On Runlength Codes

EPHRAIM ZEHAVI AND JACK K. WOLF, FELLOW, IEEE

Abstract—Several new results on binary \( (d, k) \) codes are given. First, a new derivation for the capacity of these codes based upon information theoretic principles is given. Based upon this result the spectrum of a \( (d, k) \) code is computed. Finally, the problem of computing the capacity of the binary symmetric channel under the condition that the input sequences satisfy the \( (d, k) \) constraint is considered. Lower bounds on the capacity of such a channel are derived.

I. INTRODUCTION

In many digital transmission and recording systems, considerations such as spectral shaping, self-clocking, and reduction of intersymbol interference require that the recorded sequence have special binary properties called channel constraints. A channel code is a mapping of input sequences onto channel sequences that satisfy these channel constraints. A common channel constraint used in the magnetic recording channel is the so-called \( (d, k) \) constraint for binary sequences. The codes are also called runlength-limited (RLL) codes. Binary sequences satisfy a \( (d, k) \) constraint if the number of consecutive zeros between adjacent ones is greater than or equal to \( d \) and less than or equal to \( k \).

This paper addresses two problems related to \( (d, k) \) codes. Let us define an optimal \( (d, k) \) code as a \( (d, k) \) code having maximum information rate. The first problem addressed is the calculation of the capacity and the spectral density of the optimal \( (d, k) \) codes. In Section II we derive some probabilistic properties of optimal codes. Then the spectral density of an optimal \( (d, k) \) code is derived. This result was first obtained by Foley 44 years ago [1] using a different approach. The second problem treated is the capacity of the binary symmetric channel (BSC) with crossover probability \( p \) under the condition that the input sequences satisfy \( (d, k) \) constraints. In Section III a lower bound on the capacity of such a channel is derived using some information theoretic inequalities.

II. OPTIMAL \( (d, k) \) CODES

A \( (d, k) \) code is a set of binary sequences that satisfy the \( (d, k) \) constraint. Any binary sequence, denoted by \( \{c_n\} \), of finite or infinite length can be parsed uniquely into a concatenation of phrases, each phrase ending in a single "1" and beginning with a string of none or one or more "0's." For example, the binary sequence

\[ 0101001011101 \cdots \]

would parse as

\[ 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, \cdots \]

where commas are used to separate the phrases. A binary sequence is said to satisfy a \( (d, k) \) constraint if and only if all its phrases contain no less than \( (d + 1) \) digits and no more than \( (k + 1) \) digits. Thus "\( d \)" refers to the minimum number of 0's in a phrase and "\( k \)" refers to the maximum number of 0's in a phrase.

Let \( X_i \) be a random variable describing the number of binary digits in the \( i \)th phrase of the parsed sequence. Then the information rate of a random sequence of infinite length is defined as

\[ R = \lim_{n \to \infty} H(X_1, X_2, \ldots, X_n) / E(X_1 + X_2 + \cdots + X_n). \tag{1} \]

Here \( H(X_1, X_2, \ldots, X_n) \) is the joint entropy of the first \( n \) phrases and \( E(X_i) \) is the expectation of \( X_i \).

An optimal \( (d, k) \) sequence is a \( (d, k) \) sequence having maximum information rate; the capacity of the code, denoted by \( C \), is this maximum rate. We first determine the information rate (or capacity) for an optimal \( (d, k) \) code. This problem has already been solved by Shannon [2] by finding an expression for the number of binary sequences of length \( n \) satisfying the \( (d, k) \) constraints. Here we use a probabilistic approach to compute the capacity. The new derivation gives us a better understanding of the structure of \( (d, k) \) sequences and allows us to find upper and lower bounds for \( C \). As was shown by Shannon, we find that the capacity of the code, \( C \), is given as the base-two logarithm of the largest real root of the equation

\[ 1 + Z^{k+2} - Z^{k+1} - Z^{k+1-d} = 0. \tag{2} \]

We first prove the following theorem.

Theorem 1: The code that achieves the maximum information rate has the following properties.

