On Synchronous Variable Length Coding for Discrete Noiseless Channels

P. A. Franaszek

IBM Watson Research Center, Yorktown Heights, New York

A method is presented for finding the shortest variable length codes with a given bit per symbol ratio for discrete noiseless channels. The central feature of the procedure is a dynamic programming algorithm for determining the optimal code paths. Bounds are found for the channel capacity.

LIST OF SYMBOLS

$\{s_i\} = S$ Set of channel states
$\{\alpha_k\}$ Set of channel symbols
$D$ Skeleton state transition matrix
$C$ Channel capacity
$N$ Basic word length
$\alpha/N$ Bit per symbol ratio
$NM$ Maximum word length
$\psi(s_i, S^*)$ Number of bits which can be transmitted with words starting at state $s_i$ and terminating in states $s_j \in S^*$.
$A$ State transition matrix
$P$ Code path

I. INTRODUCTION

In many digital transmission and recording systems, considerations such as spectral shaping, self-timing and limitations on intersymbol interference require that, before modulation, the data be mapped onto a sequence with special properties. Such suitable sequences often define discrete noiseless channels of the type considered by Shannon (1948), in which restrictions are represented by finite state sequential machines. In practice, the encoding of the binary data onto the channel sequence must
usually be synchronous. By this is meant that the bit per symbol ratio is constant over each word, thus eliminating the need for buffers.

Examples of such codes, designed for specific applications, were presented by Sipress (1965) and Gabor (1967). Properties of synchronous codes were studied by Franaszek (1968). In this paper, a construction procedure for the shortest variable (or fixed) length codes with a given bit per symbol ratio is described. The central feature of the method is a dynamic programming algorithm for finding the optimal code paths. An example is included to illustrate the method and to demonstrate the advantage of varying the word length. In addition, bounds are found for the channel capacity.

II. THE CHANNEL AND ITS CAPACITY

Let \( S \) denote the set of channel states. To each \( \sigma_i \in S \), there corresponds a set of allowable channel symbols \( \{a_k \sigma_i\} \). The transmission of a symbol takes the channel to a new state which is a function of the previous state and the transmitted symbol.

It is convenient to define a channel skeleton transition matrix as follows:

\[
d^n_{ij} \triangleq \text{the number of sequences of } n \text{ channel symbols which take the channel from } \sigma_i \text{ to } \sigma_j.
\]

Let

\[
D \triangleq [d^1_{ij}] = [d_{ij}],
\]

where the brackets denote a matrix. Then, as described in Gill (1962),

\[
d^n_{ij} = [D^n]_{ij}, \quad n = 1, 2, 3 \cdots,
\]

where \([D^n]_{ij}\) denotes the \(ij\)th entry of the \(n\)th power of \(D\).

The capacity is defined as the number of bits per symbol that can be carried by the channel with codes that are arbitrarily long. The capacity of such channels was determined by Shannon (1948) and may be expressed as the logarithm to the base two of the largest real root of the equation

\[\det [d_{ij}Z^{-1} - \delta_{ij}] = 0,\]

where \(\delta_{ij}\) is the Kronecker delta.
SYNCHRONOUS VARIABLE LENGTH CODING

THEOREM. The channel capacity $C$ is bounded by

$$\frac{\log_2 \left[ \sum_{ij} d_{ij}^n \right] - 2 \log_2 R}{n + R - 1} \leq C \leq \frac{\log_2 \left[ \sum_{ij} d_{ij}^n \right]}{n},$$

(4)

for $n \geq 1$. The $R$ is the number of channel states. It is assumed that the channel states are strongly connected.

Proof. Counting all possible sequences with no constraints on initial or final states, it is clear that

$$C \leq \log_2 \left[ \sum_{ij} d_{ij}^n \right] \quad n = 1, 2, \ldots$$

(5)

Similarly

$$C \geq \log_2 \left[ d_{ii}^n \right] \quad n = 1, 2, \ldots \quad i = 1, 2, \ldots, R.$$  

(6)

The number of states in $R$. This implies that

$$\sum_{i=0}^{R-1} d_{ij}^{M-j} \geq \sum_{k=1}^{R} d_{ik}^{M-(R-1)}, \quad M \geq R,$$

(7)

as it is possible to reach $\sigma_i$ from $\sigma_k$ with at most $(R - 1)$ symbols. Since

$$\log_2 \left[ \frac{1}{R} \sum_{i=0}^{R-1} d_{ij}^{M-j} \right] \leq C,$$

(8)

$$\log_2 \left[ \frac{1}{R} \sum_{k=1}^{R} d_{ik}^{M-(R-1)} \right] \leq C,$$

(9)

for $i = 1, 2 \ldots R$. Hence

$$\log_2 \left[ \frac{1}{R^2} \sum_{i,k=1}^{R} d_{ik}^{M-(R-1)} \right] \leq C,$$

