Full Decodable Code-Word Sets*

M. P. SCHÜTZENBERGER† AND R. S. MARCUS‡

Summary– This paper considers further how the decodability condition imposes restrictions on a set of code words. A generating function is defined that describes the composition of the code words. The relation between the generating function and a "full" set of code words is found. This relation shows that the sum of arbitrary probabilities associated with the words of a full set must be one. A full set of code words is one to which no code word can be added and still keep the set decodable. It is also shown that a full set is "completable." For a completable set of code words any string of symbols can be made into a sentence by adding a suitable prefix and a suffix.

Introduction

SEVERAL authors have considered the restrictions that are imposed on the set of code words by the decodability condition.[1-5] (A code-word set is decodable if no string of symbols can be broken up into code words in more than one way.) Most of the results thus far have had to do with the *lengths* of the code words. This paper includes some conclusions relating to the more detailed *composition* of the code words.

It is important to consider the composition of the code words, as well as their lengths, when the symbols are not of the same cost. For example, in the Morse code the dot is shorter in time duration than the dash. The less costly dot, therefore, should be used more frequently for efficiency of information transmission.

In particular, this paper defines a generating function that describes the composition of the code words. The relation between this function and a "full" set of code words is found. A full set of code words is one to which no code words can be added and still keep the set decodable. It is also shown that a full set is "completable." For a completable set of code words, any string of symbols can be made into a sentence by adding a suitable prefix and a suffix.

* Manuscript received by the PGIT, July 25, 1958. This work was supported in part by the U. S. Army (Signal Corps), the U. S. Air Force (Office of Sci. Res., Air Res. and Dev. Com.), and the U. S. Navy (Office of Naval Research).

† Faculté des Sciences de Poitiers, France; formerly with Res. Lab. of Electronics, Mass. Inst. Tech., Cambridge, Mass.

‡ Res. Lab. of Electronics, Mass. Inst. Tech., Cambridge, Mass.

[1] A. A. Sardinas and G. W. Patterson, "A necessary and sufficient condition for unique decomposition of coded messages," 1953 IRE National Convention Record, pt. 8, pp. 104–108.

[2] B. Mandelbrot, "On recurrent noise limiting coding," *Proc. Symp. on Information Networks*, New York, N. Y.; 1954.

[3] B. McMillan, "Two inequalities implied by unique decipherability," IRE Trans. on Information Theory, vol. IT-2, pp. 115–116; December, 1956.

[4] M. P. Schützenberger, "On an application of semi-group methods to some problems in coding," IRE Trans. on Information Theory, vol. IT-2, pp. 47–60; September, 1956.

[5] R. S. Marcus, "Discrete noiseless coding," S. M. thesis, Dept. Elec. Eng., M. I. T., Cambridge, Mass.; January, 1957.

Statement of the Problem

Let us consider an information-carrying channel with D symbols, d_j, $j = 1, \cdots, D$. For any given string of symbols, s, we write $|s|_j \equiv$ the number of occurrences of symbol d_j in s, and $|s| \equiv$ the total number of symbols in s. Thus, $|s| = \sum_j |s|_j$. A *code word*, w_k, is a particular s. The *code-word set*, P_0, is a set of M code words. *Sentences* are strings of words and they form the infinite set $P = \{P_0\}$. It is always understood that the lengths of the code words are bounded. Without this hypothesis, the conclusions are somewhat different.

It is convenient to associate with the set $\{d_j\}$ an arbitrary set of probabilities, p_j ($\sum p_j = 1$, $p_j > 0$, $j = 1, \cdots, D$). Then we write $Pr(s) = \prod_j p_j^{|s|_j}$. We may now define the *generating function of the words*, $\phi_{P_0}(t)$:

$$\phi_{P_0}(t) = \sum_k Pr\,(w_k) t^{|w_k|} = \sum_{i=1}^{n_m} a_i t^i, \tag{1}$$

where

$$a_i = \sum_{|w_k| = i} Pr\,(w_k)$$

$$n_m = \max\{|w_k|\}.$$

Similarly, we define the *generating function of the sentences*, $\Phi_P(t)$:

$$\Phi_P(t) = \sum_{s \in P} \nu(s)\, Pr\,(s) t^{|s|} = \sum_n A_n t^n, \tag{2}$$

where

$$A_n = \sum_{|s| = n} \nu(s)\, Pr\,(s)$$

$\nu(s)$ = number of decompositions of s into words.