1) The random variables \( X_1, X_2, \ldots \) are statistically independent and identically distributed.

2) The probability distribution of \( X \) is given by

\[ P(X = i) = 2^{-ic}, \quad i = d + 1, d + 2, \ldots, k = 1 \tag{3} \]

with \( C \) such that \( \Sigma_{i=d+1}^{k} 2^{-ic} = 1. \)
The only solution to this equation under the constraint
where the supremum is taken over all joint probabilities
$P_n(X)$ for the $n$ random variables $X_1, X_2, \ldots, X_n$. By using
the well-known inequality [3]
$H(X_1, X_2, \ldots, X_n) \leq H(X_1) + H(X_2) + \cdots + H(X_n)$,
we obtain
$R \leq \lim \sup_{n \to \infty} \sup_{P_n(X)} H(X_1) + E(X_1) + \cdots + E(X_n)$
with equality if and only if the $X_i$ are statistically independent. Note that the supremum is now taken over the product of the marginal probabilities. Thus
$R \leq \lim \sup_{n \to \infty} \frac{H(X_1, X_2, \ldots, X_n)}{E(X_1) + E(X_2) + \cdots + E(X_n)}$.

The last inequality follows from applying induction to the
inequality
$A + B \leq C + D \leq \max \left\{ \frac{A}{C}, \frac{B}{D} \right\}, \quad A, B, C, D > 0$.

$R$ is equal to the right side of (5) if the $X_i$ are statistically independent and identically distributed. For the code of maximum information rate, let $C$ and $L$ be this maximum rate and the average length of a phrase, respectively. Denote by $P^*(X = i)$ the maximizing probability (that a phrase $X$ has length $i$). Then using the technique of Lagrange multipliers, this problem can be treated as the problem of maximizing
$$F(P, \lambda) = \frac{H(P(X))}{E(P(X))} + \lambda \sum_{i=d+1}^{k+1} P(X = i)$$
where $\lambda$ is obtained from the constraint equation
$$\sum_{i=d+1}^{k+1} P(X = i) = 1.$$

Therefore, the $P^*$ which maximizes the code rate must satisfy the equation
$$\frac{\partial F(P^*, \lambda)}{\partial P^*(X = i)} = \frac{-(\log(e) + \log(P^*(X = i)) + iC)}{L} + \lambda = 0.$$

The only solution to this equation under the constraint
$\sum_{i=d+1}^{k+1} P(X = i) = 1$ is
$P^*(X = i) = 2^{-iC}, \quad i = d+1, d+2, \ldots, k+1$
where $C$ is the solution to the equation
$$\sum_{i=d+1}^{k+1} 2^{-iC} = 1.$$
Define \( q_j = 1 - p_j \) where
\[
p_j = P(X = j + 1|X > j) = \frac{P(X = j + 1, X > j)}{P(X > j)}
\]
\[
= \frac{P(X = j + 1)}{1 - P(X \leq j)}. \quad (7)
\]

We now have the following lemma.

**Lemma 1:** 1) The optimal code has the following stationary probabilities:
\[
P(S_n = j) = P(S_n = 0) \prod_{i=0}^{j-1} q_i = P(S_n = 0) P(X > j),
\]
\[
P(S_n = 0) = 1/L, \quad (8)
\]
where \( L \) is the average length of a pulse corresponding to a single phrase. 2) The transition probability matrix \( F \) for the state diagram is given by
\[
F = \begin{bmatrix}
0 & 0 & \cdots & p_d & p_{d+1} & \cdots & p_{k-2} & p_{k-1} & 1 \\
1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & \cdots & q_d & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & \cdots & 0 & q_{d+1} & \cdots & 0 & 0 & 0 \\
& & & & \cdots & & & & \cdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & q_{k-2} & 0 & q_{k-1} \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & q_{k-1} & 0
\end{bmatrix}
\]

**Proof:** The finite-state machine for the optimal code is described by an irreducible aperiodic finite Markov chain such that all states are ergodic with stationary probability distribution \( P(S_n) \), where \( P(S_n) \) is the reciprocal of the mean recurrence time of state \( S_n \) [5, p. 356]. The mean recurrence of state 0 is the average length of a pulse. Since there is a one-to-one mapping of phrases to state transitions (see Fig. 2), the transition-probability matrix follows directly from the probability distribution of \( X \).

