UDC 621.391.15:519.2

BOUNDS OF THE PROBABILITY AND OF THE NUMBER OF CORRECTABLE ERRORS FOR NONBLOCK CODES

M. S. Pinsker

Problemy Peredachi Informatsii, Vol. 3, No. 4, pp. 58–71, 1967

Upper bounds are derived for the number of correctable errors for nonblock codes. Explicit expressions are given for the upper limits of the power of a code as the decoding delay $\tau \to \infty$ for the case where a fixed number t of errors may occur in the interval τ, and for the case where the number of errors $t = \alpha\tau$ increases linearly with τ. These bounds are similar to the Hamming and Elias bounds for block codes. The upper limit of the error probability is also calculated for nonblock transmission over a binary symmetric memoryless channel: the exponential term of this limit is the same as the exponential term for block codes of length τ.

It is shown that for transmission rates tending to zero, the exponential term of the error probability converges to the exponential term of the error probability for block transmission at zero rate.

It is shown that many estimates also remain valid when feedback is present.

§1. Introduction

Transmission methods using nonblock encoding and decoding occupy an important place in theoretical investigations and their applications. Among these are the well-known books of Wozencraft and Reiffen [1], and Massey [2]; however, it is still not clear how far nonblock transmission methods enable an improvement to be made in transmission characteristics such as the number of correctable errors, the error probability, and the transmission rate for fixed error probability. The main result of this paper is that for a fixed transmission delay nonblock encoding and decoding do not improve these characteristics.

To simplify the discussion we consider in detail only the case of q-ary codes and a binary symmetric memoryless channel, although the methods and results remain in force in extremely general situations, in particular, for any discrete memoryless channel.

§2. Fundamental Concepts and Notation

We recall some concepts and definitions used in coding theory.

By a message we understand a random variable ξ with a probability distribution p_ξ, which assumes values on the set $\{\bar{x}_1, \ldots, \bar{x}_{|\bar{X}|}\} = \bar{X}$ with the finite number of elements $|\bar{X}|$. The elements of the set $\bar{X}$ are called the values of the message. Usually a message is given as a sequence of random variables $(\xi_1, \ldots, \xi_m) = \xi$, and the values of the message as a sequence of symbols $(x_1, \ldots, x_m) = \bar{x}$, $\bar{x} \in \bar{X}$, where $x_i \in X$, $i = \overline{1, n}$; X is a finite set, $\bar{X} = X \times \ldots \times X = X^m$.

Let Y^n be the set of all possible sequences $\bar{y} = (y_1, \ldots, y_n)$ of length n, where y_i, $i = \overline{1, n}$ are elements of the finite field $G = Y$, $|Y| = q$ is the number of elements of the field. The set Y^n and the conditional probability distribution $p(\bar{y}'/\bar{y})$, $\bar{y}$, $\bar{y}' \in Y^n$ defined on it, form a channel.

In this paper we consider the case where $p(\bar{y}'/\bar{y}) = p(\bar{y} - \bar{y}') = p(\bar{z})$, $\bar{z} = \bar{y} - \bar{y}' \in B$ ($\bar{y}' - \bar{y}$ is the componentwise difference of the sequences $\bar{y}'$ and $\bar{y}$), where $B \subseteq Y^n$ and $p(\bar{z}) = p_\zeta(\bar{z})$ is the distribution of the random variable ζ (noise) defined on B; B is called the set of noise values.

We define the code A of size $|A| = N$ as any subset $\{\bar{y}_1, \ldots, \bar{y}_{|N|}\}$ of elements of the set Y^n.

The unique mapping of the set $\bar{X}$ onto the code A is called encoding: $\mathbf{a} = (y_1, \ldots, y_n) = (f_1(\bar{x}), \ldots, f_n(\bar{x})) = f(\bar{x}) \in A$, $\bar{x} \in \bar{X}$. The unique mapping of the set Y^n onto $\bar{X}$ is called decoding: $\bar{x}' = \varphi(y_1', \ldots, y_n') = \varphi(\bar{y}')$, $\bar{y}' \in Y^n$, $\bar{x} \in \bar{X}$. We call the pair of functions (f, φ) defining encoding and decoding, the transmission method (or simply the transmission); the number n is the length of the transmission. The random variable $\eta = f(\xi)$ is called the input signal, and the random variable η', which assumes values in Y^n and has with η the joint distribution $p_{\eta\eta'}(\bar{y}, \bar{y}') = p_\eta(\bar{y})p(\bar{y}'/\bar{y})$ is called the output signal: in the case where $p(\bar{y}'/\bar{y}) = p_\zeta(\bar{y}' - \bar{y})$, we have $\eta' = \eta + \zeta$, where η is independent of the noise ζ. The random variable $\xi' = \varphi(\eta')$ is called the reproducible message.

Encoding maps the message ξ onto the input signal η; under the action of the noise ζ the input signal η is transformed into the output signal η'; decoding reproduces from the output signal η' the message ξ'. The function $\bar{x} = \varphi(\bar{y})$ thereby defines a partition of the space Y^n into the subsets $Y^{1,n}, \ldots, Y^{|M|,n}$, $\varphi(Y^{i,n}) = \bar{x}_i$.

In block transmission with delay τ the message value $\bar{x} = (x_1, \ldots, x_m)$ is partitioned into blocks of length l; $(x_{kl+1}, \ldots, x_{(k+1)l}) = \bar{x}_k$, $k = 0, 1, 2, \ldots, m/l - 1$; each of these blocks is encoded into a code block of length τ: $(y_{k\tau+1}, \ldots, y_{(k+1)\tau}) = f(x_{kl+1}, \ldots, x_{(k+1)l})$. Decoding amounts to the restitution of the blocks $\bar{x}_k = (x_{kl+1}, \ldots, x_{(k+1)l})$ from the output blocks $(y'_{k\tau+1}, \ldots, y'_{(k+1)\tau})$.

