On the Reliability Function for a BSC with Noiseless Feedback at Zero Rate

M. V. Burnashev

Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia
email: burn@iitp.ru

Abstract — We consider the transmission of nonexponentially many messages through a binary symmetric channel with noiseless feedback. We obtain an upper bound for the best decoding error exponent. Combined with the corresponding known lower bound, this allows to find the reliability function for this channel at zero rate.

Key words: reliability function, noiseless feedback.

1 Introduction and Main Results

The binary symmetric channel $\text{BSC}(p)$ with crossover probability $0 < p < 1/2$, $q = 1 - p$, and noiseless feedback is considered. We consider the case when the overall transmission time $n$ and $M_n = 2^{Rn}$, $0 < R < 1$, equiprobable messages $\{\theta_1, \theta_2, \ldots, \theta_{M_n}\}$ are given. After the moment $n$ the receiver makes a decision $\hat{\theta}$ on the true message $\theta_{\text{true}}$ transmitted.

Define the minimal possible decoding error probability

$$P_e(M_n, n, p) = \min_{\theta_i} \frac{1}{M_n} \sum_{i=1}^{M_n} P(e|\theta_i),$$

where $P(e|\theta_i)$ — conditional decoding error probability for a transmission method used, provided $\theta_i$ is the true message $\theta_{\text{true}}$, and minimum is taken over all transmission methods of length $n$.

Denote by $F(R, p)$, $0 < R < 1$, the best decoding error exponent for $M_n = e^{Rn}$ codewords over BSC($p$) with noiseless feedback, i.e.

$$F(R, p) = \lim_{n \to \infty} \sup_n \frac{1}{n} \ln \frac{1}{P_e(M_n, n, p)}, \quad M_n = e^{Rn},$$

where $P_e(M_n, n, p)$ is defined in (1). Clearly, the function $F(R, p)$ is non-increasing in $R$.

Introduce also the limiting value $F(0, p)$

$$F(0, p) = \lim_{R \to 0} F(R, p), \quad 0 < p < 1/2.$$

The reported study was funded by RFBR according to the research project 19-01-00364.
The limit in (3) is well-defined, since the function $F(R, p)$ is bounded and non-increasing in $R$.

Equivalently, the function $F(0, p)$ is defined by (2), if the number of messages $M_n$ is such that $M_n \to \infty$, but $\log M_n = o(n)$ as $n \to \infty$.

Similarly, define by $F_K(p)$, $K = 2, 3, \ldots$, the best error exponent for $K$ codewords over BSC(p) with noiseless feedback, i.e.

$$F_K(p) = \limsup_{n \to \infty} \frac{1}{n} \ln \frac{1}{P_e(K, n, p)}, \quad (4)$$

where $P_e(K, n, p)$ – minimal possible decoding error probability (for all transmission methods of length $n$). It was shown in the paper [1], that

$$F_3(p) = F_4(p) = \ldots = F(0, p), \quad (5)$$

and therefore for investigation of the function $F(0, p)$ it is sufficient to find the value $F_3(p)$.

Denote by $E_k(p)$, $k \geq 2$, the best error exponent for $k$ codewords over BSC(p). Clearly,

$$E_2(p) = F_2(p) = \frac{1}{2} \ln \frac{1}{4pq}.$$ 

It is also clear that $E_3(p)$ is defined by $n$-simplex code $(x_1, x_2, x_3)$ (i.e. the code, for which $d(x_i, x_j) \approx 2n/3$ for all $i \neq j$), and therefore

$$E_3(p) = \frac{1}{3} \ln \frac{1}{4pq}.$$ 

Clearly, we have $E_3(p) \leq F_3(p) \leq E_2(p)$.

The next result was proved in [1].