![Fig. 2. Transition probabilities of \((d, k)\) code.](image)

The correlation function between channel symbols \( \{ y_n \} \) can now be computed by using the sequence \( \{ c_n \} \). It is easy to see that the autocorrelation of \( \{ y_n \} \), denoted by \( R_N \), is given by
\[
R_N = E(y_n y_{n+N}) = \Pr(\text{sequence } \{ c_i \}, r < i \leq r + N \\
\text{has an even number of ones})
\]
\[
- \Pr(\text{sequence } \{ c_i \}, r < i \leq r + N \\
\text{has an odd number of ones}), \quad N \geq 1. \quad (9)
\]

By definition, \( R_0 = 1 \). In computing (9) one uses the stationary probabilities for the initial state \( S_0 \). Note that for \((d, k)\) codes \( c_i = 1 \) always bring the finite-state machine back to state zero. Therefore, in the calculation of \( R_N \) it is sufficient to enumerate the number of times that a sequence of length \( N \) passes through the zero state an even and an odd number of times. To simplify the computation, we will insert a dummy variable \( D \) in those entries of the transition matrix \( F \) corresponding to a state transition to the zero state. The matrix \( F(D) = \{ F_{ij}(D) \} \) is thus defined as
\[
F(D) = \begin{bmatrix}
0 & 0 & \cdots & p_d D & p_{d+1} D & \cdots & p_{k-2} D & p_{k-1} D & p_k D \\
1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & \cdots & 0 & q_d & 0 & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & q_{k-2} & 0 & q_{k-1} \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & q_{k-1} & 0
\end{bmatrix}
\]

where \( q_j \) and \( p_j \) were defined in (7). The entries of \( \{ F(D) \}^N \), which describe sequences of length \( N \), are polynomials in \( D \). By setting \( D \) equal to \(-1\), one then assigns a positive sign to sequences with an even number of visits to state 0 (since such sequences are characterized by an even power of \( D \)) and assigns a negative sign to sequences with an odd number of visits to state 0 (since such sequences are characterized by an odd power of \( D \)). Thus we can give the correlation function of the optimal \((d, k)\) codes by using the matrix equation
\[
R_N = A \{ F(-1) \}^N P^T \quad (10)
\]
where \( P = (P(S = 0), P(S = 1), \cdots, P(S = k)) \) are the stationary probabilities of the states, \( A \) is a row vector of all ones of length \( k + 1 \), and \( F(-1) \) is \( F(D) \) with \( D = -1 \).

The spectrum of the optimal \((d, k)\) codes can now be written using the definition
\[
S_d(f) = \sum_{n=-\infty}^{\infty} R_n e^{-j2\pi fn} = -1 + \sum_{n=0}^{\infty} R_n e^{-j2\pi fn} + \sum_{n=-\infty}^{0} R_n e^{-j2\pi fn}
\]

Defining \( Z = \exp(-j2\pi f) \), we have
\[
S_d(f) = -1 + \sum_{n=0}^{\infty} R_n Z^n + \sum_{n=-\infty}^{0} R_n Z^n = -1 + A \left( \sum_{n=0}^{\infty} [ZF(D)]^n \right) P^T
\]
\[
= -1 + A \left( \{ I - ZF(D) \}^{-1} \right) P^T |_{D = -1}. \quad (11)
\]

The inverse matrices in (11) exist since it is shown in Appendix II that for \( k > d \), \( ZF(-1) \) has all its eigenvalues inside the unit circle, and therefore the geometric matrix series are summable.
Let $y(t)$ be the transmitted waveform given by
\[
y(t) = \sum_{n} y_n (t - nT)
\]
where $s(\cdot)$ is a rectangular pulse of duration $T$ time units and amplitude 1. Then the spectral density of the sequence $\{y_n\}$ and the waveform $y(t)$ is given in the following lemma.

