Nothing prevents the use of a nonlinear encoder, i.e., an arbitrary bijective map g from F_q^K into V. If g is such a map and g^{-1}: $V \to F_q^K$ is its inverse, then the corresponding symbol error rate is

$$P_s(g) = \frac{1}{K} \sum_{j=1}^{M-1} \mu_j \nu_j(g)$$

where

$$\nu_j(g) = \frac{1}{M} \sum_{i=0}^{M-1} \omega\left(g^{-1}(y_i + y_j) - g^{-1}(y_i)\right).$$

Finding a g that will minimize $P_s(g)$ will involve a detailed study of the function $\nu_j(g)$. We conjecture that in general the best encoder g_0 will outperform the best linear encoder G_0.

Acknowledgment

The author wishes to thank the reviewers for their useful criticisms, which have helped to improve the correspondence. The author also wishes to thank Hélène Kourilko for her professional typing of the manuscript.

References

[1] L. A. Dunning and W. E. Robbins, "Optimal encodings of linear block codes for unequal error protection," *Inform. Contr.*, vol. 37, pp. 150–177, May 1978.

[2] W. J. Van Gils, "Two topics on linear unequal error protection codes: Bounds on their length and cyclic codes classes," *IEEE Trans. Inform Theory*, vol. IT-29, pp. 866–876, Nov. 1983.

[3] F. J. MacWilliams and M. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1978, pp. 20–21.

[4] D. J. A. Welsh, *Matroid Theory*. New York: Academic, 1976, p. 358.

An Entropy Maximization Problem Related to Optical Communication

ROBERT J. McELIECE, FELLOW, IEEE, EUGENE R. RODEMICH, AND LAIF SWANSON

Abstract—**Motivated by a problem in optical communication, we consider the general problem of maximizing the entropy of a stationary random process that is subject to an average transition cost constraint. Using a recent result of Justesen and Hoholdt, we present an exact solution to the problem and suggest a class of finite state encoders that give a good approximation to the exact solution.**

I. Introduction

In several recent studies of energy-efficient direct detection optical communication systems, the use of pulse-position modulation (PPM) has been shown to be optimal, or nearly so [1], [10], [12]. In a proposed system for NASA applications [5], [8] 256-ary PPM combined with Reed–Solomon coding achieves an energy efficiency of about 2.5 b/photon at a decoded bit error probability of 10^{-6}.

Manuscript received October 18, 1984; revised June 3, 1985. This work was supported in part by the National Aeronautics and Space Administration under Contract NAS7-100. This work was partially presented at the 17th Annual Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore, MD, March 1983.

R. McEliece is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125.

E. Rodemich and L. Swanson are with the Jet Propulsion Laboratory 238-420, 4800 Oak Grove Drive, Pasadena, CA 91109.

IEEE Log Number 8406625.

Unfortunately, PPM is characterized by a large peak-to-average power ratio requirement: for M-ary PPM the needed ratio is exactly M. If the optical power source is a semiconductor laser, it is difficult to achieve such large values of the peak-to-average power ratio. One possible way to reduce this ratio while at the same time maintaining energy efficiency is to use more than one optical frequency (color). In fact, if N distinguishable colors are used with M-ary PPM, then for a model similar to that in [10] the channel capacity is $\log(M) + \log(N)$ nat/photon with a peak-to-average power ratio of M. For example, the proposed monochromatic NASA system is in principle equivalent to a 16-color 16-ary PPM system whose peak-to-average power ratio is only 16, instead of 256.

Several difficulties need to be overcome before multicolor laser PPM can become a reality, however. One of these noticed by Katz [6] is that additional transmitter energy is needed in order to implement the color frequency modulation. Katz showed that for a voltage-controlled color-modulation scheme in which each color is represented by a different voltage level, the power required to jump from voltage V_i to voltage V_j is proportional to the square of the voltage difference $(V_i - V_j)^2$.

In a monochromatic system no voltage change is necessary, and so no additional power is consumed. However, in an N-color system the additional modulation power required could significantly reduce the overall system efficiency. We shall show in this correspondence that if the color modulation is done intelligently, this loss can be kept to a minimum. The idea is to ensure that large voltage jumps occur relatively infrequently, while at the same time ensuring that the diversity represented by the color shifting gives a large increase in transmitted information over a monochromatic system. The key to our results is the study of the entropy of the sequence of transmitted colors.