**Proposition.** For $P_e(3, n, p)$ the upper bound holds (see (4), (5))

$$P_e(3, n, p) \leq \left(\frac{q}{p}\right)^{1/3} \left(p^{1/3}q^{2/3} + p^{2/3}q^{1/3}\right)^n. \quad (6)$$

It follows from (3), (5) and (6) that

$$F_3(p) \geq F_{fb}(p), \quad (7)$$

where

$$F_{fb}(p) = -\ln\left(p^{1/3}q^{2/3} + p^{2/3}q^{1/3}\right) = -\ln\left[p^{1/3}q^{2/3}(1 + z^{1/3})\right] \geq 0 \quad (8)$$

and

$$p + q = 1, \quad z = z(p) = p/q.$$ 

In particular, the formula (7) followed also from subsequent papers [2, 3] (where by other methods the whole reliability function $F(R)$ was investigated).
Moreover, it was also claimed in [1] that the opposite to (7) inequality holds
\[ F_3(p) \leq F_{lb}(p), \]  
and if the formula (9) is correct, then from (7) the equality would follow
\[ F_3(p) = F_{lb}(p). \]  

However, there was no rigorous proof of formula (9) in [1]. Later, in [4] one more attempt to establish formula (9) was done (using general Bellman’s equation), but later it was found that the proof is also incorrect.

Below in the paper formula (9) is proved, and therefore formula (10) holds.

We describe all possible transmission methods of one of three messages over BSC with noiseless feedback. Note that any reasonable transmission strategy has the following form.

At each instant \( k, k = 1, \ldots, n \), based on previous channel outputs \( y_{k-1} \), the receiver selects some message \( \theta_{i_0} \), and asks the transmitter whether \( \theta_{i_0} \) is the true message \( \theta_{true} \). Here the noiseless feedback is important! If the true message \( \theta_{true} \) coincides with \( \theta_{i_0} \), i.e. \( \theta_{true} = \theta_{i_0} \), then the signal \( x_k = 0 \) is transmitted. If \( \theta_{true} \neq \theta_{i_0} \), then the signal \( x_k = 1 \) is transmitted. After the instant \( n \) the decision is made in favor of the most probable message \( \theta_i \).

The transmission strategy used in papers [1, 2, 3] is quite natural: at each instant \( k \) as \( \theta_{i_0}(k) \) the most probable message \( \theta_i \), provided output \( y^{k-1} \) is selected. It seems that such transmission strategy provides the best decoding error exponent \( F_3(p) \) (that statement needs to be proved, what was not done in [1, 2, 3]).

The main result of the paper is as follows.

**Theorem 1.** For \( P_e(3, n, p) \) the lower bound is valid
\[ P_e(3, n, p) \geq \frac{1}{2} (p^{1/3} q^{2/3} + p^{2/3} q^{1/3})^n = \frac{1}{2} [p^{1/3} q^{2/3} (1 + z^{1/3})]^n, \quad z = p/q, \]  
and therefore for \( F_3(p) \) the formula holds (see (8))
\[ F_3(p) = F(0, p) = F_{lb}(p). \]  

Note that,
\[ E_2(p) = F_2(p) > F_3(p) > E_3(p), \quad 0 < p < 1/2. \]  

For BSC output denote \( y^k = y^k_1 = (y_1, \ldots, y_k), \) \( k = 1, \ldots, n, \) \( y_k \in \{0, 1\}. \)

**Remark 1.** Let us explain why for three messages noiseless feedback can help to improve the decoding error probability. Indeed, assume that at a time instant \( i \) we have
\[ p(y^i | x_1) \approx p(y^i | x_2) \gg p(y^i | x_3), \]
i.e., the message \( \theta_3 \) is much less probable than messages \( \theta_1, \theta_2 \) (and due to noiseless feedback it is known at the transmitter!). Then for time instants \( t > i \) we may mainly test remaining messages \( \theta_1 \) and \( \theta_2 \) (for example, using for that purpose opposite code blocks, as for two
messages). Since \( E_2(p) > E_3(p) \) (see (13)), then such coding would allow to decrease the decoding error probability.

**Remark 2.** The right-hand side of the formula (8) has the following useful interpretation (not recorded earlier). Let \((x_1, x_2, x_3) - n\)-simplex code (i.e., the code, such that \(d(x_i, x_j) \approx 2n/3 \) for all \(i \neq j\)). Then the formula holds (see proof in Appendix)

\[
P\{\mathcal{E}_n\} \sim e^{-E_{fb}(p)n},
\]  

(14)

where

\[
\mathcal{E}_n = \{y^n: p(y^n|\theta_1) \approx p(y^n|\theta_2) \approx p(y^n|\theta_3)\}.
\]  

(15)

In other words, the event \(\mathcal{E}_n\) defines the exponent of the decoding error probability.

