An Easy-to-Implement Coding Scheme for Multifrequency PPM

R. J. McEliece and L. Swanson
Communications Systems Research Section

In implementing multifrequency PPM, a naturally arising question is: Let P be a fixed number; among all integer valued processes $X_1, X_2, X_3, \ldots$ with $E(|X_{n+1} - X_n|^2) \leqslant P$, which has the largest entropy? Earlier work by McEliece and Rodemich answered this question, but there is no obvious way to use this process to implement a code for multifrequency PPM. The present article describes an easy-to-implement process $X_1, X_2, \ldots$, with $E(|X_{n+1} - X_n|^2) \leqslant P$, whose entropy is nearly as great as that of the McEliece-Rodemich process.

I. Introduction

McEliece and Rodemich, in a study of the feasibility of using multifrequency PPM for optical communication (Ref. 1), were led to consider the following problem. Among all stationary random processes $\{\ldots, X_{-1}, X_0, X_1, \ldots\}$ taking values in the set $\{1, 2, \ldots, N\}$ and subject to the constraint

$$E(|X_n - X_{n+1}|^2) \leqslant P \tag{1}$$

how large can the entropy be? McEliece and Rodemich showed that the maximum entropy is achieved by a Markov chain and gave an exact formula for the transition probabilities of the maximizing chain (Ref. 1). In order to apply these results to the original problem in optical communication, however, it would be necessary to encode a given binary data stream into a sequence of symbols from $\{1,2, \ldots, N\}$ which closely resembles a "typical sequence" from the optimal Markov chain. Given the complex form of the solution given in Ref. 1, such an encoding would very likely be difficult to implement. In this article we shall examine some approximations to the solution given in Ref. 1 which are not far from optimal (about 0.25 bit) and which suggest practical ways to implement energy-efficient multicolor PPM.

II. Approximations

In this section we attempt to motivate the construction to be given in Section III, as follows. We consider a simplified version of the problem mentioned in the Introduction, viz, we enlarge the state space from $\{1,2, \ldots, N\}$ to $\mathbf{Z}$, the set of all integers. For a given value of P this will increase the maximum possible entropy, but if $P \ll N^2$ this increase will be small. In any case our object in this section is to compare the entropy vs. P relationship for three possible Markov chains on $\mathbf{Z}$ subject to Eq. (1): the optimal Markov chain and two others. In Section III we will modify one of the suboptimal chains discussed here to devise a practical modulation scheme for multicolor optical communication, whose performance is (in a sense to be given below) within 0.25 bits of the theoretical optimum.

With the state set $\mathbf{Z}$, the maximum entropy Markov chain subject to Eq. (1) is a Markov chain for which the increments

$$\Delta_n = X_{n+1} - X_n \tag{2}$$

are independent and identically distributed. The entropy of the resulting chain is just the entropy of the random variable

Δ_n, and so to solve the original problem on $\mathbf{Z}$ we just need to maximize $H(\Delta)$ subject to the condition $E(\Delta^2) \leqslant P$. This can be done using straightforward variational techniques (Ref. 2, Problem 1.8), and the maximum entropy is given parametrically as follows:

For each $\lambda > 0$, define

$$m(\lambda) = \sum_{n=-\infty}^{\infty} e^{-n^2\lambda} \tag{3}$$

so that

$$-m'(\lambda) = \sum_{n=-\infty}^{\infty} n^2 e^{-n^2\lambda} \tag{4}$$

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

$$H_{\max} = \log m(\lambda) - \lambda m'(\lambda)/m(\lambda) \tag{5}$$

where

$$P = -\frac{m'(\lambda)}{m(\lambda)} \tag{6}$$

For small values of λ the sums in Eqs. (3) and (4) are well approximated by the following integrals:

$$m(\lambda) \approx \int_{-\infty}^{\infty} e^{-n^2\lambda}\, dx = \sqrt{\pi/\lambda} \tag{7}$$

$$-m'(\lambda) \approx \int_{-\infty}^{\infty} x^2 e^{-n^2\lambda}\, dx = \frac{1}{2\lambda}\sqrt{\pi/\lambda} \tag{8}$$

Using these approximations in Eqs. (5) and (6), we obtain the approximation

$$H_{\max} \approx \frac{1}{2}\log P + \frac{1}{2}\log 2\pi e, \quad P \text{ Large} \tag{9}$$

which is extremely good even for small values of P, as exhibited in Table 1. In Table 1 we have tabulated, for a range of λ's, the corresponding value of P calculated from Eq. (6), the exact value of $H_{\max}$ calculated from Eq. (5), and the approximate value of $H_{\max}$ from Eq. (9). We conclude that for $P \geqslant 1$, there is no significant difference between the exact value of $H_{\max}$ given by Eq. (5) and the approximation given by Eq. (9).

