Journal of Statistical Physics, Vol. 3, No. 3, 1971

Transmission Time and Bandwidth in Pulse Code Modulation (Application of the Statistical Thermodynamic Formalism)

C. Huang[1] and H. Reiss[1]

Received January 18, 1971

Employing the "statistical thermodynamic formalism" developed in an earlier paper, it is possible to determine "compact" sets of transmission times for the words of PCM (pulse code modulation) messages. In particular, we deal with pulses of zero or unit heights. These compact signals, which lead to shorter message times and eliminate redundancy even when successive words are correlated (Markov source), may, however, require additional bandwidth. We examine two simple cases where autocorrelation functions, and therefore power spectra, can be evaluated. In one case, that of the Markov source, it proves possible to accomplish *both* shorter transmission time and narrower bandwidth (half-width of the power spectrum), showing that optimization of transmission times can be very worthwhile. Techniques for deriving autocorrelation functions are discussed at length.

KEY WORDS: Pulse code modulation; transmission time; bandwidth; statistical mechanics; power spectrum.

1. INTRODUCTION

In a recent paper,[1] the authors developed a connection between the methods of statistical thermodynamics (especially the methodology associated with the many-body problem) and certain problems of coding, in information theory. In that paper, a particularly simple concrete example was treated, representing a case of pulse code modulation (PCM). It may be described as follows. A continuous signal is sampled,[2] say at time intervals of $1/2\omega$, where ω is the bandwidth, and the samples are converted into binary numbers for transmission by pulse code modulation (PCM)[3] over a noisy channel. In order to combat noise, check digits[4] may be added to each binary

[1] Department of Chemistry, University of California, Los Angeles, California.

code word; and to gain a measure of compactness in advance of binary coding, some procedure, such as Huffman[5] or Fano[6] coding, may be employed. As a final generalization, the same procedure may be used for the transmission of discrete, as well as continuous, messages. For example, the binary numbers, instead of representing samples of a continuous message, may correspond to the letters of the alphabet which appear in the sequence of some text being transmitted. In any event, the pulses are reconstituted into the original signal at the receiver end of the system.

The transmitted message then consists of a sequence of zeros and ones. This sequence will have a set of statistics generated by the constraints in the original message, those implicit in Huffman or Fano coding, and in the method of assigning check digits. The chance that a given digit will be zero or one depends, in some way, on the preceding digits. This correlation amounts to a redundancy which can still be squeezed out of the message to be transmitted. Even in the absence of correlation, there may be redundancy implicit in unequal frequencies of appearance of zeros and ones. We assume that the statistics of the message can be determined by a suitable investigation.

Usuaully in PCM the pulses representing zeros and ones are of equal duration. By assigning different transmission times to different pulses, depending upon the statistics, it is possible to increase the rate of transmission. Of course, this is possible by merely shortening the duration of each pulse in scale; but for this, one pays the price of greater bandwidth. An important question, however, is the following. Can one, by choosing pulse transmission times of various magnitudes, decrease mean transmission time *without* simultaneously increasing bandwidth?

In the present paper, we concentrate on examining certain aspects of this question. We are not able to show that this is generally possible, but have worked out two nontrivial cases in which the above-mentioned goals can and cannot be achieved, respectively. Assuming that we are presented with a set of source probabilities, we make use of the statistical thermodynamics formalism outlined in Ref. 1 in order to choose pulse transmission times which match the code to the source and make it compact.[2] In order to examine the bandwidth question, we evaluate the autocorrelation functions[7] of the transmitted signals (as they depend on message statistics and assigned pulse transmission times), and, then, in accordance with the requirements of Wiener theory,[8] we derive the exact power spectrum[9] by Fourier transformation. With the power spectrum in hand, it is possible to estimate bandwidth. In each

[2] The term "compact" in the present case should really be replaced by "compact *relative* to scale." For example, suppose we are confronted with a source having unequal probabilities of word emission. Intuitively, one would expect the least probable (least frequent) word to be assigned the largest transmission time. There may be an additional requirement that this longest time be no shorter than a certain minimum. Then, subject to *this* requirement, the code can be made "compact" (made to have the shortest mean transmission time per word); the scale being fixed by the requirement of fixed *largest* transmission time. A practical example in which such a requirement exists is illustrated by the case when bandwidth is limited by a certain *maximum*. Since short transmission times are almost certain to be accompanied by large bandwidth, the above-mentioned "maximum" in bandwidth will force the scale of transmission times to be larger, and, in particular, will set a lower limit on the largest transmission time for a compact or optimized code.

instance, the bandwidth is determined for the case in which all pulses are of the same length regardless of the statistics, and for the case in which transmission times have been optimized by the above-mentioned matching procedure. Comparisons are then made to see if narrowing is possible while yet achieving minimum transmission time. In this way, the application of the statistical thermodynamic formalism to the investigation of this possibility and, when it is possible, to the selection of otpimized transmission times, is illustrated.

In the low, white Gaussian noise case, the noisy channel coding theorem[10]

$$I = \omega T \log_2[1 + (P/N)] \tag{1}$$

(where I is the maximum amount of information in bits which can be transmitted in time T through a channel of bandwidth ω, having average signal and noise powers P and N, respectively) places little or no restriction on the product ωT. Therefore, the possibility of maintaining I fixed while reducing both ω and T is very real. Stated another way, for the case of low noise one will in most cases be operating below channel capacity.

2. THE WIENER–KHINCHINE THEOREM

The Wiener–Khinchine theorem[8] states that the autocorrelation function $\phi(\tau)$ and the power spectrum $G(\omega)$ are a Fourier transform pair:

$$G(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau}\phi(\tau)\, d\tau \tag{2}$$

$$\phi(\tau) = (1/2\pi) \int_{-\infty}^{\infty} G(\omega)\, e^{i\omega\tau}\, d\omega \tag{3}$$

in which ω is the angular frequency and τ is the quantity defined in connection with Eq. (4). For signals generated by complex mechanisms (for example, sources using English grammar) in which repeated experiments performed under similar conditions produce results having the same statistical properties, the correlation function may be written as an ensemble average,

$$\phi(\tau) = \int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{\infty} dy_2\, y_1 y_2 p(y_1, y_2, \tau) \tag{4}$$

where y_1 and y_2 are amplitudes of the signal taken at times an interval τ apart, and $p(y_1, y_2, \tau)$ is the joint probability density function for the process. The double integration represents the ensemble average. Thus,

$$\phi(\tau) = \langle y_1 y_2 p(y_1, y_2, \tau)\rangle_E \tag{5}$$

In the cases which we study, only ones and zeros will be transmitted, and Eq. (5) simplifies to

$$\phi(\tau) = \langle p(1, 1, \tau)\rangle_E \tag{6}$$

3. MEMORYLESS SOURCE WITH UNEQUAL EMISSION PROBABILITIES

Assume that the PCM signal consists of ones and zeros whose appearances in the messages are completely uncorrelated, i.e., the probability of emission of a one or zero is completely independent of whether the symbol emitted previously was one or zero. However, we assume that in the overall message, ones and zeros appear with unequal frequencies. For example, the fractions or probabilities of ones and zeros could be

$$P_1 = 0.618, \qquad P_0 = 0.382 \tag{7}$$

respectively. These unequal probabilities represent information redunancy because the maximum information in a message composed of a fixed number of digits is transmitted when p_1 and p_0 are equal. It is therefore possible to adjust the relative transmission times for ones and zeros so as to minimize the overall transmission time of the message, except for a scale factor.[1] Suppose, however, we ignore this optimization step and arbitrarily assign the following transmission times:

$$t_1 = t_0 = \sigma \tag{8}$$

A portion of a typical message might then look like Fig. 1. In order to evaluate the autocorrelation function, we proceed as follows. Let

$$\tau = N\sigma + \epsilon \tag{9}$$

where $\epsilon \leqslant \sigma$ and N is an integer. When $N = 0$ or $\tau = \epsilon \leqslant \sigma$, there are two types of situations. There are illustrated in Fig. 2. In Fig. 2(a), the end points of τ are both within the same unit pulse, while in Fig. 2(b), they lie in different pulses. As $\tau \to 0$, the probability of the configuration in Fig. 2(a) will exceed that in Fig. 2(b), while as $\tau \to \sigma$, that in Fig. 2(b) will become the more probable. In fact, it is easy to see that

$$\begin{aligned} &\text{probability of } \tau \text{ entirely within a unit } \sigma \text{ is } 1 - (\tau/\sigma) \\ &\text{probability of } \tau \text{ having ends in two different units of } \sigma \text{ is } \tau/\sigma \end{aligned} \tag{10}$$

In order for τ to contribute to the right side of Eq. (6), both ends must fall in regions occupied by ones. Therefore, for $\tau \leqslant \sigma$,

$$\phi_0(\tau) = \langle P(1, 1, \tau)\rangle_\epsilon = (\tau/\sigma)\, P_1{}^2 + [1 - (\tau/\sigma)]P_1 \tag{11}$$

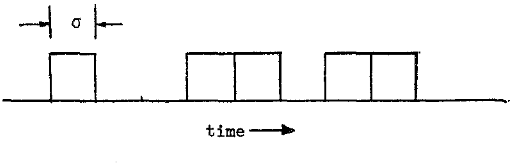

Fig. 1. Typical message for a zeroth-order Markovian source $t_1 = t_0 = \sigma$.

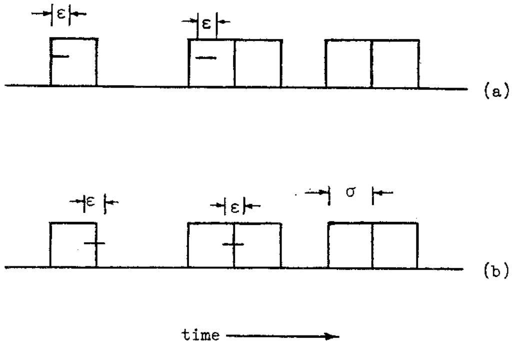

Fig. 2. Possible configurations for $\tau = \epsilon \leqslant \sigma$.

in which P_1 is the *a priori* probability for the occurrence of a one. As we might expect, when $\tau \rightarrow 0$,

$$\phi_0(\tau) \rightarrow P_1 \tag{12}$$

and for $\tau \rightarrow \sigma$,

$$\phi_0(\tau) \rightarrow P_1^2 \tag{13}$$

When $\tau \geqslant \sigma$, or $N = 1, 2, 3, \ldots$, it is clear that

$$\phi_N(\tau) = P_1^2 \tag{14}$$

since the probability that the ends of the interval τ fall in unit pulses is simply the product of the independent chances such that unit pulses exist. Thus, the autocorrelation function for messages from this source has the form illustrated in Fig. 3.