A code-word set, P_0, is then uniquely decodable or, let us say, just *decodable* (d), if $\nu(s) = 1$ for all s in P. (Of course, $\nu(s) = 0$, if s is not in P.) P_0 is said to be *full* (F) if no word can be added to P_0 to form a code-word set that is decodable. P_0 is said to be *completable* (C) if any string, s, can be made to fit in P by adding some suitable prefix and suffix. (Symbolically, we write: P_0 is C if $\forall\, s\ \exists\, x$ and y ϶ $xsy \in P$.)[6]

The four theorems that will be presented show that the four following statements are equivalent for decodable code-word sets:

I. P_0 is full.

II. P_0 is completable.

III. $\phi_{P_0}(1) = 1$ for some particular p_j set. (3)

IV. $\phi_{P_0}(1) \equiv 1$ for all p_j sets.

[6] The symbols xsy denote the string x, followed by the string s, followed by the string y. Here x and y may vary for different s's. $\forall$ means for all; $\exists$ means there exists; ϶ means with the property that; $\in$ means belonging to.

THEOREMS

Theorem I: If P_0 is C, then $\phi_{P_0}(1) = 1$.

Method of Proof: Since the sentences are defined recursively, the A_n are given by a difference equation and are the sums of roots to the nth power, as shown in section 1). For P_0 completable, we show that the A_n cannot become vanishingly small, as shown in sections 2)–4). But for P_0 decodable, the A_n cannot become larger than one. Thus the root of minimum modulus, the real root, must be one.

Proof:

1) $$A_n = \sum_{i=1}^{n_m} B_i T_i^{-n}, \tag{4}$$

where T_i are the roots of $\phi_{P_0}(t) = 1$

B_i are constants.

Eq. (4) is true, since A_n is given by

$$A_n = \sum_{i=1}^{n_m} a_i A_{n-i}. \tag{5}$$

The solution of the difference (5) is given by

$$A_n = \sum_{i=1}^{n_m} B_i \rho_i^n, \tag{6}$$

where ρ_i are roots of $\rho^{n_m} - a_1\rho^{n_m-1} - \cdots - a_{n_m} = 0$

B_i are constants.

Letting $T = \rho^{-1}$, we have

$$T^{-n_m} - a_1 T^{1-n_m} - a_2 T^{2-n_m} - \cdots - a_{n_m} = 0$$

$$= 1 - a_1 T - a_2 T^2 - \cdots - a_{n_m} T^{n_m}$$

$$= 1 - \phi_{P_0}(T).$$

This proves (4).

2) If P_0 is C, then the number of symbols in any prefix and suffix that is needed to make s in P is bounded. More specifically, we have

$$|x| + |y| \leq L = 2n_{m-1}. \tag{7}$$

This is obviously true, since if $|x| > n_m$ we could break up x into words and a string x' with the property that $|x'| < n_m$. This x' could serve as a suitable prefix; similarly for y.

3) If P_0 is C, then $\sum_{n=\alpha}^{\alpha+L} A_n > C_1 > 0$; (for any α). (8)

To prove this, let u_i be the $D^\alpha s$ ∋ $|s| = \alpha$. (See Fig. 1.)

Let u_i' be one $xu_iy \in P$ ∋ $|x| + |y| \leq L$. Hence,

$$\alpha \leq |u_i'| \leq \alpha + L.$$

Some of the u_i' may be the same but we can pick a set of distinct u_i', say v_j, with the property that each u_i can be expanded into at least one v_j. Let $u_{i,j}$ be the set of u_i that can be expanded into a given v_j.

Let

$$\sum_{u_i \in u_{i,j}} Pr(u_i) \equiv Pr(u_{i,j}).$$

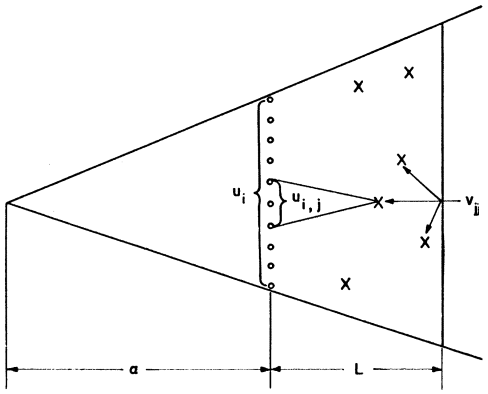

Fig. 1—Abstraction from code-word tree.

Then

$$\sum_j Pr(u_{i,j}) \geq \sum_i Pr(u_i) = 1.$$

Now from each $u_{i,j}$ pick the u_i (call it w_j) with the maximum $Pr(u_i)$. The maximum number of u_i in any $u_{i,j}$ is $|v_j| - \alpha + 1$. An upper bound on this number is $(\alpha + L) - \alpha + 1 = L + 1$. Thus $Pr(w_j) \geq [1/(L+1)]\, Pr(u_{i,j})$. But $Pr(v_j) = Pr(x)\, Pr(w_j)\, Pr(y) \geq p_{\min}^L\, Pr(w_j)$, where $p_{\min} = \min\{p_j\}$. Hence,

$$\sum_{n=\alpha}^{\alpha+L} A_n \geq \sum_j Pr(v_j) \geq p_{\min}^L \sum_j Pr(w_j)$$

$$\geq \frac{p_{\min}^L}{L+1} \sum_j Pr(u_{i,j}) \geq \frac{p_{\min}^L}{L+1} \equiv C_1 > 0$$

This proves (8).

