markov sources as lower bounds to the rate of the best codes.

The construction is based on making the additional assumption that the LMMSE predictor for $\bar{X}$ coincides with the conditional expectation

$$\hat{X}_t = E[X_t | s_k] = \sum_j p_{kj} x_{kj}. \quad (47)$$

It is not clear when this extra condition can be satisfied or how a process with this property compares with the maxentropic sequence. However, if it is possible to construct a source satisfying (47), then it has a rational power spec-

trum and, since the best linear predictor has a correct form, it follows from (14) that the sequence has the desired power spectrum up to a constant factor. The construction of a finite-state Markov source satisfying (47) is possible only if $\hat{X}_t$ assumes values from a finite set.

If $S_X(\omega) = |G(e^{-i\omega})|^{-2}$, where G is a polynomial of degree N, then the linear predictor has only N terms

$$\hat{X}_t = \sum_{j=1}^{N} h_j X_{t-j}, \quad (48)$$

and we may attempt at constructing an Nth order Markov process as a Markov source with one state for each of the K^N values of $X_{t-1}, X_{t-2}, X_{t-3}, \cdots, X_{t-N}$. To each of these states (48) assigns a value of the linear predictor, and we should adjust the transition probabilities in such a way that (47) is satisfied. Slepian [2] constructed Markov sources with the same choice of states as discussed here. Binary sources with $N = 2$ and ternary sources with $N = 1$ must have transition probabilities that satisfy (47) in order to have the required values of $R_X(0)$, $R_X(1)$, and, for the binary sources, $R_X(2)$. Since Slepian proves that such sources have maximum entropy for these values of the correlation sequence, they are certainly maxentropic for the spectra in question. For larger N and larger alphabets, Slepian's approach leads to sources with more complicated spectra. For binary sources (47) becomes

$$P(1|s_j) - P(-1|s_j) = 1 - 2P(-1|s_j) = \zeta(s_j).$$

$$(49)$$

Thus the transition probabilities are uniquely determined by the vector $\bar{\zeta}$.

Example 3: The spectrum $S_X(\omega) = |1 - (1/2)e^{-i\omega} + (1/2)e^{-2i\omega}|^{-2} = (3/2 - (3/2)\cos \omega + \cos 2\omega)^{-1}$ may be obtained with a four-state binary Markov source. The LMMSE predictor for this spectrum is

$$\hat{X}_t = (1/2)X_{t-1} - (1/2)X_{t-2}$$

and takes values 0 and ± 1. Let the states s_0, s_1, s_2, s_3 correspond to $(X_{t-1}, X_{t-2}) = (1, 1)$, $(1, -1)$, $(-1, 1)$, $(-1, -1)$. $\hat{X}_t$ is a function $\zeta(S(t))$ with $\zeta(s_0) = 0$, $\zeta(s_1) = 1$, $\zeta(s_2) = -1$, $\zeta(s_3) = 0$.

Using (49), we get the transition probabilities

$$Q = \begin{bmatrix} 1/2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/2 \end{bmatrix},$$

and the stationary probabilities are $p_S(s_0) = p_S(s_3) = 1/3$, $p_S(s_1) = p_S(s_2) = 1/6$. The source emits runs of 1's and -1's with at least two symbols in each run. The entropy is $2/3$ bits. An encoder would encode a binary symbol each time the machine is in state s_0 or s_3. This sequence may be seen as a simplified model for delay modulation or Miller code [17] which has a similar spectrum. This code has rate $1/2$ and is used in magnetic recording. By changing a sign

to $S_X(\omega) = (3/2 + (3/2)\cos\omega + \cos 2\omega)^{-1}$ we get a source that emits runs of length two at the most.

Example 4: For a binary code, let $R_X(1) = -1/2$, $R_X(k) = 0$ for $k > 1$. In a later section we shall determine the maxentropic sequence for this spectrum. Here we shall consider a third-order Markov process which has the same values of the correlation for $k \leq 4$. Solving (8) we get the predictor

$$\hat{X}_t = -(3/4)X_{t-1} - (1/2)X_{t-2} - (1/4)X_{t-3}$$

and consequently

$$S_X(\omega) = |1 + (3/4)e^{-i\omega} + (1/2)e^{-2i\omega} + (1/4)e^{-3i\omega}|^{-2}.$$

Since (47) can be satisfied only for $|\bar{X}_t| \leq 1$, the variables $X_{t-1}, X_{t-2}, X_{t-3}$ cannot all take the same value. If the states $s_0, s_1, \cdots, s_5$ correspond to $(X_{t-1}, X_{t-2}, X_{t-3}) = (1, 1, -1), (1, -1, 1), (1, -1, -1), (-1, 1, 1), (-1, 1, -1), (-1, -1, 1)$, the transition matrix becomes

$$Q = \begin{Bmatrix} 0 & 1/4 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 3/4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3/4 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/4 & 0 \end{Bmatrix}.$$

In both of the previous examples the denominator of $S_X(\omega)$ is $\|E - e^{-i\omega}(Q' - \bar{Q}'')\|^2$.

We shall apply the same approach to sources with the spectrum $S_X^r(\omega)$ of (26). From Theorem 3, we have that $Z_{t-1} = \zeta(S(t))$ and (36) now shows that

$$\hat{X}_t = -(1 - r)\zeta(S(t)).$$

In this section, we shall let each value of Z_t correspond to a unique state. We first select a value of α, $0 \leq \alpha < 1$. For each element in the set W_α that has nonzero probability, (47) provides a linear constraint on the transition probabilities. It follows from (45) that if the transition probabilities satisfy these constraints, r is an eigenvalue of Q. We consider the most important alphabets in detail.