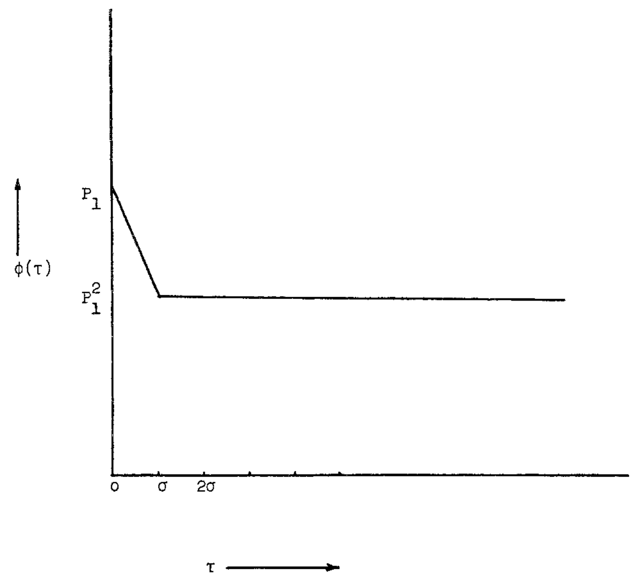

Fig. 3. Autocorrelation function for a zeroth-order Markovian source, $P_1 = 0.618$, $P_0 = 0.382$ $t_1 = t_0 = 1\sigma$.

Since $\phi(\tau)$ must be an even function of τ, the power spectrum described by Eq. (2) can be expressed in the form

$$\begin{aligned} G(\omega) &= \int_{-\infty}^{\infty} e^{-i\omega\tau}\phi(\tau)\,d\tau = \int_{-\infty}^{0} e^{-i\omega\tau}\phi(\tau)\,d\tau + \int_{0}^{\infty} e^{-i\omega\tau}\phi(\tau)\,d\tau \\ &= \int_{\infty}^{0} e^{i\omega\tau}\phi(-\tau)\,d(-\tau) + \int_{0}^{\infty} e^{-i\omega\tau}\phi(\tau)\,d\tau \\ &= \int_{0}^{\infty} e^{i\omega\tau}\phi(\tau)\,d\tau + \int_{0}^{\infty} e^{-i\omega\tau}\phi(\tau)\,d\tau \\ &= 2\,\mathrm{Re}\int_{0}^{\infty} e^{-i\omega\tau}\phi(\tau)\,d\tau \end{aligned} \tag{15}$$

Because $\phi(\tau)$ has different forms for τ greater and less than σ, Eq. (15) exhibits two terms

$$G(\omega) = 2\,\mathrm{Re}\left[\int_{0}^{\sigma} \phi_0(\tau)\,e^{-i\omega\tau}\,d\tau + \int_{\sigma}^{\infty} \phi_N(\tau)\,e^{-i\omega\tau}\,d\tau\right] \tag{16}$$

In this equation, $\phi_0(\tau)$ and $\phi_N(\tau)$ come, respectively, from Eqs. (11) and (14). Upon carrying out the indicated integrations, we obtain[3]

$$G(\omega) = [2P_1P_0(1 - \cos\omega\sigma)/\omega^2\sigma^2]\sigma + 2\pi P_1^2\,\delta(\omega) \tag{17}$$

the desired form for the power spectrum, and in which δ is the Dirac delta function, equal to infinity for $\omega = 0$ and zero otherwise.

Figure 4 is a plot of $G(\omega)$. Several points deserve comment: (1) The half-width is independent of P_1 and P_0, although the value of $G(\omega)$, at half-width, is dependent on these quantities. (2) As σ becomes smaller, the half-width increases; the power spectrum is broadened. This expected result is evident in the fact that $G(\omega)$ depends on the product $\omega\sigma$ rather than on ω alone.

Having obtained the power spectrum for the case in which transmission times for ones and zeros are identical, we next optimize these times, matching them to the source probabilities specified by Eq. (7). We define t_0 and t_1 as the transmission times going with zero and one, respectively. Since the source is without memory and we wish the code to be maximally compact, the appropriate relation between source probabilities and word transmission times is Eq. (37) of Ref. 1, with q of that equation set equal to unity in accordance with Eq. (42) of the same reference. Thus, we may write

$$t_1 = -\kappa\tau_*\ln P_1\,, \qquad t_0 = -\kappa\tau_*\ln P_2 \tag{18}$$

in which κ, the pseudo-Boltzmann constant, is $\log_2 e$ and τ_* is the information theory temperature, the star being appended in order to distinguish it from the quantity τ used in the previous equations of this paper. Substitution of the p values as specified

[3] Note $\delta(\omega) = \sigma\delta(\omega\sigma)$.[11]

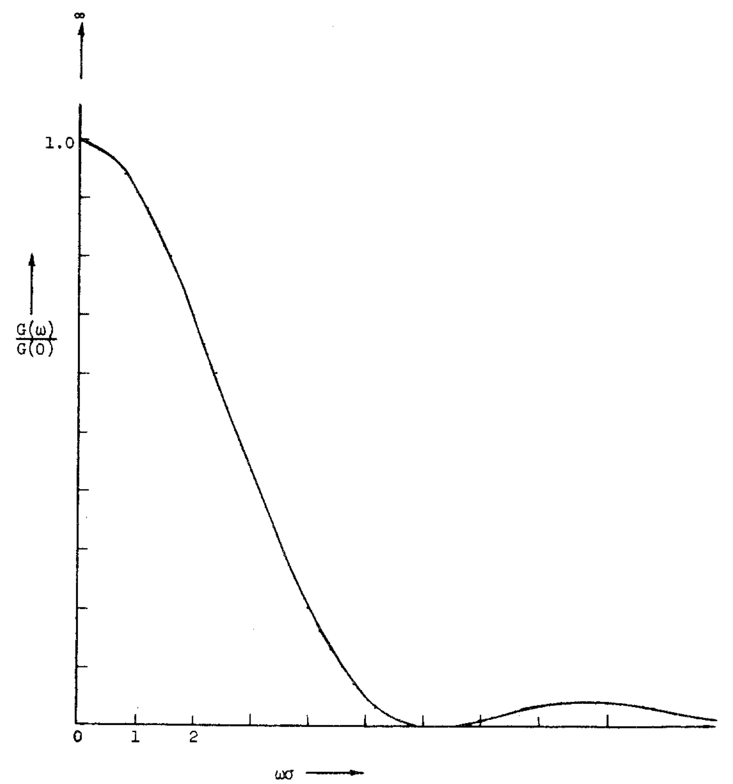

Fig. 4. Power spectrum for a zeroth-order Markovian source, $P_1 = 0.618$, $P_0 = 0.382$, $t_1 = t_0 = 1\sigma$.

by Eq. (7) into Eq. (18) indicates that the optimum t_1 is just half the optimum t_0, so that we may write

$$t_1 = \tfrac{1}{2}t_0 = \sigma \tag{19}$$

where σ is again the unit pulse length and determines the scale. As emphasized in Ref. 1, when we deal with continuous rather than block coding, compact message times are not determined absolutely but only to within a scale factor.

With this choice of transmission times, the portion of a typical message exhibited in Fig. 1 now looks like Fig. 5. Even so simple a change as that contained in the prescription given by Eq. (19) makes the determination of the autocorrelation function considerably more difficult. In fact, our choice of source probabilities has allowed

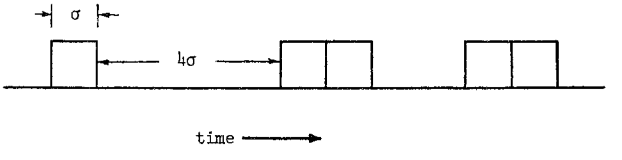

Fig. 5. Typical message of a zeroth-order Markovian source after optimization of transmission times.

the optimum transmission times to be integral multiples of a common unit, and it was for this reason that the specific choice in Eq. (7) was deliberately made. At the moment, it appears as though the analytical evaluations of autocorrelation functions, when transmission times are not related in an integral manner, would present almost insuperable difficulties. Nevertheless, if the problem is studied for source probabilities demanding a variety of integral relationships, some insight into the behavior expected under nonintegral conditions might be obtained by interpolation, although care must be taken to ensure that singular situations do not arise.

Once again we make use of Eq. (9). Since, in order for an interval τ to contribute to the right-hand side of Eq. (6), both of its ends must fall within pulses representing ones, and, in this as it was in the previous case, the transmission time for a one is σ, we have, for $N = 0$, a result similar to Eq. (11), which we now write as

$$\phi_0(\tau) = (\tau/\sigma)[P_1^{(s)}]^2 + [1 - (\tau/\sigma)]\, P_1^{(s)} \tag{20}$$

Instead of P_1 we have used $P_1^{(s)}$. This quantity may be called the *space* probability of a one as opposed to P_1 itself, which is the *time* probability of a one. In fact, the relation between $P_1^{(s)}$ and P_1 is illustrated by the following equation

$$P_1^{(s)} = P_1/(P_1 + 2P_0) \tag{21}$$

From this it can be seen that $P_1^{(s)}$ is essentially the fraction of the *space* of the message covered by pulses representing ones. The coefficient of the second term on the right in Eq. (20) is the probability that the interval τ falls entirely within a unit pulse of length σ. As we have indicated earlier, in order for it to contribute to the ensemble average leading to the autocorrelation function, it must also fall within a spatial interval occupied by a one. Since the coefficient already takes care of the probability of τ falling *entirely* within a pulse of length σ, we need only multiply it by the chance that *one* end of τ falls within the space of a one. Clearly, this probability is given by Eq. (21), since the right-hand side represents the fraction of message space occupied by ones. Since the first term on the right of Eq. (20) refers to situations in which the ends of τ fall in regions occupied by two different ones, and the source has no memory, so that ones are emitted with independent probabilities, it is $[P_1^{(s)}]^2$ which appears, multiplied by the coefficient τ/σ which gives the chance that τ does *not* fall entirely within a unit σ. For $N \geqslant 1$, the expression for the autocorrelation function is similar to that given by Eq. (20) except that each term in that equation must be modified by a factor which accounts for the fact that several different messages may be fitted into the interval between the ones in which the ends of the interval τ(now long enough to span many units of σ) fall. More precisely, these factors represents the probabilities of complete messages being fitted into the above-mentioned intervals. They are represented by the quantities $P(N)$ and $P(N - 1)$ in the following equation, which represents the autocorrelation function for the case $N \geqslant 1$:

$$\phi_N(\tau) = (\epsilon/\sigma)[P_1^{(s)}]^2\, P(N) + [1 - (\epsilon/\sigma)][P_1^{(s)}]^2\, P(N - 1) \tag{22}$$

A little thought will show that ϵ/σ is the chance that τ cover N whole units of σ, not including the two fractional units terminated by the end points of τ. As before, $[P_1^{(s)}]^2$ is the probability that both end points are ones. $P(N)$ is the probability, or normalized weight factor, that a complete message can be fitted into the interval of length $N\sigma$. This quantity was not required for the noncompact case with $t_1 = t_0 = \sigma$, since there it was always possible to fit some message, ones and zeros occupying the the same space interval. Thus, $(\epsilon/\sigma)[P_1^{(s)}]^2\, P(N)$ gives the contribution to the autocorrelation function for cases where τ "covers" N whole units. In contrast, $[1 - (\epsilon/\sigma)][P_1^{(s)}]^2\, P(N-1)$ accounts for configurations such as that exhibited in Fig. 6(a), in which one less unit is "covered."