If no constraints are put on the color sequence and there are N colors, the maximum possible entropy is $\log N$ nat/symbol. However, if constraints are put on the color sequence in order to reduce the modulation power, this value will be reduced to some value below $\log N$.

For what follows, the sequence of modulator voltages is modeled by a stationary random process $X = \{\cdots, X_{-1}, X_0, X_1, \cdots\}$, where each X_i can take values in the set $\{V_1, V_2, \cdots, V_N\}$ of possible voltages; V_i is the voltage corresponding to the ith color. The entropy $H(X)$ of the process, as defined, e.g., in [2] and [11], represents the average amount of (potential) information per unit of time transmitted by the process. Motivated by Katz's observation [6] we define the *average modulation power* consumed by the process X by

$$P = E\left(|X_i - X_{i+1}|^2\right) \tag{1.1}$$

which is, by the stationarity assumption, independent of i. Our problem is to find out how large H can be for a given value of P.

In Section II of this correspondence we will see how a recent result of Justesen and Hoholdt will allow us to solve this problem —and many similar problems—exactly. We shall see, for example, that if 16 colors are used, and if the transmitted wavelength is related linearly to the corresponding modulator voltages, one can, in principle, obtain an entropy of about $0.9 \times \log(16)$, using only about 25 percent of the power required to obtain the full entropy $\log(16)$. For 256 colors the results are even better: one can get 90 percent efficiency using only ten percent of the maximum power.

The results in Section II do not suggest an efficient way to encode, i.e., to map a sequence of binary data bits into a sequence of modulator voltages with a large entropy and small power consumption. In Section III we will address this question and describe a simple finite-state encoder whose performance (in a sense made precise there) is within 0.25 bits of optimal for all values of N. These results, besides being of interest in their own right, suggest that multicolor PPM may be a practical way to overcome the peak-to-average power problems associated with optical communication.

II. Calculation of Channel Capacity (Entropy)

In this section we will use a recent result of Justesen and Hoholdt [4] to give a parametric solution to the entropy maximization problem described in the introduction. The specific form of the cost function (1.1) seems no easier to handle than the following generalization.

Let $X = \{\cdots, X_0, X_1, X_2, \cdots\}$ be a stationary random process, taking values in the state set $S = \{1, 2, \cdots, N\}$. With each pair (k, l) from S we associate a real number D_{kl}, the *cost* of going from state k to state l. The *average transition cost* of the process X is defined as

$$P = \sum_{k,l=1}^{N} \Pr\{X_{j-1} = k, X_j = l\} D_{kl}. \tag{2.1}$$

This quantity is independent of j since X is stationary. The question we ask is this: among all stationary processes $\{X_n\}$ with average transition cost $\leq P_0$, say, what is the largest possible entropy? Notice that the problem stated in Section I corresponds to the special case $D_{kl} = (k-l)^2$ of this problem.

As a first step toward a solution we note that we may restrict our attention to stationary Markov chains. Suppose that $\{X_n\}$ is a stationary process, not necessarily a Markov chain. Then if $\{X_n'\}$ is that stationary Markov chain whose transition probabilities are the same as those for $\{X_n\}$, i.e.,

$$\Pr\{X_j' = l \mid X_{j-1}' = k\} = \Pr\{X_j = l \mid X_{j-1} = k\},$$

then plainly $\{X_n\}$ and $\{X_n'\}$ will have the same value of P. On the other hand, the entropy of the process $\{X_n\}$ is less than or equal to $H(X_1 \mid X_0)$ [2, theorem 3.5.1], which is the same as $H(X_1' \mid X_0')$, and this is in turn the entropy of the Markov chain $\{X_n'\}$. Thus the stationary Markov chain $\{X_n'\}$ has the same value of P as $\{X_n\}$, and at least as large an entropy. Therefore we may safely restrict our search for the largest possible entropy to the set of stationary Markov chains.