A binary source can satisfy (47) only if $|\hat{X}_t| \leq 1$. Since the numerical value of Z_t is limited by this constraint, there must be states s_0 and s_N with

$$P[1|s_0] = P[-1|s_N] = 0,$$

and

$$\zeta(s_0) = (1 - r)^{-1}, \quad \zeta(s_N) = -(1 - r)^{-1}$$

since otherwise the numerical value of Z_t could increase further. Thus the construction is possible only if $1 - r$ divides 2. For negative values of r there is no binary source since $R_X(1)$ and $R_X(2)$ cannot have the required values [2]. We have not found any finite-state Markov source with spectrum $S_X^r(\omega)$ for values of r other than 0, 1/3, 1/2, 3/5, 2/3, $\cdots$. For these values of r, (47) is satisfied by a unique source with $N + 1 = 1 + 2/(1 - r)$ states, stationary probability distribution, transition probabilities,
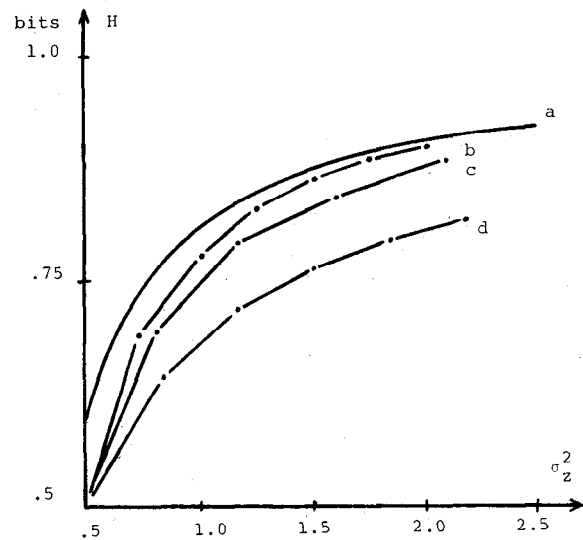

Fig. 4. Comparison of binary dc-free codes. a) Upper bound for codes with first-order spectra. b) Lower bound for first-order spectra. c) Sequences with limited digital sum variation. d) Balanced block codes.

and entropy given by

$$p_S(s_k) = 2^{-N}\binom{N}{k}$$

$$p_{k, k-1} = k/N$$

$$H_N = 2^{-N} \sum_{k=0}^{N} \binom{N}{k} \mathcal{H}(k/N), \qquad (50)$$

where $\mathcal{H}(p)$ is the binary entropy function.

As was discussed in Section III, the bound of Theorem 2 can be evaluated explicitly for binary codes

$$H \leq \mathcal{H}(1/2 - \sigma_z/2) \qquad (51)$$

and for first-order spectra

$$H \leq \mathcal{H}\left(1/2 - (1/2)\sqrt{(1 - r)/2}\right). \qquad (52)$$

For a small number of states, (52) is a rather loose bound, and we shall derive a stronger result in the following section. For large N, the difference between the entropy given by (50) and the upper bound (52) decreases. Both functions are plotted in Fig. 4.

It is often useful to adjust the probabilities of the encoded strings by splitting certain states in the source. We use the concept of expanding a Markov chain [8, pp. 140–145] such that a transition in the original chain is replaced by a state in the expanded chain. Thus, if there are transitions from states s_i and s_j to s_k, then s_k is split into s_k', which can be reached from s_i, and s_k'', which can be reached from s_j. In a later section, we shall use this method to obtain a prescribed correlation function. In the following example, the splitting of certain states is used to facilitate the encoding. An optimum binary variable length code can be constructed when the probabilities of the source strings are inverse powers of two [5, pp. 43–48]. We shall obtain an encoding of a Markov source by adjusting

the transition probabilities in such a way that all strings that start and end in a certain subset of the states have probabilities of the required form.

Example 5: For $r = 1/2$, we obtain a five-state source with transition matrix and sum vector

$$Q = \begin{Bmatrix} 0 & 1/4 & 0 & 0 & 0 \\ 1 & 0 & 1/2 & 0 & 0 \\ 0 & 3/4 & 0 & 3/4 & 0 \\ 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 1/4 & 0 \end{Bmatrix}$$

$$\bar{\zeta} = \begin{Bmatrix} 2 \\ 1 \\ 0 \\ -1 \\ -2 \end{Bmatrix}.$$

This sequence has entropy $H_4 = (3/8)\mathcal{H}(1/2) + (1/2)\mathcal{H}(1/4) \simeq 0.781$ bit. By splitting states s_1, s_2, and s_3 and adjusting the transition probabilities we obtain the following transition matrix and sum vector:

$$Q' = \begin{Bmatrix} 0 & 1/4 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 2/3 & 0 & 0 & 2/3 & 0 & 0 \\ 0 & 3/4 & 0 & 0 & 0 & 0 & 3/4 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 1/4 & 0 \end{Bmatrix}$$

$$\bar{\zeta} = \begin{Bmatrix} 2 \\ 1 \\ 1 \\ 0 \\ 0 \\ -1 \\ -1 \\ -2 \end{Bmatrix}.$$

All strings that start and terminate in states s_1' or s_3'' can be encoded by an optimum prefix code. The state sequence s_1', s_2'', s_1'', s_0, s_1' which produces the output $-1, 1, 1, -1$ has probability $3/4 \cdot 1/2 \cdot 1/3 \cdot 1 = 1/8$. Thus, this output can be used as an encoding of three binary data symbols. In Table I we give a prefix code for the strings that start in s_1'. The strings from s_3' are encoded similarly.

This code has $\sigma_z^2 = 15/14$ and entropy $H = 11/14 \simeq 0.786$ bit. A block code with block length four and rate $3/4$ has been suggested for optical communication [14]. Such a code can be selected to have the same distribution of the sum as the Markov source with $r = 1/2$.

Franklin and Pierce [15] discussed simple binary dc-free block codes. They suggested a code consisting of all blocks of length $2n$ with an equal number of 1's and -1's. The entropy of such a code is

$$H = \frac{1}{2n} \log \binom{2n}{n} \simeq 1 - \frac{1}{4n} \log \pi n. \tag{53}$$

TABLE I
VARIABLE LENGTH ENCODING OF BINARY CODE WITH $r = 1/2$

Data	Codeword
11	+ −
101	− + + −
100	− − − +
010	− + − +
011	− + − −
001	− − + +
000	− − + −

The spectrum of the encoded sequence was derived in [15] and a table of the entropies and cut-off frequencies of several codes was given. It is easier to determine the sum variance. As noted in [15], the correlation between two positions in the same block satisfies

$$E[X_k X_l] = \frac{-1}{2n - 1}, \qquad k \neq l,$$

and thus the correlation function for the encoded sequence is

$$R_X(k) = \begin{cases} 1, & k = 0, \\ \dfrac{k - 2n}{2n(2n - 1)}, & |k| \le 2n, \\ 0, & |k| > 2n, \end{cases}$$

and from (24) we get

$$\sigma_z^2 = (2n + 1)/6. \tag{54}$$

The approximate value of ω_0 calculated from (54) using (38) is about five percent greater than the value determined directly from the spectrum for the example considered. The values plotted in Fig. 4 indicate that the cut-off frequencies of these block codes are significantly lower than those obtained with the Markov sources constructed above.