(10)

substituting $n = M - R + 1$

$$\frac{\sum_{i,k=1}^{R} d_{ik}^n}{n + R - 1} - 2 \log_2 R \leq C \leq \frac{\sum_{i,k=1}^{R} d_{ik}^n}{n},$$

(11)

which is the desired result.
Equation (4) may at times be a more convenient method for calculating the capacity than Eq. (3), especially if the number of channel states is large. It should be noted that if \( n \) is a power of two, then \( D^n \) may be obtained by \( \log_2(n) \) matrix multiplications.

### III. DISCUSSION OF SYNCHRONOUS CODES

Although it is known that codes exist with bit per symbol ratios arbitrarily close to \( C \), it is in practice necessary to determine how well one can do with rather short block lengths. This and the following sections are addressed to the problems of the existence and construction of the shortest codes with a given bit per symbol ratio.

It is convenient to provide some background from Franaszek (1968). The formulation has been revised for greater ease in deriving the new results.

Codes may be of either fixed or of variable length. For the variable length case, the requirement that the code be synchronous, coupled with the assumption that each word carries an integer number of information bits, implies that the word lengths are integer multiples of a *basic word length* \( N \), where \( N \) is the smallest integer for which the bit per symbol ratio is that of two integers. For example, if \( \alpha/N = 1.5 \), the basic word length is 2.

A necessary and sufficient condition for the existence of a code is the existence of a *principal state set* \( S_p \), with the property that from each \( \sigma_i \in S_p \) there exist a sufficient number of distinct paths to other principal states to maintain the information rate. If the words are of length \( gN \), \( g = 1, 2, \ldots, M \), then the number of words available for coding from a given state \( \sigma_i \) must satisfy the inequality

\[
\psi(\sigma_i, S_p) = L_1(\sigma_i) + 2^{-\alpha}L_2(\sigma_i) + \cdots + 2^{-\alpha[M-1]}L_M(\sigma_i) \geq 2^\alpha, \quad (12)
\]

where \( L_g(\sigma_i) \) is the number of words of length \( gN \) allowed from \( \sigma_i \) which terminate in states \( \sigma_q \in S_p \), and \( \alpha \) is the number of bits carried by a word of length \( N \).

The existence of a principal state set and hence, the existence of a code, may be determined by the following method: Let \( S^* \) be the set of states which have not been eliminated and \( \sigma_i \in S^* \) a state to be tested. If \( \psi(\sigma_i, S^*) \leq 2^\alpha, \sigma_i \) is eliminated from \( S^* \).

Starting with \( S^* = S \), the procedure is continued until either all states have been eliminated, or the routine goes through a complete cycle of
remaining states without further elimination. In the latter case, the remaining states are the principal states.

For the fixed length case, \( \psi(\sigma_i, S^*) \) is particularly easy to determine:

\[
\psi(\sigma_i, S^*) = \sum_{j \in S^*} d_{ij}^N,
\]

where \( N \) is the code word length.

If codes of variable length are permitted, the construction of a code involves the problem of discovering whether a set of paths exist which satisfy Eq. (12). In the following section, this problem is solved by finding an optimal path set.

IV. CODE PATH OPTIMIZATION

This section presents a method for finding a set of code paths with the following properties:

1. The paths are of lengths \( jN, j = 1, 2, \ldots, M \).
2. All paths start from \( \sigma_i \) and terminate in states \( \sigma_k \in S^* \).
3. States in which paths of length \( gN \) terminate are not entered with \( gN \) symbols with paths of longer length. This is equivalent to imposing the prefix property on words starting from a given state.
4. The paths maximize \( \psi(\sigma_i, S^*) \) while minimizing the sum of the code path lengths

\[
W(\sigma_i, S^*) = \sum_{i=1}^{M} gNL_0(\sigma_i).
\]

Let \( S^*[n, \sigma_i] \) be the subset of states belonging to \( S^* \) which are reachable from \( \sigma_i \) in \( n \) steps. This is the set of states for which \( d_{ij}^n \neq 0 \) with \( \sigma_i \in S^* \).

