THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 51, No. 4, April, 1972
Printed in U.S.A.

The Spectral Density of a Coded Digital Signal

By B. S. BOSIK

(Manuscript received August 23, 1971)

The stochastic process appearing at the output of a digital encoder is investigated. Based upon the statistics of the code being employed, a systematic procedure is developed by means of which the average power spectral density of the process can be determined. The method is readily programmed on the digital computer, facilitating the calculation of the spectral densities for large numbers of codes. As an example of its use, the procedure is applied in the case of a specific multi-alphabet, multi-level code.

I. INTRODUCTION

In recent years, increased interest has been focused on more complex multi-alphabet, multi-level codes.[1-4] Such codes are designed to produce a digital pulse train with specific spectral properties making it suitable for transmission over digital repeatered lines. These properties generally include the absence of a dc component and a strong spectral component from which timing can be extracted. This paper presents a method for calculating the spectral composition of the pulse trains resulting from the use of these codes. The procedure is applicable to a wide variety of codes.

A code may be defined as a set of mappings from a set of input symbols (or words) to a set of codewords. Each mapping is called an alphabet. The code may use different alphabets depending upon the state of the coded signal.[1] It is desirable for unique decipherability that the set of mappings be one-to-one, i.e., that no matter to how many alphabets a codeword belongs, it always corresponds to the same input symbol. However, this restriction will not be imposed here.

In general, when the code is applied to a sequence of input symbols, the resulting encoded signal is a stochastic process, the statistics of which depend on the input symbol sequence statistics and the code statistics. For convenience, a random input symbol sequence will be assumed so that the input symbols are equally likely. Even if the symbols are not equiprobable, the procedure for calculating the spectral density outlined in Section III remains valid, although the methods for calculating the required signal statistics described in Section IV must be modified. The spectral density derivation only becomes inapplicable when the statistics of the input symbols vary with time, so that the ergodicity and stationarity assumptions of Sections II and III are no longer valid.

The codes to be considered in this paper will have N states, each state corresponding to a single alphabet. The alphabet assignment need not be unique, i.e., more than one state can correspond to the same alphabet. The codes will have a block length L and the number of codeword symbol values (levels) will be M.

II. THE CODED SIGNAL

The codeword symbols are, in general, transmitted on some standard pulse shape $g(t)$ at intervals of duration T. The signal, then, may be expressed by

$$x(t) = \sum_{n=-\infty}^{\infty} a_n g(t - nT) \tag{1}$$

where a_n is the codeword symbol value for the time slot $nT \leqq t \leqq (n + 1)T$. The values which $\{a_n\}$ assume are determined by the code and the input symbols which are to be coded. The discrete parameter random process formed by the sequence of codeword symbols $\{a_n\}$ has an autocorrelation function $R(k) = E\{a_n a_{n+k}\}$, and, thus, is assumed to be wide-sense stationary. The autocorrelation function of the coded signal $x(t)$ is, then,

$$R_x(t + \tau, t) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} R(m - n) g(t + \tau - nT) g(t - mT) \tag{2}$$

which is, in general, a function of both t and τ. The coded signal is not, therefore, wide-sense stationary. However, it is easily shown that $R_x(t + \tau, t)$ is periodic in t with period T. The coded signal is, then, a cyclostationary process.

III. THE POWER SPECTRAL DENSITY OF THE SIGNAL

Since the coded signal is not a wide-sense stationary random process, the Fourier transform relationship between the autocorrelation function and the power spectral density cannot be invoked to find its power spectrum. However, the average power spectral density of a cyclostationary process of the form described above has been derived by W. R. Bennett.[5] Under the assumption that the process is ergodic, the spectral density is

$$w_x(f) = \frac{1}{T} \mid G(f) \mid^2 \left[R(0) + 2 \sum_{k=1}^{\infty} R(k) \cos 2\pi k f T \right] \tag{3}$$

where $G(f)$ is the Fourier transform of $g(t)$. The determination of the spectral density, then, requires the calculation of the autocorrelation function $R(k)$.