An example of nonblock transmission with delay τ is the transmission of a binary message by means of a binary tree code at the rate $R = \frac{k_0}{\tau_0} \leqslant 1$ with the delay τ. A complete and precise description of such a code can be found in [1, 3, 4]. We need only note that the code sequence $\bar{y} = (y_1, \ldots, y_n)$ is formed from the sequences $\bar{y}_{r+1} = (y_{r\tau_0+1}, \ldots, y_{(r+1)\tau_0}) = \bar{f}_r(x_1, \ldots, x_{(r+1)k_0})$ $r = 0, 1, \ldots, \frac{n}{\tau_0} - 1$ of length τ_0, called arcs of the code. The code is a tree with a node on each arc from which 2^{k_0} branches emerge, corresponding to the k_0 binary symbols of the message $(x_{rk_0+1}, \ldots, x_{(r+1)k_0})$. Decoding is defined in such a way that after receiving the subsequence $(y_1', \ldots, y'_{(r+1)\tau_0})$ the k_0 message symbols $\bar{x}'_{r-s+1} = (x'_{(r-s)k_0+1}, \ldots, x'_{(r-s+1)k_0})$ are reproduced, where $s = \tau/\tau_0$, $k_0 = R\tau_0$, $k_0\left(\frac{n}{\tau_0} - s\right) = m$.

We now introduce the general definition of the concept of transmission with the decoding delay τ. We call the symbol y_{j_i} of the code sequence $\bar{y}(y_1, \ldots, y_{j_i}, \ldots, y_n)$ initial for the i-th message symbol, if the values of the symbols $y_l = f_l(\bar{x}) = f_l(x_1, \ldots, x_i, \ldots, x_m)$, $l < j_i$ are not dependent on the values of the symbols $x_i, \ldots, x_m$ while a pair of message values $\bar{x} = (x_1, \ldots, x_m)$ and $\bar{x}^* = (x_1^*, \ldots, x_m^*)$, $x_s = x_s^*$, $s = \overline{1, i-1}$ can be found such that $f_{j_i}(\bar{x}) \neq f_{j_i}(\bar{x}^*)$. We have $j_i \leq j_{i+1}$. The transmission (f, φ) proceeds with decoding delay τ, if

$$x_i' = \varphi_i(y_1', \ldots, y'_{j_i+\tau-1}, \ldots, y_n') = \varphi_i(y_1', \ldots, y'_{j_i+\tau-1}),$$

where y_{j_i} is the initial symbol for x_i. The block and tree methods of transmission indicated above are transmissions with decoding delay τ.

We also need the following concept. We call the mapping

$$y_j = f_j(\varphi(y_1', \ldots, y_n')) = f_j(x_1', \ldots, x_m') = \psi_j(y_1', \ldots, y_n')$$

the decoding of the code A. We say that the decoding of the code symbol y_j proceeds with delay τ, if $y_j = f_j(\varphi(y_1', \ldots, y_j', \ldots, y_n'))$ does not depend on the values of the symbols $y'_{j+\tau}, \ldots, y'_n$. If for all $j = \overline{1, n}$ the decoding of y_j proceeds with delay τ, we call τ the decoding delay of the code A. It is obvious that in transmission with decoding delay τ the decoding of the code A also proceeds with delay τ. In the case of block codes the delay in decoding a code symbol y_j depends on the position of this symbol in the block and may oscillate from 1 to τ: in the case of a tree code this delay depends on the position of this symbol on the arc and may oscillate from $\tau - \tau_0 + 1$ to τ. In this section we will be interested in properties of codes determined only by the delay τ of decoding the code.

The code distance $\rho(\bar{y}, \bar{y}^*)$ between two code sequences $\bar{y} = (y_1, \ldots, y_n)$, $\bar{y}^* = (y_1^*, \ldots, y_n^*)$ with the decoding delay τ is defined as the number of nonzero symbols of the sequence

$$\overset{*}{y}_{j+1} - y_{j+1}, \ldots, \overset{*}{y}_{j+\tau'} - y_{j+\tau'},$$

where

$$y_i^* = y_i \quad \text{if } i \leqslant j, \quad \overset{*}{y}_{j+1} \neq y_{j+1};$$

$$\tau' = \begin{cases} \tau & \text{if } j + \tau \leqslant n, \\ n - j & \text{if } j + \tau > n. \end{cases}$$

Then y_{j+1} is called the point of branching of the pair of sequences $\bar{y}$ and $\bar{y}^*$.

We define the code distance $\rho(A)$ with decoding delay τ as the number

$$\rho(A) = \min_{\bar{y}, \bar{y}^* \in A, \bar{y} \neq \bar{y}^*} \rho(\bar{y}, \bar{y}^*). \tag{1}$$

We define the weight of the sequence $\bar{z} = (z_1, \ldots, z_n) \in Y^n$ with decoding delay τ as the number

$$\mu(\bar{z}) = \max_{0 \leqslant j \leqslant n-\tau} \mu(z_{j+1}, \ldots, z_{j+\tau}), \tag{2}$$

where $\mu(z_{j+1}, \ldots, z_{j+\tau})$ is the number of nonzero symbols of the sequence $(z_{j+1}, \ldots, z_{j+\tau})$. We will say that t errors are corrected in the transmission (f, φ) with decoding delay τ, if $\varphi(f(\bar{x}) + \bar{z}) = \bar{x}$ for any sequence $\bar{z}$ with $\mu(\bar{z}) \leq t$. Similarly the code A corrects t errors if a method of decoding the code $\varkappa(\bar{y}') = \bar{y} \in A$, $\bar{y}' \in Y^n$ with delay τ exists such that

$$\bar{y} = \dot{\varphi}(\bar{y} + \bar{z}) \tag{3}$$

for $\bar{y} \in A$ and any sequence z with $\mu(\bar{z}) \le t$.

In the case of block codes with block length τ the weight of the sequence $\bar{z}$ is defined as

$$\mu_{bl}(\bar{z}) = \max \mu(z_{\tau l+1}, \ldots, z_{\tau(l+1)}), \quad l = \overline{0, \frac{n}{\tau} - 1}, \tag{4}$$

and the code distance $\rho_{bl}(A)$ is defined by the relation

$$\rho_{bl}(A) = \min_{\bar{y}, \bar{y}^* \in A, \bar{y} \ne \bar{y}^*} \rho(\bar{y}, \bar{y}'),$$

where $\rho(\dot{y}, \bar{y}^*) \ne 0$ is the number of nonzero symbols of the sequence $\dot{y}_{\tau l+1} - y_{\tau l+1}, \ldots, \dot{y}_{\tau(l+1)} - y_{\tau(l+1)}$, and $y_i^* = y_i$ if $i \le \tau l$.

In the case of tree codes with arc length τ_0:

$$\mu_{\tau_0}(\bar{z}) = \max \mu(z_{\tau_0 l+1}, \ldots, z_{\tau_0(l+s)}), \quad s = \frac{\tau}{\tau_0} \quad l = \overline{0, \frac{n}{\tau_0} - s}, \tag{5}$$

if s = 1 the tree code is a block code.