The problem of finding a maximal-entropy Markov chain is solved in the paper of Justesen and Hoholdt [4]. They show that if *any* Markov chains with $P \leq P_0$ exist, then a unique maximal one also exists. If $h(P_0)$ denotes the largest possible entropy, they show that $h(P_0)$ is a monotonic increasing function of P_0 and reaches its maximum, $\log N$, when the extremal Markov chain is in fact a sequence of independent identically distributed (i.i.d.) uniform random variables, and the average transition cost is given by

$$P_1 = \frac{1}{N^2} \sum_{k,l} D_{kl}.$$

For all $P_0 \geq P_1$, $h(P_0) = \log N$.

The precise description of the extremal process for $P_0 \leq P_1$ is as follows. For each $\lambda \geq 0$ define the matrix

$$G = G(\lambda) = (e^{-\lambda D_{kl}}).$$

Then G has unique (up to a change of scale) left and right eigenvectors with positive components corresponding to the same eigenvalue $m = m(\lambda)$. Call these eigenvectors

$$\gamma = \gamma(\lambda) = (\gamma_1, \gamma_2, \cdots, \gamma_N) \qquad \text{right}$$
$$\xi = \xi(\lambda) = (\xi_1, \xi_2, \cdots, \xi_N) \qquad \text{left.}$$

Then the optimizing Markov process has steady-state probabilities

$$p_k = \frac{\gamma_k \xi_k}{\sum_l \gamma_l \xi_l}, \qquad k = 1, 2, \cdots, N \tag{2.2}$$

and transition probabilities

$$a_{kl} = \frac{1}{m} \frac{\gamma_l}{\xi_k} e^{-\lambda D_{kl}}. \tag{2.3}$$

The value of the entropy is

$$H = \lambda P_0 + \log m \tag{2.4}$$

while the transition cost is given by

$$P = -\frac{1}{m} \frac{dm}{d\lambda} \tag{2.5}$$

$$= \left(\sum_{k,l} \xi_k \gamma_l e^{-\lambda D_{kl}} D_{kl} \right) \Big/ \left(\sum_{k,l} \xi_k \gamma_l e^{-\lambda D_{kl}} \right). \tag{2.6}$$

Given these results, one can usually compute the function $h(P_0)$ without much trouble. This is done for the special case $D_{kl} = (k-l)^2$ with $N = 2^n$, $n = 1, 2, \cdots, 8$. The results are plotted in Fig. 1. Note that for large values of N the curves are essentially log linear over a large portion of their span. This means, as mentioned in Section I, that a large information rate can be obtained from a relatively small amount of expended modulation power.

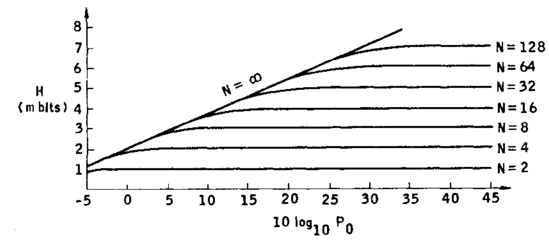

Fig. 1. Exact solution to entropy maximization problem for $D_{kl} = (k-l)^2$ for finite and infinite N.

III. Description of an Efficient Encoder when $D_{kl} = (k-l)^2$

In order to apply the results of Section II to the optical communication problem described in Section I, it would be necessary to encode a given binary data stream into a sequence of symbols from $\{1, 2, \cdots, N\}$, which closely resembles a typical sequence from the optimal Markov chain for $D_{kl} = (k-l)^2$. Given the complex form of this chain, such an encoding would be very difficult to implement. In this section we shall examine some *approximations* to the optimal solution that are not far from optimal (about 0.25 b when P_0 is large) and that suggest ways to implement energy-efficient multicolor PPM. We begin by studying a simplification of the problem.

A. Approximations

The maximum-entropy problem of Section II is modified as follows: the state set is enlarged from $\{1, 2, \cdots, N\}$ to $\mathbb{Z}$, the set of all integers.

Theorem 1: If the cost function D_{kl} depends only on the difference between k and l, then the maximum entropy stationary process subject to $P \leq P_0$ is a Markov chain for which the increments

$$\Delta_n = X_{n+1} - X_n \tag{3.1}$$

are i.i.d.

