P. A. Franaszek

On Future-Dependent Block Coding for Input-Restricted Channels

Consider a restricted channel whose constraints may be characterized by a finite state machine model. Conventional coding techniques for such channels result in codes where the choice of a word to be transmitted is only a function of the current state and the information to be represented by this word. This paper develops techniques for constructing codes where the code word choice may also depend on future information to be transmitted. It is shown that such future-dependent codes exist for channels and coding rates where no conventional code may be constructed.

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 [1] in which restrictions are represented by finite-state sequential machines. In practice, a common additional requirement is that the encoding of the binary data onto the channel sequence must be synchronous. By this is meant that the bit per symbol ratio (the coding rate) is constant over each word.

An extensive body of literature exists on such codes; references [2, 3] are survey papers. Although there have been techniques proposed for creating a direct mapping between data and code words [4-7], the majority of channel codes encountered in practice employ table look-up techniques; construction of such codes requires finding a set of code words or paths which correspond to state transitions in the model for the channel constraints. Information to be coded is then associated with these paths [8, 9].

The requirement that the code be synchronous implies that there exists a *basic code word length* W [9] of which all word lengths in a given code are multiples. A code word of length MW then represents M letters drawn from an alphabet B where $\alpha = \log_2 |B|$. That is, the basic word length is associated with α bits (for simplicity, only transmission of binary information is considered), and the coding rate is α/W. Two classes of codes described in earlier papers are as follows:

1. Fixed length (FL) codes [9]. Here all code words are of the same length W, each representing one of the alphabet of 2^α letters. The word used to represent a letter may be a function of the state occupied by the channel at the beginning of a code word.
2. Variable length (VL) codes [9-11]. Here code word lengths are integer multiples of the basic word length W. The code, as in the fixed length case, may be state dependent.

In general, the size of the coding table grows exponentially with the word length. Error propagation tends to be proportional to this length. Thus it is desirable to obtain, for a given rate, α/W, the shortest possible code. The generation of a set of paths for an optimal (shortest) fixed length code may be done by a simple recursive procedure [9]. In the case of a variable length code, the generation of a set of paths may be done by a recursive procedure each step of which involves the solution of a dynamic programming problem [10]. An interesting property of variable length codes is that the maximal word length for the shortest existing code with a given rate (bit per symbol ratio) may be considerably smaller than the word length required for a fixed length code [10].

Consider the above coding process. Information sequences of length M are mapped onto code words of length MW channel symbols, as a function of the channel state occupied at the beginning of a code word. Another way to look at this process is that, for each state entered at the end of a code word, there exists an M for which a code word may be found to map M information letters onto MW channel symbols. An optimal VL code minimizes the maximum value of M. But suppose no such M exists. Is coding then not possible? It is shown below that in fact there are cases where such "infinite" codes may be constructed by incorporating future dependency or look-ahead into the coding process. That is, coding of the next N letters may always require knowledge of $N + M$. Such codes will be termed future-dependent (FD).

FL and VL codes may be regarded as special cases of FD codes. As an example, consider a variable length code with words of length W and $2W$. One way of looking at the coding process for this code is that it is done one letter at a time, but the path (W channel symbols) chosen may depend on the following or previous letter as well as the current and preceding channel state. An FL code of length $2W$ may be viewed in the same manner. Examples of specific codes incorporating future dependency may be found in the literature [2, 12], but these are often equivalent to some FL or VL code with different assignments of letters to coding paths.

This paper considers a special class of FD codes. Let $\{\sigma_i\}$, $i = 1, 2, \cdots, N$, denote the states for the channel constraint model. Let L_j, $j = 1, 2, \cdots$, be the sequence of letters to be transmitted. Each L_j is drawn from an alphabet containing 2^α distinct letters.

Definition: An FD(KW, K, Q) code of rate α/W is a mapping

$$\mathcal{M}\colon (\sigma_i;\, L_{rK+1}, L_{rK+2}, \cdots, L_{(r+1)K};$$
$$L_{(r+1)K+1}, \cdots, L_{(r+1)K+Q}) \rightarrow \omega, \qquad r = 1, 2, \cdots$$

defined over a set of *terminal* states $\{\sigma_i\}$. The code word ω is a sequence of KW symbols which take the channel from σ_i to some other terminal state σ_j. The mapping is such that $L_{rK+1}, \cdots, L_{(r+1)K}$ may be recovered, given σ_i and ω (i.e., the code is instantaneously decodable). For simplicity, the codes considered will be restricted to the special case where letters used for coding from each terminal state σ_i are independent of the path by which this state is entered. This may be contrasted with the general case where the continuation of a particular path entering σ_i could be restricted to a subset of the mapping $\mathcal{M}$.