**Remark 3.** From the viewpoint of reliability functions behavior the channel BSC(p) and Gaussian channel \(G(A)\) with limitation on average power \(A\) are mainly similar to each other [5] [6] [7]. But the same channels with noiseless feedback show also an essential difference. In particular, for BSC(p) we have \(F_3(p) < E_2(p)\) (see (13)), while for Gaussian channel \(G(A)\) we have \(F_3(A) = E_2(A)\) (see [7]). That difference is based on the feature that for Gaussian channel \(G(A)\) at some instants it is possible to transmit very strong signals, while it is impossible for the channel BSC(p).

Next result describes the best transmission method in the case of three messages.

**Theorem 2.** At each instant \(k, k = 1, \ldots, n\), the best partition of messages \(\{\theta_1, \theta_2, \theta_3\}\) (minimizing the decoding error probability \(P_{e}(n)\)) has the form: the most probable message (provided output \(y^{k-1}\)) versus two remaining messages.

The paper is organized as follows. For a completeness purpose, in § 2 a short and elegant proof of formula (13) from [11] is presented (it seems that such proof is available only in the thesis [11] and it was not published in other more available sources). In § 3 Theorem 2 is proved. In § 4 the Markov diagram for the decoder of the optimal transmission strategy is introduced and described. In § 5 using that diagram Theorem 1 is proved.

## 2 Proof of Proposition

By \(d(y, x)\) we denote the Hamming distance between vectors \(y\) and \(x\). For each instant \(k, k = 1, \ldots, n\), by \(\theta^{(1)}(k), \theta^{(2)}(k), \theta^{(3)}(k)\) we denote the ordering of messages \(\theta_1, \theta_2, \theta_3\) provided \(y^k\), such that

\[
p(y^k|\theta^{(1)}(k)) \geq p(y^k|\theta^{(2)}(k)) \geq p(y^k|\theta^{(3)}(k)).
\]  

(16)

By \(x^{(1)}(k), x^{(2)}(k), x^{(3)}(k)\) we denote the corresponding ordering of codewords used. Then (16) is equivalent to the ordering

\[
d(y^k, x^{(1)}(k)) \leq d(y^k, x^{(2)}(k)) \leq d(y^k, x^{(3)}(k)).
\]  

We call \(d^{(i)}(k) = d(y^k, x^{(i)}(k))\) as the number of “negative votes” against \(\theta^{(i)}(k)\) during the time \(k\). Denote also \(d_i = d_i(n)\).
Denote by \(d_{1,3}(k), k = 1, \ldots, n\), the average number of “negative votes” against all messages during the time \(k\), i.e.,

\[
d_{1,3}(k) = \frac{1}{3} \sum_{i=1}^{3} d^{(i)}(k).
\]

(17)

We use the strategy, when at each instant \(k\) the most probable message \(\theta^{(1)}(k)\) is selected, and the transmitter answers whether \(\theta^{(1)}(k)\) is the true message \(\theta_{\text{true}}\). If \(\theta_{\text{true}} = \theta^{(1)}(k)\), then the transmitter sends the signal \(x_k = 0\), while if \(\theta_{\text{true}} \neq \theta^{(1)}(k)\), then the signal \(x_k = 1\) is sent.

Therefore, if the output signal \(y_k = 1\), then the message \(\theta^{(1)}(k)\) gets one additional negative vote, while remaining two messages \(\{\theta^{(2)}(k), \theta^{(3)}(k)\}\) do not get additional negative votes. If the output signal \(y_k = 0\), then the message \(\theta^{(1)}(k)\) does not get additional negative votes, while each of remaining messages \(\theta^{(2)}(k)\) and \(\theta^{(3)}(k)\) gets one additional negative vote. As a result, if \(y_k = 1\), then the value \(d_{1,3}\) from (17) increases by \(1/3\). If \(y_k = 0\), then the value \(d_{1,3}\) increases by \(2/3\). If \(m\) zeros and \(n - m\) ones were received on the output during the total time \(n\), then \(d_{1,3}(n) = (n + m)/3\). There are \(\binom{n}{m}\) ways to set \(m\) zeros on \(n\) positions.

