# THE METRIC STRUCTURE OF CODES FOR THE BINARY SYMMETRIC CHANNEL

A. J. THOMASIAN
UNIVERSITY OF CALIFORNIA, BERKELEY

## 1. Introduction and summary

The main result of this paper, theorem 4, is that there exists an infinite sequence of binary digits which can be used to generate sliding parity check codes for the binary symmetric channel (BSC), and the error probability for such codes is close, in a certain sense, to minimum *uniformly* in all of the parameters: block length, probability of error for a single digit, and rate not too much less than capacity. This result is obtained by studying the metric structure which can be required of codes, and then relating this to the error probability. The paper is essentially self-contained.

Let $B^n$ be the set of $2^n$ ordered $n$-tuples from $B = B^1 = \{0, 1\}$ where an element of $B$ is called a binary digit or bit. For any $n \geqq 1$ we define a metric on $B^n$ by $\overline{ab}$ = the number of coordinates in which $a$ and $b$ differ. An $n$-code $C_n$ is a nonempty subset of $B^n$; an element $c \in C_n$ is called a codeword. A code is any set which is an $n$-code for some $n$. The code $C_n$ in a sequence of codes $C_1, C_2, \cdots$ is always an $n$-code. The probability law of a BSC is defined as follows: for each $a \in B^n$ there is a probability distribution on $B^n$ given by

$$P\{b|a\} = p^{\overline{ab}}q^{n-\overline{ab}}, \tag{1}$$

where $0 < p < 1/2$ and $q = 1 - p$. A statement which is made for all $p$ means for all $p$ such that $0 < p < 1/2$. Here $P\{b|a\}$ is the probability that the channel output is $b$ given that the channel input is $a$. When $n = 1$ we have $P\{0|0\} = P\{1|1\} = q$, $P\{1|0\} = P\{0|1\} = p$ so that the probability $p$ of an error in a single digit does not depend on what that digit is, hence the "symmetric" designation.

A decoder for an $n$-code $C_n$ is a set $D$ of disjoint subsets of $B^n$ and a 1 to 1 correspondence between $C_n$ and $D$. Assume for the moment that the elements of $C_n$ and $D$ are ordered by indices so that $C_n = \{c_1, \cdots, c_s\}$, $D = \{D_1, \cdots, D_s\}$ and $c_k$ corresponds to $D_k$ under the 1 to 1 correspondence. In application, the sender and receiver first decide on $C_n$ and $D$ and then use these repeatedly. The

This research was supported by the Information Systems Branch of the Office of Naval Research under Contract Nonr-222(53).