4) Hence, $\lim_{n\to\infty} A_n > C_1/(L+1) \equiv C_2 > 0$ (if the limit exists).

5) Hence, $|T_1| \leq 1$, where $|T_1|$ is the minimum modulus. If $|T_1| > 1$, $A_n \to 0$ as $n \to \infty$.

6) A_n must be bounded. If A_n were not bounded, then $\nu(s)$ would be greater than one for some s because

$$A_n = \sum_{|s|=n} \nu(s)\, Pr(s) \quad \text{and} \quad \sum_{|s|=n} Pr(s) = 1.$$

This would mean that P_0 is not d, contrary to the hypothesis that P_0 is C.

7) Hence, $|T_1| \geq 1$. Otherwise A_n would be unbounded.

8) Since all the coefficients of $\phi_{P_0}(t)$ are positive, $\phi_{P_0}(t)$ is monotonic and $\phi_{P_0}(t) = 1$ has one real root, and no other root has a modulus smaller than this.[7]

9) Hence, $|T_1| = T_1 = 1$.

10) Hence, $\phi_{P_0}(1) = 1$ (and this is true for all p_i sets).

Theorem II: If $\phi_{P_0}(1) = 1$ and P_0 is d, then P_0 is F.

Proof: If we add a word to P_0 to give P_0', then $\phi_{P'_0}(1) > 1$, and T_1', the real root of $\phi_{P'_0}(t) = 1$, is less than one. But by Theorem I, section 6), and Theorem I, section 7), this implies that P_0' is not d. Thus P_0 is F.

[7] The fact that the real root has the minimum modulus follows from Cauchy's theorem. Cf. Morris Marden, "The Geometry of the Zeros of a Polynomial in a Complex Variable," Mathematical Surveys No. III, American Mathematical Society, New York, N. Y.; 1949. See especially Theorem (27.1), p. 95.

Theorem III: If P_0 is d and $\phi_{P_0}(1) = 1$ for a given p_i, then P_0 is C.

Proof: Suppose P_0 is not C. Then $\exists\ s_0 \ni \forall x, y\ xs_0y \notin P$. Since no s with s_0 as a prefix is in P, those strings in that part of the tree that "grows" from s_0 can be eliminated as possible s in P. This means that for $n \geq |s_0|$,

$$A_n \leq 1 - Pr(s_0).$$

Of those strings that do not begin with s_0, we can eliminate that fraction whose second $|s_0|$ symbols are s_0. Thus

$$A_n \leq [1 - Pr(s_0)]^2 \quad \text{for} \quad n \geq 2 \mid s_0 \mid.$$

Similarly,

$$A_n \leq [1 - Pr(s_0)]^m \quad \text{for} \quad n \geq m \mid s_0 \mid.$$

Hence, $A_n \to 0$ as $n \to \infty$. But for given p_i, $T_1 = 1$ and $A_n > C_3 > 0$ for some $n > N$ for any N. Hence, we have a contradiction and P_0 is C.

Theorem IV: If P_0 is F, then P_0 is C.

Method of Proof: Assuming that P_0 is not completable, we consider the string, u, which cannot be completed. If we add u as a word to P_0, we obtain a new set, $\bar{P}_0$, which cannot be decodable. We then show that this implies that u has the same string of symbols in its beginning as at its end, as shown in section 14). But this leads to a contradiction.

Proof:

1) Assume that P_0 is F but not C.

2) Hence, $\exists\ u \ni \forall\ x, y\ xuy \notin P = \{P_0\}$.

3) Consider $\bar{P}_0 = P_0 \cup u$ and $\bar{P} = \{\bar{P}_0\}$.

4) Since $\bar{P}_0$ is not decodable, $\exists\ v$ with two decompositions in $\bar{P}$.

5) Choose v as a minimal doubly decomposable string (minimal d.d. string); that is, a string that cannot remain d.d. if any symbols are removed from its beginning and/or end.

6) Since P_0 is d and $\bar{P}_0$ is not, one of the decompositions of v must contain u as a word. Thus $v = x_1uy_1$, where $x_1, y_1 \in \bar{P}$.

7) Since u is not completable in P, $v \notin P$.

8) But $v \in \bar{P}$.

9) Hence, the second decomposition of v also contains u, *i.e.*, $v = x_2uy_2$.

10) Assume that $|x_1| \leq |x_2|$. If this is not so, reverse designations.

11) $|x_2| \neq |x_1|$. If $|x_2| = |x_1|$, then $x_2 = x_1$, and for v to be d.d. either $x_1 = x_2$ is d.d. or $y_1 = y_2$ is d.d., contrary to the hypothesis that v is a minimal d.d. string.