Proof: The argument given in Section II shows that the maximizing process is Markov. Thus if we write the cost function D_{kl} as

$$D_{kl} = D(k-l), \tag{3.2}$$

the problem is to

$$\begin{aligned}&\text{maximize} && H(X|Y)\\ &\text{subject to} && E(D(X-Y)) \le P_0,\end{aligned} \tag{3.3}$$

X and Y identically distributed. However, it is easy to see that $H(X|Y) \le H(Y-X)$:

$$\begin{aligned}H(X) + H(Y|X) &= H(X,Y)\\ &= H(X, Y-X)\\ &\le H(X) + H(Y-X).\end{aligned}$$

These steps follow from results in [9, ch. 1]. Hence the maximum in (3.3) is less than or equal to the maximum in (3.4):

$$\begin{aligned}&\text{maximize} && H(Z)\\ &\text{subject to} && E(D(Z)) \le P_0.\end{aligned} \tag{3.4}$$

On the other hand, if $\{\cdots, X_{-1}, X_0, X_1, \cdots\}$ is a stationary Markov chain with i.i.d. increments, the common distribution being the solution to (3.4), the entropy of the chain is $H(Z)$.

According to Theorem 1, to solve our problem on $\mathbb{Z}$ with $D(k-l) = (k-l)^2$ it is sufficient to maximize $H(\Delta)$ subject to $E(\Delta^2) \le P_0$. This can be done using straightforward variational techniques [9, problem 1.8], and the maximum entropy is given parametrically as follows.

For each $\lambda > 0$ define

$$m(\lambda) = \sum_{n=-\infty}^{\infty} e^{-n^2\lambda} \tag{3.5}$$

so that

$$-m'(\lambda) = \sum_{m=-\infty}^{\infty} n^2 e^{-n^2\lambda}. \tag{3.6}$$

Then among all random variables satisfying $E(\Delta^2) \le P_0$ the maximum entropy $H_{\max}$ is given by

$$H_{\max} = \log m(\lambda) - \lambda m'(\lambda)/m(\lambda) \tag{3.7}$$

where

$$P = -m'(\lambda)/m(\lambda). \tag{3.8}$$

For small values of λ the sums in (3.5) and (3.6) are well approximated by the following integrals:

$$m(\lambda) \approx \int_{-\infty}^{\infty} e^{-\lambda x^2}\,dx = \sqrt{\pi/\lambda} \tag{3.9}$$

$$-m'(\lambda) \approx \int_{-\infty}^{\infty} x^2 e^{-\lambda x^2}\,dx = \frac{1}{2\lambda}\sqrt{\pi/\lambda}. \tag{3.10}$$

Using these approximations in (3.7) and (3.8), we obtain the approximation

$$H_{\max} \approx \tfrac{1}{2}\log P_0 + \tfrac{1}{2}\log 2\pi e, \qquad P_0 \text{ large} \tag{3.11}$$

which is extremely good, even for small values of P_0 as exhibited in Table I. In Table I we have calculated the corresponding value of P_0 for a range of λ's from (3.8), the exact value of $H_{\max}$ calculated from (3.7), and the approximate value of $H_{\max}$ from (3.11). We conclude that for $P_0 \ge 1$ no significant difference exists between the exact value of $H_{\max}$ given by (3.7) and the approximation given by (3.11). Both (3.7) and (3.11) appear as the envelope ($N = \infty$) in Fig. 1.

The optimal distribution Δ_n is unfortunately ill-suited for adaptation to a practical modulation scheme. The exact distribution is in fact

$$\Pr\{\Delta = k\} = e^{-k^2\lambda}/m(\lambda), \tag{3.12}$$

a nonuniform distribution on a countable set $\{0, \pm 1, \pm 2, \cdots\}$ of values. However, we can get a surprisingly large entropy by considering instead of (3.12) a much simpler random variable $\Delta^{(L)}$ which is uniformly distributed on $\{-L, -L+1, \cdots, L-1\}$:

$$\Pr\{\Delta^{(L)} = k\} = \begin{cases}\dfrac{1}{2L}, & \text{if } -L \le k \le L-1\\ 0, & \text{otherwise}\end{cases}. \tag{3.13}$$