**Lemma 2:** a) The spectral density of the sequence $\{y_n\}$ corresponding to an optimal $(d,k)$ code is given by
\[
S_d(f) = \frac{1 - |G(\exp(j2\pi f))|^2}{L \sin^2(\pi f) \| G(\exp(j2\pi f)) \|^2}
\]
where
\[
G(Z) = \sum_{i=0}^{k} P(X = i + 1) Z^{i+1}.
\]
b) The spectral density of the transmitted waveform $y(t)$ is
\[
S_y(f) = \frac{1 - |G(\exp(j2\pi f/T))|^2}{LT(f\pi)^2 \| G(\exp(j2\pi fT)) \|^2}.
\]

The proof is given in Appendix I.

This is the same formula obtained by Foley [1]. In Figs. 3–9 the power spectra $S_d(f)$ are given for various $(d,k)$ constraints. We compare the codes in one of two ways. Either we hold the data rate constant (Figs. 3–6), in which case the different codes have different minimum pulse durations, or we hold the minimum pulse duration constant (Figs. 7–9) in which case the data rates differ for the different codes. The spectral density at zero frequency is given as $TE(X-L)^2/L$, where $E(X-L)^2$ is the variance of the length of a phrase. This ratio has a maximum value of $1/T$ achieved by the code $(0, \infty)$. As $d$ increases, the spectral density at zero frequency decreases. Furthermore, the peak in the spectrum increases and shifts in the direction of higher frequencies. As $k$ increases, the spectral density at zero frequency increases, and the spectrum becomes more flat.
III. THE CAPACITY OF THE BSC CHANNEL UNDER 
\((d, k)\) CONSTRAINT

To this point we have analyzed the entropy and spectra of sequences that satisfy the \((d, k)\) constraints. In most data storage systems, however, the entropy is of less interest than the probability that the source letters are incorrectly reproduced. This error probability was the subject of the major theorem of information theory, the channel coding theorem. For a broad class of sources and channels, this theorem states that if the source entropy (per unit of time) is less than the channel capacity (per unit of time), then the error probability can be reduced to any desired level by using a sufficiently complex encoder and decoder. In this section we are interested in the capacity of the
binary symmetric channel (BSC) under the constraint that the input sequences are \((d, k)\) sequences. Although the channel is memoryless, each letter in the output of the channel depends both on the corresponding input and on past inputs and outputs.

Let \(Y_N = (Y_1, Y_2, \ldots, Y_N)\) denote a sequence of binary digits satisfying the \((d, k)\) constraints. Furthermore, let us denote by \(S_N = (S_1, S_2, \ldots, S_N)\) the sequence of \(N\) states of the finite-state machine which produced the sequence \(Y_N\). The channel is assumed to be a binary symmetric channel (BSC) with crossover probability \(p\). The output of the channel is denoted by the sequence \(R_N = (r_1, r_2, \ldots, r_N)\). The channel capacity is equal to

\[
C = \lim_{N \to \infty} \sup_{P(r_N)} \frac{I(Y_N : R_N)}{N} = \lim_{N \to \infty} \sup_{P(s_N)} \frac{I(S_N : R_N)}{N}
\]

(14)

where the supremum is either taken over all probabilities \(P(Y_N)\) for the sequence \(Y_N\) satisfying the \((d, k)\) constraint or over all probabilities \(P(S_N)\) for the sequence of states \(S_N\) of the finite-state machine corresponding to the \((d, k)\) sequence. The last equality in (14) results from the one-to-one mapping from a sequence \(S_N\) to the corresponding state transitions sequence \(S_N\), for \(N > k\). In the previous section we showed that for \(p = 0\), capacity was achieved for a Markovian finite-state machine. However, for \(p > 0\), we do not know this to be the case. To obtain a lower bound on the capacity, we impose the restriction that the supremum be taken only over probabilities \(P(Y_N)\) which are stationary and Markovian and which satisfy the \((d, k)\) constraint.