A method of calculating $R(k)$ can be derived as follows. Let the probability of being in state i during time slot n be

$$P(s_n = S_i) = P(S_i); \quad i = 1, 2, \cdots, N \tag{4}$$

and let the probability that the symbol a_n assumes the value A_l, $l = 1, 2, \cdots, M$, and a_{n+k} assumes the value A_m, $m = 1, 2, \cdots, M$, given that $s_n = S_i$ be

$$P(a_n = A_l, a_{n+k} = A_m \mid s_n = S_i) = P_k(A_l, A_m \mid S_i). \tag{5}$$

Then $R(k)$ can be expressed as

$$R(k) = \sum_{i=1}^{N} P(S_i) \sum_{l=1}^{M} \sum_{m=1}^{M} A_l A_m P_k(A_l, A_m \mid S_i). \tag{6}$$

But in time slot n, there is a probability of $1/L$ of being in the jth symbol, a_j, of a codeword, $j = 1, 2, \cdots, L$. Thus,

$$P_k(A_l, A_m \mid S_i) = \sum_{j=1}^{L} \frac{1}{L} P_k(A_l, A_m \mid S_i, j). \tag{7}$$

Substituting eq. (7) in eq. (6), we obtain

$$R(k) = \frac{1}{L} \sum_{i=1}^{N} P(S_i) \sum_{j=1}^{L} \sum_{l=1}^{M} \sum_{m=1}^{M} A_l A_m P_k(A_l, A_m \mid S_i, j). \tag{8}$$

Now define

$$\begin{aligned} R_{S_i}(j, j+k) &= \sum_{l=1}^{M} \sum_{m=1}^{M} A_l A_m P_k(A_l, A_m \mid S_i, j) \\ &= E_{S_i}\{a_j a_{j+k}\}; \qquad j = 1, 2, \cdots, L. \end{aligned} \tag{9}$$

Thus,

$$R(k) = \frac{1}{L} \sum_{i=1}^{N} P(S_i) \sum_{j=1}^{L} R_{S_i}(j, j+k). \tag{10}$$

The equation by which $R_{S_i}(j, j + k)$ is calculated is a function of both j and k. The values of j and k determine the relative positions of the codewords to which the symbols a_j and a_{j+k} belong. This knowledge combined with a knowledge of the coded signal statistics allows the calculation of $R_{S_i}(j, j+k)$ as follows:

(*i*) $k = 0$:

For $k = 0$, a_j and a_{j+k} are in the same position of the same codeword. Thus,

$$R_{S_i}(j, j) = \sum_{l=1}^{M} A_l^2 P(A_l \mid S_i, j) \tag{11}$$

where

$$P(A_l \mid S_i, j) = P(a_j = A_l \mid s_j = S_i). \tag{12}$$

(*ii*) $k = 1, 2, \cdots, L$:

For this range of k, a_{j+k} is in the same codeword as a_j for $j \leqq L - k$, and is in the next codeword for $j > L - k$. Thus,

$$R_{S_i}(j, j+k) = \begin{cases} \displaystyle\sum_{l=1}^{M} \sum_{m=1}^{M} A_l A_m P_k(A_l, A_m \mid S_i, j), & j \leqq L - k \\ \displaystyle\sum_{l=1}^{M} \sum_{m=1}^{M} A_l A_m \sum_{n=1}^{N} P(A_l, S_n \mid S_i, j) & \\ \qquad \cdot P(A_m \mid S_n, j + k - L), & j > L - k \end{cases} \tag{13}$$

where

$$P(A_l, S_n \mid S_i, j) = P(a_j = A_l, s_{j+L} = S_n \mid s_j = S_i). \tag{14}$$

(*iii*) $k = L + 1, L + 2, \cdots, 2L$:

For this range of k, a_{j+k} is in the codeword immediately following the codeword containing a_j for $j \leqq 2L - k$, and is two codewords away for $j > 2L - k$. Thus,

$$R_{S_i}(j, j+k) = \begin{cases} \displaystyle\sum_{l=1}^{M} \sum_{m=1}^{M} A_l A_m \sum_{n=1}^{N} P(A_l, S_n \mid S_i, j) & \\ \qquad \cdot P(A_m \mid S_n, j + k - L), & j \leqq 2L - k \\ \displaystyle\sum_{l=1}^{M} \sum_{m=1}^{M} A_l A_m \sum_{n=1}^{N} P(A_l, S_n \mid S_i, j) \sum_{p=1}^{N} P_1(S_p \mid S_n) & \\ \qquad \cdot P(A_m \mid S_p, j + k - 2L), & j > 2L - k \end{cases} \tag{15}$$

where

$$P_1(S_p \mid S_n) = P(s_{j+L} = S_p \mid s_j = S_n). \tag{16}$$

(*iv*) $k = QL + 1, QL + 2, \cdots, (Q + 1)L$ for $Q \geqq 2$:

For this range of k, a_{j+k} is in Qth codeword following the codeword containing a_j for $j \leqq QL - k$, and is $Q + 1$ codewords away for $j > QL - k$. Thus,

$$R_{S_i}(j, j + k)$$
$$= \begin{cases} \sum_{l=1}^{M} \sum_{m=1}^{M} A_l A_m \sum_{n=1}^{N} P(A_l, S_n \mid S_i, j) \sum_{p=1}^{N} P_{Q-1}(S_p \mid S_n) \\ \qquad \cdot P(A_m \mid S_p, j + k - QL), \qquad j \leqq QL - k \\ \sum_{l=1}^{M} \sum_{m=1}^{M} A_l A_m \sum_{n=1}^{N} P(A_l, S_n \mid S_i, j) \sum_{p=1}^{N} P_Q(S_p \mid S_n) \\ \qquad \cdot (A_m \mid S_p, j + k - (Q + 1)L), \qquad j > QL - k \end{cases} \tag{17}$$

where

$$P_Q(S_p \mid S_n) = \sum_{i=1}^{N} P_{Q-1}(S_i \mid S_n) P_1(S_p \mid S_i), \qquad Q \geqq 2 \tag{18}$$

and can be calculated recursively.

In summary, then, the procedure described in eqs. (11) through (18) is used to calculate $\{R_{S_i}(j, j + k)\}$ which is substituted in eq. (10) to obtain $\{R(k)\}$. The spectral density $w_x(f)$ can then be obtained from $\{R(k)\}$ by means of eq. (3).

It is easily seen that for any code other than the most trivial, the calculation of the spectral density is a formidable task. However, it is a relatively straightforward procedure to program a digital computer to perform the above calculations; and this is the most profitable use of the procedure.

IV. THE CODED SIGNAL STATISTICS

The calculation of the spectral density described above requires the knowledge of numerous probabilities concerning the code. These statistics are, in general, readily obtainable from the code by merely counting the number of occurrences of the phenomenon involved, or are determined from simple calculations involving previously obtained probabilities. The procedure for obtaining each of the necessary probabilities follows:

(*i*) The State Transition Probabilities:

$$P_1(S_p \mid S_n) = P(s_{j+1} = S_p \mid s_j = S_n); \quad p, n = 1, 2, \cdots, N. \tag{19}$$

This transition probability is obtained by counting the number of codewords in the alphabet used when in state S_n, whose next state is S_p. (Given any codeword in any state, the next state is uniquely defined at the end of that codeword.) The resulting sum is divided by the number of codewords in the alphabet. The probability of a transition from state S_n to S_p in Q steps, $P_Q(S_p \mid S_n)$, can be calculated recursively from $P_1(S_p \mid S_n)$ via eq. (18).

(*ii*) The State Probabilities:

$$P(S_i) = P(s_n = S_i); \quad i = 1, 2, \cdots, N. \tag{20}$$

The probability of being in state S_i is calculated from $\{P_1(S_p \mid S_n)\}$. $\{P(S_i)\}$ are the solutions to the set of simultaneous linear equations

$$\sum_{k=1}^{N} P(S_k)P(S_j \mid S_k) = P(S_j); \qquad j = 1, 2, \cdots, N. \tag{21}$$

However, only $N - 1$ of these equations are linearly independent. An additional equation must be used:

$$\sum_{j=1}^{N} P(S_j) = 1. \tag{22}$$

(*iii*) The Symbol Probabilities:

$$P(A_l \mid S_i, j) = P(a_j = A_l \mid s_j = S_i); \qquad \begin{array}{l} i = 1, 2, \cdots, N \\ j = 1, 2, \cdots, L \\ l = 1, 2, \cdots, M. \end{array} \tag{23}$$

These probabilities are determined by counting the number of occurrences of symbol A_l in the jth position of the codewords in the alphabet used when in state S_i. This sum is divided by the number of codewords in the alphabet.

(*iv*) The Symbol Combination Probabilities:

$$\begin{aligned} &P_k(A_l, A_m \mid S_i, j) \\ &\quad = P(a_j = A_l, a_{j+k} = A_m \mid s_n = S_i); \qquad \begin{array}{r} i = 1, 2, \cdots, N \\ j = 1, 2, \cdots, L \\ l, m = 1, 2, \cdots, M. \\ k \leqq L - j \end{array} \end{aligned} \tag{24}$$

This is the probability of the occurrence of the symbols A_l and A_m, in time slots j and $j + k$ respectively, within the codewords of the alphabet used in state S_i (i.e., $k \leqq L - j$). Thus, for position j of the alphabet of S_i, the number of times $a_j = A_l$ and $a_{j+k} = A_m$ in the same codeword are counted, and divided by the number of codewords in the alphabet.