For the sequence $\{\cdots, X_{-1}, X_0, X_1, \cdots\}$ whose increments $X_{n+1} - X_n$ are i.i.d. with common distribution $\Delta^{(L)}$, a simple calculation gives

$$P = E(X_{n+1} - X_n)^2 = \frac{L^2}{3} + \frac{1}{6} \tag{3.14}$$

$$H = \log(2L). \tag{3.15}$$

Thus for this particular Markov chain the relationship between the entropy H and the power P_0 is

$$H = \tfrac{1}{2}\log P_0 + \tfrac{1}{2}\log 12 + \tfrac{1}{2}\log(1 - 1/6P_0). \tag{3.16}$$

Comparing (3.11) and (3.16), we see that the difference in entropy between the optimal distribution of increments (3.12) and the suboptimal distribution (3.13) is approximately $(1/2)\log(\pi e/6) = 0.255$ b when P_0 is large. This result is shown graphically in Fig. 2.

TABLE I
COMPARISON OF MAXIMUM ENTROPY TO APPROXIMATION

λ	P	$H_{\max}$ (Exact)	$H_{\max}$ (From (3.11))
1.00	0.499	1.0715	1.0713
0.95	0.526	1.0974	1.0974
0.90	0.5551	1.1247	1.1247
0.85	0.5880	1.1534	1.1534
0.80	0.6249	1.1838	1.1838

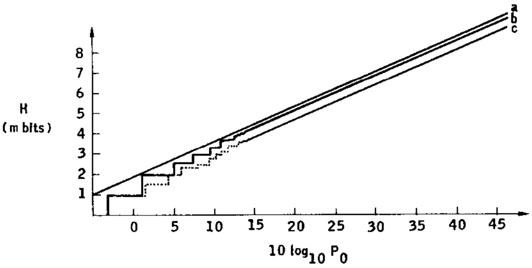

Fig. 2. Exact solution for $N = \infty$ versus two approximations. a: Exact, from Section II. b: From (3.16). c: From (3.19).

Note that this factor $1/2\log(\pi e/6)$ occurred in [3] in a similar but apparently not identical context. There the problem was to *minimize* the entropy of a process whose variance is *above* a certain value. The minimum entropy was shown to be 0.255 b/sample above the rate distortion function of the source.

As a comparison we consider the Markov chain $\{\cdots, X_{-1}, X_0, X_1, \cdots\}$ in which the *components* X_i are i.i.d. uniformly distributed on $\{1, 2, \cdots, L\}$. In this case it is easy to calculate (cf. (3.14) and (3.15))

$$P_0 = \frac{L^2}{6} - \frac{1}{6} \tag{3.17}$$

$$H = \log L \tag{3.18}$$

from which follows

$$H = \frac{1}{2}\log P_0 + \frac{1}{2}\log 6 + \frac{1}{12P_0} + 0\left(P_0^{-2}\right). \tag{3.19}$$

Comparing (3.19) and (3.16), we see that for a given value of P_0, a Markov chain with uniform and independent increments gives about one half bit more entropy than one whose components are independent and uniform. Stated in another way, for a fixed value of H, the distribution (3.13) requires about 1.53 dB more power than the optimal distribution, whereas a uniform distribution on the X's requires 4.53 dB more power.

Motivated by these results, we now introduce an encoding process that maps a sequence of zeros and ones into a sequence of elements from $\{-L, -L+1, \cdots, L-1\}$ that closely resembles a typical sequence from the Markov chain whose increments are described by (3.13).

B. An Encoding Algorithm

Motivated by the aforementioned results we propose a method of encoding a random stream of zeros and ones, say $U_1, U_2, U_3, \cdots$, into a sequence $X_1, X_2, \cdots$ of elements from the set

$$\{-N/2, -N/2+1, \cdots, 0, 1, \cdots, N/2-1\}$$

such that $E(|X_n - X_{n+1}|^2) \le P_0$ is satisfied. For notational convenience in this section we assume that N is even, and let the state set be as described, instead of the $\{1, 2, \cdots, N\}$.