It is obvious that

$$\mu(\bar{z}) \geqslant \mu_{\tau_0}(\bar{z}) \geqslant \mu_{bl}(\bar{z}), \quad \rho_{bl}(A) \leqslant \rho(A). \tag{6}$$

We will only stop to consider the weights $\mu(\bar{z})$ and $\mu_{bl}(\bar{z})$.

We introduce some notation: $A_{n,t}$, $A_{bl,n,t} \subseteq Y^n$ are respectively the codes of greatest length with code distance $\rho(A_{n,t}) \ge 2t + 1$, $\rho_{bl}(A_{bl,n,t}) \ge 2t + 1$; $B_{n,t}$ and $B_{bl,n,t}$ are the sets of all the sequences $\bar{z}$ of length n of weight $\mu(\bar{z}) \le t$ and $\mu_{bl}(\bar{z}) \le t$. By reason of the inequalities (6) we have inequalities for the number of elements of the sets.

$$|B_{bl,n,t}| \geqslant |B_{n,t}|, \quad |A_{bl\,n,t}| \leqslant |A_{n,t}|.$$

§3. Upper Bounds for the Size of the Code

Theorem 1. For the code A to correct t errors, it is necessary and sufficient that $\rho(A) \ge 2t + 1$.

The proof is obvious. By Theorem 1

$$|A_{n,t}| \leqslant \frac{q^n}{|B_{n,t}|}. \tag{7}$$

We have

$$|B_{bl,n,t}| = \left(\sum_{i=0}^{t} (q-1)^i C_\tau^i\right)^{n/\tau} \tag{8}$$

and

$$|A_{bl,n,t}| \leqslant \frac{q^n}{\left(\sum_{i=0} (q-1)^i C_\tau^i\right)^{n/\tau}}. \tag{9}$$

Formulas (8) and (9) are well known in the case where $\tau = n$ (the transmission of one block).

Formula (9) is called Hamming's bound for the code length.

We now estimate the powers of the sets $|B_{n,t}|$ and $|A_{n,t}|$. For this we consider the random variable $\zeta = (\zeta_1, \ldots, \zeta_n)$ with values on the set $B_{n,t}$. It is easy to see that*

$$\ln |B_{n,t}| \geqslant H(\zeta),$$

where $H(\zeta)$ is the entropy of the random variable ζ. The inequality (10) becomes an equation if and only if ζ is uniformly distributed on $B_{n,t}$. If ζ is uniformly distributed, the sequence of random variables $\zeta_1, \ldots, \zeta_n$ forms a Markov chain of complexity $\tau - 1$. Then

$$\ln |A_{n,t}| \leqslant I(\eta, \eta') = H(\eta) - H(\zeta) = Cn, \tag{11}$$

where η is a random variable with a uniform distribution on a set of sequences $\bar{y} \in Y^n$ independent of ζ, $\eta' = \eta + \zeta$, $I(\eta, \eta')$ is the amount of information in a pair (η, η'), C is the capacity of a channel with noise ζ. Formula (11) may be obtained as follows. We denote by $\tilde{\eta}$ a random variable with a uniform distribution on the set $A_{n,t}$; then the values $\tilde{\eta}$ are uniquely defined by the values of the random variable $\tilde{\eta}' = \tilde{\eta} + \zeta$, hence $\ln |A_{n,t}| = H(\tilde{\eta}) = I(\tilde{\eta}, \tilde{\eta}')$; on the other hand $I(\tilde{\eta}, \tilde{\eta}') \le I(\eta, \eta') = Cn$. Determining an exact expression for the entropy $H(\zeta)$ is an extremely difficult problem. However, it happens that an estimate of the entropy of the simpler Markov chains enables fairly good estimates of $|B_{n,t}|$ to be obtained.

We consider a Markov chain $\ldots, \zeta_{-1}, \zeta_0, \zeta_1, \ldots$ of complexity $\tau - 1$ with transition probabilities

$$p(z_j = l \mid z_{j-1}, \ldots, z_{j-\tau+1}) = \begin{cases} p_0, & \text{if} \quad \mu(z_{j-1}, \ldots, z_{j-\tau+1}) < t, \\ 0, & \text{if} \quad \mu(z_{j-1}, \ldots, z_{j-\tau+1}) = t, \; l = \overline{1, q-1}, \end{cases} \tag{12}$$

defined on infinite two-sided sequences $(\ldots, z_{-1}, z_0, z_1, \ldots)$. This Markov chain is homogeneous and ergodic and a stationary distribution of states $p(z_1, \ldots, z_{\tau-1})$ is defined for it, and consequently, a stationary Markov chain with transition probabilities (12) exists. We denote by $\bar{H}$ its entropy per step. We define the random variable $\zeta = (\zeta_1, \ldots, \zeta_n)$ as a segment of this Markov chain of length n. We have

$$H(\zeta) \geqslant n\bar{H} \tag{13}$$

and

$$\frac{1}{n} \ln |B_{n,t}| \geqslant \bar{H} =$$

$$= -\sum_{z_\tau} \sum_{\mu(z_1, \ldots, z_{\tau-1}) < t} p(z_1, \ldots, z_\tau) \ln p(z_\tau \mid z_1, \ldots, z_{\tau-1}) =$$

$$= -\sum_{\mu(z_1, \ldots, z_{\tau-1}) < t} p(z_1, \ldots, z_{\tau-1}) [(q-1) p_0 \ln p_0 + (1 - (q-1) p_0) \times$$

$$\times \ln(1 - (q-1) p_0))] = \sum_{\mu(z_1, \ldots, z_{\tau-1}) < t} p(z_1, \ldots, z_{\tau-1}) [H((q-1) p_0) +$$

$$+ (q-1) p_0 \ln(q-1)]. \tag{14}$$

To estimate $\bar{H}$ we need to estimate the probability

$$\sum_{\mu(z_1, \ldots, z_{\tau-1}) < t} p(z_1, \ldots, z_{\tau-1}) = \tilde{p}_{si}.$$

