CODE FIXING SINGLE ASYMMETRIC ERRORS, UDC 621.391.154

R. R. Varshamov, G. M. Tenengolts
(Moscow)

Abstract

A binary corrective code is considered that corrects single asymmetric errors.

I. Introduction

In communication technology, it is of interest to study asymmetric coding systems, i.e., systems with an asymmetric channel. Asymmetric is a channel with unequal damage probabilities various elementary parcels (impulses). In the binary case, for example, this means that the probability of the character "1" going to "0" in code word when passing through the channel is significantly less than the probability of transition "0" to "1". In such a situation, one usually neglects the smallest probabilities and tries to protect the working signals only from some partial errors, which in this case have the highest probability. This raises the question of constructing corrective codes capable of correcting any $r > 0$ random and independent asymmetric errors. The task of constructing such a code [1] is equivalent to the following mathematical problem: Select from the G_n, n-dimensional vector space over the field G of residues modulo 2, the maximum (in terms of the number of elements) subset of K, any distinct pair of vectors of which would satisfy condition

$$\rho(x, y) = |x - y| + ||x| - |y|| \geq 2r + 1, \tag{1}$$

where $|x| = \sum_{i=1}^{n} x_i$ is the norm (weight) of the vector x.

It is easy to see that condition (1) is weaker than the requirement imposed by to the working signals of the correcting code that corrects the same number r of random and independent symmetric errors, namely:

$$|x - y| \geq 2r + 1. \tag{2}$$

Therefore, a priori one could assume that the maximum possible transmission rate of a message of a system with an asymmetric channel, generally speaking, should be greater than the transmission rate of the corresponding system with a symmetric channel.

However, as shown in [1], in the case of linear coding, the asymmetric code does not have the expected advantages and is almost always identical to the symmetric one. (Here and below, when comparing codes their main parameters, such as the length of the signal n and the number of corrected errors r are assumed to be the same in both cases). In this article, we consider an asymmetric code (generally speaking, non-linear), correcting single errors, the power (number of signals) of which is greater than the power of the corresponding maximum symmetric code, as well as the asymmetric code known in the literature Freiman - Kim.

II. Correction code, correcting single asymmetric errors

Let it be required to construct a code with the correction of asymmetric single errors for a given length of code words n. Consider the comparison

$$W = \sum_{i=1}^{n} i\alpha_i \equiv a \qquad (\text{mod } n + 1\ ), \tag{3}$$

where α_i are binary numbers, a is an arbitrary integer satisfying the relation $0 \leq a < n + 1$. The code that corrects single asymmetric errors is represented by the set K_a all possible binary sequences of the form $a = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, where $\alpha_1, \alpha_2, \ldots, \alpha_n$ is the solution of comparison (3). For example, in the case of $n = 8$, one of the solutions to (3) for $a = 0$ will be:

$$\alpha = (1, 0, 0, 1, 1, 0, 0, 1),$$

since $1+4+5+8 \equiv 0 \pmod 9$. Accordingly, the code word has the form: 10011001. Let's consider the construction of the code on one particular example.

Example. Let, as before, $n = 8$. Setting $a = 0$, comparison (3) can be written in the form

$$\sum_{i=1}^{8} i\alpha_i \equiv 0 \qquad (\text{mod } 9)$$

all solutions of which form a code containing 30 different signals, having the form:

00000000	00001110	11011100
10000001	11000011	01010111
01000010	10100101	00111011
00100100	10010110	11110001
00011000	10011001	11101010
11000100	01011010	10111101
10101000	01101001	01111110
01110000	01100110	11100111
00100011	00111100	11011011
00010101	10001111	11111111

TABLE I

III. Single error correction

For definiteness, we will further assume that the probability the transition of the symbol "1" to "0" is significantly greater than the probability of transition "0" in "1", i.e., errors of the form $0 \to 1$ are neglected. This assumption is not essential and, as shown in [1], a code capable of correcting non-symmetric errors of the form $1 \to 0$ is also suitable for error correction like $0 \to 1$.

$$-l \equiv W' - a^1 \qquad (\text{mod n + 1}) \tag{4}$$

Let's illustrate the above with an example.

Let one of the possible signals of the system (Table 1) be transmitted, namely: $\alpha = 11101010$. However, as a result of its passage through the channel, it suffers distortion and is perceived at the receiving end in the form: $\alpha' = 11001010$ (an error occurred on third signal position). According to the definition for the code word a, the relation

$$W \equiv 0 \qquad (\text{mod } 9)$$

where W is the "generalized weight" of the signal a, determined by formula (3). Meanwhile, in the case of a single error of the form $1 \to 0$, the generalized weight of the distorted signal α' is determined by the relation

$$W' = W - l,$$

where l is the position number of the signal α, in which the error occurred. That's why

$$W = W' + l = 0 \qquad (\text{mod } 9).$$

Where

$$l = -W' \qquad (\text{mod } 9).$$

In our case

$$W' = \sum_{i=1}^{8} i\alpha_i = 1 + 2 + 5 + 7 = 15.$$

Consequently,

$$l \equiv -15 \qquad (\text{mod } 9)$$

or

$$l = 3.$$

This means that the error occurred in the third position (which actually happened). Thus, the corrected signal looks like: 11101010.