The following is a synopsis of the paper. Section 2 reviews methods for generating paths between channel states. Section 3 develops techniques for generating FD (W, 1, 1) codes. Equations are presented the solutions of which are a necessary condition for the existence of a code. The solution of these, coupled with techniques for assigning letters to the available paths, produces a code. Section 4 gives some examples. It is shown that an FD(W, 1, 1) code may be constructed for a set of channel constraints and coding rate which admit no conventional code of finite length. Section 5 considers the case of FD(QW, Q, K) codes, where $Q \neq K$. It is shown that whenever such a code may be constructed, it is possible to form an FD(W', 1, 1) code, where $W' = QSW$ and S is the ceiling of K/Q. Section 6 summarizes the results.

2. Coding paths and channel constraints

Let $S = \{\sigma_i\}$, $i = 1, 2, \cdots, N$, denote the states for the channel constraint model. To each $\sigma_i \in S$, there corresponds a set of allowable channel symbols $\{V_k\}_i$. 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:

$$\mathbf{D} = \{d_{ij}\}, \tag{1}$$

where $d_{ij} \triangleq$ the number of symbols which take the channel from σ_i to σ_j.

Let $d^n_{ij} \triangleq$ the number of paths of length n symbols which take the channel from σ_i to σ_j. Then

$$d^n_{ij} = [\mathbf{D}]^n_{ij}. \tag{2}$$

The channel capacity, defined as the maximum bit per symbol rate permitted by the channel constraints, is given [1] by the base-two logarithm of the largest real root of

$$\det\,[d_{ij}Z^{-1} - \delta_{ij}] = 0. \tag{3}$$

Symbols associated with paths of length W may be obtained from the Wth power of the channel transition matrix

$$\mathbf{A} = \{a_{ij}\}, \tag{4}$$

where a_{ij} represents the disjunction, +, of sequence symbols which take the channel from σ_i to σ_j. If no such symbol exists, then $a_{ij} = \varnothing$. Powers of $\mathbf{A}$ are formed by the operations of disjunction, +, and concatenation. The concatenation of $\varnothing$ with any symbol results in $\varnothing$.

3. FD(W, 1, 1) codes

As mentioned above, an FD(KW, K, K) code maps K letters on channel sequences whose basic word length is KW, with the knowledge, at the time that a code word is chosen, of the current as well as the next K letters to be transmitted. It is easy to see that this is equivalent to an FD(W', 1, 1), where $W' = KW$.

Let L_t, $t = 1, 2, \cdots$, be the letters encoded at time t, $t = 1, 2, \cdots$. Each letter L is chosen from an alphabet B containing 2^α. The channel states occupied at the beginning of each code word are denoted by $\sigma_{k(i)}$, $i = 1, 2, \cdots$. The ith code word transmitted will be represented by W_i.

Definition: The *terminal* or *coding* states for a code are those entered at the end of code words.

Consider the transmission of L_1. The path chosen, W_1, takes the channel from state $\sigma_{k(1)}$ to $\sigma_{k(2)}$. At $t = 1$, the time of transmission, L_2 is known, but L_3 is not. Thus $\sigma_{k(2)}$ must be such that L_2 in particular and any L_3 can be transmitted.

Definition: An *independent path set* of length NW (termed an independent path for short) is a path or set of paths which can be used to represent N symbols.

It can easily be seen that each terminal state must have associated with it an integral number of independent paths of length W. In the above example, L_2 is known at $t = 1$, but L_3 is not. Thus, given that the channel resides in $\sigma_{k(1)}$ at $t = 1$, only one state $\sigma_{k(2)}$ can be associated with the transmission of L_2.

Definition: The *weight* Φ_i of a state σ_i is the number n_i of independent paths of length W from σ_i times $2^{-\alpha}$.