We put $\mu(z_1, \ldots, z_{\tau-1}) = t$ and consider the state $(z_{\tau+s}, \ldots, z_{2\tau+s-1})$, $s \ge 0$, of a Markov chain such that $\mu(z_1, \ldots, z_{\tau-1}) = \mu(z_{\tau+s}, \ldots, z_{2\tau+s-1}) = t$ and $\mu(z_{\tau+l}, \ldots, z_{2\tau+l-1}) < t$, $0 \le l < s$, and let $\bar{s}$ be the mean value of s, and $\bar{m}_t$ the mean number of states of weight less than t between the states $(z_1, \ldots, z_{\tau-1})$ and $(z_{\tau+s}, \ldots, z_{2\tau+s-1})$. We have $\tilde{p}_{si} = \frac{\bar{m}_t}{\tau + \bar{s} - 1}$. Also we denote by $\tilde{m}$ the mean number of zeros between two successive nonzero symbols of the Bernoulli scheme $\zeta_1', \zeta_2', \ldots$ with $p(z) = p_0$, $z = \overline{1, q-1}$, that is,

$$\tilde{m} = \frac{1 - (q-1) p_0}{(q-1) p_0}.$$

We will show that

$$\bar{m}_t \geqslant \tilde{m}t. \tag{15}$$

To prove (15) we notice that the Markov chain with the transition probabilities (12) can be obtained from the one-sided Markov chain $\ldots, \zeta_{-2}, \zeta_{-1}, \zeta_0$ and the sequence of random variables $\zeta_1', \zeta_2', \ldots$ as follows:

* Everywhere in this paper where nothing is said to the contrary, logarithms to the natural base are considered.

$$\zeta_1 = \begin{cases} \zeta_1', & \text{if} \quad \mu(\zeta_{-\tau+2},\dots,\zeta_0) < t, \\ 0, & \text{if} \quad \mu(\zeta_{-\tau+2},\dots,\zeta_0) = t, \end{cases}$$

$$\zeta_i = \begin{cases} \zeta_i', & \text{if} \quad \mu(\zeta_{i-\tau+1},\dots,\zeta_{i-1}) < t, \\ 0, & \text{if} \quad \mu(\zeta_{i-\tau+1},\dots,\zeta_{i-1}) = t. \end{cases} \quad (i = 2, 3, \dots)$$

The inequality (15) follows from the fact that if $\zeta_i' = 0$, then $\zeta_i = 0$. By the definition of $\bar{s}$:

$$\bar{m}_t \geqslant \bar{s}. \tag{16}$$

Therefore, by (15) and (16),

$$\bar{m}_t \geqslant \max\left(\frac{t[1-(q-1)]p_0}{(q-1)p_0}, \bar{s}\right),$$

and consequently,

$$\tilde{p}_{si} = \frac{\bar{m}_t}{\tau + \bar{s} - 1} \geqslant \max\left[\frac{\frac{t[1-(q-1)p_0]}{(q-1)p_0}}{\tau+\bar{s}-1}, \frac{\bar{s}}{\tau+\bar{s}}\right] \geqslant$$

$$\geqslant \frac{\frac{t(1-(q-1)p_0)}{(q-1)p_0}}{\tau + \frac{t(1-(q-1)p_0)}{(q-1)p_0}} = \frac{t(1-(q-1)p_0)}{\tau(q-1)p_0 + t(1-(q-1)p_0)}.$$

Putting $p_0 = \frac{\ln \tau}{(q-1)\tau}$, from the last relation and from (14) we obtain

$$\frac{1}{n}\ln|B_{n,t}| \geqslant \bar{H} \geqslant$$

$$\geqslant -\frac{t\left(1-\frac{\ln\tau}{\tau}\right)}{\tau\frac{\ln\tau}{\tau} + t\left(1-\frac{\ln\tau}{\tau}\right)}\left[\frac{\ln\tau}{\tau}\ln\frac{\ln\tau}{(q-1)\tau} + \left(1-\frac{\ln\tau}{\tau}\right)\ln\left(1-\frac{\ln\tau}{\tau}\right)\right].$$

From this as $\tau \to \infty$ we obtain

$$|B_{n,t}| \geqslant \left[\frac{\tau^t(q-1)^t}{\ln^t\tau}(1+o(1))\right]^{n/\tau}. \tag{17}$$

We now prove the converse inequality

$$|B_{n,t}| \leqslant \left[\frac{\tau^t(q-1)^t}{\ln^t\tau}(1+o(1))\right]^{n/\tau}. \tag{18}$$

In the case $t = 1$, $q = 2$ an expression for $|B_{n,t}|$ was found by Vershubskii and Mikhelev [5]

$$|B_{n,1}| = \sum_{i\geqslant 0} C^i_{n-(i-1)(\tau-1)}, \quad C_a^b = 0 \quad \text{if } b > a.$$

This relation is easily generalized to the case $q > 2$:

$$|B_{n,t}| = \sum_{i\geqslant 0} C^i_{n-(i-1)(\tau-1)}(q-1)^i = \left[\frac{\tau(q-1)}{\ln\tau}(1+o(1))\right]^{n/\tau}, n, \tau\to\infty. \tag{19}$$

The inequality (18) follows from the relation (19) and the inequality

$$|B_{n,t}| \leqslant |B_{n,1}|\,|B_{n,t-1}|.$$

In order to obtain the last inequality, it is sufficient to notice that if in the sequence of the set $B_{n,t}$ the τs-th nonzero symbols, $s = 1, 2, \dots$, are replaced by zeros, the sequence of the set $B_{n,t-1}$ is obtained, but if only the τs-th symbols remain nonzero, the sequence of $B_{n,1}$ is obtained. It is also easy to show that

$$|B_{n,t}| \leqslant |B_{t,1}|\,|B_{n,t-1}|.$$

Therefore we have the following theorem.

Theorem 2.

$$|B_{n,t}| = \left[\frac{\tau^t(q-1)^t}{\ln^t\tau}(1+o(1))\right]^{n/\tau}, \ t = \text{const}, \ n\to\infty, \ \tau\to\infty. \tag{20}$$

It is obvious that the expressions in brackets on the right sides of (8) and (20) differ by the factor $\frac{t!}{\ln^t\tau}(1+o(1))$.