The following lemma enables us to find a lower bound on the capacity of the channel. For convenience, we define \(S_0 = \emptyset\).

**Lemma 3:** For a stationary and Markovian source and memoryless channel, we have

a) \(H(Y_N | R_N) = H(S_N | R_N) = \sum_{i=1}^{N} H(S_i | S_{i-1}, r_1, \ldots, r_N),\)

for \(N > k\) (15)

b) \(H(S_i | S_{i-1}, r_1, \ldots, r_N) \geq H(S_i | S_{i-1}, r_1, \ldots, r_N),\)

for \(i > j \geq 2\). (16)

**Proof:** a) By definition,

\[
H(S_N | R_N) = H(S_1 | r_1, \ldots, r_N) + H(S_2 | S_1, r_2, \ldots, r_N)
+ H(S_3 | S_1, S_2, r_3, \ldots, r_N)
+ \cdots + H(S_N | S_1, \ldots, S_{N-1}, r_1, \ldots, r_N).
\]

By using the assumptions that the source is Markovian and the channel is memoryless, we obtain (15).

b) We have

\[
H(S_i | S_{i-1}, r_1, \ldots, r_N) = H(S_i | S_1, r_2, \ldots, r_{N-j+2})
\geq H(S_1 | S_1, r_2, \ldots, r_{N-j+2})
\cdots
= H(S_{i-1} | r_1, \ldots, r_N).
\]

The two equalities follow from the assumption that the process is stationary and Markov, while the inequality is due to [3, theorem 3.5.1] which states that conditioning cannot increase the entropy.

We are now ready to derive the first lower bound on the capacity.

**Lemma 4:** The capacity of a memoryless channel under the constraint that the source is stationary and Markovian is lower-bounded by

\[
C \geq \sup_{P(S_1, S_2)} I(S_2 : r_2 | S_1).
\]

**Proof:** If the source is stationary and Markovian, then by using [3, theorem 3.5.1] we obtain

\[
\lim_{N \to \infty} \frac{H(S_N)}{N} = \lim_{N \to \infty} \frac{H(S_N | S_1, S_2, \ldots, S_{N-1})}{N}
= \lim_{N \to \infty} H(S_N | S_{N-1}) - H(S_2 | S_1).
\]

Therefore, it remains to find a bound on the conditional entropy \(H(S_N | R_N) / N\). By combining (15) and (16), we obtain the following:

\[
\frac{H(S_N | R_N)}{N} \leq \frac{H(S_1, r_2, \ldots, r_N)}{N} + \frac{H(S_1, r_1, \ldots, r_N)}{N}.
\]

(18)

Since the inequality is for any \(N\), taking the limit of (18) as \(N \to \infty\) yields the following upper bound on the conditional entropy:

\[
\lim_{N \to \infty} \frac{H(S_N | R_N)}{N} \leq H(S_1 | S_2, r_2).
\]

(19)

Substituting these results into (14) yields (17) directly.

We are now ready to derive a lower bound on the capacity of a BSC channel under \((d, k)\) constraints.

**Lemma 5:** The capacity of the BSC channel under \((d, k)\) constraints of the source is lower-bounded as

\[
C \geq C_1 = \max_{j=0}^{k} \sum_{j=0}^{k} P(S = j) h(p + p_j - 2pq_j) - h(p)
\]

(20)

where \(p_j\) is the conditional probability of a transition from state \(j\) to state 0. The maximum is taken over all possible codeword length distributions that satisfy the \((d, k)\) constraints and \(h(x) = -x \log x - (1-x) \log (1-x)\).
Proof: By definition,
\[
I(S_2 : r_2 | S_1) = H(S_2 | S_1) + H(r_2 | S_1) - H(S_2, r_2 | S_1)
\]
\[
= \sum_{j=0}^{k} P(S = j) \left[ h(p_j) + h(p + p_j - 2p_j) - [h(p_j) + h(p)] \right].
\]
Now the capacity of the channel is lower-bounded by the maximum of \(I(S_2 : r_2 | S_1)\) over all possible source distributions which satisfy the \((d, k)\) constraint.