(v) The Conditional State Transition Probabilities:

$$P(A_l, S_n \mid S_i, j) = P(a_j = A_l, s_{j+L} = S_n \mid s_j = S_i); \quad i, n = 1, 2, \cdots, N; \quad j = 1, 2, \cdots, L; \quad l = 1, 2, \cdots, M. \quad (25)$$

This is the probability that a codeword in the alphabet of state S_i has the symbol A_l in the jth position and has the state S_n as its next state. Thus, the number of times that a codeword is in the alphabet of state S_i, whose next state is S_n and whose symbol level is A_l in the jth position, are counted, and divided by the number of codewords in the alphabet.

These five sets of statistics are all that are required to perform the calculation of the spectral density. Although following the procedures for obtaining these probabilities is a very straightforward task, it is, again, a tedious one, especially for any reasonably complex code. Here again, the digital computer can be used to good advantage.

V. AN EXAMPLE—THE FRANASZEK MS-43 CODE

The Franaszek MS-43 code[1] is a ternary, 4-state, 3-alphabet code of word length 3. It is one of a family of codes designed to produce a digital pulse train with specific desirable properties. These properties include the absence of a dc component, a bounded sum of previous digits, and a strong spectral component from which the signaling frequency can be derived. The code is shown in Table I. Alphabet R_1 is used when in state S_1, alphabet R_2 is used in state S_2 or S_3, and alphabet R_3 is used in state S_4. The state is determined at the end of a codeword by summing all previous digital symbols. This sum is inherently restricted to be 1, 2, 3, or 4 corresponding to states S_1, S_2, S_3, and S_4. Tables II through VI list the statistics necessary for calculating the spectral density as determined by the procedures described in Section IV. The digital computer was utilized to perform the spectral

TABLE I—THE MS-43 CODE

Binary Input Words	R_1	R_2	R_3
0000	+++	−+−	−+−
0001	++0	00−	00−
0010	+0+	0−0	0−0
0100	0++	−00	−00
1000	+−+	+−+	−−−
0011	0−+	0−+	0−+
0101	−0+	−+0	−0+
1001	00+	00+	−−0
1010	0+0	0+0	−0−
1100	+00	+00	0−−
0110	−+0	−+0	−+0
1110	+−0	+−0	+−0
1101	+0−	+0−	+0−
1011	0+−	0+−	0+−
0111	−++	−++	−−+
1111	++−	+−−	+−−

TABLE II—STATE TRANSITION PROBABILITIES FOR MS-43 CODE

$P_1(S_p|S_n)$

	P			
n	1	2	3	4
1	6/16	6/16	3/16	1/16
2	5/16	6/16	5/16	0
3	0	5/16	6/16	5/16
4	1/16	3/16	6/16	6/16

TABLE III—STATE PROBABILITIES FOR MS-43 CODE

$P(S_i)$

i	$P(S_i)$
1	5/28
2	9/28
3	9/28
4	5/28

density calculation. The resulting normalized spectrum is plotted in Fig. 1. The result is consistent with the expected properties of the coded signal spectrum, i.e., zero dc component and periodicity with period $1/T$.

VI. CONCLUSION

A general procedure for determining the average power spectral density of a coded digital signal has been presented. The procedure is long, but straightforward and readily programmable on the digital computer. With the aid of the computer, the spectral content of signals resulting from the implementation of large numbers of codes can be obtained.

VII. ACKNOWLEDGMENTS

The author is indebted to W. R. Bennett whose advice and encouragement made this work possible. The author also had the benefit

TABLE IV—SYMBOL PROBABILITIES FOR MS-43 CODE

$P(A_m|S_i, j) = P(A_m|S_i), \forall j = 1, 2, 3$

A_m	S_i			
	S_1	S_2	S_3	S_4
+	8/16	5/16	5/16	3/16
−	3/16	5/16	5/16	8/16

TABLE V—SYMBOL COMBINATION PROBABILITIES FOR MS-43 CODE

$P_k(A_l, A_m|S_i, j)$

$j = 1, k = 1$

$A_l A_m$	S_i			
	S_1	S_2	S_3	S_4
++	3/16	0	0	0
+−	2/16	3/16	3/16	2/16
−+	2/16	3/16	3/16	2/16
−−	0	0	0	3/16