The function $P(N)$ is evaluated in Appendix A. The result is

$$P(N) = [(P_1 - 1)^{N+1} - 1]/(P_1 - 2) \tag{23}$$

since

$$\epsilon = \tau - N\sigma \tag{24}$$

Substitution of Eqs. (23) and (24) into (22) yields

$$\phi_N(\tau) = [P_1^{(s)}]^2 \left[(P_1 - 1)^N \frac{\tau}{\sigma} - N(P_1 - 1)^N + \frac{(P_1 - 1)^N}{P_1 - 2}\right] - \frac{[P_1^{(s)}]^2}{P - 2} \tag{25}$$

where $N = 1, 2, 3, \ldots$. From Eq. (15), we find

$$\begin{aligned} G(\omega) &= 2 \operatorname{Re} \int_0^\infty e^{-i\omega\tau} \phi(\tau)\, d\tau \\ &= 2 \operatorname{Re} \left[\int_0^\sigma \phi_0(\tau)\, e^{-i\omega\tau}\, d\tau + \sum_{N=1}^{\infty} \int_{N\sigma}^{(N+1)\sigma} \phi_N(\tau)\, e^{-i\omega\tau}\, d\tau\right] \end{aligned} \tag{26}$$

in which $\phi_0(\tau)$ and $\phi_N(\tau)$ are given by Eqs. (20) and (25), respectively. The first term on the right of Eq. (26) can be treated as before, but the second term must be integrated between $N\sigma$ and $(N + 1)\sigma$, and the integrated results summed over N as a geometric

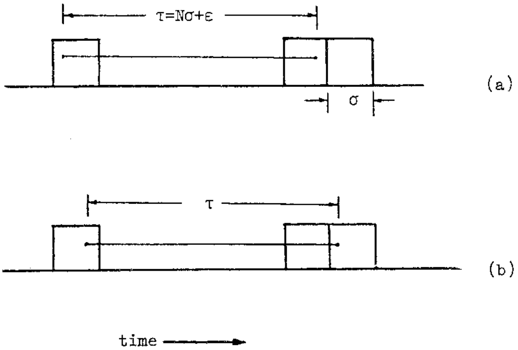

Fig. 6. Two configurations that contribute to the autocorrelation function.

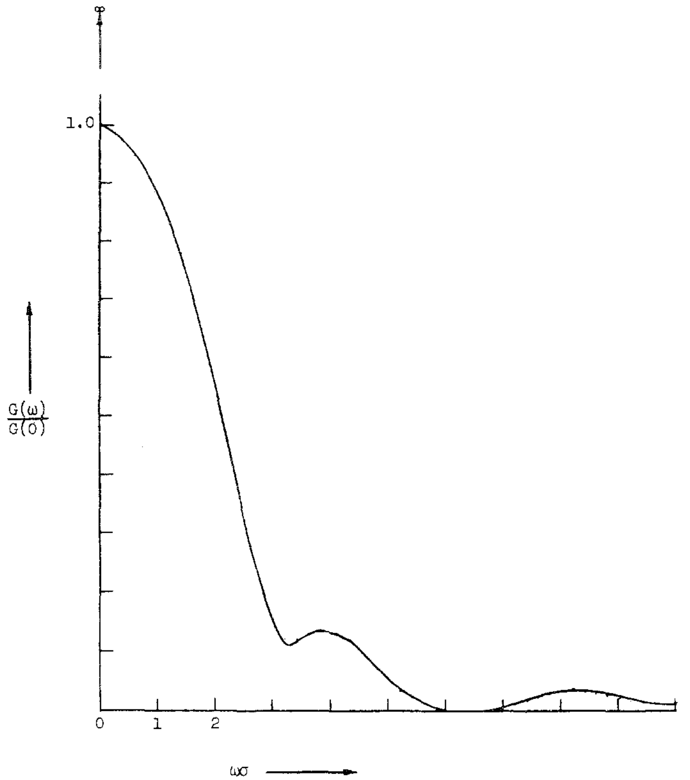

Fig. 7. Power spectrum for a zeroth-order Markovian source after optimization of transmission times, $P_1 = 0.618$, $P_0 = 0.382$, $t_1 = 1\sigma$, $t_0 = 2\sigma$.

series to obtain the final result in closed form. This procedure, which is straightforward but somewhat lengthy, leads to the following expression:

$$G(\omega) = \frac{2(1 - P_1)[P_1^{(s)}]^2}{\omega^2\sigma}\left[\frac{\cos\omega\sigma - \cos 2\omega\sigma - (P_1 - 1)(1 - \cos\omega\sigma)}{P_1 - 2(P_1 - 1)\cos\omega\sigma}\right]$$
$$- \frac{2[P_1^{(s)}]^2(P_1 - 1)\sin\omega\sigma}{(P_1 - 2)\,\omega} + \frac{2[P_1^{(s)}]^2}{P_1 - 2}\left[\frac{\sin\omega\sigma}{\omega} - \pi\delta(\omega)\right]$$
$$+ \frac{2P_1^{(s)}(1 - P_1^{(s)})}{\omega^2\sigma}(1 - \cos\omega\sigma) + \frac{2[P_1^{(s)}]^2\sin\omega\sigma}{\omega} \tag{27}$$

Figure 7 shows a plot of $G(\omega)$ as a function of $\omega\sigma$. The half-width now occurs at $\omega\dot{\sigma} = 2.11$, compared with the previous value of $\omega\sigma = 2.78$, and the bandwidth has been narrowed by a factor fo 1.32 by the optimization of transmission times. However, with σ unaltered, the average transmission time

$$\langle t\rangle = (P_1 + 2P_2)\sigma = 1.382\sigma \tag{28}$$

is increased by a factor of 1.382. If we wish to do no more than eliminate this increase, to say nothing of reducing $\langle t \rangle$ to a value less than σ so that a reduction in transmission time is achieved by optimization, σ will have to be reduced by a factor of 1/1.382. This means that the width of the power spectrum, at half-width, will have to be

$$\omega = 2.11/(\sigma/1.382) = 2.92/\sigma \tag{29}$$

compared with

$$\omega = 2.78/\sigma \tag{30}$$

for the previous case in which transmission times were not optimized. Thus, in the present case it is not possible to reduce bandwidth and message time simultaneously. If one wishes to reduce message time by reducing σ by a factor larger than the 1/1.382 which demands the bandwidth shown in Eq. (29) (and which already exceeds that of the nonoptimized case), the bandwidth will have to exceed the value given in Eq. (3 by an even larger amount. In such a case, no advantage will be gained through use of a compact code unless bandwidth is not a critical consideration.

In the next section, we study the optimization and bandwidth problem for a source having memory; to be exact, a first-order Markovian source in which memory extends only to the previously emitted word. Here, we shall discover that it is possible to reduce both message time and bandwidth simultaneously so that even when bandwidth *is* a consideration, transmission time optimization still represents an advantage.

4. A FIRST-ORDER MARKOVIAN SOURCE

The application of the statistical thermodynamic formalism to the optimization of a first-order Markovian source emitting ones and zeros was discussed in Section 7 of Ref. 1. In that section, it was shown that the relation between t_{ij}, the optimum transmission time for the jth word emitted by the source after the ith word (both i and j can be zero or one) and P_{ij}, the probability that the ith and jth words are emitted in sequence [probability of the digraph (ij)] is given by

$$P_{ij} = (S_{ij}/\lambda)(\partial\lambda/\partial S_{ij})_{S_{i'j'}}, \qquad i' \neq i, \quad j' \neq j \tag{31}$$

where

$$S_{ij} = e^{-t_{ij}/\kappa T}. \tag{32}$$

and λ is the maximum eigenvalue of a secular equation originating from the application of the matrix method to the evaluation of partition functions. With the evaluation of this eigenvalue and substitution in Eq. (31), it can be shown that

$$\begin{aligned} P_{11} &= \tfrac{1}{2}S_{11}\{1 + [(S_{11} - S_{00})/C]\}, \\ P_{00} &= \tfrac{1}{2}S_{00}\{1 + [(S_{00} - S_{11})/C]\}, \\ P_{10} &= P_{01} = S_{10}S_{01}/C \end{aligned} \tag{33}$$

where

$$C = [(S_{11} - S_{00})^2 + 4S_{10}S_{01}]^{1/2} \tag{34}$$

Equations (33) and (34) can be inverted to yield

$$\begin{aligned} S_{11} &= 2P_{11}/(1 - P_{00} + P_{11}), \qquad S_{00} = 2P_{00}/(1 - P_{11} + P_{00}) \\ S_{10}S_{01} &= (1 - P_{11} - P_{00})^2/(1 - P_{11} + P_{00})(1 + P_{11} - P_{00}) \end{aligned} \tag{35}$$

Equations (31)–(35) constitute the recipe for optimizing transmission times if the message statistics are known, or vice versa. We use them in the present section.

We now introduce the *conditional* probability p_{ij}, which represents the chance that the jth word will be emitted given that the ith word was emitted previously. This is not to be confused with the digraph probability P_{ij} introduced earlier. If the *a priori* probabilities for emission of a one or zero are P_1 and P_0, respectively, then the relations between the conditional and digraph probabilities are

$$P_1p_{11} = P_{11}, \qquad P_1p_{10} = P_{10}, \qquad P_0p_{00} = P_{00}, \qquad P_0p_{01} = P_{01} \tag{36}$$

and

$$P_1 = (1 - p_{00})/(2 - p_{00} - p_{11}), \qquad P_0 = (1 - p_{11})/(2 - p_{00} - p_{11}) \tag{37}$$

Some additional useful relations follow:

$$P_1 + P_0 = 1 \tag{38}$$

$$p_{00} + p_{01} = 1, \qquad p_{11} + p_{10} = 1 \tag{39}$$

$$P_{11} + P_{10} = P_1, \qquad P_{00} + P_{01} = P_0, \qquad P_{10} = P_{01} \tag{40}$$

$$\begin{aligned} p_{11} &= 2P_{11}/(1 - P_{00} + P_{11}), \qquad p_{00} = 2P_{00}/(1 - P_{11} + P_{00}) \\ p_{10} &= (1 - P_{00} - P_{11})/(1 - P_{00} + P_{11}), \qquad p_{01} = (1 - P_{00} - P_{11})/(1 - P_{11} + P_{00}) \end{aligned} \tag{41}$$

From these equations, it is clear that once p_{00} and p_{11} are given, the message statistics are unambiguously defined.