For each instant \(k\) the following inequalities hold

\[
d^{(1)}(k) \leq d^{(2)}(k) \leq d^{(3)}(k) \leq d^{(2)}(k) + 1.
\]

(18)

Only the last one of inequalities (18) should be explained. Indeed, it is true for \(k = 1\) (i.e., after getting an output \(y_1\)). Further, for \(k \geq 2\) for the strategy used messages \(\theta^{(2)}(k)\) and \(\theta^{(3)}(k)\) always fall in one group, and therefore the condition \(d^{(3)}(k) \leq d^{(2)}(k) + 1\) remains valid (although messages \(\theta^{(2)}(k)\) and \(\theta^{(3)}(k)\) themselves may change).

From (17) and (18) the inequality follows

\[
d^{(2)}(k) \geq d_{1,3}(k) - 1/3.
\]

(19)

Note that each realization of an output \(y^n\) with \(e\) errors has the probability \(p^e q^{n-e}\). Since the true message gets \(e\) negative votes, then for decoding error it is necessary to have \(d^{(2)}(n) = e\) or \(d^{(3)}(n) = e\). In either case, by (19) we need \(e \geq d_{1,3}(n) - 1/3\), and therefore it is necessary to have

\[
p^e q^{n-e} \leq \left(\frac{q}{p}\right)^{1/3} p^{d_{1,3}(n)} q^{n-d_{1,3}(n)}.
\]

(20)

Condition (20) bounds the probability of any erroneous path via the value \(d_{1,3}(n)\). Note that if \(m\) – the number of zeros, received at output during all time \(n\), then the value \(d_{1,3}(n) = (n + m)/3\), \(m = 0, 1, \ldots, n\) corresponds to every erroneous path. Since there are \(\binom{n}{m}\) ways to distribute \(m\) zeros on \(n\) positions, then by (20) we get

\[
P_e(3, n, p) \leq \left(\frac{q}{p}\right)^{1/3} \sum_{m=0}^{n} \binom{n}{m} p^{(n+m)/3} q^{(2n-m)/3} = \left(\frac{q}{p}\right)^{1/3} \left(p^{1/3} q^{2/3} + p^{2/3} q^{1/3}\right)^n,
\]

from where (6) follows. △
3 Proof of Theorem 2

We consider transmission of three equiprobable messages \(\{\theta_1, \theta_2, \theta_3\}\). After each instant \(k\) we find posterior message probabilities \(\pi_i(k), i = 1, 2, 3\), based on received block \(y^k = y_1^k = (y_1, \ldots , y_k), k = 1, \ldots , n\). Transmission at instant \(k + 1\) depends only on probabilities \(\{\pi_i(k)\}\) (since they constitute a sufficient statistics). We may assume that at instant \(k + 1\) we start transmission, but using prior probabilities \(\{\pi_i(k)\}\).

We denote by \(d_i(k) = d_i(y^k) = d(y^k, x_i(k))\) the total number of “negative votes” against \(\theta_i\) during the time \([1, k]\). Denote also \(d_i = d_i(n)\).

All information the decoder has at an instant \(k, k = 1, \ldots , n\) after receiving an output \(y^k\), are posterior probabilities \(\pi_i(y^k)\) of messages \(\theta_i, i = 1, 2, 3\) (or, equivalently, the set of distances \(d_i(y^k), i = 1, 2, 3\). Denote by \(i_0(y^k)\) \(\in \{1, 2, 3\}\), the index providing the maximal value to \(\pi_i(y^k)\) (or, equivalently, the minimal value to \(d_i(y^k)\), i.e.

\[
\pi_{i_0(y^k)}(y^k) = \max_i \pi_i(y^k), \quad d_{i_0(y^k)}(y^k) = \min_i d_i(y^k), \quad k = 1, \ldots , n. \tag{21}
\]

From the decoder viewpoint the value \(\pi_i(y^n)\) is the posterior probability of the event \(\{\theta_i = \theta_{\text{true}}\}\). Therefore, best (from the decoding error probability viewpoint) is to make the decision in favor of the message \(\theta_{i_0}(y^n)\) with the maximal posterior probability \(\pi_{i_0}(y^n)(y^n)\). Then we have by (21)

\[
P_e(n) = P\{\theta_{i_0(y^n)} \neq \theta_{\text{true}}\} = 1 - E I_{\theta_{i_0(y^n)} = \theta_{\text{true}}} = 1 - E \pi_{i_0(y^n)}. \tag{22}
\]

By \(A_k \in \{\theta_1, \theta_2, \theta_3\}, k = 1, \ldots , n\), we denote the message, selected by the receiver at instant \(k\), on which it asks the question, whether \(A_k\) is the true message \(\theta_{\text{true}}\).

