On the Existence of Binary Simplex Codes

H. Taylor
Communications Systems Research Section

Using a very simple combinatorial construction, we prove the existence of a binary simplex code with m codewords for all $m \geqslant 1$. The problem of the shortest possible length is left open.

I. Introduction

In a variety of communications problems it is necessary to have a set of digital sequences, i.e., codewords, where mutual correlations are as small as possible. In Ref. 1, Theorem 4.1, it is shown that if $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m\}$ is any set of m binary (components ± 1) codewords of length n, and if the correlation between $\mathbf{x}_i = (x_{i1}, \cdots, x_{iN})$ and $\mathbf{x}_j = (x_{j1}, \cdots, x_{jN})$ is defined as $\rho(\mathbf{x}_i, \mathbf{x}_j) = (\Sigma_{k=1}^{N} x_{ik}x_{jk})/n$, then

$$\max_{i \neq j} \rho(\mathbf{x}_i, \mathbf{x}_j) \geqslant \begin{cases} \dfrac{-1}{m-1}, & \text{if } m \text{ is even} \\ -\dfrac{1}{m}, & \text{if } m \text{ is odd} \end{cases} \tag{1}$$

A set of m binary codewords achieving the bound (1) is called a *simplex* code. In Ref. 1 it is shown that simplex codes exist for an infinite number of values of m. In this article we will show that simplex codes exist for *all* m, using a very simple combinatorial construction, which bypasses the question of making the codewords as short as possible.

II. Construction

First we note that if $\mathbf{x}_1, \cdots, \mathbf{x}_m$ is a simplex code with $m = 2n$ then $\{\mathbf{x}_1, \cdots, \mathbf{x}_{m-1}\}$ is a simplex code with $m = 2n - 1$, so we focus our attention on even values of m.

Form the incidence matrix of all the n-subsets of a $2n$-set. It will be a matrix of 0's and 1's having $2n$ rows and $\binom{2n}{n}$ columns. Each row corresponds to an element of the $2n$-set, and in each column the n 1's mark the elements of the corresponding n-subset. A good definition for the binomial coefficient $\binom{b}{a}$ for nonnegative integers a, b is that $\binom{b}{a}$ is the number of a-subsets of a b-set. For example, when $n = 3$, the incidence matrix, easily written in binary lexicographic order, will be:

1	1	1	1	1	1	1	1	1	1										
1	1	1	1							1	1	1	1	1	1				
1				1	1	1				1	1	1				1	1	1	
	1			1			1	1		1			1	1		1	1		1
		1			1		1		1		1		1		1	1		1	1
			1			1		1	1			1		1	1		1	1	1

In this example, with $m = 6$, we can calculate, for any two rows:

$$A = \text{number of agreements} = 8$$

$$D = \text{number of disagreements} = 12$$

Thus, $(A - D)/(A + D) = -1/5$.

In general each element will belong to $\binom{2n-1}{n-1}$ of the n-sets. Thus, each codeword will have weight $\binom{2n-1}{n-1}$. Each pair of elements will be a 2-subset of $\binom{2n-2}{n-2}$ of the n-sets. Any two distinct codewords will overlap (*both* have 1's) in $\binom{2n-2}{n-2}$ of the columns. The interpretation in terms of sets has made it easy to evaluate:

$$D = 2\binom{2n-1}{n-1} - 2\binom{2n-2}{n-2}$$

$$A = \binom{2n}{n} - D$$

By straightforward calculation:

$$\frac{A-D}{A+D} = \frac{\binom{2n}{n} - 4\binom{2n-1}{n-1} + 4\binom{2n-2}{n-2}}{\binom{2n}{n}}$$

$$= \frac{\frac{2n(2n-1)}{n(n-1)} - 4\frac{(2n-1)}{(n-1)} + 4}{\frac{2n(2n-1)}{n(n-1)}}$$

$$= \frac{2n(2n-1) - 4n(2n-1) + 4n(n-1)}{2n(2n-1)}$$

$$= \frac{(2n-1) - 2(2n-1) + 2(n-1)}{(2n-1)}$$

$$= \frac{-1}{2n-1} = \frac{-1}{m-1}$$

III. Comments

These codewords are at least twice as long as they need to be – making them shorter will be the subject of a later article. Meanwhile, this is a good place to observe that these matrices have complete symmetry with respect to incidence. The codewords all have the same weight. Any two of them overlap in the same number of places. Any three of them overlap in the same number of places. And so on. And all those overlap numbers are easy to calculate.

Reference

1. Golumb, S. W., Baumert, L. D., Easterling, M. F., Stiffler, J. J., and Viterbi, A. J., *Digital Communication with Space Applications*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1964.