As in the previous example dealing with a memoryless source, we calculate the autocorrelation function in terms of the conditional probabilities for both the noncompact and compact cases, respectively. To simplify our discussion without losing the point, we deal with a source with the following conditional probabilities:

$$p_{01} = p_{10} = 6.18, \qquad p_{11} = p_{00} = 3.82 \tag{42}$$

We begin with the noncompact case, and instead of optimizing transmission times we choose

$$t_{11} = t_{10} = t_{01} = t_{00} = \sigma \tag{43}$$

where, as indicated earlier, t_{ij} is the transmission time for the jth word following the ith word. As in Eq. (9), we write $\tau = N\sigma + \epsilon$. For $N = 0$, we define ϕ as ϕ_0, and it follows that

$$\phi_0(\tau) = [1 - (\tau/\sigma)]\, P_1 + (\tau/\sigma)\, P_1p_{11} \tag{44}$$

Again, the factors $1 - (\tau/\sigma)$ and τ/σ are the respective probabilities that the interval

τ falls entirely within a unit σ, or bridges two units. The unit within which τ falls or the units bridged must be the loci of ones in order that a contribution to $\phi(\tau)$ be made, i.e., the ends of τ must fall in ones. Thus, in Eq. (44), the factors mentioned must be multiplied by the probabilities that ones are in the called-for locations. In the first case, since we deal with a *single* one, within which τ lies, the appropriate factor is the *a priori* probability P_1 for the occurrence of a one, while in the case that τ bridges two ones, we require the digraph probability $P_1 p_{11}$. These mutually exclusive probabilities are added to give ϕ_0 in Eq. (44).

For $N \geqslant 1$,

$$\phi_N(\tau) = (\epsilon/\sigma)\, P_1(N+1)\, P_1 + [1 - (\epsilon/\sigma)]\, P_1(N)\, P_1 \tag{45}$$

The origin of this equation is similar to that of Eq. (22) with, however, some modifications. The factors ϵ/σ and $1 - (\epsilon/\sigma)$ are the probabilities that τ spans N or $N - 1$ whole units of σ (including the units within which the ends lie; this would be $N + 2$ or $N + 1$ units, respectively). $P_1(N)$ is the probability that a message of *exactly* N units has its first unit *following* a one, while its *last* unit *is* a one. Thus, consider the the second term on the right of Eq. (45). The first factor gives assurance that τ spans $N - 1$ whole units as described previously. The last factor measures the probability that the unit within which τ begins is a one; while the middle factor is a *conditional* probability, given the first unit is a one, that the following message of N units ends with a one so that this τ makes a nonzero contribution to ϕ. The first term in Eq. (45) is constructed in a similar fashion, and deals with the mutually exclusive situation in which τ spans N whole units ($N + 2$ units if those within which the ends lie are included). The quantity $P_1(N)$ is evaluated in Appendix B, and may be expressed as

$$P_1(N) = [(p_{11} - 1)/(p_{11} + p_{00} - 2)](p_{11} + p_{00} - 1)^N + [(p_{00} - 1)/(p_{11} + p_{00} - 2)] \tag{46}$$

The reader may satisfy himself concerning the validity of Eq. (46) by examining certain special cases where the answer is easily arrived at by inspection. Useful points at which to make such a check are $\tau = 0$, $\tau = \sigma$, $\tau = \sigma N$ with $\epsilon = \sigma$, $\tau = (N + 1)\sigma$ with $\epsilon = 0$. It also turns out that as $N \to 0$, Eq. (45) reduces automatically to Eq. (44).

Upon substitution of Eq. (46) and $\epsilon = \tau - N\sigma$ into Eq. (45), one obtains

$$\begin{aligned}\phi_N(\tau) = {} & \frac{P_1(1 - p_{00})}{2 - p_{11} - p_{00}} + \frac{P_1(1 - p_{11})}{2 - p_{11} - p_{00}}\,(p_{11} + p_{00} - 1)^N \\ & + N(1 - p_{11})\, P_1(p_{11} + p_{00} - 1)^N \\ & - \frac{\tau}{\sigma}\, P_1(1 - p_{11})(p_{11} + p_{00} - 1)^N, \qquad N = 0, 1, 2, \ldots, \infty\end{aligned} \tag{47}$$

The power spectrum $G(\omega)$ may be written as

$$\begin{aligned}G(\omega) &= 2\,\mathrm{Re} \int_0^\infty e^{-i\omega\tau} \phi(\tau)\, d\tau \\ &= 2\,\mathrm{Re} \sum_0^\infty \int_{N\sigma}^{(N+1)\sigma} \phi_N(\tau)\, e^{-i\omega\tau}\, d\tau\end{aligned} \tag{48}$$

As before, the integration may be carried out, the geometric series summed, and the final result expressed as

$$G(\omega) = \left[\frac{2P_1(1 - p_{11})(p_{11} + p_{00})(1 - \cos \omega\sigma)}{1 - 2(p_{00} + p_{11} - 1) \cos \omega\sigma + (p_{11} + p_{00} - 1)^2} \frac{1}{\omega^2\sigma^2}\right] \sigma + 2\pi\delta(\omega) \times \frac{P_1(1 - p_{00})}{2 - p_{11} - p_{00}} \tag{49}$$

This curve is illustrated in Fig. 8. It is an interesting feature that the factor

$$(p_{00} + p_{11} - 1)$$

can be negative, because p_{00} and p_{11} can lie anywhere between zero and one independently. Under this circumstance, it is entirely within the realm of possibility that it may lie very close to -1, say it is $-1 + h$, where h is a small positive number. Then

$$\begin{aligned}
&[1 - 2(p_{00} + p_{11} - 1) \cos \omega\sigma + (p_{00} + p_{11} - 1)^2] \\
&= 1 - 2(-1 + h) \cos \omega\sigma + (1 - 2h) \\
&= 2(1 - h)(1 + \cos \omega\sigma)
\end{aligned} \tag{50}$$

Thus, as $\cos \omega\sigma \to -1$, the denominator in Eq. (49) becomes small enough to, dominate the $1/\omega^2\sigma^2$ term and $G(\omega)$ becomes very large. As a result, we do not have a simple bell-shaped curve in Fig. 8.

The case where $(L_{00} + p_{11} - 1) \to 1 - h$ is not as serious, because the numerator in Eq. (49) contains a factor $1 - \cos \omega\sigma$ which also becomes small, i.e.,

$$[1 - 2(p_{00} + p_{11} - 1) \cos \omega\sigma + (p_{00} + p_{11} - 1)^2] = 2(1 + h)(1 - \cos \omega\sigma) \tag{51}$$

$$G(\omega) = [P_1(1 - p_{11})(p_{11} + p_{00})/(1 + h)](1/\omega^2\sigma^2) \tag{52}$$

The lesson to be learned is of course that $G(\omega)$ is not a simple curve, under the circumstances in which are interested. Hence, when making comparisons between power spectra, the half-width provides only a semiquantitiative indication of relative bandwidth. This is especially true when one of the two curves being compared is bell-shaped and the other is not.

As we let $p_{00} \to P_0$ and $p_{11} \to P_1$, so that the source no longer has memory, Eq. (49) reduces to

$$G(\omega) = [2P_1(1 - P_1)(1 - \cos \omega\sigma)/\omega^2\sigma^2]\sigma + 2\pi\delta(\omega) P_1^2 \tag{53}$$

which is identical with Eq. (17), as it should be.

Now, we examine the case in which transmission times are optimized. Analyzing Eqs. (31)–(35), together with (36) and (37), we find that the t_{ij} going with the p_{ij} prescribed by Eq. (42) are

$$t_{01} = t_{10} = \sigma, \qquad t_{00} = t_{11} = 2\sigma \tag{54}$$

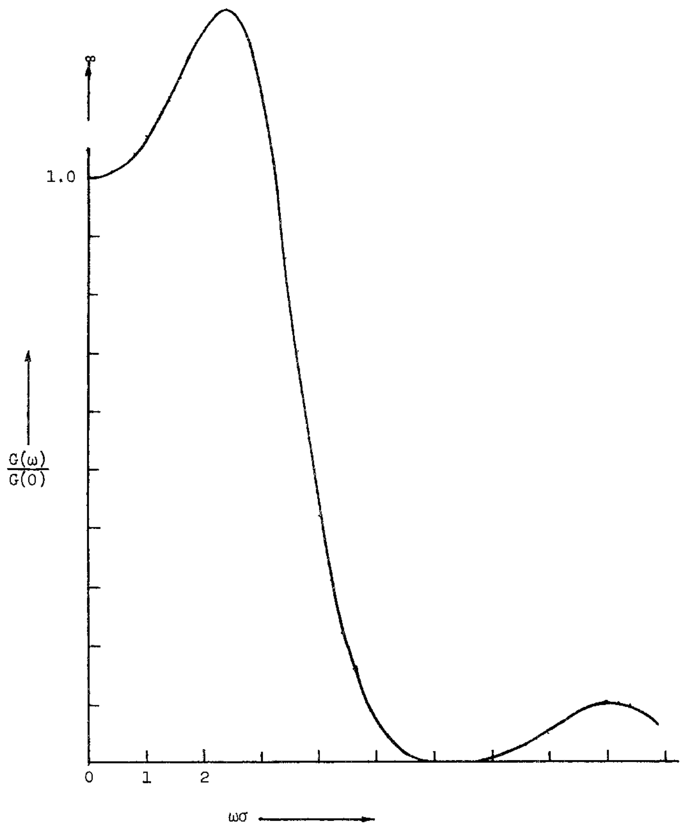

Fig. 8. Power spectrum for a first-order Markovian source, $p_{11} = p_{00} = 0.382$, $t_{11} = t_{00} = t_{10} = t_{01} = 1\sigma$.

We begin by examining a portion of what might be a typical message having the transmission times prescribed in Eq. (54). This is illustrated in Fig. 9. The alert reader may have already determined, by inspection, that if a segment of a message exactly M units of σ long follows a unit occupied by all or part of a one, and if M is odd, that message can only end with a zero; while if M is even, it must end with a one. This and many other constraints interact to render the evaluation of the autocorrelation function both detailed and tedious. Nevertheless, it can be evaluated precisely. Perhaps the best way to explain the process is to concentrate on a unified group of terms in $\phi(\tau)$, show how, in this particular case, the constraints interact and the result is formulated, and then merely write down the final result for the complete $\phi(\tau)$. The reader should then be able, using exactly the same process, to evaluate the remaining terms which have not been explained in detail.