Let $\{P^r_{ij}\}$, $P^r_{ij} = 1$, be associated with the distinct paths of length W which lead from σ_i to σ_j. That is, for each path of length W from σ_i to σ_j, there is a P^r_{ij}. Let $\{C^r_{ij}\}$, $C^r_{ij} =$ 0 or 1, be a set of choices associated with the distinct paths. That is, if a given path is included in the code, then its associated C^r_{ij} is 1.

Proposition 1

Suppose an FD(W, 1, 1) code can be constructed for a given rate α/W. Let Φ_i be the weight of a terminal state σ_i after the paths not used in coding have been eliminated. Then,

$$\Phi_i = \sum_{r,i,j} C^r_{ij} P^r_{ij} \Phi_j 2^{-\alpha}. \qquad (5)$$

Proof

The number of independent paths of length $2W$ leading from σ_i is

$$2^{2\alpha}\Phi_i = \sum_{r,i,j} C^r_{ij} P^r_{ij} n_j. \qquad (6)$$

But this is just enough to transmit any of

$$n_i = \sum_{r,i,j} C^r_{ij} P^r_{ij} n_j 2^{-\alpha} \qquad (7)$$

letters from σ_i, followed by any other letter. □

Proposition 2

A necessary condition for the existence of an FD(W, 1, 1) code is that there exists a set of choices $\{C^r_{ij}\}$ for which Eqs. (5) have a set of solutions such that

1. $\Phi_i \le 1$

and

2. the Φ_j are integer multiples of $2^{-\alpha}$.

Proof

Conditions 1 and 2 follow from Proposition 1. Weight $\Phi_i > 1$ means that not all the paths from σ_i can be utilized. The $\{\Phi_j\}$ must be integer multiples of $2^{-\alpha}$ since they represent integer numbers of independent paths. □

Given that an appropriate solution has been found to Eqs. (5), the problem arises of assigning letters to the coding paths. Note that if all nonzero Φ_i are 1, then an FL code can be constructed. Otherwise, two further conditions are necessary:

3. Suppose n_i distinct letters have been assigned for transmission from a terminal state σ_i. Then each coding path from σ_i must be associated with a single letter, since the code is to be instantaneously decodable.
4. The letters assigned to terminal state σ_i are independent of the path used to reach this state.

Proposition 3

The following two conditions ensure the possibility of satisfying 3.

a. No terminal state σ_i has more than a single code word leading to each state σ_i of fractional weight.
b. The terminal states reachable via code words from σ_i may be partitioned into subsets of total partial weight one.

Proof

Each path leading to a state σ_i with $\Phi_i = 1$ may be used to represent a single letter. The set of paths leading to a group of states of combined partial weight 1 may likewise be assigned a single letter. □

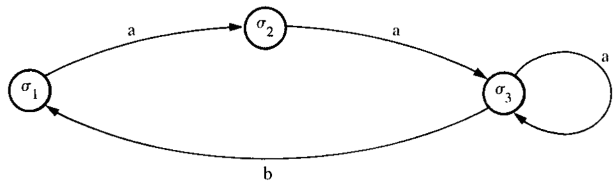

Figure 1 State transition diagram for a set of channel constraints.

Table 1 Letter assignments.

Initial state	*Information*	*Code word*	*New state*
σ_3	0	aa	σ_3
σ_3	(1/0)	ab	σ_1
σ_3	(1/1)	ba	σ_2
σ_2	0	aa	σ_3
σ_1	1	aa	σ_3

Table 2 Optimal variable length code.

Initial state	*Information*	*Code word*	*New state*
σ_3	0	aa	σ_3
σ_3	10	abaa	σ_3
σ_3	11	baaa	σ_3

Condition 4 is related to the problem of state-independent decoding for FL codes [9, 11]. Consider a state σ_i of fractional weight, reachable via paths of length W from a set of terminal states $\{\sigma_j\}$. The state σ_i must be assigned letters which are the same for all states in $\{\sigma_j\}$. This can be attempted by a process of exhaustive search.