Example 6: A code that has four 1's in each block of eight symbols has entropy $H = (1/8)\log 70 \simeq 0.766$ bit. The cut-off frequency is approximately $\omega_0 \simeq \sigma_z^{-2}/2 = 1/3$. The true cut-off frequency of the block code is 0.321 [15]. However, with $r = 1/2$ the sequence of Example 5 has slightly better entropy and about 50 percent better cut-off frequency.

Chien [17] considered the problem of determining sequences of maximum entropy for given digital sum variation. The sequence is a Markov source with one state for each value of the sum. In order to determine the transition probabilities that maximize the entropy of an N-state source, one must determine the eigenvector $\bar{p}_a$ that satisfies

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 1 & 0 & 1 & 0 & \cdots \\ 0 & 1 & 0 & 1 & \cdots \\ \cdots\cdots\cdots\cdots\cdots \end{bmatrix} \cdot \bar{p}_a = \lambda \bar{p}_a,$$

where $\lambda = 2\cos \pi/(N + 1)$ is the maximal eigenvalue of the N by N adjacency matrix. The entropy is $\log \lambda$. The transition probabilities are then given by

$$p_{ij} = \lambda^{-1} p_{aj}/p_{ai}. \tag{55}$$

It may be readily verified that the transition matrix Q with these transition probabilities has eigenvector

$$\left(p_{a1}^2, \, p_{a2}^2, \, \cdots \right)$$

with eigenvalue one. Thus, the stationary probability distribution becomes

$$\bar{p}_S = \left\{ \frac{2}{N+1} \sin^2 \frac{\pi k}{N+1} \right\}, \qquad k = 1, 2, \cdots, N, \quad (56)$$

and the sum variance is

$$\sigma_z^2 = \frac{2}{N+1} \sum_{k=1}^{N} \left(\frac{N+1}{2} - k \right)^2 \sin^2 \frac{\pi k}{N+1}. \quad (57)$$

Again several values of the entropies of these sequences are plotted in Fig. 4. It is clear from these data that although sequences with a small digital sum variation have fairly good spectra, better results may be obtained by allowing a larger number of states in the Markov sources.

Example 7: If the sum is allowed to take values $0, \pm 1, \pm 2$, the maximal entropy is obtained by considering the matrix

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

which has maximal eigenvalue $\lambda = \sqrt{3}$. Thus, the entropy becomes $H = (1/2) \log 3 \simeq 0.793$ bit. The eigenvector is

$$\left(\frac{1}{2}, \frac{\sqrt{3}}{2}, 1, \frac{\sqrt{3}}{2}, \frac{1}{2} \right).$$

The transition matrix for the source with greatest entropy is

$$Q = \begin{bmatrix} 0 & 1/3 & 0 & 0 & 0 \\ 1 & 0 & 1/2 & 0 & 0 \\ 0 & 2/3 & 0 & 2/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 1/3 & 0 \end{bmatrix}$$

with stationary distribution $\bar{p} = (1/12, 1/4, 1/3, 1/4, 1/12)$. The sum variance becomes $\sigma_z^2 = 7/4$. Thus, the sequence is somewhat inferior to the Markov sources with $r = 1/2$ and $r = 3/5$.

It may be noted that the three classes of binary sequences considered here have the same first member. This is the biphase sequence which may be described either as balanced blocks of length two or as a three-state Markov source with spectrum $S_X^0(\omega) = 1 - \cos \omega$.

For ternary sequences with spectra $S_X^r(\omega)$, one finds

$$P\left(1 \mid s_j\right) - P\left(-1 \mid s_j\right) = -(1 - r) z_j. \quad (58)$$

In addition, the variance of the sequence must be specified by one of the equations

$$\sum_j p_S(s_j) z_j^2 = \sigma_z^2$$

$$\sum_j p_S(s_j) P(0 \mid s_j) = 1 - \sigma_x^2. \quad (59)$$

Equation (58) and one part of (59) imply the other part of (59). Ternary Markov sources satisfying (58) and (59) may be constructed for all values of r, $-1 < r < 1$. For negative r, the construction is possible only for $\sigma_x^2 \leq (1 - r)/2$, and in this case there is a unique solution with two states. The case $r = 0$ is of special interest and is considered in the following example.

Example 8: A ternary source with spectrum $S_X(\omega) = \sigma_x^2(1 - \cos \omega)$ may be constructed as a Markov source with three states, $z_0 = 1, z_1 = 0, z_2 = -1$, and transition matrix

$$Q = \begin{bmatrix} 0 & p & 0 \\ 1 & 1-2p & 1 \\ 0 & p & 0 \end{bmatrix}.$$

Here the variance is $\sigma_x^2 = 4p/(1 + 2p)$ and the entropy $H = (2p + \mathcal{H}(2p))/(1 + 2p)$. The entropy reaches its maximum value of 1 bit for $p = 1/4$ where $\sigma_x^2 = 2/3$. Alternatively, a two-state source with $z_0 = p$, $z_1 = p - 1$ and transition matrix

$$Q = \begin{bmatrix} 1-p & 1-p \\ p & p \end{bmatrix}$$

may be considered. This sequence has variance $\sigma_x^2 = 2p(1 - p)$ and entropy $\mathcal{H}(p)$. The greatest entropy occurs for $p = 1/2$, where the sequence is the bipolar sequence considered in Example 1. For $0.40 \leq \sigma_x^2 \leq 0.50$ this sequence has greater entropy than the three-state source with the same variance. Several practical modifications of the bipolar code have variances exceeding 1/2 [3], but rate 1 bit. The upper bound of Theorem 2 is $H < 1.1547$ bit for $\sigma_x^2 = 2/3$, suggesting that entropies greater than 1 bit may be possible. We shall return to this case in Section VII.