The problem of constructing the code length

$$q^n \Big/ \left[\frac{d\tau^t(q-1)^t}{\ln^t\tau}\right]^{n/\tau}, \quad d = \text{const}$$

is extremely interesting. For $t = 1$ and $q = 2$, Ash and Wyner [6] constructed convolutional tree codes of length

$$|A_{n,t}| = \left(\frac{d2^\tau}{\tau/\log_2\tau}\right)^{n/\tau}.$$

We will now estimate $|B_{n,t}|$ in the case where $t = \alpha\tau$. We again consider a Markov chain with transition probabilities (12). In this case $1 - \tilde{p}_{si}$ will be conveniently estimated by the following inequality:

$$1 - \tilde{p}_{si} \leqslant p\{\zeta_1' + \dots \zeta_{\tau-1}' \geqslant t\}, \tag{21}$$

where ζ_i', $i = \overline{1, \tau-1}$ are independent binary random variables, and $p\{\zeta_i' = 1\} = (q-1)p_0$. The inequality (21) obviously follows from the proof of the inequality (15). In order to select the value of p_0 minimizing the right side of (14), we use the estimate (see, for example, [12])

$$p\{\zeta_1' + \dots + \zeta_{\tau-1}' + (q-1)p_0(\tau-1) \geqslant x\sigma\sqrt{\tau}\} \leqslant \frac{\varepsilon_1}{\sqrt{\tau}} + \Phi(x)$$

$$\varepsilon_1 \leqslant \frac{1.33\mathrm{M}(\zeta_i' - (q-1)p_0)^3}{\sigma^3} \leqslant 1.33/\sigma; \quad \sigma^2 = (q-1)p_0(1-(q-1)p_0);$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-x^2/2}\,dx. \tag{22}$$

Putting $(q-1)p_0 = \frac{t}{\tau} - \delta$, where $\delta = \sqrt{\frac{\sigma^2\ln\tau}{\tau}}$ we obtain from (22)

$$1 - \tilde{p}_{si} \leqslant p\{\zeta_1' + \dots + \zeta_{\tau-1} \geqslant t\} =$$

$$= p\left\{\zeta_1' + \dots + \zeta_{\tau-1}' - (q-1)p_0\tau \geqslant \frac{\delta\sqrt{\tau}}{\sigma}\sigma\sqrt{\tau}\right\} \leqslant$$

$$\leqslant \frac{\varepsilon_1}{\sqrt{\tau}} + \frac{1}{\delta/\sigma\sqrt{2\pi\tau}}e^{-\left(\frac{\delta}{\sigma}\sqrt{\tau}\right)^2/2} = \frac{\varepsilon_1}{\sqrt{\tau}} + \frac{1}{\sqrt{2\pi\tau\ln\tau}}. \tag{23}$$

Substituting $\frac{t}{\tau} - \delta = (q-1)p_0$ in (14) and using the inequality $H\left(\frac{t}{\tau}-\delta\right) \geqslant H\left(\frac{t}{\tau}\right) + \delta\ln\frac{\tau - t + \tau\delta}{t - \tau\delta}$, $\delta > 0$, and also the relation (23), we obtain:

Theorem 3.

$$\frac{1}{n}\ln|B_{n,t}| \geqslant H\left(\frac{t}{\tau}\right) + \frac{t}{\tau}\ln(q-1) + (1+o(1))\sqrt{\frac{\sigma^2\ln\tau}{\tau}} \times$$

$$\times\left[\ln\frac{\frac{t}{\tau}}{1-\frac{t}{\tau}} - \ln(q-1)\right], \tag{24}$$

where $H(p_0) = -p_0\ln p_0 - (1-p_0)\ln(1-p_0)$, $t = \alpha\tau$, $\alpha < 1$, $n\to\infty$, $\tau\to\infty$.

For $q = 2$, formula (24) assumes the form

$$\frac{1}{n}\ln|B_{n,t}| \geqslant H\left(\frac{t}{\tau}\right) + (1+o(1))\sqrt{\frac{\sigma^2 \ln\tau}{\tau}}\ln\frac{\frac{t}{\tau}}{1-\frac{t}{\tau}}. \qquad (25)$$

This expression differs from the corresponding expression for a block code by the presence of the second term of orde

$$O\left(\sqrt{\frac{\ln\tau}{\tau}}\right).$$

Theorem 4. (Elias bound, see [8, 9]). For binary codes

$$R = \frac{1}{n}\ln|A_{n,t}| \leqslant \ln 2 - H\left(\frac{1-\sqrt{1-4\frac{t}{\tau}}}{2}\right) + O\left(\sqrt{\frac{\ln\tau}{\tau}}\right). \qquad (26)$$

We prove the theorem by means of the following lemmas.

Let D be a set of sequences of n symbols with the j-initial symbols in common: we call the number of branch the set D on the segment [j + 1, j + l], the number of different subsequences of the sequences of D on the segment [j + l].

Lemma 1. If $(r+l)^2 > \frac{1}{2}(2(r+l)-d)(l+\tau)$ the subset D of the set $B_{n,r}$ of sequences with code distance $\rho(D) \geq d = 2t+1$ and j common initial symbols, has not more than

$$\frac{{}^1/_2 d(l+\tau)(r+l)}{(r+l)^2 - {}^1/_2(2(r+l)-d)(l+\tau)} \qquad (27)$$

branches on the segment [j + 1, j + l].

The proof of Lemma 1 is similar to the proof of the lemma in [9]. We denote by $|A_{n,t,r}|$ the maximal numbe elements of the subset of the set $B_{n,r}$ with code distance d = 2t + 1. It follows from Lemma 1 that

$$|A_{n,t,r}| \leqslant \left[\frac{{}^1/_2 d(l+\tau)(r+l)}{(r+l)^2 - {}^1/_2(2(r+l)-d)(l+\tau)}\right]^{n/l}. \qquad (28)$$

Lemma 2.

$$|A_{n,t}|\,|B_{n,r}| \leqslant 2^n |A_{n,t,r}|. \qquad (29)$$

This formula is similar to formula (5) of [9]. Putting

$$r+l = \frac{1}{2}(\tau+l)\left(1-\sqrt{1-\frac{4t}{\tau+l}}\right), \quad l = \sqrt{\tau}\ln\tau, \qquad (30)$$

we obtain from (28)

$$|A_{n,t,r}| \leqslant (d(r+\sqrt{\tau}\ln\tau))^{\frac{n}{\sqrt{\tau}\ln\tau}}. \qquad (31)$$

Comparing (29) and (31), we find that

$$|A_{n,t}| \leqslant \frac{2^n (d(r+\sqrt{\tau}\ln\tau))^{\frac{n}{\sqrt{\tau}\ln\tau}}}{|B_{n,r}|}. \qquad (32)$$

By (25),

$$\frac{1}{n}\ln|B_{n,r}| \geqslant H\left(\frac{r}{\tau}\right) + O\left(\sqrt{\frac{\ln\tau}{\tau}}\right) = H\left(\frac{1-\sqrt{1-4\frac{t}{\tau}}}{2}\right) + O\left(\sqrt{\frac{\ln\tau}{\tau}}\right). \qquad (33)$$

The result (24) follows from (33) and (32). The theorem is proven.