A chain of lower bounds on the capacity will now be derived by using the same arguments as in Lemmas 4 and 5. Note that in (18) the conditional entropy is decreased whenever more information exists about future values of the channel outputs. Therefore, for any given \(N\) there are at least \(N - 1 - m\) elements in the sum of entropies that are conditioned on at least \(m\) values \((r_1, \ldots, r_m)\). Now for any given value of \(m\), the conditional entropy is upper-bounded by
\[
\lim_{N \to \infty} \frac{H(S_2 | S_1, r_2, \ldots, r_m)}{N} \leq H(S_2 | S_1, r_2, \ldots, r_m)
\]
\[
\leq H(S_2 | S_1, r_2, \ldots, r_4) \leq H(S_2 | S_1, r_2, r_4).
\]
Therefore, we obtain a set of lower bounds on the channel capacity
\[
C_1 \leq C_2 \leq \cdots \leq C_m \leq \cdots \leq C
\]
where \(C_m\) is
\[
C_m = \sup_{P(S_1, \ldots, S_m)} \{ H(S_2 | S_1) - H(S_2 | S_1, r_1, \ldots, r_m) \}.
\]

The first lower bound on the capacity \(C_1\) was computed for various values of \(d\) and \(k\). In Figs. 10 and 11 this bound is shown for various \((d, k)\) constraints as a function of \(p\), the crossover probability of the channel.

In practical systems many factors influence the maximum recording density. One of these factors is the maximum duration of a pulse that can be recorded and reliably recovered from the media. Here we consider how this factor indicates a choice of \(d\) in a \((d, k)\) code when the received signal is corrupted by noise.

Let us first ignore the effects of the noise. Let \(T_{\text{min}}\) seconds be the duration of the shortest pulse that can be recorded. If no \((d, k)\) code is used (that is, a \((0, \infty)\) code is used), then in the absence of noise, information will be recorded at the density of 1 bit/\(T_{\text{min}}\) seconds. If a \((d, k)\) code is used, then \((d + 1)\) code bits can be recorded in \(T_{\text{min}}\) seconds, so information can be recorded at a density of \((d + 1)C(d, \infty)\) bits/\(T_{\text{min}}\) seconds. The bounds for \(C(d, \infty)\) given in (6) thus reveal that by letting \(d\) approach infinity, in the absence of noise we can store an infinite number of information bits in \(T_{\text{min}}\) seconds. When we take into account noise, however, we find that this is no longer the case. Furthermore, we find that for a given signal-to-noise ratio, there will be a best value of \(d\) in the sense of maximizing the information density.

Most magnetic recording systems today use a "peak detector." Here we choose to compare the performance of \((d, k)\) codes in the presence of noise using a detector more familiar to communication engineers—a "matched filter detector." Our reasons for doing this are that the performance of the matched filter detector is superior to that of the peak detector (at least for additive white Gaussian...
noise) and that equations for the performance of the
matched filter detector are well-known. Care should be
exercised by the reader in applying the results to situations
other than those which satisfy the assumptions made.

We compare the performance of \((d, k)\) codes in noise
under the following set of assumptions.

1) The only noise in the system is additive white Gaussian
noise of zero mean and power spectral density \(N_0/2\).
2) There is no intersymbol interference from pulse to
pulse.
3) A matched filter detector is utilized followed by a
hard decision. This leads to a crossover probability
\(p\) of

\[
Q(p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy.
\]

4) The density of recording in bits/T_{\text{min}} seconds is taken
as \((d + 1)C_1\), where \(C_1\) is given in (20).