Consider a path \( P \) of length \( (M - 1)N \) leading from \( \sigma_i \) to \( \sigma_k \in S^*[(M - 1)N, \sigma_i] \). If \( P \) is terminated at \( \sigma_k \), then its contribution to \( \psi(\sigma_i, S^*) \) is \( 2^{-\alpha(M-1)} \). If \( P \) is continued beyond \( \sigma_k \), then the maximum contribution to \( \psi(\sigma_i, S^*) \) is

\[
\Delta\psi(\sigma_i, S^*) = 2^{-\alpha(M-1)} \sum_{\sigma_j \in S^*[(M-1)N, \sigma_i]} d_{kj}^N.
\]

Clearly, it does not pay to continue \( P \) beyond \( \sigma_k \) if

\[
\sum_{\sigma_j \in S^*[(M-1)N, \sigma_i]} d_{kj}^N \leq 2\alpha.
\]

Termination of \( P \) at \( \sigma_k \) when the above condition holds with an equality that insures that \( W(\sigma_i, S^*) \) is minimized for the maximum \( \psi(\sigma_i, S^*) \).
The set of states which are members of \( S^*[(M - 1)N, \sigma_i] \) for which Eq. (16) holds will be denoted by \( T[(M - 1)N, \sigma_i] \).

Similarly, one may define terminal state sets \( T[(M - V)N, \sigma_i] \subset S^*[(M - V)N, \sigma_i], V = 1, 2, \ldots, M - 1 \), at which it is desirable to terminate paths of length \((M - V)N\). The sets \( T[(M - V)N, \sigma_i] \) may be recursively computed from \( T[(M - V + j)N, \sigma_i], j = 1, 2, \ldots, V - 1 \). That is, a state is tested for inclusion in a terminal state by forming the optimal paths beyond that point. A state \( \sigma_j \in S^*[(M - V)N, \sigma_i] \) is a member of \( T[(M - V)N, \sigma_i] \) if

\[
\left( \sum_{\sigma_k \in T[(M - V + 1)N, \sigma_i]} d_{jk}^N \right) + 2^{-\alpha} \left( \sum_{\sigma_k \in T[(M - V + 1)N, \sigma_i]} \sum_{\sigma_m \in T[(M - V + 2)N, \sigma_i]} d_{jk}^N \right) \cdot d_{lm}^N \\
\vdots \\
+ 2^{-\alpha[V-1]} \left( \sum_{\sigma_k \in T[(M - V + 1)N, \sigma_i]} \cdots \right) \cdot d_{jk}^N \cdots d_{qr}^N \leq 2^\alpha.
\]

The maximum of \( \psi(\sigma_i, S^*) \) is given by

\[
\max \psi(\sigma_i, S^*) = \left( \sum_{\sigma_h \in T[(M, \sigma_i)]} d_{ih}^N \right) + 2^{-\alpha} \left( \sum_{\sigma_h \in T[(M, \sigma_i)]} \sum_{\sigma_m \in T[(2M, \sigma_i)]} d_{ih}^N \right) \cdot d_{em}^N \\
\vdots \\
+ 2^{-\alpha[M-1]} \left( \sum_{\sigma_h \in T[(M, \sigma_i)]} \cdots \right) \cdot \sum_{\sigma_r \in T[(M, \sigma_i)]} d_{ih}^N \cdots d_{qr}^N.
\]

The code paths corresponding to \( \max \psi(\sigma_i, S^*) \) may be obtained from the state transition matrix \( A \),

\[
A = \{a_{ij}\},
\]

where \( a_{ij} \) is the disjunction (+) of channel outputs corresponding to a transition from \( \sigma_i \) to \( \sigma_j \). Powers of \( A \) are formed by combining the above (+) operation with concatenation. The concatenation of a sequence of symbols and the null symbol (\( \emptyset \)) results in the null symbol. Let \( P_r(\sigma_i) \)}
be the set of words of length $rN$ available from state $z$. Then,

$$P_1(z) = \sum_{q \in T[N,z]} a_{iq}^N$$

$$P_2(z) = \sum_{q \in T[N,z]} a_{iq}^N a_{qj}^N$$

$$\vdots$$

$$P_M(z) = \sum_{q \in T[N,z]} a_{iq}^N a_{qj}^N \cdots a_{mn}^N$$

where

$$a_{iq}^N = [A^N]_{iq}.$$

V. EXAMPLE

Consider a digital magnetic recorder in which intersymbol interference is controlled by requiring that each transition between saturation levels be separated by at least two baud intervals. The state-transition diagram for this constrained sequence is shown in Fig. 1. The $D$ matrix is given by

$$D = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}. \quad (21)$$

The capacity of the channel is

$$C \approx .551 \text{ bits/digit} \quad (22)$$

Fig. 1
It appears reasonable to assume that fairly short codes exist which carry one bit for every two channel symbols. Setting $M = 1$ in the optimization procedure, one may show that the shortest fixed length code with this efficiency has a channel block length of 14.