We begin by choosing an even integer $L \le N$ and a sequence $\Delta_1, \Delta_2, \cdots$ of i.i.d. random variables, each uniformly distributed in the set $\{-L/2, -L/2+1, \cdots, L/2-1\}$. If we define the Markov chain $\{X'_m\}_{m \ge 0}$ by

$$X'_0 = 0$$

$$X'_{m+1} = X'_m + \Delta_{m+1},$$

it follows from the foregoing results that

$$E\left(|X'_m - X'_{m+1}|^2\right) = \frac{L^2}{12} + \frac{1}{6}$$

whereas the entropy of the chain is $\log L$. This chain is not yet satisfactory, since X'_m may not lie in the set $\{-N/2, -N/2+1, \cdots, N/2-1\}$. The following modification remedies this:

$$X_0 = 0$$

$$X_{m+1} = \begin{cases} \min(X_m, B) + \Delta_{m+1}, & \text{if } X_m \ge 0 \\ \max(X_m, -B-1) - \Delta_{m+1}, & \text{if } X_m < 0 \end{cases} \tag{3.20}$$

where

$$B = \frac{N}{2} - \frac{L}{2}.$$

The number B is the largest value for X'_m that guarantees $X'_{m+1} \le N/2 - 1$. Similarly, $-B-1$ is the smallest value for X'_m that guarantees $X'_{m+1} \ge -N/2$. It follows that the chain $\{X_m\}$ defined by (3.20) will lie in the desired range $-N'/2 \le X_m < N/2$. The entropy of $\{X_m\}$ is still $\log L$ bits, since $H(X_{m+1}|X_m) = \log L$ for all m. However, the value of $E(|X_m - X_{m+1}|^2)$ will be somewhat larger than the corresponding value for $\{X'_m\}$, since when $X_m > B$ or $X_m < -B-1$ the difference $X_{m+1} - X_m$ will no longer be uniformly distributed on $\{-L/2, \cdots, L/2-1\}$. However, since $E(\Delta_m) = -1/2$, (3.20) causes the chain to be attracted to zero and unlikely to lie near the boundaries. We make this precise in the Appendix and show that in fact

$$E\left(|X_{m+1} - X_m|^2\right) \le \frac{L^2}{12}(1 + r^B) + \frac{1}{6} \tag{3.21}$$

where r is the unique solution in $(0,1)$ to the equation

$$\sum_{k=-L/2+1}^{L/2} Z^k = L.$$

Table II gives the value of r for $L = 2^k$, $k = 2, 3, \cdots, 7$.

TABLE II
SOLUTION TO $\sum_{k=-(L/2)+1}^{L/2} Z^k = L$

L	r
4	0.414214
8	0.823408
16	0.953817
32	0.988325
64	0.997073
127	0.999268

In Fig. 3 we have plotted (for $N = 256$) $H = \log L$ versus $P_0 = E(|X_m - X'_m|^2)$ for $L = 2, 4, 6, \cdots, 256$, together with the exact solution to the entropy-maximization problem given in Section II. Asymptotically, the two curves are indeed 0.255 b apart as predicted.

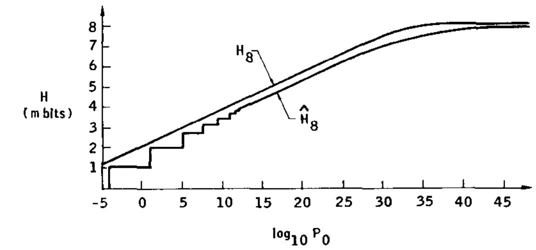

Fig. 3. Exact solution H_∞ for $N = 256$ versus approximation $\hat{H}_8$.

Here is our proposed encoding algorithm. We choose an integer L as before but now require that L be a power of two, say $L = 2^H$ where H is an integer. The sequence $\{X_m\}$ is again defined by (3.20), but now the increments Δ_m are controlled by the data stream $U_1, U_2, U_3, \cdots$. In particular, Δ_m is determined by the $(m+1)$st block of H data bits, viz. $(U_{Hm+1}, \cdots, U_{H(m+1)})$. If the U's are i.i.d and equally likely to be zero or one, the Δ's will be i.i.d. as required, and the performance of the encoding will follow the lower curve in Fig. 3 except that only integral values of H will occur. We conclude with a simple numerical example.

Example: Let $N = 64$, $P = 23$. Choose $L = 16$, $H = 4$, $B = 24$. The data sequence $u = (1000\ 0110\ 0111\ 0110\ 1100 \cdots)$ yields the increment sequence $\Delta_1 = -8$, $\Delta_2 = 6$, $\Delta_3 = 7$, $\Delta_4 = 6$, $\Delta_5 = -4$, and so by (3.20) we have $X_0 = 0$ and $(X_1, X_2, \cdots) = (-8, -14, -21, -27, -21, \cdots)$.