4. Estimation of the Error Probability

In investigating the error probability we restrict ourselves to the case of a binary symmetric memoryless channel; we consider the message to be given as a sequence of equidistributed independent binary random variables: $\xi = (\xi_1, \ldots, \xi_m)$, $P_{\xi_i}(1) = P\{\xi_1 = 1\} = 1/2$, $i = \overline{1, n}$. The transmission rate along the channel is $R = \frac{m}{n}\ln 2$.

Encoding defines the random variable $\eta = f(\xi)$, the input signal; the transition probability of the channel p_0 defines the random variable η', the output signal, such that

$$p_{\eta\eta'}(\bar{y}, \bar{y}') = p_{\eta\eta'}(y_1, \ldots, y_n; y_1', \ldots, y_n') = p_\eta(y_1, \ldots, y_n) \times$$
$$\times\, p(y_1', \ldots, y_n' | y_1, \ldots, y_n) = p_\eta(y_1, \ldots, y_n)\prod_{i=1}^{n} p(y_i'|y_i),$$

finally, decoding defines the random variable $\xi' = (\xi_1', \ldots, \xi_m') = \varphi(\eta')$. The sequence ξ, η, η', ξ' forms a Markov chain.

We define the error probability in the transmission (f, φ) as the number

$$p_e(f,\varphi) = \frac{1}{m}\sum_{i=1}^{m} p\{\xi_i' \neq \xi_i\} = \frac{1}{m}\sum_{i=1}^{m}\sum_{x_i \neq x_i'} p_{\xi_i \xi_i'}(x_i, x_i'). \qquad (34)$$

If (f, φ) is a block transmission with decoding delay τ, we know [7] the lower and upper limits of the lower bound $p_{e,\mathrm{opt}}(R) = \inf p_e(f,\varphi)$ of the error probability $p_e(f, \varphi)$, which are identical apart from the factor for (f, φ)

$$C \geqslant R \geqslant R_{\mathrm{crit}} = 1 - H\left(\frac{\sqrt{p_0}}{\sqrt{p_0}+\sqrt{1-p_0}}\right). \qquad (35)$$

The question arises of whether it is possible to decrease the error probability by using a nonblock method of transmission. The answer to this question is given by the following theorem.

Theorem 5. In transmission with delay τ along a binary symmetric memoryless channel

$$p_{e,\mathrm{opt}}(R) \equiv p_{e,\mathrm{opt}}(R,\tau) \geqslant$$
$$\geqslant \frac{p_0(1-\rho)}{4(1-p_0)\rho\sqrt{2\pi\tau\rho(1-\rho)}}\, e^{-\tau\left(\rho\ln\frac{\rho}{p_0} + (1-\rho)\ln\frac{1-\rho}{1-p_0}\right)} \equiv p_e(R,\tau) \qquad (36)$$

for

$$R = \frac{m}{n}\ln 2 \geqslant \frac{\ln 2 - H(\rho) - \sqrt{\frac{\rho(1-\rho)\ln\tau}{\tau}}\ln\frac{p_0}{1-p_0}}{1-\left[H\left(\frac{\varepsilon_1}{\sqrt{\tau}}\right) - \frac{1}{\sqrt{2\pi\tau}\ln\tau}\ln\frac{\varepsilon_1}{\sqrt{\tau}-\varepsilon_1}\right]\frac{1}{\ln 2}}, \quad \varepsilon_1 \leqslant 1.33/\sigma \qquad (37)$$

where ρ is a parameter, $p_0 + \sqrt{\frac{\sigma^2\ln\tau}{\tau}} \leqslant \rho < \frac{1}{2}$, $\sigma^2 = \sqrt{p_0(1-p_0)}$.

Proof. It has to be proved that for any method of transmission with decoding delay

$$p_e(f,\varphi) = \frac{1}{m}\sum_{i=1}^{m} p(\xi_i \neq \xi_i') \geqslant p_e(R,\tau), \qquad (38)$$

where $\xi_i' = \varphi_i(\eta_1', \ldots, \eta_{j_i+\tau-1}')$, $\eta_{j_i} = y_{j_i}$ is the initial symbol of the i-th symbol of the x_i-message. The essence of the proof reduces to the following. From the value of the error probability $p_e(f, \varphi)$ for the transition probability p_0 of the channel, the quantity of information $I(\xi, \eta')$ is estimated for the transmission probability $\tilde{p}_0 \geq p_0$ and for the same transmission method (f, φ). The inequality (36) is obtained from the condition that $\frac{1}{n}I(\xi, \eta')$ cannot be greater than the transmission capacity of the channel $C(p_0)$ for a transition probability $\tilde{p}_0 \geq p_0$. We will now deduce the corresponding bounds. Since the noise $\tilde{\zeta} = (\zeta_1, \ldots, \zeta_n) = (\eta_1' - \eta_1, \ldots, \eta_n' - \eta_n)$ is independent of the message $\tilde{\xi} = (\xi_1, \ldots, \xi_m)$, the sequence of random variables

$$\xi_i, (\eta_1, \ldots, \eta_{j_i-1}, \eta_{j_i}', \ldots, \eta_{j_i+\tau-1}'), (\eta_1', \ldots, \eta_{j_i-1}, \eta_{j_i}', \ldots, \eta_{j_i+\tau-1}') \tag{39}$$

forms a Markov chain, and therefore on decoding for the maximum a posteriori probability defined by the expression

$$\tilde{\xi}_i' = \tilde{\varphi}_i(\eta_1, \ldots, \eta_{j_i-1}, \eta_{j_i}', \ldots, \eta_{j_i+\tau-1}'), \tag{40}$$

where

$$p\{\xi_i = \tilde{\xi}_i' | \eta_1, \ldots, \eta_{j_i-1}, \ldots, \eta_{j_i}', \ldots, \eta_{j_i+\tau-1}'\} \geqslant$$
$$\geqslant p\{\xi_i \neq \xi_i' | \eta_1, \ldots, \eta_{j_i-1}, \eta_{j_i}', \ldots, \eta_{j_i+\tau-1}'\} \tag{41}$$

we have

$$p\{\xi_i \neq \xi_i'\} \geqslant p\{\xi_i \neq \tilde{\xi}_i'\}. \tag{42}$$

Also we denote by $A_i(y_1, \ldots, y_{j_i+\tau-1})$ the set of all subsequences $z_{j_i}, \ldots, z_{j_i+\tau-1}$ such that if