Consider changing of the value \(\pi_{i_0(y^k)}\) from (21), (22) depending on a choice of the message \(A_{k+1}\). For that purpose it is sufficient to consider changing of the value

\[
\sum_{j \neq i_0} z^{d_j(k) - d_{i_0}(k)},
\]

where \(z = p/q\), i.e., changing of the value \(1/\pi_{i_0(y^k)}\) (see formulas (37), (38)).

Two cases are possible:

1. There exists a unique index \(i_0(y^k)\), such that \(d_j(y^k) - d_{i_0(y^k)}(y^k) \geq 1\) for all \(j \neq i_0(y^k)\). Then \(i_0(y^{k+1}) = i_0(y^k)\) for all \(y_{k+1}\). In that case the most probable message \(\theta_{i_0(y^k)}\) at instant \(k\) remains the same for instant \(k + 1\) for any output \(y_{k+1}\).

2. There are two different indices \(i_0(y^k)\) and \(i_1(y^k)\), such that \(d_{i_0(y^k)}(y^k) = d_{i_1(y^k)}(y^k)\) and \(d_j(y^k) - d_{i_0(y^k)}(y^k) \geq 1\) for the third index.

It is clear that in the third possible case (when all distances \(d_j(y^k), j = 1, 2, 3\), are equal) due to symmetry any choice of \(A_{k+1}\) (i.e. any partition of messages) leads to the same result.
Consider first the case 1. Denote for short (where $z = p/q$)

$$
a_j = a_j(y^k) = z^{d_j(y^k) - d_{i_0}(y^k)},
\delta_j = \delta_j(y_{k+1}) = d_j(y_{k+1}) - d_{i_0}(y^k)(y_{k+1}),
B(k, y^k) = \sum_{j \neq i_0} z^{d_j(y^k) - d_{i_0}(y^k)} = \sum_{j \neq i_0} a_j,
$$

(23)

$$
B(k + 1) = \sum_{j \neq i_0} z^{d_j(y^{k+1}) - d_{i_0}(y^{k+1})} = \sum_{j \neq i_0} z^{d_j(y^{k+1}) - d_{i_0}(y^k)} = \sum_{j \neq i_0} a_j \delta_j(y_{k+1}).
$$

Denote also

$$
B_j(k + 1) = B(k + 1) \quad \text{if} \quad A_{k+1} = \theta_j, \quad j = 1, 2, 3.
$$

(24)

Note that values $\delta_j, j = 1, 2, 3$, take on only values 0, 1 and -1. Without loss of generality we may assume that $i_0(y^k) = 1$, and therefore $i_0(y^{k+1}) = 1$. Then we have $\delta_1(y_{k+1}) = 0$ and

$$
B(k) = a_2 + a_3, \quad B(k + 1) = a_2 z^{\delta_2(y_{k+1})} + a_3 z^{\delta_3(y_{k+1})},
\pi_1(k) = \frac{1}{1 + B(k)}, \quad \pi_1(k + 1) = \frac{1}{1 + B(k + 1)}.
$$

(25)

Consider distributions of the random variables $B_j(k + 1), j = 1, 2, 3$, provided $i_0(y^k) = 1$. For $j = 1$, i.e. $A_{k+1} = \theta_1$, we have

$$
\delta_2 = \delta_3 = \begin{cases} 1 & \text{with probability } \pi_1(k)q + (1 - \pi_1(k))p = p + (q - p)\pi_1(k), \\ -1 & \text{with probability } q - (q - p)\pi_1(k), \end{cases}
$$

(26)

and then

$$
B_1(k + 1) = \begin{cases} (a_2 + a_3)z = B(k)z & \text{with probability } p + (q - p)\pi_1(k), \\ (a_2 + a_3)/z = B(k)/z & \text{with probability } q - (q - p)\pi_1(k). \end{cases}
$$

(27)

For $j = 2$, i.e., $A_{k+1} = \theta_2$, we have

$$
\delta_3 = 0, \quad \delta_2 = \begin{cases} 1 & \text{with probability } \pi_1(k)q + (1 - \pi_1(k))p = p + (q - p)\pi_1(k), \\ -1 & \text{with probability } q - (q - p)\pi_1(k), \end{cases}
$$