$j = 1, k = 2$

$A_l A_m$	S_i			
	S_1	S_2	S_3	S_4
++	3/16	1/16	1/16	0
+−	2/16	2/16	2/16	2/16
−+	2/16	2/16	2/16	2/16
−−	0	1/16	1/16	3/16

$j = 2, k = 1$

$A_l A_m$	S_i			
	S_1	S_2	S_3	S_4
++	3/16	3/16	1/16	0
+−	2/16	2/16	2/16	2/16
−+	2/16	2/16	2/16	2/16
−−	0	1/16	1/16	3/16

TABLE VI—CONDITIONAL STATE TRANSITION PROBABILITIES FOR MS-43 CODE

$P(A_t, S_n|S_i, j)$

		S_i											
		S_1			S_2			S_3			S_4		
		j											
A_t	S_n	1	2	3	1	2	3	1	2	3	1	2	3
+	S_1	2/16	2/16	2/16	1/16	1/16	0	0	0	0	0	0	0
+	S_2	3/16	3/16	3/16	2/16	2/16	2/16	1/16	1/16	0	0	0	0
+	S_3	2/16	2/16	2/16	2/16	2/16	3/16	2/16	2/16	2/16	1/16	1/16	1/16
+	S_4	1/16	1/16	1/16	0	0	0	2/16	2/16	3/16	2/16	2/16	2/16
−	S_1	2/16	2/16	2/16	2/16	2/16	3/16	0	0	0	1/16	1/16	1/16
−	S_2	1/16	1/16	1/16	2/16	2/16	2/16	2/16	2/16	3/16	2/16	2/16	2/16
−	S_3	0	0	0	1/16	1/16	0	2/16	2/16	2/16	3/16	3/16	3/16
−	S_4	0	0	0	0	0	0	1/16	1/16	0	2/16	2/16	2/16

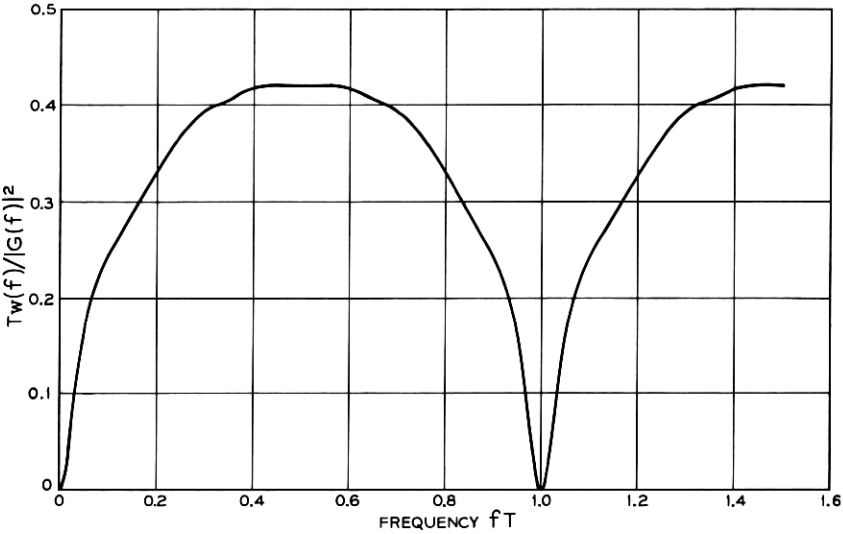

Fig. 1—Power spectral density for the Franaszek MS-43 code.

of a private communication from A. Fromageot discussing the problem of calculating the spectral density of the Franaszek MS-43 code.

REFERENCES

1. Franaszek, P. A., "Sequence-State Coding for Digital Transmission," B.S.T.J., *47*, No. 1 (January 1968), pp. 143–157.
2. Sipress, J. M., "A New Class of Selected Ternary Pulse Transmission Plans for Digital Transmission Lines," IEEE Trans. Commun. Tech, *COM-13*, (September 1965), pp. 366–372.
3. Neu, W., and Kündig, A., "Project for a Digital Telephone Network," IEEE Trans. Commun. Tech, *COM-16*, (October 1968), pp. 633–648.
4. Johannes, V. I., Kaim, A. G., and Walzman, T., "Bipolar Pulse Transmission with Zero Extraction," IEEE Trans. Commun. Tech., *COM-17*, No. 2 (April 1969), pp. 303–310.
5. Bennett, W. R., "Statistics of Regenerative Digital Transmission," B.S.T.J., *37*, No. 6 (November 1958), pp. 1501–1542.