$$D^{14} = \begin{bmatrix} 41 & 28 & 60 \\ 60 & 41 & 88 \\ 88 & 60 & 129 \end{bmatrix}. \quad (23)$$

All three states are principal states. The alphabet size required for coding from each state is $2^7 = 128$ words.

$$\sum_{j=1}^{3} d_{1j}^{14} = 129 \quad (24)$$
$$\sum_{j=1}^{3} d_{2j}^{14} = 189 \quad (25)$$
$$\sum_{j=1}^{3} d_{3j}^{14} = 277. \quad (26)$$

It is of interest to see how well one can do with a variable block length code. The bit per symbol ratio is $1/2$, so that $N = 2$. It will be shown that a variable length code with a maximum word length of four exists which has the above ratio of bits per symbol.

The quantities $\alpha$, $M$ and $N$ are given by

$$\alpha = 1 \quad (27)$$
$$M = 2 \quad (28)$$
$$N = 2. \quad (29)$$

From

$$D^4 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix}, \quad (30)$$

it can be seen that, starting with $S^* = S$,

$$S[4, \sigma_1] = S[4, \sigma_2] = S[4, \sigma_3] = (\sigma_1, \sigma_2, \sigma_3) = S. \quad (31)$$
SYNCHRONOUS VARIABLE LENGTH CODING

Starting with \( \sigma_1 \), and \( S^* = S \), the quantity \( \psi(\sigma_1, S^*) \) will be maximized.

\[
D^2 = \begin{bmatrix}
0 & 0 & 1 \\
1 & 0 & 1 \\
1 & 1 & 1
\end{bmatrix},
\]

so that

\[
S^*[2, \sigma_1] = \sigma_3
\]

moreover

\[
\sum_{j=1}^{3} d_{3j}^2 = 3 > 2;
\]

so that \( T[2, \sigma_1] \) is empty. Hence

\[
\max \psi(\sigma_1, S^*) = 2^{-1} \sum_{j=1}^{3} d_{1j}^2 = \frac{3}{2} < 2,
\]

so that \( \sigma_1 \) is eliminated from \( S^* \). Consider state \( \sigma_2 \).

\[
S^*[4, \sigma_2] = (\sigma_2, \sigma_3),
\]

and

\[
S^*[2, \sigma_2] = \sigma_3
\]

\( T[2, \sigma_2] \) consists of \( \sigma_3 \), since

\[
2^{-\alpha} \sum_{\sigma_j \in S^*} d_{3j}^2 = 1.
\]

The maximum of \( \psi(\sigma_2, S^*) \) is given by

\[
d_{23}^2 + 2^{-1} \sum_{j \neq 3} \sum_{k \neq 1} d_{2j}^2 = \frac{3}{2} < 2,
\]

so that \( \sigma_2 \) is eliminated from \( S^* \), which now contains only \( \sigma_3 \).

\[
2^{-\alpha} d_{33}^2 = \frac{1}{2},
\]

so that

\[
\sigma_3 \in T[2, \sigma_3].
\]

The maximum of \( \psi(\sigma_3, S^*) \) is given by

\[
d_{33}^2 + 2^{-1} \sum_{j \neq 3} d_{3j}^2 d_{j3}^2 = 2,
\]
so that $\sigma_3$ is not eliminated from $S^*$. Moreover, $\sigma_3$ is the only remaining state. Hence, it must be a principal state. The available code alphabet is given below:

<table>
<thead>
<tr>
<th>Word 1.</th>
<th>0 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word 2.</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Word 3.</td>
<td>1 0 0 0</td>
</tr>
</tbody>
</table>

A code may be constructed as follows:

<table>
<thead>
<tr>
<th>Binary Block</th>
<th>Code Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 0</td>
</tr>
<tr>
<td>1 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>1 1</td>
<td>1 0 0 0</td>
</tr>
</tbody>
</table>

The coder could operate by observing whether the received bit was either a one or a zero. In the latter case, 0 0 would be transmitted. Otherwise, the coder would transmit a 0 1 0 0 or a 1 0 0 0 depending on whether the next symbol was a one or a zero. The scheme may be implemented by including two-symbol delays in coding and decoding.

The above example illustrates the advantage of variable length block codes. It shows how a code with a 128 word dictionary may be replaced by one with only three words. Thus, the extra effort involved in considering variable-length codes is at times worthwhile.

VI. CONCLUSION

A method has been presented for the construction of the shortest variable length synchronous codes for discrete noiseless channels. The optimal paths are found by a dynamic programming algorithm. It was shown that varying the word length at times results in a significant decrease in code complexity. Bounds were found for the channel capacity.

Received: April 7, 1969

REFERENCES