Example 9: Franaszek [18] has described a ternary block code with rate 4/3, which is an improvement over earlier block codes with the same rate. The spectrum of the coded sequence was computed by Bosik [19] and by Cariolaro and Tronca [20].[1] The sum variance can be calculated directly from the definition of the code, and in this way the cut-off frequency is found to be approximately $\omega_0 \simeq 0.35$. This result agrees well with the figures in [19] and [20]. A sequence with the same entropy can be constructed as a Markov source satisfying (58) and (59) with $r = 1/2$. Thus significantly better dc suppression is possible. If we choose a symmetric source with four states, $z_0 = 3/2, z_1 = 1/2, z_2 = -1/2, z_3 = -3/4$, there is a unique transition matrix

$$Q = \begin{Bmatrix} 1/4 & 15/76 & 0 & 0 \\ 3/4 & 27/76 & 17/38 & 0 \\ 0 & 17/38 & 27/76 & 3/4 \\ 0 & 0 & 15/76 & 1/4 \end{Bmatrix}.$$

The entropy $H \simeq 1.365$ is comparable to the rate of the Franaszek code if it is taken into consideration that the all-zero sequence is eliminated from the block code in order to facilitate synchronization. As in Example 5, we

---

[1] The complicated expression for the power spectrum in [20] contains an error since it does not give $S(0) = 0$.

may modify the transition probabilities and split certain states to obtain a sequence which can be divided into segments with probabilities of the form 2^{-m}. It is then possible to obtain a variable length code that converts independent data into a string with the desired probability distribution. A transition matrix with this property and the associated sum vector are

$$Q = \begin{bmatrix} 1/4 & 1/4 & 1/6 & 0 & 0 & 0 \\ 0 & 1/4 & 1/3 & 1/2 & 0 & 0 \\ 3/4 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/4 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 3/4 \\ 0 & 0 & 0 & 1/4 & 1/6 & 1/4 \end{bmatrix}$$

$$\bar{\zeta} = \begin{Bmatrix} 3/4 \\ 1/4 \\ 1/4 \\ -1/4 \\ -1/4 \\ -3/4 \end{Bmatrix}.$$

This sequence has variance $\sigma_x^2 = 13/18$, sum variance $\sigma_z^2 = 25/36$ and thus cut-off frequency $\omega_0 \simeq 0.52$. The entropy is $4/3$. The upper bound of Theorem 2 for $\sigma_x^2 = 2/3$ and $r = 1/2$ is $H \leq 1.391$. In practical applications of ternary codes, long runs of zeros must be avoided in order to ensure proper synchronization. We shall not discuss the effect of this extra constraint, but an alternative approach might be the use of sequences with four symbols. Such sequences have received little attention, but we note the following example.

Example 10: A sequence with alphabet V_4 and $r = 0$ is defined by the transition probabilities and the sum vector

$$Q = \begin{Bmatrix} 0 & 0 & 0 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 1/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/4 & 0 & 1/2 & 0 \\ 1 & 0 & 1/2 & 0 & 1/2 & 0 & 1 \\ 0 & 1/2 & 0 & 1/4 & 0 & 1/2 & 0 \\ 0 & 0 & 1/4 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 & 0 \end{Bmatrix}$$

$$\bar{\zeta} = \begin{Bmatrix} 3 \\ 2 \\ 1 \\ 0 \\ -1 \\ -2 \\ -3 \end{Bmatrix}.$$

An encoding method is given in Table II. The entropy of this sequence is $4/3$.

Sequences with larger alphabets V_{2K-1} and spectra $A(1 - \cos \omega)$ may be obtained by passing independent symbols from V_K through a discrete linear system with

TABLE II
ENCODING OF A QUATERNARY CODE WITH $r = 0$*

State	Data	Symbols
$s_0(s_6)$	∅	+3
$s_1(s_5)$	0	+1
	1	+3
$s_2(s_4)$	0	-1
	10	+1
	11	+3
s_3	00	+3
	01	+1
	10	-1
	11	-3

* ∅ denotes the empty string. In states s_4, s_5, and s_6 the opposite sign is used.

transfer function $(1 - e^{-i\omega})/2$. For equiprobable symbols, the rate is obviously $\log K$. Higher rates and lower cut-off frequencies may be obtained by the following construction [21]: one or more of the extreme symbols of V_{2K-1} are not used while the remaining symbols represent independent data. For $Z_t > 0$, some of the positive symbols are excluded; for $Z_t < 0$, the corresponding negative symbols are not used. The bipolar code may be interpreted as the simplest example of both encoding methods. In the final example of this section we consider codes with alphabets V_5.

Example 11: A sequence from V_5 with spectrum $(4/3)(1 - \cos \omega)$ and rate $\log 3 \simeq 1.585$ bits may be obtained by filtering equiprobable ternary symbols. However, the same spectrum is obtained with the Markov source defined by the transition matrix and sum vector

$$Q = \begin{Bmatrix} 0 & 5/38 & 5/38 & 0 \\ 1/2 & 7/19 & 7/19 & 1/2 \\ 1/2 & 7/19 & 7/19 & 1/2 \\ 0 & 5/38 & 5/38 & 0 \end{Bmatrix}$$

$$\bar{\zeta} = \begin{Bmatrix} 3/2 \\ 1/2 \\ -1/2 \\ -3/2 \end{Bmatrix}.$$

This sequence has entropy $H \simeq 1.658$ bits. Better rates can be obtained if the variance is increased, but it follows from Theorem 2 that the rate cannot reach 2 bits with $r = 0$. A code with rate 2 bits can be obtained by encoding two binary symbols as $\{-1, 0, 1, 2\}$ when the sum is negative and $\{-2, -1, 0, 1\}$ when the sum is positive. In order to avoid ambiguity for $Z_t = 0$ and asymmetry, we let Z_t take values from the set $W_{1/2}$. The stationary distribution of the sum is