Following this plan, we will derive, in some detail, the contributions to $\phi(\tau)$, going with those values of τ expressible as $\tau = N\sigma + \epsilon$, or values of τ bounded in the following way:

$$N\sigma \leqslant \tau \leqslant (N + 1)\sigma \tag{55}$$

which of course implies $\epsilon < \sigma$. A little though will show that a τ of this magnitude can span either N *whole* units of σ or alternatively only $N - 1$ *whole* units. This means that in the first case, τ contacts $N + 2$ units; the N whole units which it spans plus the two units on either side of these N units within which its end points lie, neither of which is completely spanned by τ. In the second case, by the same line of reasoning, τ will contact $N + 1$ units, the two end units again being incompletely spanned. In order to arrive at the respective probabilities for each case, one can think of "throwing down" a straight line of length τ onto a typical message. In what fraction of the total space of the message can one of the end points (which of course determines the location of the entire τ interval) fall so that N complete units will be spanned? Again, inspection will show this fraction to be

$$f_N = \epsilon/\sigma \tag{56}$$

For the second case (τ spanning $N - 1$ complete units), the appropriate fraction is

$$f_{N-1} = 1 - (\epsilon/\sigma) \tag{57}$$

These factors will have to appear in the probabilities which are used to determine the contribution to the ensemble average, for $\phi(\tau)$, coming from the interval τ. Reference to Fig. 10 will be helpful in understanding the process.

Figure 10(a) treats the case where τ covers N complete units. The probability factor f_N, described by Eq. (56), must therefore be used, and is in fact shown at the bottom of the diagram in Fig. 10(a). We work out a case where N is odd. There are only three possible message situations for this case, and these are exhibited on different levels in Fig. 10(a). The unit σ is shown in the upper left-hand corner. Notice that in each of the three cases, the first visible portion of the message sequence (on the left) is all or part of a one. Thus, on the first level, it entirely precedes the N units under consideration. It remains an open question as to whether it is the last part of a

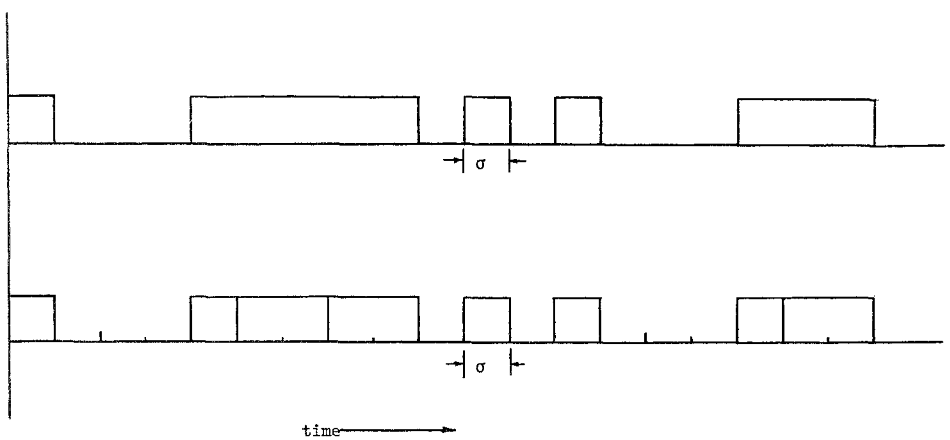

Fig. 9. Typical message for a first-order Markovian source after optimization of transmission times, $t_{10} = t_{01} = 1\sigma$, $t_{11} = t_{00} = 2\sigma$.

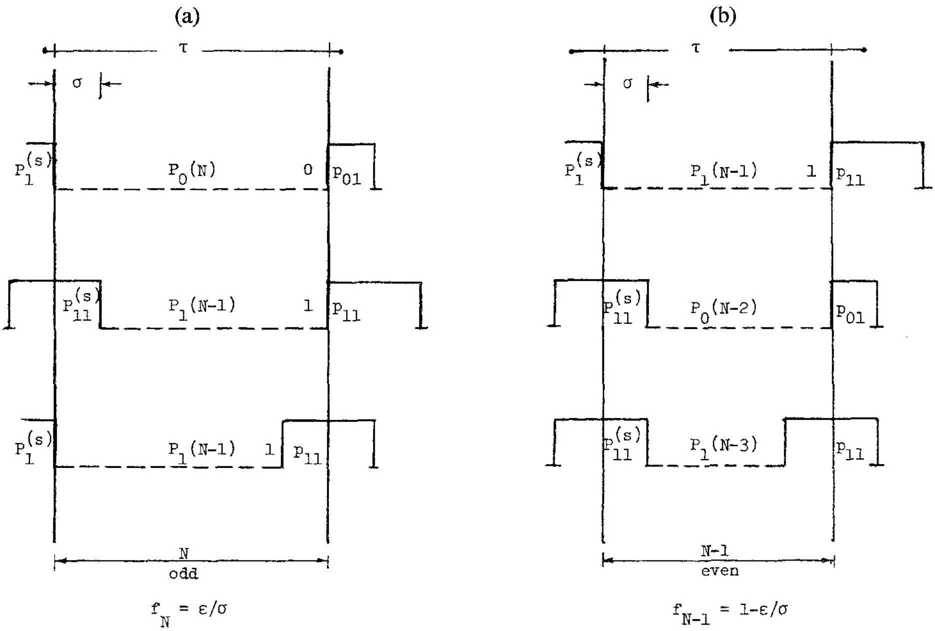

Fig. 10. Terms in autocorrelation function.

pulse, 2σ in length (a one following a one), or all of a pulse, σ in length (a one following a zero). The probability $P_1^{(s)}$ represents the *space* probability, as opposed to the *time* probability. We work out below the exact form for $P_1^{(s)}$. Here, however, we note that the symbol $P_1^{(s)}$ has a different meaning than that given previously during the discussion of a source without memory. No confusion should result from this, and it avoids the introduction of an even longer list of symbols than we already have.

On the second level in Fig. 10(a), the one at the left is in fact a pulse 2σ units long, and it invades, to the extent of one σ unit, the $N\sigma$ units under consideration. The space probability for the second half of such a pulse covering the first of the N units is denoted by $P_{11}^{(s)}$. We also derive this quantity below. On the third level, the first unit on the left is, again, either the second half of a pulse two units long or a single pulse one unit long, and again we use the probability $P_1^{(s)}$.

Returning to the first level, since N is odd, the intermediate message N units long (which follows a one) can only end (from what has been said above) in a zero. The probability of finding just such a message to fill the N units in question is $P_0(N)$, a function derived in Appendix C. Actually, because we know that the last symbol must be a zero when N is odd, it is really unnecessary to append the subscript zero in $P_0(N)$; and in Appendix C we do in fact drop it. For clarity in the present exposition, however, we shall retain it as well as the subscript 1 when the last symbol must be a one. On the first level, we show a 0 in the last place to emphasize the fact that the message N units long must end in a zero.

On the second level in Fig. 10(a), it can be seen that we deal with an intermediate message exactly $N - 1$ units long. In this case, it must end in a one, and we have to use the probability $P_1(N-1)$. The same is true on the third level, where again we deal with an intermediate message $N-1$ units long.

Since the final symbol within which the trailing end of τ falls must also be a one in order for there to be a contribution to the autocorrelation function, we must multiply the previously mentioned probabilities by the probability, in each case, that the final symbol is indeed a one. On the first level, since we know that the last symbol in the intermediate message is a zero, this probability is the conditional probability p_{01}. The appropriate pulse is shown. On the second and third levels, the appropriate conditional probability is clearly p_{11}, and it, together with its associated pulse, is shown.

Each configuration contributes a term, unity, to the autocorrelation function, each unity weighted by the probability of the configuration. The latter probability is determined, clearly, by multiplying $f_N = \epsilon/\sigma$ by the three probabilities shown on each level. Thus, the first level in Fig. 10(a) makes a contribution $(\epsilon/\sigma)\, P_1^{(s)} P_0(N)\, p_{01}$. The second level contributes $(\epsilon/\sigma)\, P_{11}^{(s)} P_1(N-1)\, p_{11}$, while the third level contributes $(\epsilon/\sigma)\, P_1^{(s)} P_1(N-1)\, p_{11}$.

The contribution coming from τ includes the situations shown in Fig. 10(b). Here, of course, $N-1$ is even; and furthermore, the f probability factor is $1-(\epsilon/\sigma)$, as indicated at the bottom of the diagram. The reader can easily work out the contributions from Fig. 10(b) in a manner entirely similar to that used in connection with Fig. 10(a).

In order to see that no other configurations can contribute to either Fig. 10(a) or 10(b), the reader should try those which he may have in mind, applying all the requisite constraints. Thus, suppose one wished to have a situation such as in the first level of Fig. 10(a), but in which the last factor was p_{11}. This would require that the last symbol in the intermediate message be a one, an impossibility when N is odd.

We now turn to the evaluation of $P_1^{(s)}$ and $P_{11}^{(s)}$. Assume that the total message consists of W symbols. The message space occupied by ones following a one, zeros following a zero, ones following a zero, and zeros following a one, is $2\sigma WP_1p_{11}$, $2\sigma wP_0p_{00}$, σWP_0p_{01}, and σWP_1p_{10}, respectively. These, added together, correspond of course to the total space. Now, consider only the portion of message space occupied by ones following one. As indicated above, this is $2\sigma WP_1P_{11}$. Only half of this space, however, corresponds to the second half of a pulse of length 2σ, associated with a one following a one. Thus, the space probability of the second half of such a pulse is obtained by dividing $2\sigma WP_1p_{11}$ by the four quantities listed above (the total space of the message) and dividing the whole by 2. Similarly, the space probability of a pulse representing a one following a zero (of length σ) is obtained by dividing σWP_0p_{01} by the sum of the four quantities. $P_1^{(s)}$ is the sum of these two probabilities, and we obtain

$$P_1^{(s)} = (P_1p_{11} + P_0p_{01})/(2P_1p_{11} + 2P_0p_{00} + P_0p_{01} + P_1p_{10}) \tag{58}$$

which can of course be modified and re-expressed through use of Eqs. (36) and (37).