(28)

and therefore

$$
B_2(k + 1) = \begin{cases} a_2 z + a_3 & \text{with probability } p + (q - p)\pi_1(k), \\ a_2/z + a_3 & \text{with probability } q - (q - p)\pi_1(k). \end{cases}
$$

(29)

Similarly, for $j = 3$, i.e., $A_{k+1} = \theta_3$, we have

$$
\delta_2 = 0, \quad \delta_3 = \begin{cases} 1 & \text{with probability } p + (q - p)\pi_1(k), \\ -1 & \text{with probability } q - (q - p)\pi_1(k), \end{cases}
$$

(30)
and then
\[ B_3(k + 1) = \begin{cases} a_3 z + a_2 & \text{with probability } p + (q - p)\pi_1(k), \\ a_3 z/a_2 + a_2 & \text{with probability } q - (q - p)\pi_1(k). \end{cases} \] (31)

As a result, we have for \( i_0(y^k) = 1 \)
\[
E_1 = E[\pi_{i_0}(k + 1) | y^k, A_{k+1} = \theta_1] = E\left[ \frac{1}{1 + B_1(k + 1)} \right] y^k, A_{k+1} = \theta_1
\]
\[ = \frac{p + (q - p)\pi_1(k)}{1 + (a_2 + a_3)z} + \frac{q - (q - p)\pi_1(k)}{1 + (a_2 + a_3)/z}. \] (32)

Similarly we have
\[
E_2 = E[\pi_{i_0}(k + 1) | y^k, A_{k+1} = \theta_2] = E\left[ \frac{1}{1 + B_2(k + 1)} \right] y^k, A_{k+1} = \theta_2
\]
\[ = \frac{p + (q - p)\pi_1(k)}{1 + a_2 z + a_3} + \frac{q - (q - p)\pi_1(k)}{1 + a_2/z + a_3}. \] (33)

and
\[
E_3 = E[\pi_{i_0}(k + 1) | y^k, A_{k+1} = \theta_3] = E\left[ \frac{1}{1 + B_3(k + 1)} \right] y^k, A_{k+1} = \theta_3
\]
\[ = \frac{p + (q - p)\pi_1(k)}{1 + a_2 + a_3 z} + \frac{q - (q - p)\pi_1(k)}{1 + a_2 + a_3/z}. \] (34)

We shall show that \( E_1 \geq \max\{E_2, E_3\} \), what means that best is to use \( A_{k+1} = \theta_1 = \theta_{i_0} (y^k) \). Due to symmetry it is sufficient to show that \( E_1 \geq E_2 \). Indeed, by (32) and (33) we have
\[
E_1 - E_2 = \left[ \frac{p + (q - p)\pi_1(k)}{1 + (a_2 + a_3)z} - \frac{1}{1 + a_2 z + a_3} \right] + \left[ \frac{q - (q - p)\pi_1(k)}{1 + (a_2 + a_3)/z} - \frac{1}{1 + a_2/z + a_3} \right]
\]
\[ = \frac{[p + (q - p)\pi_1(k)]a_3(1 - z)}{[1 + (a_2 + a_3)z][1 + a_2 z + a_3]} + \frac{[q - (q - p)\pi_1(k)]a_3(1 - 1/z)}{[1 + (a_2 + a_3)/z][1 + a_2/z + a_3]}
\]
\[ = a_3(1 - z) \left\{ \frac{p + (q - p)\pi_1(k)}{[1 + (a_2 + a_3)z][1 + a_2 z + a_3]} - \frac{q - (q - p)\pi_1(k)}{z[1 + (a_2 + a_3)/z][1 + a_2/z + a_3]} \right\}
\]
\[ = qa_3(1 - z) \left\{ \frac{z + (1 - z)\pi_1(k)}{[1 + (a_2 + a_3)z][1 + a_2 z + a_3]} - \frac{1 - (1 - z)\pi_1(k)}{(z + a_2 + a_3)(1 + a_2/z + a_3)} \right\}. \] (35)

It is sufficient to show, that
\[
\frac{z + (1 - z)\pi_1(k)}{[1 + (a_2 + a_3)z][1 + a_2 z + a_3]} - \frac{1 - (1 - z)\pi_1(k)}{(z + a_2 + a_3)(1 + a_2/z + a_3)} \geq 0,
\]
or, equivalently (after a standard algebra using the formula $\pi_1(k) = 1/(1 + a_2 + a_3)$),

$$a_2(1 - z) \geq 0. \quad (36)$$

The relation (36) holds, if $z \leq 1$ (i.e., if $p \leq 1/2$). By (35) and (36) we have $E_1 \geq E_2$. Similarly we get $E_1 \geq E_3$. Therefore, $E_1 \geq \max\{E_2, E_3\}$, what means that best is to use $A_{k+1} = \theta_1 = \theta_{i_0(y^k)}$. It completes considering of the case 1).