$$P[Z_t = \pm 1/2] = 1/2(\sqrt{2} - 1)$$

$$P[Z_t = \pm(k - 1/2)] = (\sqrt{2} - 1)^k, \quad k \geq 2.$$

From this distribution we get $\sigma_z^2 = 9/4 + \sqrt{2} \simeq 3.66$ and then $\omega_0 \simeq 0.205$. However, the same entropy can be ob-

tained with a seven-state Markov source with $r = 1/2$ and transition matrix

$$\begin{bmatrix} 1/8 & 1/16 & 1/20 & 0 & 0 & 0 & 0 \\ 1/4 & 3/16 & 7/50 & 17/60 & 0 & 0 & 0 \\ 5/8 & 7/16 & 17/50 & 1/12 & 27/100 & 0 & 0 \\ 0 & 5/16 & 1/5 & 4/15 & 1/5 & 5/16 & 0 \\ 0 & 0 & 27/100 & 1/12 & 17/50 & 7/16 & 5/8 \\ 0 & 0 & 0 & 17/60 & 7/50 & 3/16 & 1/4 \\ 0 & 0 & 0 & 0 & 1/20 & 1/16 & 1/8 \end{bmatrix}.$$

Both codes have symbol probabilities $(1/8, 1/4, 1/4, 1/4, 1/8)$ but the finite-state source has $\omega_0 \simeq 0.5$ and entropy $H \simeq 2.031$ bits. The upper bound for this distribution is $H \leq 2.051$ bits.

We conclude that all but the simplest codes that have been suggested for dc suppression have lower cut-off frequencies than the best codes with the same rates. For the type of power spectra that are of greatest practical interest we have, in most cases, obtained close agreement between the upper bound and the entropy of the Markov sources constructed in this section. Significant differences between the upper and lower bounds occur for binary codes with large cut-off frequencies and for ternary codes with large variance and large cut-off frequencies. These cases will be considered in detail in the last two sections.

VI. MARKOV SOURCES WITH GIVEN SUM VARIANCE AND MAXIMAL ENTROPY

Since it is difficult to construct a maxentropic sequence for a given power spectrum, we shall consider the easier problem of maximizing the entropy of a dc-free sequence when σ_x^2 and σ_z^2 are given. One motivation for this approach is the observation in Section IV that the cut-off frequencies of sequences with equal values of these variances are approximately equal. In addition, the construction considered here will provide a new upper bound to the rate of dc-free codes. If the power spectrum is given, we can determine σ_z^2 from (24), and the maximum value of the entropy cannot decrease when a constraint on the power spectrum is replaced by a constraint on the sum variance. In some cases, the bound of Theorem 2 is not very tight. This is due to a significant difference between the distribution $P(x_t, y)$ and possible distributions $P(x_t, \hat{x}_t)$. In such cases we may obtain better results with the bound to be developed in this section. In particular we shall prove that, for first-order spectra $S_X^r(\omega)$, we always get a bound that is tighter than Theorem 2.

We shall construct a sequence $\overline{X}$ with alphabet V_K, variance σ_x^2, and a sum sequence $\overline{Z}$ with variance σ_z^2, mean value $E[Z_t] = 0$, where

$$Z_t = Z_{t-1} + X_t \tag{60}$$

in such a way that $H(X)$ is maximized. It follows immediately from (60) that $H(Z) = H(X)$, and the problem

can be stated entirely in terms of $\overline{Z}$ if (60) is replaced by

$$Z_t - Z_{t-1} \in V_K$$

$$E\left[(Z_t - Z_{t-1})^2\right] = \sigma_x^2. \tag{61}$$

Assume that the probability distribution $P_Z(z_i)$ and the transition probabilities $P[Z_t = z_k \mid Z_{t-1} = z_j]$ are known. It then follows from the convexity of the entropy function that we get the largest entropy by letting

$$P\left[Z_t = z_{i_0} \mid Z_{t-1} = z_{i_1}, Z_{t-2} = z_{i_2}, \cdots \right]$$

$$= P\left[Z_t = z_{i_0} \mid Z_{t-1} = x_{i_1}\right]. \tag{62}$$

Thus $\overline{Z}$ should be a Markov chain. The alphabet of $\overline{Z}$ is the set W_α.

In the analysis of this Markov chain we shall use the simplified notation for the stationary distribution and transition probabilities

$$P[Z_t = \alpha + j] = P_j, \qquad j = 0, \pm 1, \pm 2, \cdots,$$

$$P[Z_t = \alpha + k + j \mid Z_{t-1} = \alpha + j] = p_j(k), \qquad k \in V_K.$$

Considering the transitions between those subsets of states for which $Z_t \leq \alpha + j$ and $Z_t > \alpha + j$ respectively, we get

$$\sum_{k>0}^{k-1} \sum_{i=0} \left[P_{j-1}p_{j-1}(k) - P_{j+k-i}p_{j+k-i}(-k) \right] = 0,$$

$$j = 0, \pm 1, \pm 2, \cdots. \tag{63}$$

For binary and ternary sequences, (63) implies that the probability distribution of Z_t, Z_{t+1} satisfies

$$P(z_{i_1}, z_{i_2}) = P(z_{i_2}, z_{i_1}) \tag{64}$$

and by the Markov property (62), the distribution of $Z_t, Z_{t+1}, \cdots, Z_{t+n-1}$ has the symmetry

$$P(z_{i_1}, z_{i_2}, \cdots, z_{i_n}) = P(z_{i_n}, z_{i_{n-1}}, \cdots, z_{i_1}). \tag{65}$$

It follows from (65) that the distribution of $X_t, X_{t+1}, \cdots, X_{t+n-1}$ satisfies

$$P(v_{i_1}, v_{i_2}, \cdots, v_{i_n}) = P(-v_{i_n}, -v_{i_{n-1}}, \cdots, -v_{i_1}). \tag{66}$$