4. Examples

Two examples of FD(W, 1, 1) codes are derived in this section. The first is a rate 1/2 code, with $W = 2$, for the constraints illustrated in Fig. 1. These correspond to a digital magnetic recording channel in which intersymbol interference is controlled by requiring that each transition between saturation levels be separated by at least two baud intervals [9]. The **D** matrix for this example is given by

$$\mathbf{D} = \begin{bmatrix} 0\,1\,0 \\ 0\,0\,1 \\ 1\,0\,1 \end{bmatrix}. \tag{8}$$

The capacity is approximately 0.55 bits per channel symbol.

Consider the set of inequalities

$$\Phi_1 \le \frac{1}{2}(P_{13}\Phi_3),$$

$$\Phi_2 \le \frac{1}{2}(P_{23}\Phi_3 + P_{21}\Phi_1),$$

$$\Phi_3 \le \frac{1}{2}(P_{31}\Phi_1 + P_{32}\Phi_2 + P_{33}\Phi_3). \tag{9}$$

Choosing $P^*_{21} = 0$, $P^*_{ij} = 1$ otherwise, where $P^*_{ij} = C_{ij}P_{ij}$, one obtains

$$\Phi_1 = \frac{1}{2}(\Phi_3),$$

$$\Phi_2 = \frac{1}{2}(\Phi_3),$$

$$\Phi_3 = \frac{1}{2}(\Phi_1 + \Phi_2 + \Phi_3), \tag{10}$$

which is satisfied by $\Phi_1 = \Phi_2 = 1/2$, $\Phi_3 = 1$. In other words, σ_1 and σ_2 each have a single independent path of length 2, while Φ_3 has two. Note that the paths satisfy the conditions stipulated in Proposition 3.

Let the alphabet A be (0, 1). The two independent paths from σ_3 imply that either 0 or 1 may be transmitted from this state. The other states each have only a single independent path, corresponding to either 0 or 1. One way of assigning letters to the paths is shown in Table 1, where (1/0) indicates that 1 is to be transmitted with the knowledge that 0 is to be the next letter.

It is interesting to note that this is equivalent to an optimal variable length code, derived in reference [7], with a maximum word length of 2, as shown in Table 2.

Figure 2 illustrates another set of channel constraints; for these the channel capacity is one bit per symbol.

Proposition 4

There exist sets of channel constraints for which no fixed or variable length code can be constructed with a rate of α bits per W channel symbols, but which admit an FD(W, 1, 1) code with this rate.

Proof

The channel constraints are those shown in Fig. 2. The existence of an FD(W, 1, 1) code with a rate of one bit per symbol is shown in Table 3. The proof that no fixed or variable length code can be constructed is given in the appendix. □

An FD(1, 1, 1) code will now be constructed. Consider the inequalities

$$\Phi_1 \le \frac{1}{2}(P_{11}\Phi_1 + P_{12}\Phi_2 + P_{13}\Phi_3),$$

$$\Phi_2 \le \frac{1}{2}(P_{21}\Phi_1),$$

$$\Phi_3 \le \frac{1}{2}(P_{32}\Phi_2 + P_{33}\Phi_3). \tag{11}$$

Letting $P^*_{ij} = 1$, the corresponding equations are satisfied by $\Phi_1 = 1$, $\Phi_2 = \Phi_3 = 1/2$. Thus two independent paths of length 1 lead from σ_1, and one each from σ_2 and σ_3.

A code can be constructed as in Table 3.

The coding paths are illustrated by the trellis shown in Fig. 3.

5. Fixed length codes with general future dependency

This section considers the general class of FD(KW, K, MK) codes with $M \neq 1$. It is shown that whenever such a code exists for integer $M > 1$, then an FD(MKW, MK, MK) code exists. Clearly, if a code exists for $M < 1$, then an FD(KW, K, K) code exists. In other words, it is sufficient to restrict attention to FD(W, 1, 1) codes at the expense of possibly increased complexity and coding delay.

Suppose a FD(KW, K, MK) code can be constructed. Coding is done for K letters at a time, with knowledge of the next MK. Then it is certainly possible to code with the knowledge of the next $(2M - j)K$ letters, $j = 1, 2, \cdots, M$. But this would be an FD(MKW, MK, MK) code.