In Fig. 12 we compare the performance of \((0,7)\), \((1,7)\),
\((2,7)\), and \((3,7)\) codes under this set of assumptions. We
plot the density of recording (as defined earlier) versus
\(E_b/No\) for each of these codes. We see that the higher
the value of \(E_b/No\), the higher the value of “d” which
maximizes this density. This agrees with our intuition that at
high signal-to-noise ratios one can use a smaller decision
window than at low signal-to-noise ratios. Although we are
confident that similar results apply under a wide range of
assumptions, the reader is reminded of the previous cau-
tions.

**IV. SUMMARY**

We have presented a probabilistic approach to the un-
derstanding of \((d, k)\) codes. This approach provided a new
method for deriving some well-known results and yielded
several new results. We cite, in particular, a new derivation
of the capacity of \((d, k)\) codes for the case of no noise, a
derivation of the spectra of these codes, and a lower bound
to the capacity of \((d, k)\) codes used in conjunction with a
binary symmetric channel.

**ACKNOWLEDGMENT**

The authors are indebted to the referees for their per-
serverance and for their contributions in improving the
contents of this paper.

**APPENDIX I**

**PROOF OF LEMMA 2**

Referring to (11), we define \(C(Z)\) as

\[
C(Z) = A \left[ \sum_{n=0}^{\infty} [2ZF(D)]^n P^n \right]_{D=-1} \tag{1.1}
\]

\(C(Z)\) is the total flow (from the initial state \(S_i\) to the final state
\(S_f\) in the signal flow diagram shown in Fig. 13. This flow will be
computed by Mason’s formula [6]. It is convenient to define three
new quantities \(g(Z), Q(Z),\) and \(M_i(Z)\) as

\[
g(Z) = \sum_{i=0}^{d} P(X > i) Z^i \tag{1.2}
\]

\[
Q(Z) = 1 + \sum_{i=0}^{d} P(X = i + 1) Z^{i+1} \tag{1.2}
\]

\[
M_i(Z) = P(X > n) \left[ 2 \sum_{r=0}^{n} Z^r - g(Z) \right] \tag{1.3}
\]
It is easy to verify that
\[ Q(Z) = 2 + (Z - 1)g(Z). \]  
(1.4)

The total flow from \( S_i \) to \( S_j \) can be written as the summation of the flows \( C_i(Z) \) which pass through state \( i \) to state \( S_j \) for \( i = 0, 1, \ldots, k \). From flow graph considerations one can write the recursion
\[ C_i(Z) = C_{i-1}(Z) = Z + P(S = i), \]
\[ i = 1, 2, \ldots, k. \]  
(1.5)

Computing \( C_0(Z) \) from Mason's gain rule,
\[ C_0(Z) = \frac{1}{Q(Z)} \left( P(S = 0) - \sum_{i = 1}^{k} P(S = i) \right) \left( 1 - \sum_{i = 1}^{k} P(X > i) Z^i \right) \]
\[ = \frac{P(S = 0)}{Q(Z)} \left( 1 - \sum_{i = 1}^{k} P(X > i) Z^i \right) \]
\[ = \frac{P(S = 0)}{Q(Z)} \left( 2 - g(Z) \right) \]
\[ = \frac{P(S = 0) M_0(Z)}{Q(Z)}. \]

Using (1.5), one can prove by induction that
\[ C_i(Z) = \frac{P(S = 0) M_i(Z)}{Q(Z)}, \quad i = 1, 2, \ldots, k. \]  
(1.6)

Now we are ready to obtain an explicit expression for the spectral density of \((d, k)\) codes. From (11), (1.1), (1.3), and (1.6) we obtain
\[ S_d(f) = -1 + \left( \sum_{i = 0}^{k} M_i(Z)/Q(Z) \right) \left( \sum_{i = 0}^{k} M_i(1/Z)/Q(1/Z) \right) P(S = 0). \]  
(1.7)

Let us compute the first summation in (1.7):
\[ \sum_{i = 0}^{k} M_i(Z) - \sum_{i = 0}^{k} P(X > i) \left( 2 \sum_{j = 0}^{i} (1 - Z)^{-1} \right) g(Z). \]