The optimal distribution Δ_n is unfortunately not well suited for adaptation to a practical modulation scheme. The exact distribution is in fact

$$Pr\,\{\Delta = k\} = e^{-n^2\lambda}/m(\lambda) \tag{10}$$

a nonuniform distribution on a countable set $\{0, \pm 1, \pm 2, \ldots\}$ of values. However, we can get a surprisingly large entropy by considering instead of Eq. (10) a much simpler random variable $\Delta^{(L)}$, which is uniformly distributed on $\{-L, -L+1, \ldots, L-1\}$:

$$\begin{aligned} Pr\,\{\Delta^{(L)} = K\} &= \frac{1}{2L} \quad \text{if } -L \leqslant K \leqslant L-1 \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{11}$$

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

$$\begin{aligned} P &= E\,(X_{n+1} - X_n)^2 \\ &= E\,(\Delta^{(L)\,2}) \\ &= \frac{L^2}{3} + \frac{1}{6} \end{aligned} \tag{12}$$

$$H = \log(2L) \tag{13}$$

Thus for this particular Markov chain, the relationship between the entropy H and the "power" P is

$$H = \frac{1}{2}\log P + \frac{1}{2}\log 12 + \frac{1}{2}\log\left(1 - \frac{1}{6P}\right) \tag{14}$$

Comparing Eqs. (9) and (14), we see that the difference in entropy between the optimal distribution of increments (Eq. 10) and the suboptimal distribution (Eq. 11) is approximately $1/2 \log \pi e/6 = 0.255$ bits.

As a comparison, we consider the Markov chain $\{\ldots, X_{-1}, X_0, X_1 \ldots\}$ in which the *components* X_i are i.i.d., uniformly distributed on $\{1, 2, \ldots, L\}$. In this case it is easy to calculate [cf. Eqs. (12) and (13)],

$$P = \frac{L^2}{6} - \frac{1}{6} \tag{15}$$

$$H = \log L \tag{16}$$

from which follows

$$H = \frac{1}{2}\log P + \frac{1}{2}\log 6 + \frac{1}{12P} + 0\,(P^{-2}) \qquad (17)$$

Comparing Eqs. (17) and (14), we see that for a given value of P, a Markov chain with uniform and independent increments gives one-half bit more entropy than one whose components are independent and uniform. For a fixed value of H, the distribution (Eq. 11) 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, in the next section we introduce an encoding process which maps a sequence of 0's and 1's into a sequence of elements from $\{-L, -L+1, \ldots, L-1\}$ that closely resembles a typical sequence from Markov chain whose increments are described by Eq. (11).

III. An Encoding Algorithm

Motivated by the results of Section II, we propose a method of encoding a random stream of 0's and 1's, say U_1, U_2, U_3, ... into a sequence X_1, X_2, ... of elements from the set $\{-N/2, -N/2+1, \ldots, 0, 1, \ldots, N/2-1\}$, such that Eq. (1) is satisfied.[1]

First choose L to be the largest integer such that

$$\frac{L^2}{3} + \frac{1}{6} < P \qquad (18)$$

and H to be the largest integer such that

$$2^{H-1} \leqslant L \qquad (19)$$

In our encoding algorithm the symbol X_{m+1} will be determined by X_m and the $(m+1)$-st block of H data bits, viz $[U_{Hm+1}, \ldots, U_{H(m+1)}]$. These H bits are in fact used to determine an integer Δ_{m+1} in the range $[-2^{H-1}, 2^{H-1}-1]$ via two's complement arithmetic. For example,

$$\Delta_1 = \sum_{k=1}^{H} U_k\, 2^{H-k} - U_1\, 2^H \qquad (20)$$

[1]For notational convenience, in this section we assume N is even, and let the state set be as described, instead of $\{1, 2, \ldots, N\}$.