12) Hence, $|x_1| < |x_2|$.

13) Let us so choose the second decomposition that $|x_2| < |x_1| + |u|$. (See Fig. 2.)

Otherwise, x_2 contains u and must be decomposed as $x_2 = x_3uy_3$ by the same reasoning that led to section 9). Thus we could have chosen to consider the first u as the word u in the second decomposition of v.

14) Thus $u = x_4u_2 = u_2y_4$. (See Fig. 3.)

15) We can find (as we shall show) a u' for which the equation of section 14) cannot be satisfied. Hence, the assumption that P_0 is F, but not C, which leads to this conclusion, is false and the theorem is proved.

16) To find u' we consider two cases that cover all the possibilities.

Case 1): $u = a^{|u|}$.

Case 2): $u = a^kby_5$; $0 < k < |u|$, $0 \leq |y_5|$.

We have arbitrarily called the first symbol in u "a" and the first symbol in u which is not a, if such a symbol exists, "b".

17) For case 1), let $u' = ub = a^{|u|}b$.

Clearly, u' cannot satisfy

$$u' = x_5w = wy_6; \qquad |\ y_6\ | > 0,$$

since w must start with a and end with b.

18) For case 2), let $u' = ub^{|u|}$.

$|w| \leq |u|$ is clearly impossible, for then w would have to start with "a" but consist only of b's.

But if $|w| > |u|$, we can write

$$w = x_6ab^{r+|u|},$$

where $0 \leq |x_6|$; $0 \leq r < |u|$.

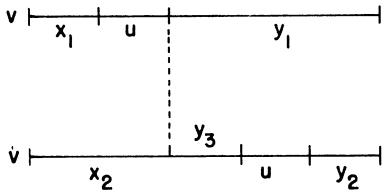

Fig. 2—Grouping of symbols in the string v.

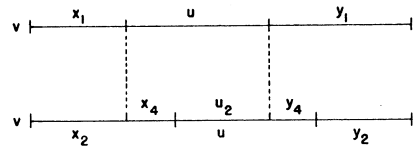

Fig. 3—Grouping of symbols in the string v.

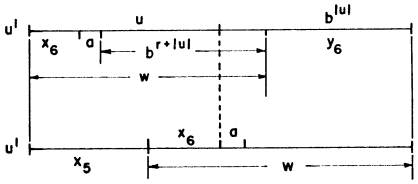

Fig. 4—Grouping of symbols in the string u'.

Then, as is apparent from Fig. 4, $u' = x_5w$ requires that the "a" in question occur in a position that must be a "b" from the fact that $u' = wy_6$. This contradiction shows that the given u' for case 2) does not satisfy section (14). Thus section 15) is proved and Theorem IV, in turn, is proved.

Conclusion

The four theorems, taken together, show the logical equivalence of the four properties of the statements of equation 3), as is indicated in Fig. 5.

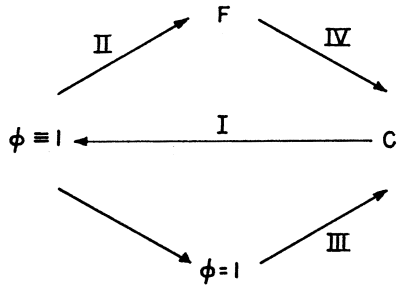

Fig. 5—Diagram showing the relations of the four theorems.

Sections 1)–5) of Theorem I then show that the probabilities associated with a full code-word set must sum at least to one. Sections 6) and 7) of Theorem I show that this sum must be no more than one if the code is decodable; that is, $\phi_{P_0}(1) \leq 1$ if P_0 is d. It can easily be shown that this inequality leads to the generalized Kraft[8] inequality

$$\sum_{k=1}^{M} 2^{-q_k} \leq 1,$$

where q_k is the normalized cost of word w_k.

Our discussion shows that the equality sign holds only when P_0 is full. This inequality was obtained by Marcus[5] by extending Mandelbrot's proof[2] for the equal-cost case. Mandelbrot used Shannon's Fundamental Theorem for Discrete Noiseless Channels[9] and pointed out that a similar inequality had been obtained previously by Szilard. McMillan[3] obtained a proof in the equal-cost case without using information-theory concepts. Note that the proofs of this paper are also independent of the Shannon theorem.

For the equal-cost case, the normalized cost is just $q_k = n_k \log D$, with $n_k = |w_k|$. Thus the inequality reads:

$$\sum_{k=1}^{M} D^{-n_k} \leq 1.$$

[8] L. G. Kraft, "A device for quantizing, grouping and coding amplitude modulated pulses," S. M. thesis, Dept. Elec. Eng., M. I. T. Cambridge, Mass.; 1949.

[9] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423; July, 1948.