Since

$$P_1p_{11} + P_0p_{01} = P_1 \tag{59}$$

a given one is either part of a 11 diagraph or not. Using Eq. (59), we may rewrite Eq. (58) as

$$P_1^{(s)} = P_1/(2P_1p_{11} + 2P_0p_{00} + P_0p_{01} + P_1p_{10}) \tag{60}$$

Now, the probability $P_{11}^{(s)}$ is clearly the quantity obtained by dropping the term P_0p_{01} in the numerator of Eq. (58). We see through comparison of Eq. (60) with that result that

$$P_{11}^{(s)} = P_1^{(s)}p_{11} \tag{61}$$

which will prove to be a useful shorthand. In fact, using this formula, together with the expressions already derived for the contributions coming from Figs. 10(a) and 10(b), we are able to write, for that portion of $\phi(\tau)$,

$$\begin{aligned}[\phi(\tau)]_{\text{odd}} &= (\epsilon/\sigma)\{P(N)p_{01} + P(N-1)\,p_{11} + P(N-1)\,p_{11}^2\}\,P_1^{(s)} \\ &\quad + [1-(\epsilon/\sigma)]\{P(N-1)\,p_{11} + P(N-2)\,p_{11}p_{01} + P(N-3)\,p_{11}^2\}\,P_1^{(s)}, \\ &\qquad N = 3, 5, 7,\ldots, \quad \tau = N\sigma + \epsilon \end{aligned} \tag{62}$$

Similarly, when τ has such a value that N is even, we may construct, by an entirely analogous detailed process, the contribution

$$\begin{aligned}[\phi(\tau)]_{\text{even}} &= (\epsilon/\sigma)\{P(N)p_{11} + P(N-1)\,p_{01}p_{11} + P(N-2)\,p_{11}^2\}\,P_1^{(s)} \\ &\quad + [1-(\epsilon/\sigma)]\{P(N-1)\,p_{01} + P(N-2)\,p_{11} + P(N-2)\,p_{11}^2\}\,P^{(s)}, \\ &\qquad N = 2, 4, 6,\ldots, \quad \tau = N\sigma + \epsilon \end{aligned} \tag{63}$$

In these equations, we have dropped the subscripts 0 and 1 (used in connection with Fig. 10) in the symbol $P(N)$ since there is a mutual exclusion with the oddness or evenness of N. Furthermore, it will be noticed that the cases $N = 0, 1$ are not included because $P(N)$ are not defined for negative values for N. These have to be treated specially and can be worked out using the same general principles, enumerating and examining the finite number of messages which can be fitted into the short intervals involved. $P(N)$ is calculated in detail in Appendix C, with the result

$$\begin{aligned} P(N) &= [1/2(1-p_{11}p_{00})]\{(1-p_{11})\,p_{00}^{1/2}(p_{00}^{1/2}p_{11}^{1/2})^N[(p_{00}^{1/2}-p_{11}^{1/2}) + (-1)^N(p_{00}^{1/2}+p_{11}^{1/2})] \\ &\quad + (2-p_{00}-p_{11}) + (p_{11}-p_{00})(-1)^N\} \end{aligned} \tag{64}$$

In anticipation of the fact that it will later be necessary to perform a summation of an infinite series, we introduce transformations in Eqs. (62) and (63) which allow these summations to be performed with greater convenience. Thus, we define in

Eq. (62) $n = \frac{1}{2}(N-3)$ and in Eq. (63) $n = \frac{1}{2}(N-2)$. Using these transformations, together with Eq. (64), we convert (62) and (63) into

$$\begin{aligned}\phi_{\text{even}} &= \left[\frac{P_1^{(s)}(1-p_{00})(3+3p_{11}^2-4p_{11})}{1-p_{11}p_{00}}\right] \\ &\quad - (p_{11}p_{00})^{n+1}\left[\frac{(1-p_{11})\,p_{00}P_1^{(s)}(-3p_{11}/p_{00}-3+4p_{11})}{1-p_{11}p_{00}}\right] \\ &\quad + \left(\frac{\tau}{\sigma}-2n\right)\left[\frac{(1-p_{11})(1-p_{00})(p_{11}-1)\,P_1^{(s)}}{1-p_{11}p_{00}}\right] \\ &\quad \left(\frac{\tau}{\sigma}-2n\right)(p_{11}p_{00})^{n+1}\left[\frac{P_1^{(s)}(1-p_{11})\,p_{00}}{1-p_{00}p_{11}}\left(1-2p_{11}+\frac{p_{11}}{p_{00}}\right)\right] \qquad (65)\end{aligned}$$

and

$$\begin{aligned}\phi_{\text{odd}} &= \left[\frac{(1-p_{00})\,P_1^{(s)}}{1-p_{11}p_{00}}(8p_{11}-3p_{11}^2-3)\right] - (p_{11}p_{00})^{n+1} \\ &\quad \times\left[\frac{(1-p_{11})\,p_{00}P_1^{(s)}}{1-p_{11}p_{00}}(3p_{11}p_{00}+3p_{11}^2-8p_{11})\right] \\ &\quad + \left(\frac{\tau}{\sigma}-2n\right)\left[\frac{(1-p_{00})(1-p_{11})^2\,P_1^{(s)}}{1-p_{11}p_{00}}\right] \\ &\quad - \left(\frac{\tau}{\sigma}-2n\right)(p_{11}p_{00})^{n+1}\left[\frac{(2p_{11}-p_{11}p_{00}-p_{11}^2)\,p_{00}(1-p_{11})\,P_1^{(s)}}{1-p_{11}p_{00}}\right] \qquad (66)\end{aligned}$$

The cases for $N = 0$ and $N = 1$ give the following contributions:

$$\begin{aligned}\phi_0 &= \left(1-\frac{\epsilon}{\sigma}\right)P_{01}^{(s)} + P_{11}^{(s)}\left(1-\frac{\epsilon}{\sigma}\right) + P_{11}^{(s)}\frac{\epsilon}{2\sigma} + P_{01}^{(s)}p_{11}\frac{\epsilon}{\sigma} \\ &= P_1^{(s)} + \frac{\tau}{\sigma}\left(-P_1^{(s)} + \frac{P_{11}^{(s)}}{2} + P_{01}^{(s)}p_{11} + \frac{p_{11}P_{11}^{(s)}}{2}\right), \qquad 0 \leqslant \tau \leqslant \sigma \qquad (67)\end{aligned}$$

and

$$\begin{aligned}\phi_1 &= \frac{P_1^{(s)}\epsilon}{\sigma}[p_{10}p_{01}+p_{11}+p_{11}^2] + P_{01}^{(s)}p_{11}\left(1-\frac{\epsilon}{\sigma}\right) + \frac{\sigma-\epsilon}{2\sigma}(P_{11}^{(s)}+p_{11}P_{11}^{(s)}) \\ &= \frac{\tau}{\sigma}\left[(p_{10}p_{01}+p_{11}+p_{11}^2)\,P_1^{(s)} - p_{11}P_{01}^{(s)} - \frac{P_{11}^{(s)}}{2} - \frac{P_{11}^{(2)}p_{11}}{2}\right] \\ &\quad + [-(p_{10}p_{01}+p_{11}+p_{11}^2)\,p_1^{(s)} + 2P_{01}^{(s)}p_{11} + P_{11}^{(s)} + P_{11}^{(s)}p_{11}], \qquad \sigma \leqslant \tau \leqslant \sigma. \qquad (68)\end{aligned}$$

Again, it can be verified that $\phi_0(\sigma) = \phi_1(\sigma)$, $\phi_1(2\sigma) = \phi(N=2, \epsilon=0)$,

$$\phi(N, \epsilon=\sigma) = \phi(N+1, \epsilon=0).$$

Using the portions of the autocorrelation function derived in this way for the various integrals of τ, we may write the Fourier transform, or the power spectrum,

$$
\begin{aligned}
G(\omega) &= 2 \operatorname{Re} \int_0^\infty \phi(\tau)\, e^{-i\omega\tau}\, d\tau \\
&= 2 \operatorname{Re} \Big[\int_\sigma^\sigma \phi_0(\tau)\, e^{-i\omega\tau}\, d\tau + \int_\sigma^{2\sigma} \phi_1(\tau)\, e^{-i\omega\tau}\, d\tau \\
&\quad + \sum_{N=2,4,6\ldots}^{\infty} \int_{N\sigma}^{(N+1)\sigma} \phi_e(N, \tau)\, e^{-i\omega\tau}\, d\tau + \sum_{N=3,5,7\ldots}^{\infty} \int_{N\sigma}^{(N+1)\sigma} \phi_0(N, \tau)\, e^{-i\omega\tau}\, d\tau \Big] \\
&= 2 \operatorname{Re} \Big[\int_0^\sigma \phi_0(\tau)\, e^{-i\omega\tau}\, d\tau + \int_\sigma^{2\sigma} \phi_1(\tau)\, e^{-i\omega\tau}\, d\tau \Big] \\
&\quad + 2 \operatorname{Re} \sum_{n=0}^{\infty} \Big[\int_{(2n+2)\sigma}^{(2n+3)\sigma} \phi_e(n, \tau)\, e^{-i\omega\tau}\, d\tau + \int_{(2n+3)\sigma}^{(2n+4)\sigma} \phi_0(n, \tau)\, e^{-i\omega\tau}\, d\tau \Big]
\end{aligned} \tag{69}
$$

or

$$G(\omega) = G_1 + G_2 + G_3 + G_4 \tag{70}$$

where we have subdivided $G(\omega)$ into four parts. For G_3 and G_4, the integrated results will have to be summed as a geometric series. The final results are lengthy but straight forward. They are summarized in the following equations[4]:

$$
\begin{aligned}
G_1(\omega) &= \frac{2(A_z + B_z)}{\omega} \sin \omega\sigma + \frac{2B_z}{\omega^2\sigma} (\cos \omega\sigma - 1) \\
G_2(\omega) &= \frac{2D_z \sin \omega\sigma}{\omega} [2 \cos \omega\sigma - 1] \\
&\quad + \frac{2C_z}{\omega^2\sigma} [\cos 2\omega\sigma - \cos \omega\sigma + 2\omega\sigma \sin 2\omega\sigma - \omega\sigma \sin \omega\sigma] \\
G_3(\omega) &= \frac{A_E}{\omega} \left[\frac{\sin 3\omega\sigma - \sin 2\omega\sigma - \sin \omega\sigma}{1 - \cos 2\omega\sigma} \right] \\
&\quad - \frac{2B_E p_{00} p_{11}}{\omega} \left[\frac{\sin 3\omega\sigma - \sin 2\omega\sigma - p_{11} p_{00} \sin \omega\sigma}{1 - 2p_{11} p_{00} \cos 2\omega\sigma + (p_{11} p_{00})^2} \right] \\
&\quad + \frac{C_E}{\sigma\omega^2} \\
&\quad \times \{(3\omega\sigma \sin 3\omega\sigma + \cos 3\omega\sigma - 2\omega\sigma \sin 2\omega\sigma - \cos 2\omega\sigma \\
&\quad - 3\omega\sigma \sin \omega\sigma - \cos \omega\sigma + 1) \\
&\quad \div (1 - \cos 2\omega\sigma)\} \\
&\quad - \frac{2D_E p_{00} p_{11}}{\omega^2\sigma} \\
&\quad \times \{(3\omega\sigma \sin 3\omega\sigma + \cos \omega\sigma - 2\omega\sigma \sin 2\omega\sigma - \cos 2\omega\sigma - 3\omega\sigma p_{11} p_{00} \sin \omega\sigma \\
&\quad - p_{11} p_{00} \cos \omega\sigma + p_{11} p_{00}) \\
&\quad \div [1 + (p_{11} p_{00})^2 - 2p_{11} p_{00} \cos 2\omega\sigma]\}
\end{aligned} \tag{71}
$$

---

[4] Note that here we did not include the contributions due to delta functions at $\omega\sigma = 0$.