The entropy of this chain is $H = 4$ b, and from (3.21) and Table II

$$E\left(|X_{m+1} - X_m|^2\right) \le \frac{64}{3}\left(1 + (0.953817)^{24}\right) + \frac{1}{6} = 28.36.$$

In fact, an exact calculation of the steady-state probabilities for this chain shows that $P = 22.6$. As a comparison, note that (3.11) shows that the largest possible entropy for a Markov chain with $P = 22.6$ is 4.3 b. The performance of our algorithm in this example is very close to the predicted 0.255-b loss.

APPENDIX

This appendix uses Kingman's bound to bound the value of $E(X_{m+1} - X_m)^2$, where $X_1, X_2, \cdots$ is the process of Section III-B. This is possible because Kingman's bound gives a maximum possible value to $P(X_m \ge l)$ for $|l| > B$, while we have $E((X_{m+1} - X_m)^2 \mid |l| \le B) = (2^{2H-2}/3) + (1/6)$. Throughout this appendix K will mean 2^{H-1}, and N is an even integer larger than $2K$.

Lemma 1: Let $Y_1, Y_2, \cdots$ be i.i.d. random variables, $P(Y_n = l) = K/2$, $-K \le l < K$. Let $W_0 = 0$, and let $W_{n+1} = \max(0, W_n + Y_{n+1})$. Let r, $0 < r < 1$, be a solution of $\sum_{j=-K+1}^{K} z^j = 2K$. Then $P(X_m \ge l) \le r^l$.

Lemma 1 is a special case of Kingman's bound [7, p. 44].

Lemma 2: Let $Y_1, Y_2, \cdots$ be as in Lemma 1. Let $T_0 = 0$ and $B = N/2 - K$, and define $T_1, T_2, \cdots$ by

$$T_{n+1} = \begin{cases} \min(T_n, B) + Y_{n+1}, & \text{if } T_n \geq 0 \\ \max(T_n, -B-1) - Y_{n+1}, & \text{if } T_n < 0 \end{cases}.$$

Let r satisfy $0 < r < 1$ and

$$\frac{1}{2K} \sum_{j=-K+1}^{K} r^j = 1.$$

Then for all n and all l, we have $P(T_n \geq l) \leq r^l$.

Proof: The statement is trivial for $l \leq 0$. For any sequence $Y_1, Y_2, \cdots, T_n \leq W_n$, where W_n is the process of Lemma 1, and so the statement is true for $l \geq 0$.

Theorem 2: For the process $X_1, X_2 \cdots$ described in Section III the steady-state probability P satisfies

$$P(X_m \geq l) \leq \frac{(N/2 - l)}{K} \cdot r^B$$

for $l \geq B$, where $0 < r < 1$, and

$$\frac{1}{2K} \sum_{j=-K+1}^{K} r^j = 1.$$

Proof: Given that $X_0 = 0$, $X_1, X_2, \cdots$ has exactly the same distribution as $T_1, T_2, \cdots$ in Lemma 2, thus $P(X_m \geq B) \leq r^B$.

Separately, for $B \leq l < N/2$,

$$P(X_{m+1} = l) = \frac{1}{2K} \sum_{j=l-K+1}^{N/2-1} P(X_m = j)$$

so $P(X_m = B) \geq P(X_m = B+1) \geq \cdots \geq P(X_m = N/2 - 1)$. Therefore,

$$P(X_m \geq l) \leq \frac{(N/2 - l)}{K} \cdot r^B.$$

Corollary: $E(X_{m+1} - X_m)^2 \leq K^2/3 + 1/6 + (4K^2/3)r^B$.