$$\eta_1 = y_1, \ldots, \eta_{j_i+\tau-1} = y_{j_i+\tau-1}, \quad \eta_{j_i}' = y_{j_i} + z_{j_i}, \ldots, \eta_{j_i+\tau-1}' = y_{j_i+\tau-1} + z_{j_i+\tau-1},$$

then $\xi_i = \tilde{\xi}_i$. We have

$$p\{\xi_i \neq \xi_i'\} = 1 - \sum_{(y_1, \ldots, y_n) \in A} p_\eta(y_1, \ldots, y_{j_i-1}, y_{j_i}, \ldots, y_{j_i+\tau-1}) \times$$
$$\times \sum_{(z_{j_i}, \ldots, z_{j_i+\tau-1}) \in A_i(y_1, \ldots, y_{j_i+\tau-1})} \frac{p_{\eta\eta'}(y_1, \ldots, y_{j_i+\tau-1}, y_{j_i}', \ldots, y_{j_i+\tau-1}')}{p_\eta(y_1, \ldots, y_{j_i+\tau-1})} \tag{43}$$

and for the transition probability $\tilde{p}_0$:

$$\sum_{(z_{j_i}, \ldots, z_{j_i+\tau-1}) \in A_i(y_1, \ldots, y_{j_i+\tau-1})} \frac{p_{\eta\eta'}(y_1, \ldots, y_{j_i+\tau-1}, y_{j_i}', \ldots, y_{j_i+\tau-1}')}{p_\eta(y_1, \ldots, y_{j_i+\tau-1})} =$$
$$= \sum_{(z_{j_i}, \ldots, z_{j_i+\tau-1}) \in A_0(y_1, \ldots, y_{j_i+\tau-1})} \tilde{p}_0^{\,\mu(z_{j_i}, \ldots, z_{j_i+\tau-1})} (1 - \tilde{p}_0)^{\tau - \mu(z_{j_i}, \ldots, z_{j_i+\tau-1})}. \tag{44}$$

It is easy to verify that the right side of (44) is majorized by the sum over the spherical domain $A_{sp}(y_1, \ldots, y_{j_i+\tau-1})$ containing all possible sequences $(z_{j_i}, \ldots, z_{j_i+\tau-1})$ of weight $s \leq t$, and not containing sequences of weight $t + 2$, such that if $\tilde{p}_0 = p_0$ this sum equals the right side of (44). Similarly (43) is majorized by the sum over the sets (spheres) $A_{i\,sp}(y_1, \ldots, y_{j_i+\tau-1})$ containing all possible sequences $(z_{j_i}, \ldots, z_{j_i+\tau-1})$ of weight t and not containing sequences of weight $t + 2$, where t is independent of the code sequence $(y_1, \ldots, y_n)$ and of the subscript $i = \overline{1, n}$.

Therefore the inequality

$$\bar{p}_e = \frac{1}{m} \sum_{i=1}^{m} p\{\xi_i \neq \tilde{\xi}_i'\} \leqslant$$
$$\leqslant 1 - \sum_{l=0}^{t} C_\tau^l \tilde{p}_0^l (1 - \tilde{p})^{\tau - l} = \sum_{l=t+1}^{\tau} C_\tau^l \tilde{p}_0^l (1 - \tilde{p}_0)^{\tau - l} \tag{45}$$

is valid for any values $\tilde{p}_0 \geq p_0$, if it is valid for $\tilde{p}_0 = p_0$.

We estimate the quantity of information $I(\xi, \eta')$ if $\tilde{p}_0 \geq p_0$. For this we consider the random variable $\beta = (\beta_1, \ldots, \beta_m)$

$$\beta_i = \begin{cases} 1, & \text{if} \quad \tilde{\xi}_i' \neq \xi_i, \\ 0, & \text{if} \quad \xi_i = \xi_i. \end{cases} \tag{46}$$

Obviously,

$$\frac{1}{m} \sum_{i=1}^{m} p\{\beta_i = 1\} = \frac{1}{m} \sum_{i=1}^{m} p\{\xi_i \neq \tilde{\xi}_i'\} = \bar{p}_e, \quad H(\beta) \leqslant mH(\bar{p}_e). \tag{47}$$

It is easy to verify that the random variable ξ is a function of the pair of random variables (η', β). This implies that

$$m \ln 2 = H(\xi) = I(\xi, (\eta', \beta)) = I(\xi, \eta') + MI(\xi, \beta | \eta') \leqslant I(\xi, \eta') + H(\beta). \tag{48}$$

By the definition of the transmission capacity of the channel

$$I(\xi, \eta') \leqslant I(\eta, \eta') \leqslant nC(\tilde{p}_0) = n(1 - H(\tilde{p}_0)). \tag{49}$$

Comparing (47)-(49), we find

$$m \ln 2 - mH(\bar{p}_e) \leqslant nC(\tilde{p}_0). \tag{50}$$

We put that $\tilde{p}_0 = \dfrac{t}{\tau} - \sqrt{\dfrac{\tilde{\sigma}^2 \ln \tau}{\tau}}$, $\tilde{\sigma}^2 = \tilde{p}_0(1 - \tilde{p}_0)$, and $\tilde{\zeta}_1, \ldots, \tilde{\zeta}_n$ are independent binary random variables with $p\{\tilde{\zeta}_i = 1\} = \tilde{p}_0$, $i = \overline{1, n}$. Using the inequality (23), we obtain

$$\bar{p}_e \leqslant \sum_{l=t+1}^{\tau} C_\tau^l \tilde{p}_0^l (1 - \tilde{p}_0)^{\tau - l} = p\left\{\sum_{i=1}^{\tau} \tilde{\zeta}_i > t\right\} \leqslant$$
$$\leqslant \frac{\varepsilon_1}{\sqrt{\tau}} + \frac{1}{\sqrt{2\pi\tau \ln \tau}}.$$

Moreover, from this

$$H(\bar{p}_e) \leqslant H\left(\frac{\varepsilon_1}{\sqrt{\tau}}\right) - \frac{1}{\sqrt{2\pi\tau \ln \tau}} \ln \frac{\frac{\varepsilon_1}{\sqrt{\tau}}}{1 - \frac{\varepsilon_1}{\sqrt{\tau}}}, \tag{51}$$

$$C(\tilde{p}_0) = \ln 2 - H(\tilde{p}_0) \leqslant \ln 2 - H\left(\frac{t}{\tau}\right) - \sqrt{\frac{\tilde{\sigma}^2 \ln \tau}{\tau}} \ln \frac{p_0}{1 - p_0}. \tag{52}$$