For larger alphabets and symbol distributions satisfying

$$p_X(k) = p_X(-k)$$

we may use Lagrange multipliers to determine the transition probabilities that maximize the entropy for given values of P_j, the constraints (63) and

$$\sum_j \left(P_j p_j(k) + P_j p_j(-k) \right) = 2p_X(k).$$

These transition probabilities satisfy (64) and thus also (65) and (66). The maxentropic sequence studied by Slepian [2] satisfy the relations

$$P(v_{i_1}, v_{i_2}, \cdots, v_{i_n}) = P(v_{i_n}, v_{i_{n-1}}, \cdots, v_{i_1})$$

and

$$P(v_{i_1}, v_{i_2}, \cdots, v_{i_n}) = P(-v_{i_1}, -v_{i_2}, \cdots, -v_{i_n}).$$

However, for dc-free codes these relations are satisfied only if

$$P(z_{i_1}, z_{i_2}, \cdots, z_{i_n}) = P(-z_{i_1}, -z_{i_2}, \cdots, -z_{i_n}),$$

and thus we must have $\alpha = 1$ or $\alpha = 1/2$. For other values of α only the symmetry condition (66) is valid. The maximum entropy often occurs for either $\alpha = 1$ or $\alpha = 1/2$, but we shall demonstrate in an example that this is not always the case.

We have not been able to determine the transition probabilities of the Markov chain for a general alphabet V_K, but for binary sequences they must satisfy an interesting difference equation. We use the notation

$$p_j(1) = p_j, \quad p_j(-1) = 1 - p_j.$$

We want to maximize

$$H(Z) = \sum_j P_j \mathcal{K}(p_j)$$

subject to the constraints

$$\sum_j P_j = 1$$

$$\sum_j j P_j = -\alpha \qquad (67)$$

$$\sum_j j^2 P_j = \sigma_z^2 + \alpha^2 \qquad (68)$$

and the condition (63) which becomes

$$P_j p_j = P_{j+1}(1 - p_{j+1}).$$

Here (67) and (68) ensure that $E[Z_t] = 0$ and $E[Z_t^2] = \sigma_z^2$. Using Lagrange multipliers, we find the conditions

$$\mathcal{K}(p_j) + \mu + \eta j + \rho j^2 + \lambda_j p_j + \lambda_{j-1}(1 - p_j) = 0 \qquad (69)$$

and

$$\ln \frac{1 - p_j}{p_j} + \lambda_j - \lambda_{j-1} = 0. \qquad (70)$$

Solving (69) and (70), we get

$$p_j(1 - p_{j+1}) = \gamma \beta^{(j+\delta)^2}. \qquad (71)$$

This difference equation may be used to compute a set of transition probabilities for a suitable initial value p_0 and given values of the parameters β, γ, δ. A complete discussion of the solutions to (71) will appear elsewhere [22], here we shall only state some of the main properties. The parameter δ is closely related to α. Thus for $\delta = 0$ and $p_0 = \sqrt{\gamma}$, we get a symmetric chain with $\alpha = 1/2$, and for $\delta = 1/2$ and $p_0 = 1/2$ we get a symmetric solution with $\alpha = 0$. Thus, without loss of generality we may assume $0 \leq \delta \leq 1/2$. The variance of Z_t is largely determined by β, which must be positive and at most one. If we want a finite state solution with $p_N = 0$ and $p_{-M} = 1$ for given N and M, (71) may be solved for γ. However, we conjecture that the largest entropy is obtained when γ is chosen as the unique value for which the chain has infinitely many states

and

$$p_j \simeq \gamma \beta^{(j+\delta)^2}$$

for sufficiently large j. For other values of γ, the p_j cannot be transition probabilities of a stable Markov chain. A certain amount of numerical calculation is necessary in order to obtain the best chain for a given variance. The correct value of δ must be found, and the exact value of the variance depends upon all three parameters. However, since the p_j decrease very rapidly with increasing j, an excellent approximation may often be obtained by considering a small number of states.

We state the upper bound as the following theorem.

Theorem 4: The entropy of a binary dc-free sequence with sum variance σ_z^2 is upper-bounded by the maximum entropy of a Markov chain, $\bar{Z}$, with transition probabilities satisfying (71) and stationary distribution satisfying (67) and (68).

We shall now prove that, for the first-order spectra $S_X^r(\omega)$, Theorem 4 is a stronger result than Theorem 2. For an arbitrary dc-free sequence, $\bar{X}$, and the corresponding sum sequence, $\bar{Z}$, defined by (60), we consider the linear predictor

$$\tilde{X}_t = h Z_{t-1}. \qquad (72)$$

The value of h that gives the smallest error variance is obtained by solving

$$\sigma_z^2 h = E[X_t Z_{t-1}] = -\sigma_x^2/2, \qquad (73)$$

where the correlation is determined by taking the variance of both sides of (60). It follows from (73) that

$$h = -\sigma_x^2/2\sigma_z^2 \qquad (74)$$

$$\sigma_{e'}^2 = \sigma_x^2(1 - \sigma_x^2/4\sigma_z^2). \qquad (75)$$

For first-order spectra σ_z^2 is given by (37), and (74) and (75) become

$$h = -(1 - r)$$

$$\sigma_{e'}^2 = (1 + r)\sigma_x^2/2$$

which agree with (35) and (36). Thus for these spectra, $\bar{Z}$ has the same structure as the best linear predictor, it has the same variance, and the error variance has the correct value. In general, of course, $\bar{X}$ does not have the right spectrum and (72) does not define the best linear predictor. Since the error variance has the correct value, $H(Z)$ is upper-bounded by the entropy given by Theorem 2, and we have obtained a stronger bound.

Example 12: Let $r = 1/2$ and thus $\sigma_z^2 = 1$. We have calculated the entropies of the Markov chains with this variance for several values of α. For $\alpha = 0$, $p_0 = 1/2$, and $\delta = 1/2$, we get $\beta = 0.721$ and $\gamma = 0.426$. This gives a stationary distribution with $P_0 = 0.387$, $P_1 = 0.247$, $P_2 = 0.056$ while the remaining probabilities are very small. The entropy is $H = 0.793$. With $\alpha = 1/2$ and $\delta = 0$, we find a chain with the same variance for $\beta = 0.720$ and $\gamma = 0.426$

confirming that the variance is closely related to the value of β. In this case, the stationary distribution is $P_0 = 0.347$, $P_1 = 0.273$, $P_2 = 0.032$, and the entropy is approximately the same as before. Other values of α give the same entropy within the accuracy of the calculations. This entropy should be compared to the upper bound $H \leq 0.811$ obtained by application of Theorem 2 and the lower bound $H \geq 0.780$ that follows from the construction in Example 5.