$$
\begin{aligned}
G_4(\omega) = {} & \frac{A_0}{\omega}\left[\frac{\sin 4\omega\sigma - \sin 3\omega\sigma - \sin 2\omega\sigma + \sin \omega\sigma}{1 - \cos 2\omega\sigma}\right] \\
& - \frac{2B_0 p_{00} p_{11}}{\omega}\left[\frac{\sin 4\omega\sigma - \sin 3\omega\sigma - p_{11}p_{00}\sin 2\omega\sigma + p_{11}p_{00}\sin\omega\sigma}{1 - 2p_{11}p_{00}\cos 2\omega\sigma + (p_{11}p_{00})^2}\right] \\
& + \frac{C_0}{\omega^2\sigma} \\
& \times [(4\omega\sigma \sin 4\omega\sigma + \cos 4\omega\sigma - 3\omega\sigma\sin 3\omega\sigma - \cos 3\omega\sigma \\
& - 4\omega\sigma\sin 2\omega\sigma - \cos\omega\sigma + 3\omega\sigma\sin\omega\sigma + \cos\omega\sigma) \\
& \div (1 - \cos 2\omega\sigma)] \\
& - \frac{2p_{11}p_{00}D_0}{\omega^2\sigma} \\
& \times \{(4\omega\sigma\sin 4\omega\sigma + \cos 4\omega\sigma - 3\omega\sigma\sin 3\omega\sigma - \cos 3\omega\sigma - 4\omega\sigma p_{11}p_{00}\sin 2\omega\sigma \\
& - p_{11}p_{00}\cos 2\omega\sigma + 3p_{11}p_{00}\omega\sigma\sin\omega\sigma + p_{11}p_{00}\cos\omega\sigma) \\
& \div [1 + (p_{11}p_{00})^2 - 2p_{11}p_{00}\cos 2\omega\sigma]\}
\end{aligned} \tag{71}
$$

$$
\begin{aligned}
A_0 &= P_1^{(s)}(1 - p_{00})(8p_{11} - 3 - 3p_{11}^2)/(1 - p_{11}p_{00}) \\
B_0 &= P_1^{(s)}(1 - p_{11})\,p_{00}(3p_{11}^2 - 8p_{11} + 3p_{11}p_{00})/(1 - p_{11}p_{00}) \\
C_0 &= P_1^{(s)}(1 - p_{11})^2(1 - p_{00})/(1 - p_{11}p_{00}) \\
D_0 &= P_1^{(s)}(1 - p_{11})\,p_{00}(2p_{11} - p_{11}p_{00} - p_{11}^2)/(1 - p_{11}p_{00}) \\
A_E &= P_1^{(s)}(1 - p_{00})(3 - 4p_{11} + 3p_{11}^2)/(1 - p_{11}p_{00}) \\
B_E &= P_1^{(s)}p_{00}(1 - p_{11})(4p_{11} - 3 - 3p_{11}/p_{00})/(1 - p_{11}p_{00}) \\
C_E &= -C_0 \\
D_E &= P_1^{(s)}p_{00}(1 - p_{11})(1 - 2p_{11} + p_{11}/p_{00})/(1 - p_{11}p_{00}) \\
A_z &= P_1^{(s)} \\
B_z &= (-P_1^{(s)} + \tfrac{1}{2}P_{11}^{(s)} + p_{01}P_{01}^{(s)} + \tfrac{1}{2}p_{11}P_{11}^{(s)}) \\
C_z &= (p_{10}p_{01} + p_{11} + p_{11}^2)\,P_1^{(s)} - p_{11}P_{01}^{(s)} - \tfrac{1}{2}P_{11}^{(s)} - \tfrac{1}{2}p_{11}P_{11}^{(s)} \\
D_z &= -(p_{10}p_{01} + p_{11} + p_{11}^2)\,P_1^{(s)} + 2p_{11}P_{01}^{(s)} + P_{11}^{(s)} + P_{11}^{(s)}\,p_{11} \\
P_{11}^{(s)} &= 2P_{11}/(2P_{11} + 2P_{00} + P_{01} + P_{10}), \qquad P_{01}^{(s)} = P_{01}/(2P_{11} + 2P_{00} + P_{01} + P_{10})
\end{aligned} \tag{72}
$$

Figure 11 is a plot of the power spectrum according to Eq. (71). The values of the various probabilities are either given by Eq. (42) or can be calculated by Eqs. (36)–(41). A comparison of Figs. 8 and 11 shows that not only has the bandwidth been reduced, but also the curve once again has a simple bell shape. Insofar as the half-width

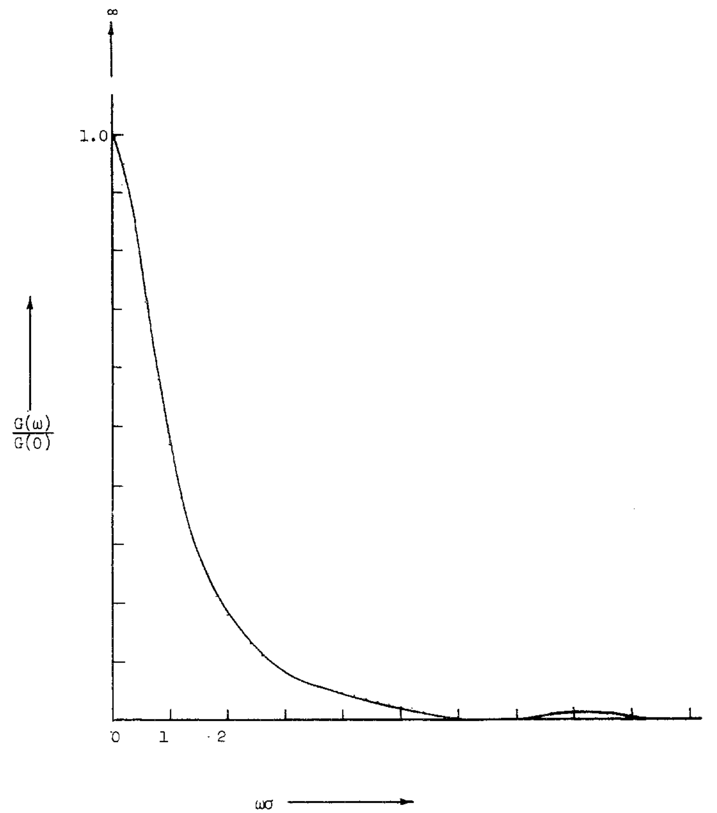

Fig. 11. Power spectrum for a first-order Markovian source after optimization of transmission times, $p_{11} = p_{00} = 0.382$, $t_{10} = t_{01} = 1\sigma$, $t_{11} = t_{00} = 2\sigma$.

can be used as an index of overall breadth, it has gone from $\omega\sigma = 3.88$ to $\omega\sigma = 0.96$. At the same time, the average time of transmission per symbol

$$\langle t \rangle = [2P_{00} + 2P_{11} + P_{01} + P_{10}]\sigma$$

has only increased by a factor of 1.382. This leaves us with a factor

$$(3.88)(1)/(0.96)(1.382) = 2.92$$

which can be used to increase the speed of transmission (reduce average transmission time by using a smaller σ) without increasing the bandwidth. Thus, we can transmit almost three times as fast without paying for it with increased bandwidth. In the case of the first-order source with memory, therefore, optimization of transmission times using the statistical thermodynamic formalism is very worthwhile.

5. CONCLUDING REMARKS

From the foregoing, it appears as though optimization of word transmission times (compact coding) by use of the statistical thermodynamic formalism does not always lead to a simultaneous reduction of message transmission time and bandwidth.

(For a given scale, it does, of course, always lead to reduction of transmission time.) On the other hand, the two cases we examined were distinguished by having source without and with memory, respectively. It may very well be that optimization of transmission times for sources with memory will always allow some narrowing of bandwidth along with a reduction of transmission time, provided that the correlation between symbols in such that it is more likely for ones to follow ones and zeros to follow zeros. In this case, the signal will contain long sequences in which the pulse height remains unchanged. These sequences might be compressed and expanded, relatively, in such a way that total time of transmission is reduced without incurring bandwidth penalties.

Since it should be clear (from the difficulties which we have encountered in deriving analytic expressions for autocorrelation functions and power spectra in relatively simple cases) that the evaluation of these functions for the most general case may be impractical, it would be useful to have a general theorem which tells us, in the absence of noise, under which circumstances both bandwidth and transmission time can be reduced simultaneously. We shall reserve the consideration of this question for a later investigation.

In any event, we have been able to show, in certain instances, that the statistical thermodynamic formalism does allow improvement of both bandwidth and transmission time, and have developed techniques for exploring the autocorrelation functions and power spectra in relevant situations in order to ascertain whether or not such improvement are possible.

APPENDIX A. SOURCE WITHOUT MEMORY WITH $t_0 = 2\sigma$, $t_1 = \sigma$

Let P_0 be the probability of source emitting a zero, P_1 the probability of source emitting a one, and $P(N)$ the probability of a message exactly N units of σ long. We can write a recursion formula for $P(N)$,

$$P(N) = P(N-2)\,P_0 + P(N-1)\,P_1 \tag{A1}$$

This difference equation has a solution of the following form[12]:

$$P(N) = a_1 X^N + a_2 \tag{A2}$$

Substitution of (A2) into (A1) yields

$$\begin{aligned} a_1 X^N + a_2 &= a_1 P_0 X^{N-2} + a_1 P_1 X^{N-1} + a_2 \\ 1 &= P_0 X^{-2} + P_1 X^{-1} \end{aligned} \tag{A3}$$

Since $P_0 + P_1 = 1$, Eq. (A3) has the following solutions for X:

$$X = 1, \quad P_1 - 1 \tag{A4}$$

We can discard $X = 1$ since we expect $P(N)$ to be a function of both N and P_1. Thus, we write

$$P(N) = a_1(P_1 - 1)^N + a_2 \tag{A5}$$

Now, we use the boundary conditions to determine a_1 and a_2. For $P(1)$, the only whole message that can fitted into a unit σ is a one, since a zero requires two units of σ. Thus,[5]

$$P(1) = P_1 \tag{A6}$$

For $P(2)$, where two units of σ are available, it is possible to have either a zero or two one's. Thus,

$$P(2) = P_1^2 + P_0 \tag{A7}$$

Combining (A5) with (A6) and (A7), we get the desired result,

$$P(N) = [(P_1 - 1)^{N+1} - 1]/(P_1 - 2) \tag{A8}$$

Notice that $P(0) = 1$, $P(\infty) = 1/(2 - P_1)$.

A simple test of (A8) with $N = 3$ yields

$$P(3) = P_1^3 + P_0 + P_0 + P_1 = [(P_1 - 1)^4 - 1]/(P_1 - 2)$$

which is satisfactory.