Putting $\rho = t/\tau$ we deduce from (50)-(52) that if the error probability

$$P_e(f, \varphi) = 1 - \sum_{l=1}^{t} C_\tau^l p_0^l (1 - p_0)^{\tau - l} \geqslant$$
$$\geqslant \frac{p_0(1 - \rho)}{4(1 - p_0)\rho\sqrt{2\pi\tau\rho(1 - \rho)}} e^{-\tau\left[\rho \ln \frac{\rho}{p_0} + (1 - \rho) \ln \frac{1 - p_0}{1 - \rho}\right]}, \tag{53}$$

the transmission rate must satisfy the relation

$$R = \frac{m}{n} \ln 2 \leqslant \frac{C(\tilde{p}_0)}{1 - \frac{H(\bar{p}_e)}{\ln 2}} \leqslant$$
$$\leqslant \frac{\ln 2 - H(\rho) - \sqrt{\frac{\rho(1 - \rho) \ln \tau}{\tau}} \ln \frac{p_0}{1 - p_0}}{1 - \left[H\left(\frac{\varepsilon_1}{\sqrt{\tau}}\right) - \frac{1}{\sqrt{2\pi\tau \ln \tau}} \ln \frac{\frac{\varepsilon_1}{\sqrt{\tau}}}{1 - \frac{\varepsilon_1}{\sqrt{\tau}}}\right] \frac{1}{\ln 2}}. \tag{54}$$

The theorem is proven.

Remark. The expression (37) for the transmission rate differs from the corresponding expression in [7] (formula (7.50)) by the presence of a term of order $\sqrt{\dfrac{\ln^2 \tau}{\tau}}$.

However, an exact bound has still not been found for the error probability of the random method of constructing nonblock codes (in particular, convolutional codes). It remains not clear whether the presence of the term $O\left(\sqrt{\frac{\ln^2\tau}{\tau}}\right)$ is caused by an imperfection of the method of obtaining the estimate, or whether this term reflects an improvement of the asymptotic error probability for nonblock transmission as $R \to C$.

As $R \to 0$ the asymptotic error probability for nonblock transmission was found in [10, 11]: it happens that the exponential term of this asymptotic function is also retained in nonblock transmission; the latter fact is implied by the following theorem.

Theorem 6. In transmission over a binary symmetric memoryless channel with decoding delay and rate $R > 0$

$$P_e(f, \varphi) \geq e^{-\frac{\tau}{4}\ln\frac{1}{4p_0(1-p_0)}+O(\ln\tau)}. \quad (55)$$

§5. Transmission with Feedback

In this section we show that many of the results established in the preceding sections are preserved if feedback is present. The definition of transmission with decoding delay in the presence of feedback is the same as the corresponding definition for the case where feedback is absent. Only the concept of coding is changed, namely, instead of $y_j = f(\bar{x})$ the functions $x_j = f_j(\bar{x}, y_1', \ldots, y_{j-1}')$ are considered, that is, the value of the j-th symbol of the code sequence is a function of the message sent and of the values of the symbols $y', \ldots, y_{j-1}'$ of the output signal.

Theorem 7. In the transmission of messages with decoding delay τ with feedback and with the correction of t errors

$$\frac{1}{n}H(\xi) = \frac{m}{n}\ln 2 \leq \frac{1}{n}\ln\frac{q^n}{|B_{n,t}|}. \quad (56)$$

(Formula (56) is similar to formula (7).) To prove the theorem it is sufficient to notice that in the transmission of the same message value $\bar{x} = (x_1, \ldots, x_m)$ to different values of "noise" $\bar{z}'$ and $\bar{z}''$ there correspond different values of the output signals $\bar{y}'$ and $\bar{y}''$, that is, the set $B_{\bar{x}}$ is such that $\varphi(B_{\bar{x}}) = \bar{x}$, $\cup B_{\bar{x}} = \bar{Y}$ has power not less than $B_{n,t}$.

Theorem 8. In transmission over a binary symmetric channel with feedback with decoding delay τ the bounds (36) and (37) are preserved.

The proof of Theorem 5 is based on bounds of the transmission capacity of a binary symmetric memoryless channel. The result of Theorem 8 easily follows from the proof of Theorem 5, if use is made of the fact that the presence of feedback does not increase the transmission capacity of a memoryless channel.

REFERENCES

1. J. M. Wozencraft and B. Reiffen, Sequential Decoding, New York, London, The M.I.T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., 1961.
2. James L. Massey, Threshold Decoding, Cambridge, Mass., The M.I.T. Press, 1963.
3. R. L. Dobrushin, "On the subject of sequential decoding by the method of Wozencraft-Reiffen," collection: Problems of Cybernetics [in Russian], vol. 12, pp. 113–123, "Nauka," Moscow, 1964.
4. J. M. Wozencraft and I. M. Jakobs, Principles of Communication Engineering, New York, London, Sydney, John Wiley and Sons, Inc., 1965.
5. V. Yu. Vershubskii and V. M. Mikhelev, "Application of special methods of encoding to increase the capacity of registers using magnetic recording," Zhurn. vychislit. matem. i matem. fiz., vol. 6, no. 3, pp. 548–554, 1966.
6. A. D. Syner and R. B. Ash, "Analysis of recurrent codes," IEEE Trans. Inform. Theory, 9, 3, pp. 143–156, 1963.
7. R. M. Fano, Transmission of Information, New York, London, The M.I.T. Press, Cambridge, Mass., John Wiley and Sons, Inc., 1961.
8. S. M. Johnson, "Improved asymptotic bounds for error-correcting codes," IEEE Trans. Inform. Theory, 9, 3, 198–205, 1963.
9. L. A. Bassalygo, "New upper bounds for error correcting codes," Problemy peredachi informatsii [Problems of Information Transmission], vol. 1, no. 4, pp. 41–44, 1963.
10. P. Elias, "Coding for noisy channels," IRE Convention Record, 4, 37, 1955.
11. Mao Shih-sung, "Asymptotic behavior of the optimal error probability on the transmission of information over a memoryless channel symmetric with respect to a pair of input states at low transmission rates," Teor veroyat. i ee primen.
12. V. M. Zolotarev, "An absolute bound of the remainder term in the central theorem," Teoriya veroyat. i ee primen., vol. 11, no. 1, pp. 99–119, 1966.

19 April 1967