It is interesting to consider what modifications to (5) are required to form a necessary condition for the existence of an FD(KW, K, MK) code. Note that each state σ_j entered at the end of a code word must have sufficient paths leading from it to represent the next (known) MK letters. But there is no knowledge of what is to follow. For example, suppose $L_1, L_2, \cdots, L_K$ are to be transmitted, taking the channel from σ_i to σ_j. At the time this path is chosen $L_{K+1}, L_{K+2}, \cdots, L_{K+MK}$ are known, but L_{K+MK+1}, $L_{K+MK+2}, \cdots$ are not. Thus paths of length MK leading from each terminal state must end in states associated with sufficient paths of length MW to represent any MK letters.

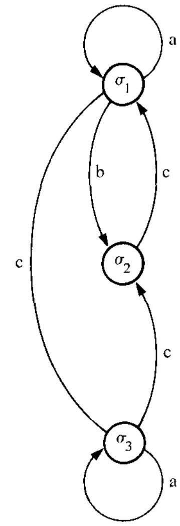

Figure 2 State transition diagram for a set of channel constraints.

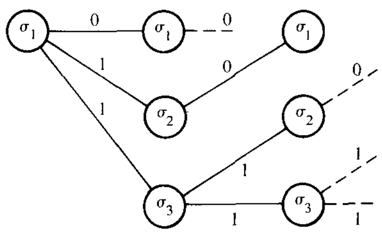

Figure 3 Example of a coding trellis for the constraints of Fig. 2.

Table 3 Constructed code.

Initial state	*Information*	*Code word*	*New state*
σ_1	0	a	σ_1
σ_1	(1/0)	b	σ_2
σ_1	(1/1)	c	σ_3
σ_2	0	c	σ_1
σ_3	(1/0)	c	σ_2
σ_3	(1/1)	a	σ_3

But this is precisely the requirement for the existence of an FD(MKW, MK, MK) code. That is, there must exist a set of chosen C_{ij} such that the equations

$$\Phi_i = \sum C^r_{ij} P^r_{ij} \Phi_j 2^{-MK\alpha} \tag{12}$$

have solutions subject to the requirements of Proposition 2.

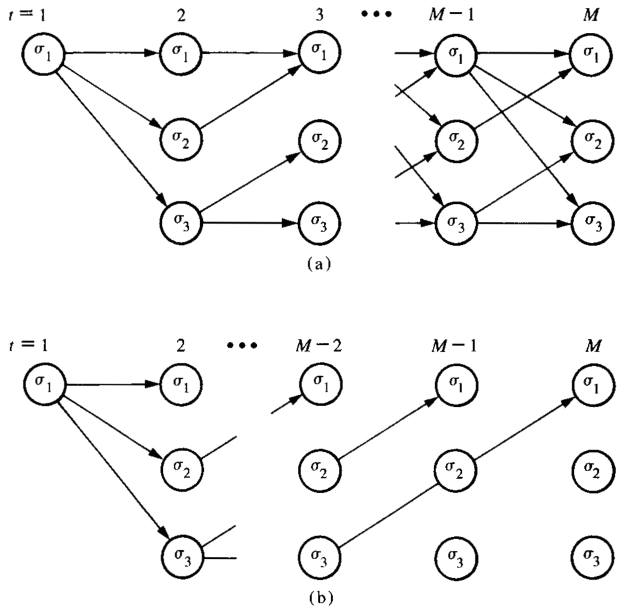

Figure 4 (a) Trellis for the constraints of Fig. 2. (b) Paths from σ_1 to σ_1 which maximize Ψ_1.

However, code words are of length KW. Thus an additional requirement on the eligible paths P_{ij} is that they enter terminal states after each KW channel symbols.

6. Conclusion

Methods were presented for the construction of FD(W, 1, 1) codes, which may be regarded as a generalization of state-dependent block codes for input-restricted channels. It was shown that FD codes exist for constraints which admit no conventional codes of finite length. It was also shown that the existence of an FD(KW, K, MK) code for integer M implies the possibility of constructing an FD(MKW, MK, MK) code.

Appendix: Proof of Proposition 4

It is shown here that no variable length code with rate 1 can be constructed for the constraints shown in Fig. 2. Much of the terminology is drawn from reference [10].

Consider a set of paths $\{W_i^n\}$ leading from σ_i, where $\{W_i^n\}$ is the ith path of length n. Let V_n be the number of distinct paths of length n. A necessary condition for the possibility of coding from σ_i is given by the Kraft inequality

$$\Psi_i(M) = \sum_{n=1}^{M} V_n^{-n\alpha} \geq 1. \tag{A1}$$

Here $\alpha = 1$ and MW is the maximum word length.