For larger alphabets, the same approach may be used, but the calculations are more complex. In the ternary case, one finds

$$p_j(0) = \gamma \beta^{(j+\delta)^2} \tag{76}$$

$$p_j(1)p_{j+1}(-1) = \epsilon p_j(0)p_{j+1}(0). \tag{77}$$

While these equations are quite similar to (71), it is much more difficult to construct a chain with given variances that satisfies (76) and (77).

VII. MAXENTROPIC SEQUENCES WITH SPECTRA $S_X(\omega) = A(1 - \cos \omega)$

In this section we shall decide the optimality of two simple codes. First we prove that the biphase sequence discussed in Section V is optimal, then we construct a sequence which demonstrates that the bipolar sequence introduced in Example 1 is not optimal. Finally, we discuss some properties of ternary sequences with greater entropy than the bipolar sequence.

For a binary sequence with spectrum $S_X(\omega) = 1 - \cos \omega$, we have $\sigma_z^2 = 1/2$, and (67) and (68) become

$$\sum_j jp_j = -\alpha \tag{78}$$

$$\sum_j j^2 p_j = 1/2 + \alpha^2. \tag{79}$$

If we take $\alpha = 0$, the only stationary distribution is $P_0 = 1/2$, $P_1 = P_{-1} = 1/4$, and we obtain the biphase sequence with entropy $1/2$ bit. However, if $\alpha = 1/2$ a Markov source with $\sigma_z^2 = 1/2$ can have a greater entropy. With four states, we get a chain with transition matrix

$$Q = \begin{cases} 0 & 1/7 & 0 & 0 \\ 1 & 0 & 6/7 & 0 \\ 0 & 6/7 & 0 & 1 \\ 0 & 0 & 1/7 & 0 \end{cases}$$

and entropy $H = (7/8)\mathcal{H}(1/7) \simeq 0.518$. However, this sequence has a rather undesirable spectrum. If at time zero the chain is in state $Z_t = -1/2$ or $Z_t = 3/2$, then it must always be in one of these states for t even, and be in $Z_t = -3/2$ or $1/2$ for t odd. Thus, the mean value is

$$E[Z_{2t}] = -1/4 \qquad E[Z_{2t+1}] = 1/4.$$

Therefore, the sequence contains a deterministic component, and the power spectrum has a line at $\omega = \pi$.

In order to prevent this situation from occurring, we replace (78) by

$$\sum_j P_{2j} = \sum_j P_{2j+1} = 1/2 \tag{80}$$

$$\sum_j 2jP_{2j} = \sum_j (2j + 1)P_{2j+1} = -\alpha/2. \tag{81}$$

Combining (79), (80), and (81), we form the sum

$$\sum_j c_j P_j = \sum_j j^2 P_j + 2\sum_j 2jP_{2j} - \sum_j P_{2j+1}$$

$$= \alpha(\alpha - 1) \leq 0. \tag{82}$$

It is readily verified that $c_0 = c_{\pm 1} = c_{-2} = 0$, $c_2 = 8$, and $c_j > 0$ for $|j| > 2$. Since $P_j \geq 0$, we conclude from (82) that $\alpha = 0$. Thus, the biphase sequence has maximum entropy for the spectrum $1 - \cos \omega$. We have not been able to prove optimality for any other dc-free code.

In the case of the ternary bipolar code, the situation is quite different. Again we follow the approach of Section VI with $\sigma_x^2 = 1/2$ and $\sigma_z^2 = 1/4$. We first note two simple cases. If $\alpha = 0$ the only possible distribution is $P_0 = 3/4$, $P_1 = P_{-1} = 1/8$. Similarly, if $\alpha = 1/2$, we must have $P_0 = P_{-1} = 1/2$. These Markov chains were found in Example 8. However, a greater entropy is possible for other values of α and this shows that the maximum does not always occur for $\alpha = 0$ or $\alpha = 1/2$. A three-state chain with $\alpha = 1/3$ and transition matrix

$$Q = \begin{cases} 0 & 1/46 & 0 \\ 1 & 14/23 & 17/25 \\ 0 & 17/46 & 8/25 \end{cases}$$

has the right variances and entropy $H \simeq 1.0083$. The spectrum is not exactly $(1 - \cos \omega)/2$, but by splitting the last state, we may obtain the desired correlation sequence (27). The transition matrix becomes

$$Q' = \begin{cases} 0 & 1/46 & 0 & 0 \\ 1 & 14/23 & 1-u & 1-v \\ 0 & 17/46 & 0 & 0 \\ 0 & 0 & u & v \end{cases}.$$

Now the general expression for the correlation sequence (45) and (46) is applied. Q' has eigenvalues $1, 0, \lambda_1, \lambda_2$, and the correlation sequence satisfies a second-order difference equation. $R_Z(1) = 0$ follows from (60) and $\sigma_z^2 = \sigma_x^2/2$. If u and v are adjusted to make $R_Z(2) = 0$, the difference equation ensures that $R_Z(k) = 0$ for $k > 2$. The solution is $u = 128/391$, $v = 7/23$. With these transition probabilities, the source has entropy $H \simeq 1.0082$ bits. Thus, we may conclude that the maxentropic sequence for the alphabet V_3 and $S_X(\omega) = (1 - \cos \omega)/2$ has entropy greater than 1 bit and that the maximum occurs for some value of α between 0 and $1/2$. Consequently, this sequence lacks some of the symmetry found in the bipolar sequence.

An improved upper bound to the entropy of a sequence with the same spectrum as the bipolar sequence may be obtained by solving (76) and (77). We have used numerical

techniques to maximize the entropy of a chain with four states. From these results, the approximate values of the parameters in (76) and (77) can be found and, finally, the difference equations may be used for calculating the best transition probabilities for a general Markov chain. The maximum occurs for $\alpha \simeq 0.354$ and, actually, a very close approximation to the maximum is obtained with only four states. In this way the upper bound $H \leq 1.01$ bits is found. The details of the maxentropic sequence appear to be of little interest. However, it would be quite feasible to split the states of the Markov chain and adjust the transition probabilities to give the spectrum $(1 - \cos \omega)/2$.