After some simple algebra this becomes
\[ \sum_{i = 0}^{k} M_i(Z) = \sum_{i = 0}^{k} \left( \frac{i + 1}{1 - Z} - \frac{Z(1 - Z)^{i+1}}{(1 - Z)^2} \right) \{- (i + 1)g(Z) \} P(X = i + 1). \]

Using (1.4), we convert this to
\[ \sum_{i = 0}^{k} M_i(Z) = \frac{1}{1 - Z} \sum_{i = 0}^{k} \left( 2(i + 1) - 2Z \frac{1 - Z^{i+1}}{1 - Z} \right) \right) P(X = i + 1), \]

which can also be written as
\[ \sum_{i = 0}^{k} M_i(Z) = (1 - Z)^{-1} \left[ 2Z(Q(Z) - 2)[1 - Z]^{-1} + LQ(Z) \right]. \]

Substituting this expression into (1.7), we obtain
\[ S_d(f) = -1 + \left[ \frac{2Z(Q(Z) - 2)}{(1 - Z)^2 Q(Z)} \right] + \frac{2ZQ(1/Z) - 2}{(1 - Z)^2 Q(1/Z) + Z - 1} \left\{ \frac{Q(Z)Q(1/Z) - Q(1/Z) - Q(1/Z)}{Q(Z)Q(1/Z)} \right\} \frac{1}{L - \exp(2k/\pi f)} \]

or
\[ S_d(f) = \frac{4Z}{L(1 - Z)^2} \frac{Q(Z)Q(1/Z) - Q(1/Z) - Q(1/Z)}{(1 - Z)^2 Q(1/Z) + Z - 1} \frac{1}{L - \exp(2k/\pi f)} \]

Substituting (1.2) into the previous expression yields (12).

The channel symbols are a series of pulses of duration \( T \) with alternating polarity. We make \( y(t) \) a stationary process by adding a random phase which is uniformly distributed over the interval [0, \( T \)]. Therefore, the power spectrum of \( y(t) \) is obtained by multiplying \( S_d(f) \) by the Fourier transform of the autocorrelation of a pulse divided by the time duration of a pulse. This yields (13).

**APPENDIX II**

**PROOF THAT THE EIGENVALUES OF ZF(−1) LIE INSIDE THE UNIT CIRCLE FOR k > d**

Let \( k \) be an eigenvalue of \( ZF(−1) \), and let \( k > d \). Then \( k \) is a root of the characteristic polynomial of \( ZF(−1) \) given by
\[ P(\lambda) = \lambda^{k+1} + \sum_{i = 0}^{k} Z^{i+1} P(x = i + 1) \lambda^{-i}. \]

Define \( x = Z/\lambda \). Since \( Z = \exp(-j2\pi f/\lambda) \), \( \lambda \) lies within the unit circle if and only if \( x \) lies outside the unit circle. Thus it is sufficient to prove that for \( k > d \), the roots of the equation
\[ \sum_{i = 0}^{k} x^{i+1} 2^{-i(k+1)} = -1 \]

cannot lie within the unit circle.
We first show that any $x$ that satisfies (II.1) must have the property that $|x| \geq 1$. The proof follows by assuming that $|x| < 1$ and proving a contradiction. Assume that $|x| < 1$. Then

$$\left| \sum_{i=d}^{k} x^{i+1} 2^{-((i+1)C)} \right| \leq \sum_{i=d}^{k} |x|^{i+1} 2^{-((i+1)C)} < \sum_{i=d}^{k} 2^{-((i+1)C)} = 1$$

where we have used the fact that $\sum_{i=d}^{k} 2^{-((i+1)C)} = 1$. Since $\left| \sum_{i=d}^{k} x^{i+1} 2^{-((i+1)C)} \right| < 1$, $x$ cannot satisfy (II.1) so we have our contradiction. Now assume that $|x| = 1$. Then since $\sum_{i=d}^{k} 2^{-((i+1)C)} = 1$, $x^{i+1} = -1$ for $i = d, d+1, \ldots, k$. Since $x = x^{d+1} / x^{d+1}$, this also leads to a contradiction.

REFERENCES