APPENDIX B. SOURCE WITH MEMORY BUT WITH $t_{00} = t_{11} = t_{10} = t_{01} = \sigma$

Let $P_i^{(N)}$ denote the probability of having exactly a whole number of messages in $N\sigma$ with the last symbol i, zero, or one. Also, let p_{ij} be the conditional probability of j following i; again, i, j can be one or zero. We can now write two coupled recursion formulas:

$$P_1(N) = P_1(N-1)\, p_{11} + P_0(N-1)\, p_{01} \tag{B1}$$

$$P_0(N) = P_1(N-1)\, p_{10} + P_0(N-1)\, p_{00} \tag{B2}$$

In order to decouple $P_1(N)$ from $P_0(N)$, Eq. (B1) is rearranged to give

$$P_0(N) = (1/p_{01})[P_1^{(N+1)} - p_{11}P_1^{(N)}] \tag{B3}$$

Substituting Eq. (B3) into (B2), we have

$$P_1^{(N+1)} - P_1(N)(p_{11} + p_{00}) - P_1^{(N-1)}(p_{10}p_{01} - p_{11}p_{00}) = 0 \tag{B4}$$

Now, the solution $P_1(N)$ has the form

$$P_1(N) = A\lambda_1^N + B\lambda_2^N \tag{B5}$$

where λ_1 and λ_2 are two independent parameters. For simplicity, substitute $P(N) = \lambda^N$ into Eq. (B4):

$$\lambda^{N+1} - \lambda^N(p_{11} + p_{00}) - \lambda^{N-1}(p_{10}p_{01} - p_{11}p_{00}) = 0$$

[5] Such a choice also makes the correlation function continuous at $t = \sigma$.

Dividing by λ^N,

$$\lambda^2 - \lambda(p_{11} + p_{00}) - (p_{10}p_{01} - p_{11}p_{00}) = 0 \tag{B6}$$

But, from Eq. (39),

$$p_{10} = 1 - p_{11}, \qquad p_{01} = 1 - p_{00}, \qquad p_{10}p_{01} - p_{11}p_{00} = 1 - p_{11} - p_{00}$$

Equation (B6) can be rewritten as

$$\lambda^2 - \lambda(p_{11} + p_{00}) + (p_{11} + p_{00} - 1) = 0 \tag{B7}$$

having roots

$$\lambda = p_{00} + p_{11} - 1, \quad 1 \tag{B8}$$

Thus,

$$P_1(N) = A(p_{00} + p_{11} - 1)^N + B \tag{B9}$$

In the determination of A and B, the proper boundary conditions are

$$P_1(1) = p_{11} \tag{B10}$$

$$P_1(2) = p_{11}^2 + p_{10}p_{01} \tag{B11}$$

Notice that $P_1(1)$ is p_{11} rather than $p_{11} + p_{01}$. This introduces the requirement that these N units are preceded by a one and that they must end in one. The same is true for $P_1(2)$.

Combining Eqs. (B10), (B11), and (B9), we have

$$A = (p_{11} - 1)/(p_{11} + p_{00} - 2) \tag{B12}$$

$$B = (p_{00} - 1)/(p_{11} + p_{00} - 2) \tag{B13}$$

or

$$P_1(N) = [(p_{11} - 1)/(p_{11} + p_{00} - 2)](p_{11} + p_{00} - 1)^N + [(p_{00} - 1)/(p_{11} + p_{00} - 2)] \tag{B14}$$

$P_0(N)$ can be calculated in the same way. But we know that $P_0(N) + P_1(N) = 1$; therefore, we have, immediately,

$$P_0(N) = [(1 - p_{11})/(p_{11} + p_{00} - 2)](p_{11} + p_{00} - 1)^N + [(p_{11} - 1)/(p_{11} + p_{00} - 2)]] \tag{B15}$$

APPENDIX C. SOURCE WITH MEMORY AND TRANSMISSION TIMES OPTIMIZED TO $t_{00} = t_{11} = 2\sigma$, $t_{10} = t_{01} = \sigma$

Here, we shall use the same notation as in Appendix B. As before, we can write

$$P_1(N) = P_1(N-2)\, p_{11} + P_0(N-1)\, p_{01} \tag{C1}$$

$$P_0(N) = P_1(N-1)\, p_{10} + P_0(N-2)\, p_{00} \tag{C2}$$

To decouple $P_0(N)$ and $P_1(N)$, we rewrite Eq. (C2) to give

$$P_1(N-1) = (1/p_{10})[P_0(N) - P_0(N-2)\,p_{00}] \tag{C3}$$

Now, substitute Eq. (C3) into (C1); after simplification, we have

$$P_0(N+1) - P_0(N-1)(1 + p_{00}p_{11}) + P_0(N-3)(p_{11}p_{00}) = 0 \tag{C4}$$

Notice that if $P_0(N)$ had been represented in terms of $P_1(N)$ and substituted into Eq. (C2), we would have obtained

$$P_1(N+1) - P_1(N-1)(1 + p_{00}p_{11}) + P_1(N-3)(p_{11}p_{00}) = 0 \tag{C5}$$

From Eqs. (C4) and (C5), it is observed that $P_0(N)$ and $P_1(N)$ satisfy the same equation. To simplify matters, define

$$P(N) = P_1(N) + P_0(N) \tag{C6}$$

Then, by combining Eqs. (C4) and (C5), we have

$$P(N+1) - P(N-1)(1 + p_{00}p_{11}) + P(N-3)(p_{11}p_{00}) = 0 \tag{C7}$$

Now, the particular solution $P(N) = \lambda^N$ is substituted into Eq. (C7):

$$\lambda^{N+1} - \lambda^{N-1}(1 + p_{00}p_{11}) + \lambda^{N+3}(p_{11}p_{00}) = 0 \tag{C8}$$

Dividing through Eq. (C8) by λ^{N-3} gives

$$\lambda^4 - \lambda^2(1 + p_{00}p_{11}) + p_{11}p_{00} = 0 \tag{C9}$$

Clearly,

$$\lambda^2 = 1, \quad p_{11}p_{00} \tag{C10}$$

and

$$\lambda = \pm 1, \quad \pm(p_{11}p_{00})^{1/2} \tag{C11}$$

The general solution is

$$P(N) = A_1(p_{11}p_{00})^{N/2} + A_2(-p_{11}p_{00})^{N/2} + A_3 + A_4(-1)^N \tag{C12}$$

There are four constants in Eq. (C12) to be determined by four boundary conditions. First, note that when $P(N)$ was used in Eqs. (62) and (63), it was always implied that one preceded this train of N units of σ. Therefore, when only one σ unit is involved, the only message which fits is a zero. Thus

$$P(1) = p_{10} = P_0(1), \qquad P_1(1) = 0 \tag{C13}$$

there is no way to fit a one following a one into a unit of length σ.

For $P(2)$, either a one or a zero followed by a one can be fitted. This gives

$$P(2) = p_{11} + p_{10}p_{01} = P_1(2), \qquad P_0(2) = 0 \tag{C14}$$

Similarly,

$$P(3) = p_{11}p_{10} + p_{10}p_{01}p_{10} + p_{10}p_{00} = P_0(3), \qquad P_1(3) = 0 \tag{C15}$$

$$P(4) = p_{10}p_{01}p_{10}p_{01} + p_{10}p_{01}p_{11} + p_{10}p_{00}p_{01} + p_{11}p_{11} + p_{11}p_{10}p_{01} = P_1(4), \quad P_0(4) = 0 \tag{C16}$$

Equations (C12)–(C16) enable us to write

$$\begin{aligned} A_1 &= [(1 - p_{11})/2(1 - p_{11}p_{00})][p_{00} - (p_{11}p_{00})^{1/2}] \\ A_2 &= [(1 - p_{11})/2(1 - p_{11}p_{00})][p_{00} + (p_{11}p_{00})^{1/2}] \\ A_3 &= (p_{11} - p_{00})/2(1 - p_{11}p_{00}), \qquad A_4 = (p_{11} - p_{00})/2(1 - p_{11}p_{00}) \end{aligned} \tag{C17}$$

Substitution of Eq. (C17) into (C12) yields

$$\begin{aligned} P(N) &= \frac{1 - p_{11}}{2(1 - p_{11}p_{00})} [p_{00} - (p_{00}p_{11})^{1/2}](p_{00}p_{11})^{N/2} \\ &\quad + \frac{1 - p_{11}}{2(1 - p_{11}p_{00})} [p_{00} + (p_{00}p_{11})^{1/2}][-(p_{11}p_{00})^{N/2}] \\ &\quad + \frac{2 - p_{11} - p_{00}}{2(1 - p_{11}p_{00})} + \frac{p_{11} - p_{00}}{2(1 - p_{11}p_{00})} (-1)^N \\ &= \frac{1}{2(1 - p_{11}p_{00})} ((1 - p_{11})(p_{11}p_{00})^{N/2}\{p_{00} - (p_{00}p_{11})^{1/2} \\ &\quad + (-1)^N[p_{00} + (p_{00}p_{11})^{1/2}]\} + (2 - p_{11} - p_{00}) + (p_{11} - p_{00})(-1)^N) \end{aligned} \tag{C18}$$

For even N, or $N = 2n$, Eq. (C18) becomes

$$P^{(2n)} = [1/(1 - p_{11}p_{00})]\{1 - (p_{00}p_{11})^{n+1} - p_{00}[1 - (p_{00}p_{11})^n]\} \tag{C19}$$

For odd N, or $N = 2n + 1$, Eq. (C18) becomes

$$P^{(2n+1)} = [(1 - p_{11})/(1 - p_{11}p_{00})][1 - (p_{11}p_{00})^{n+1}]$$

Notice that we have dropped the subscripts 1 and 0 in connection with $P(N)$. This is because of the mutual exclusiveness with respect to odd and even N. This mutual exclusiveness (subscript 1 going with odd N and subscript 0 with even N) can be proved through application of the recursion formulas (C1) and (C2) together with (C15) and (C16).

REFERENCES

1. H. Reiss and C. Huang, Statistical thermodynamic formalism in the solution of information theory problems, *J. Stat. Phys.* 3(2):191 (1971).
2. C. E. Shannon, Communication in the presence of noise, *Proc. IRE* **37**:10 (1949).
3. K. W. Cattermole, *Principles of Pulse Code Modulation* (Elsevier, New York, 1969).
4. S. W. Golomb, *Digital Communications with Space Applications* (Prentice-Hall, New Jersey, 1964), p. 12.

5. D. A. Huffman, A method for the construction of minimum redundancy codes, *Proc. IRE* **40**:1098 (1952).
6. R. Fano, *Transmission of Information* (John Wiley and Sons, New York, 1961).
7. D. J. Sakrison, *Communication Theory: Transmission of Waveforms and Digital Information* (John Wiley and Sons, New York, 1968), p. 105.
8. N. Wiener, *Acta Math.* **55**:117 (1930); A. Khinchine, *Math. Ann.* **109**:604 (1934).
9. D. J. Sakrison, *Communication Theory: Transmission of Waveforms and Digital Information* (John Wiley and Sons, New York, 1968), p. 119.
10. D. A. Bell, *Information Theory and Its Engineering Applications* (Pitman, New York, 1962), p. 58.
11. P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, London, 1966), p. 60.
12. L. Brillouin, *Science and Information Theory* (Academic Press, New York, 1963), p. 49.