As discussed in Example 8, there is more room for improvement if the variance of the sequence is increased, and in most applications it is desirable to reduce $p_X(0)$.

Example 13: If we take $\sigma_x^2 = 2/3$ and $\alpha = 1/2$, we may construct a ternary source with four states and $\sigma_z^2 = \sigma_x^2/2$:

$$Q = \begin{Bmatrix} 0 & 1/23 & 0 & 0 \\ 1 & 8/23 & 14/23 & 0 \\ 0 & 14/23 & 8/23 & 1 \\ 0 & 0 & 1/23 & 0 \end{Bmatrix}.$$

The entropy of this chain is 1.114 bits. The power spectrum can be corrected to $S_X(\omega) = (2/3)(1 - \cos \omega)$ if the middle states are split into three states each. As discussed in Section V, the difference equation for the correlation sequence can be found as the characteristic equation of a four by four matrix. Further, since $r = 0$, 0 must be one of the eigenvalues. Thus, it is enough to ensure that $R_Z(1) = R_Z(2) = R_Z(3) = 0$. This condition leads to a system of quadratic equations which may be solved for the transition probabilities. We omit the details. The entropy of the source is now 1.111 bits. Using the same technique as in Examples 5 and 9, we can also split the same states to obtain a set of transition probabilities that can be matched by a variable-length code:

$$Q' = \begin{Bmatrix} 0 & 1/16 & 1/10 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1/3 & 2/3 & 0 & 0 & 0 \\ 0 & 5/16 & 0 & 0 & 0 & 0 & 5/8 & 0 \\ 0 & 0 & 3/10 & 0 & 0 & 3/5 & 0 & 0 \\ 0 & 0 & 3/5 & 0 & 0 & 3/10 & 0 & 0 \\ 0 & 5/8 & 0 & 0 & 0 & 0 & 5/16 & 0 \\ 0 & 0 & 0 & 2/3 & 1/3 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1/10 & 1/16 & 0 \end{Bmatrix}.$$

This source has $\sigma_x^2 \simeq 0.702$, $\sigma_z^2 \simeq 0.356$, and entropy $H \simeq 1.085$ bits.

VIII. CONCLUSION

We have presented a general upper bound to the rate of a code with a given power spectrum and a computational procedure for evaluating this bound. For dc-free codes, which is the class of greatest practical interest, we have obtained a tighter bound. The suppression of low frequencies has been related to the variance of the sum of the encoded symbols, which is a parameter that has been used earlier on more intuitive grounds. We have constructed sequences with high entropies for several combinations of alphabets and power spectra. These sequences have significantly greater entropies or cut-off frequencies than most codes suggested earlier. We have described a practical encoding method based on variable length codes for suitably modified Markov sources. It has been proved that the biphase sequence has maximum entropy, but this is the only new maxentropic sequence that we have been able to determine. The rate of the familiar AMI code is very close to the upper bound, but the maxentropic sequence is much more complicated.

REFERENCES

[1] H. Kobayashi, "A survey of coding schemes for digital data," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 1087–1100, Dec. 1971.
[2] D. Slepian, "On maxentropic discrete stationary processes," *Bell Syst. Tech. J.*, vol. 51, pp. 629–654, March 1972.
[3] A. Croisier, "Introduction to pseudoternary transmission codes," *IBM J. Res. Develop.*, vol. 14, pp. 354–367, July 1970.
[4] V. A. DiEuliis and F. P. Preparata, "Spectrum shaping with alphabetic codes with finite autocorrelation sequence," *IEEE Trans. Commun. Technol.*, vol. COM-26, pp. 474–477, April 1978.
[5] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.
[6] R. M. Gray, "Sliding-block source coding," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 357–368, July, 1975.
[7] Z. Kohavi, *Switching and Finite Automata Theory.* New York: McGraw-Hill, 1970.
[8] J. G. Kemeny and J. L. Snell, *Finite Markov Chains.* Princeton, NJ: Van Nostrand, 1960.
[9] A. P. Sage and J. L. Melsa, *Estimation Theory with Applications to Communications and Control.* New York: McGraw-Hill, 1971.
[10] I. I. Gihman and A. V. Skorohod, *The Theory of Stochastic Processes I.* Berlin: Springer, 1974.
[11] A. M. Yaglom, *An Introduction to the Theory of Stationary Random Functions.* Englewood Cliffs, NJ: Prentice-Hall, 1962.
[12] T. Berger, *Rate Distortion Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1971.
[13] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 460–473, July 1972.
[14] M. Rosseau, "Block codes for optical-fibre communication," *Electron. Lett.*, vol. 12, pp. 478–479, Sept. 1976.
[15] J. N. Franklin and J. R. Pierce, "Spectra and efficiency of binary codes without dc," *IEEE Trans. Commun. Technol.*, vol. COM-20, pp. 1182–1184, Dec. 1972.
[16] J. C. Mallison and J. W. Miller, "Optimal codes for digital magnetic recording," *Radio Electron. Eng.*, vol. 47, pp. 172–176, April 1977.
[17] T. M, Chien, "Upper bound on the efficiency of dc-constrained codes," *Bell Syst. Tech. J.*, vol. 49, pp. 2267–2288, Nov. 1970.
[18] P. A. Franaszek, "Sequence-state coding for digital transmission," *Bell Syst. Techn. J.*, vol. 47, pp. 143–157, Jan. 1968.
[19] B. S. Bosik, "The spectral density of a coded digital signal," *Bell Syst. Tech. J.*, vol. 51, pp. 921–932, April 1972.
[20] G. L. Cariolaro and G. P. Tronca, "Spectra of block coded digital signals," *IEEE Trans. Commun. Technol.*, vol. COM-22, pp. 1555–1563, Oct. 1974.
[21] H. Kaneko and A. Sawai, "Feedback balanced codes for multilevel PCM transmission," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 554–563, Oct. 1969.
[22] T. Høholdt, B. Branner-Jørgensen, and J. Justesen, "Markov chains," to be submitted 1982.