The random variables Δ_1 Δ_2, ... are i.i.d., each being uniformly distributed between -2^{H-1} and $2^{H-1}-1$. Hence if we define the Markov chain $\{X'_m\}_{m\geqslant 0}$ by

$$\begin{aligned} X'_0 &= 0 \\ X'_{m+1} &= X'_m + \Delta_{m+1} \end{aligned} \qquad (21)$$

it follows from the results of Section II that the resulting Markov chain has [cf. Eq. (12)]

$$\begin{aligned} E(|X'_m - X'_{m+1}|^2) &= \frac{(2^{H-1})^2}{2} + \frac{1}{6} \\ &\leqslant \frac{L^2}{3} + \frac{1}{6} < P \end{aligned} \qquad (22)$$

by Eqs. (18) and (19). The entropy of $\{X'_m\}$ is H bits. Thus from Section II, if P is large, the difference between the entropy of this chain and that of the optimal Markov chain is only about 0.255 bits.

However the chain $\{X'_m\}$ is not satisfactory, since there is no guarantee that X'_m will lie between $-N/2$ and $N/2-1$. The following definition remedies the situation.

$$\begin{aligned} X_0 &= 0 \\ X_{m+1} &= \min(X_m, B) + \Delta_{m+1} && \text{if } X_m \geqslant 0 \\ &= \max(X_m, -B-1) - \Delta_{m+1} && \text{if } X_m < 0 \end{aligned} \qquad (23)$$

where

$$B = \frac{N}{6} - 2^{H-1} \qquad (24)$$

The number B defined in Eq. (24) is the largest value for X'_m that guarantees $X'_{m+1} \leqslant N/2-1$. Similarly $-B-1$ is the smallest value for X'_m that guarantees $X'_{m+1} \geqslant -N/2$. It follows that the chain $\{X_m\}$ defined by Eq. (23) will lie in the desired range $-N/2 \leqslant X_m < N/2$. The entropy of $\{X_m\}$ is still H bits, since $H(X_{m+1}|X_m) = H$ for all m. 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 $\{-2^{H-1}, \ldots, 2^{H-1}-1\}$. However, since $E(\Delta_m) = -1/2$ the rule [Eq. (23)] causes the chain to be attracted to 0, and unlikely to lie near the boundaries. In the Appendix we make this precise and show that in fact

$$E(|X_{m+1} - X_m|^2) \leqslant \frac{2^{2H}}{12}(1 + r^B) + \frac{1}{6} \tag{25}$$

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

$$\sum_{k=-2^{H-1}+1}^{2^{H-1}} z^k = 2^H \tag{26}$$

Table 2 gives the value of these roots for $H = 1,2, \ldots, 7$. It follows that for a fixed P, if N is sufficiently large, Eq. (1) will be satisified by the Markov chain $\{X_m\}$.

We conclude with a simple numerical example.

Example: Let $N = 64$, $P = 23$. From Eqs. (18), (19), (24) we have $L = 8$, $H = 4$, $B = 24$. The data sequence $\mathbf{u} = (1000\ 0110\ 0111\ 0110\ 1100 \ldots)$ yields the increment sequence $\Delta_1 = -8$, $\Delta_2 = 6$, $\Delta_3 = 7$, $\Delta_4 = 6$, $\Delta_5 = -4$, and so by Eq. (23) we have $X_0 = 0$ and $(X_1, X_2, \ldots) = (-8, -14, -21, -27, -21, \ldots)$. The entropy of this chain is $H = 4$ bits, and from Eq. (25) and Table 2

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

In fact, an exact calculation of the steady state probabilities for this chain shows that $P = 22.6$. As comparison, we note that Eq. (9) shows that the largest possible entropy for a Markov chain with $P = 22.6$ is 4.3 bits. The performance of our algorithm in this example is very close to the 0.255 bit loss predicted in Section II.

References

1. McEliece, Robert J., and Rodemich, Eugene, "A Maximum Entropy Markov Chain," *Proc. 17th Annual Conference on Information Sciences and Systems*, Johns Hopkins University, Baltimore, MD (1983), pp. 245-248.