[1]It is easy to see that in the case of errors of the form $0 \to 1$, the right side of the expression (4) taken with a minus sign.

IV. Number of code words

The total number M(n) of code words of length n in the best of the proposed codes, as it is easy to show, is related by the inequality

$$M(n) \geq \frac{2^n}{n+1}. \tag{5}$$

Therefore, with a simple analysis, it can be established that the maximum number of its signals is in any case not less than the total the number of signals $M_1(n) = 2^{n+[-\log_2(n+1)]}$ one Hamming code with correction of single symmetrical errors. The function [x] is defined for any real x and is the largest integer not exceeding x. Indeed, this fact follows directly from the obvious inequality

$$\frac{2^n}{n+1} \geq 2^{n+[-\log_2(n+1)]}, \tag{6}$$

where

$$\frac{1}{n+1} \geq 2^{-\log_2(n+1)},$$

or, which is the same,

$$2^{-[-\log_2(n+1)]} \geq n+1.$$

The equal sign (i.e., $M(n) = M_1(n)$) takes place only in the case $n = 2^k - 1$.

Thus, it has been established that the maximum number of signals in the proposed code is almost always greater than the total number of signals in the symmetric Hamming code and can coincide with it only for special values n = 2k

V. Comparison of the proposed code with asymmetric corrective codes

At present, only one corrective code is known, correcting single asymmetric errors. This code, proposed by Freiman and Kim [2], is formed from the Hamming code, length whose code word is respectively equal to $n-m$, and $m = \left[\frac{n}{2}\right]$. The total number of code words of length n in the Freiman-Kim code

$$M_2(n) = (h_{n-m}+1)2^{m-1},$$

where h_p is the number of Hamming code codewords with signal length p. However, as will be shown below, the cardinality $M'(n)$ of the best of the codes we propose are much larger than the power of the Freiman-Kim code. For this purpose, we first prove the following inequality, valid for any $n > 6$:

$$\left(2^{n-[n/2]+[-\log_2(n-[n/2]+1)]}+1\right)2^{[n/2]-1} < \frac{2^n}{n+1}. \tag{7}$$

In the validity of (7) for $n = 7 \div 13$ we are convinced by the direct counting (see table).

n	Hamming code $M_1(n)$	Freiman-Kim code $M_2(n)$	Proposed code $M'(n)$	n	Hamming code $M_1(n)$	Freiman-Kim code $M_2(n)$	Proposed code $M'(n)$
3	2	2	2	10	64	80	94
4	2	4	4	11	128	144	171
5	4	6	6	12	256	288	316
6	8	12	10	13	512	544	586
7	16	12	16	14	1024	1088	1093
8	16	24	30	15	2048	1088	2048
9	32	40	52	16	2048	2176	3856

TABLE II

It remains to show that it holds for $n \geq 14$. We transform expression (7) as follows:

$$2^{n-1+[-\log_2(n-[n/2]+1)]} + 2^{[n/2]-1} < \frac{2^n}{n+1},$$

where one

$$2^{-1+[-\log_2(n-[n/2]+1)]} + 2^{[n/2]-n-1} < \frac{1}{n+1},$$

or

$$2^{-\log_2(n/2+\epsilon+1)} + 2^{n/2-\epsilon} < \frac{2}{n+1},$$

where

$$\epsilon = \binom{n}{2} - \frac{n}{2} \geq 0.$$

Then we have

$$2^{-\log_2(n/2+\epsilon+1)-\log_2\epsilon_1} + 2^{-(n/2+\epsilon)} < \frac{2}{n+1},$$

i.e.

$$2^{-\log_2(n/2+\epsilon+1)\epsilon_1} + 2^{-(n/2+\epsilon)} < \frac{2}{n+1},$$

which in turn gives

$$\frac{1}{\left(\frac{n}{2}+\epsilon+1\right)\epsilon_1} + \frac{1}{2^{n/2+\epsilon}} < \frac{2}{n+1}. \tag{8}$$

Here

$$\log_2\epsilon_1 = \left[-\log_2\left(\frac{n}{2}+\epsilon+1\right)\right] + \log_2\left(\frac{n}{2}+\epsilon+1\right) \geq 0, \quad \epsilon_1 \geq 0.$$

Let us now write an inequality stronger than (8)

$$\frac{1}{\frac{n}{2}+1} + \frac{1}{2^{n/2}} < \frac{2}{n+1} \tag{8a}$$

or

$$\frac{1}{2^{n/2}} < \frac{2}{(n+1)(n+2)}. \tag{9}$$

Inequality (9) holds for any $n \geq 14$, whence automatically (8) follows, and hence also (7). The table above characterizes the lower bound $M'(n) \leq M(n)$ of the number of signals in the best of the proposed codes, $M_2(n)$ — in the Freiman-Kim code and $M_1(n)$ in the symmetric Hamming code.

Received by the editor September 28, 1963

VI. Cited Literature

1. Varshamov R. R. Some features of linear codes correcting nonsymmetric errors. Report Academy of Sciences of the USSR, vol. 157, no. 3, 1964.
2. Kim W. H. and Freiman C. V. Single Error-Correcting Codes for Asymmetric binary channels. IRE Transactions on information theory, v. IT-5, no. 2, 'June, 1959.