Consider a code of maximum word length $MW = M$, ($W = 1$). Assume first that all states are principal (i.e., code words can terminate there). Consider Fig. 4(a).

Note that paths of length M from state σ_1, which enter states σ_2 and σ_3 at depth ($M - 1$), can be terminated in their states without decreasing $\Psi_1(M)$. This is because two paths of length M are equivalent to one of length ($M - 1$) in contributing to the right-hand side of Eq. (13). Truncating paths from σ_1 at depth ($M - 2$) in states σ_2 and σ_3 likewise does not decrease $\Psi_1(M)$, since (for σ_2) one path of length $M - 2$ is worth more than three of length M, and (for σ_3) two of length $M - 1$ are equivalent to one of length M. This process can be carried backward, resulting in a trellis whose paths entering states σ_2 and σ_3 terminate there, and which maximizes $\Psi_1(M)$. Then,

$$\max \Psi_1(M) = \sum_{n=1}^{M-1} 2 \times 2^{-n} + 3 \times 2^{-M}$$

$$= i + \sum_{n=1}^{M+1} 2^{-n} < 2. \tag{A2}$$

But

$$\max (\Psi_2(M)) \leq 2^{-1} \max (\Psi_1(M - 1)) < 1, \tag{A3}$$

so that words cannot terminate in σ_2.

Moreover,

$$\max (\Psi_3(M)) \leq 2^{-1} \max (\Psi_3(M - 1)) + 2^{-2} \max (\Psi_1(M - 2)), \tag{A4}$$

which implies that $\Psi_3(M) \geq 1$ only if $\Psi_3(M - 1) > 1$. But $\max (\Psi_3(1)) \leq 1$. Thus both σ_2 and σ_3 are eliminated as candidates for the coding states.

This means that all coding paths must start and end in state σ_1. Consider Fig. 4. Note that Ψ_1 can be maximized by terminating paths entering σ_1 at depth $M - 1$. Having done so, it pays to terminate paths in σ_2 at depth $M - 2$, since there is one extension of length 1 and one of length 2. This process can be continued backward to depth 2. The result is one potential coding path of length n, $n = 1, 2, \cdots, M$. Thus

$$\Psi_1 \leq \sum_{n=1}^{M} 2^{-n} < 1; \tag{A5}$$

no rate 1 code of length M can be constructed.

References

1. C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.* **47,** 143–157 (1948).
2. H. Kobayashi, "A Survey of Coding Schemes for Transmission or Recording of Digital Data," *IEEE Trans. Commun. Technol.* **COM-19,** 1087–1100 (1971).

3. P. A. Franaszek, "Noiseless Coding for Signal Design," *Proceedings of the Ninth Allerton Conference*, October 1971.
4. W. H. Kautz, "Fibonacci Codes for Synchronization Control," *IEEE Trans. Info. Theory* **IT-11,** 284- 292 (1965)
5. D. T. Tang, "Run-Length Limited Codes," 1969 IEEE International Symposium on Information Theory.
6. D. T. Tang and L. R. Bahl, "Block Codes for a Class of Constrained Noiseless Channels," *Info. Control* **17,** 436-461 (1970).
7. A. M. Patel, "Zero-Modulation Encoding in Magnetic Recording," *IBM J. Res. Develop.* **19,** 366-378 (1975).
8. C. V. Freiman and A. D. Wyner, "Optimum Block Codes for Noiseless Input Restricted Channels," *Info. Control* **7,** 398-415 (1964).
9. P. A. Franaszek, "Sequence-state Coding for Digital Transmission," *Bell Syst. Tech. J.* **47,** 143-157 (1968).
10. P. A. Franaszek, "On Synchronous Variable-Length Coding for Discrete Noiseless Channels," *Info. Control* **15,** 155-164 (1969).
11. P. A. Franaszek, "Sequence-state Methods for Run-length-limited Coding," *IBM J. Res. Develop.* **14,** 376- 383 (1970).
12. D. T. Tang, "Practical Coding Schemes with Run-length Constraints," *Research Report RC-2022*, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 1968.

Received July 24, 1978; revised September 11, 1978

The author is located at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598.