2. McEliece, Robert J., *The Theory of Information and Coding, Encyclopedia of Mathematics and Its Applications*, Addison-Wesley Publishing Company, 1977.

3. Kleinrock, Leonard, *Queuing Systems*, Volume 2: *Computer Applications*, John Wiley and Sons, Inc., 1976.

Table 1. Comparison of H_{max} to its approximation

λ	P	H_{max} (exact)	H_{max} [from Eq. (9)]
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

Table 2. Roots of Eq. (26)

H	r
2	0.414214
3	0.823408
4	0.953817
5	0.988325
6	0.997073
7	0.999268

Appendix

This appendix uses Kingman's bound, below, to bound the value of $E(X_{m+1} - X_m)^2$, where $X_1, X_2, \ldots$ is the process of Section III. This is possible because Kingman's bound gives a maximum possible value to $P(X_m \geqslant \ell)$ for $|\ell| > B$, while we have

$$E\left((X_{m+1} - X_m)^2 \;\middle|\; |\ell| \leqslant B\right) = \frac{2^{2H-2}}{3} + 1/6$$

Throughout this appendix, K will mean 2^{H-1}, and N is an even integer larger than $2K$.

Lemma 1: Let $Y_1, Y_2, \ldots$ be i.i.d. random variables, $P(Y_n = \ell) = 1/2K$, $-K \leqslant \ell < K$. Let $W_0 = 0$, and let $W_{n+1} = \max(0, W_n + Y_{n+1})$. Let r, $0 < r < 1$, be a root of

$$\sum_{j=-K+1}^{K} z^j - 2K$$

Then $P(X_m \geqslant \ell) \leqslant r^\ell$. Lemma 1 is a special case of Kingman's bound (Ref. 3, p. 44).

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

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

Let r, again, 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 $P(T_n \geqslant \ell) \leqslant r^\ell$.

Proof: The statement is trivial for $\ell \leqslant 0$. For any sequence $Y_1, Y_2, \ldots, T_n \leqslant W_n$, where W_n is the process of lemma 1, and so the statement is true for $\ell \geqslant 0$.

Theorem: For the process $X_1, X_2, \ldots$ described in Section III, the steady state probability P satisfies, for $\ell \geqslant B$,

$$P(X_m \geqslant \ell) \leqslant \frac{(N/2 - \ell)}{K} \cdot r^B$$

where $0 < r < 1$ and

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

Proof: Given that $X_0 = 0, X_1, X_2, \ldots$ has exactly the same distribution as $T_1, T_2, \ldots$ in lemma 2. Thus $P(X_m \geqslant B) \leqslant r^B$.

Separately, for $B \leqslant \ell < N/2$,

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

So

$$P(X_m = B) \geqslant P(X_m = B + 1) \geqslant \ldots \geqslant P(X_m = N/2 - 1)$$

Therefore

$$P(X_m \geqslant \ell) \leqslant \frac{(N/2 - \ell)}{K} \cdot r^B$$

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

Proof: As shown in Section II,

$$E\left((X_{m+1} - X_m)^2 \;\middle|\; -B - 1 \leqslant X_m \leqslant B\right) = \frac{K^2}{3} + 1/6$$

Therefore

$$E(X_{m+1} - X_m)^2 < \frac{K^2}{3} + \frac{1}{6}$$

$$+ 2 \sum_{\ell=B}^{N/2-1} P(X_m = \ell)$$

$$\times E\left((X_{m+1} - X_m)^2 \mid X_m = \ell\right)$$

But

$$2 \sum_{\ell=B}^{N/2-1} P(X_m = \ell) \, E((X_{m+1} - X_m)^2 \mid X_m = \ell)$$

$$\leqslant 2 \sum_{\ell=B}^{N/2-1} \frac{r^B}{K} \cdot \left(\frac{1}{2K} \sum_{i=B-K}^{N/2-1} (i-\ell)^2 \right)$$

$$= \frac{r^B}{K^2} \sum_{\ell=B}^{N/2-1} \sum_{i=B-K}^{N/2-1} (i-\ell)^2$$

$$= \frac{(2K+1)(2K-1)}{3} \, r^B$$

$$< \frac{4K^2}{3} \, r^B$$