Proof: As in (3.14), $E((X_{m+1} - X_m)^2 | -B-1 \leq X_m \leq B) = (K^2/3) + 1/6$. Therefore

$$E(X_{m+1} - X_m)^2 < \frac{K^2}{3} + 1/6 + 2 \sum_{l=B}^{N/2-1} P(X_m = l) \cdot E\big((X_{m+1} - X_m)^2 | X_m = l\big).$$

However,

$$\begin{aligned} &2 \sum_{l=B}^{N/2-1} P(X_m = l)\, E\big((X_{m+1} - X_m)^2 | X_m = l\big) \\ &\leq 2 \sum_{l=B}^{N/2-1} \frac{r^B}{K} \cdot \left(\frac{1}{2K} \sum_{i=B-K}^{N/2-1} (i - l) \right) \\ &= \frac{r^B}{K^2} \sum_{l=B}^{N/2-1} \sum_{i=B-K}^{N/2-1} (i-l)^2 \\ &= \frac{(2K+1)(2K-1)}{3} r^B < \frac{4K^2}{3} r^B. \end{aligned}$$

References

[1] M. M. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inform. Theory*, vol. IT-26 pp. 710–715, 1980.

[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[3] H. Gish and J. N. Pierce, "Asymptotically efficient quantization," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 676–683, Sept. 1968.

[4] J. Justenson and T. Hoholdt, "Maxentropic Markov chains," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 665–666, July 1984.

[5] J. Katz, "2.5 bit/detected photon demonstration program: Phase II and III experimental results," Jet Propulsion Lab. TDA Progress Rep. 42-70, 1982, pp. 95–104.

[6] J. Katz, private communication.

[7] L. Kleinrock, *Queuing Systems, Vol. 2: Computer Applications*. New York: Wiley, 1976.

[8] J. R. Lesh "Optical communications research to demonstrate 2.5 bits/detected photon," *IEEE Commun. Mag.*, pp. 35–37, Nov. 1982.

[9] R. J. McEliece, *The Theory of Information and Coding*. Reading, MA: Addison-Wesley, 1977.

[10] ——, "Practical codes for photon communication," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 393–398, 1981.

[11] M. Rosenblatt, *Random Processes*. New York: Oxford Univ. Press, 1962.

[12] D. L. Snyder and I. B. Rhodes, "Some implications of the cutoff rate criterion for coded direct detection optical communications systems," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 337–338, 1980.

A Simple Derivation of Lloyd's Classical Result for the Optimum Scalar Quantizer

NEIL JUDELL AND LOUIS SCHARF, FELLOW, IEEE

***Abstract*—The classical result of Lloyd for the optimum scalar quantizer in the asymptotic case of fine quantization is derived from first principles. The derivation is offered as a simple alternative to Lloyd's original and elegant piece of analysis, and the result is used to derive the optimum compander. We then show why a compander that presents uniformly distributed random variables to the quantizer is not a good idea.**

I. Introduction

In this correspondence we derive Lloyd's classical result for the optimum scalar quantizer in the asymptotic case of fine quantization. A related derivation was begun in [2, pp. 136 *et. seq.*], but constraints were mysteriously introduced and the development was left uncompleted. We offer the derivation as a simple alternative to Lloyd's original derivation and use the result to derive the optimum compander. We then show why a compander that presents uniformly distributed random variables to the quantizer is not a good idea.

II. Derivation

Let X denote a continuous scalar random variable that takes real values x according to the density $p(x)\,dx$. Let $\hat{x} = Q(x)$ represent a quantized value of x. Assume $\hat{x}$ lies in the discrete set of representation values $\{\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_L\}$. As shown by Lloyd [1] and Max [3] the representation values $\hat{\hat{x}}_k$ that minimize mean-squared quantization error are determined as the conditional means of the random variable X, given that X lies in an interval I_k

$$\begin{aligned} \hat{\hat{x}}_k &= E[X/X \in I_k] \\ I_k &= \{x: x_k \leq x < x_{k+1}\} \\ x_k &= (\hat{\hat{x}}_k + \hat{\hat{x}}_{k-1})/2. \end{aligned}$$

Refer to Fig. 1.

Manuscript received October 17, 1984; revised April 8, 1985. This work was supported in part by the Office of Naval Research, Statistics and Probability Branch, Arlington, VA, under Contract N00014-82-K-0256.

N. Judell is with the Electrical Engineering Department, University of Rhode Island, Kingston, RI 02881.

L. Scharf is with the Department of Electrical and Computer Engineering, Engineering Center OT-2-32, University of Colorado, Boulder, CO 80309.

IEEE Log Number 8406616.