Consider now the case 2), when there are two different indices $i_0(y^k)$ and $i_1(y^k)$, such that $d_{i_0(y^k)}(y^k) = d_{i_1(y^k)}(y^k)$ and $d_j(y^k) - d_{i_0(y^k)}(y^k) \geq 1$ for the third index. Without loss of generality we may assume that $i_0(y^k) = 1$ and $i_1(y^k) = 3$. Then $E_1 = E_3$, and it remains to show that $E_1 \geq E_2$, and then best is to use $A_{k+1} = \theta_1 = \theta_{i_0(y^k)}$ (or $A_{k+1} = \theta_3 = \theta_{i_1(y^k)}$). Note that for any $y_{k+1}$ one of distances $d_{i_0(y^{k+1})}(y^{k+1})$ or $d_{i_1(y^{k+1})}(y^{k+1})$ remains the same as earlier for the instant $k$. Remaining calculations essentially coincide with (23)-(36) (in fact, they are even simpler) and we omit them. It completes the proof of Theorem 2. $\triangle$

4 Markov diagram of the optimal strategy decoder

Introduce the Markov chain describing the decoder evolution in time. Denote by $d_i(k) = d(y^k, x_i(k))$ the total number of “negative votes” against $\theta_i$ during the transmission period $[1,k]$. Denote also $d_i = d_i(n)$. Then ($z = p/q < 1$)

$$\pi_i(k) = \frac{z^{d_i(k)}}{\sum_{j=1}^{3} z^{d_j(k)}} = \frac{1}{1 + \sum_{j \neq i} z^{d_j(k) - d_i(k)}} \quad \pi_i(n) = \frac{1}{1 + \sum_{j \neq i} z^{d_j(n) - d_i(n)}}. \quad (37)$$

Note that,

$$\frac{\pi_i(k)}{1 - \pi_i(k)} = \sum_{j \neq i} \frac{z^{d_i(k)}}{z^{d_j(k)}} = \sum_{j \neq i} \frac{1}{z^{d_j(k) - d_i(k)}}. \quad (38)$$

For each instant $k$ and each output $y^k$ define for a message $\theta_i$ the metrics $m_i(k, y^k)$ as follows:

$$m_i(k, y^k) = d(y^k, x_i(k)) - \min_j d(y^k, x_j(k)) = d_i(k) - \min_j d_j(k), \quad i = 1, 2, 3. \quad (39)$$

Clearly, $m_i(k, y^k) \geq 0$ and $\min m_i(k, y^k) = 0$. The set $\{m_i(k, y^k)\}$ is a sufficient statistics, since it defines posterior probabilities $\{\pi_i(k)\}$ (see (37)–(39)).

Denote by $S_{ij\ell} = S_{ij\ell}(k) = S_{ij}(k, y^k)$ the chain state with $i = m_1(k, y^k), j = m_2(k, y^k), \ell = m_3(k, y^k)$.

As a result, the whole diagram looks like an “octopus” with nine “tentacles” (see Fig. 1). For example, one of such “tentacles” is $(S_{011}, S_{022}, S_{033}, \ldots)$. We call $S_{000}$ the main state and six states $\{S_{011}, S_{100}, S_{101}, S_{010}, S_{110}, S_{001}\}$ – basic states. Remaining states are on “tentacles”. 
For the decoding error probability $P_e(n)$ we have

$$P_e(n) \geq \frac{2}{3} P_0(n),$$

where

$$P_0(n) = P\{S_{000}(0) \Rightarrow S_{000}(n)\}. \quad (41)$$

We describe transitions among states for the optimal strategy. Without loss of generality we may assume that $\theta_{\text{true}} = \theta_1$.

If at instant $k$ the decoder is in the state $S_{000}(k)$, then the set $A(k+1)$ is chosen equiprobably among three possible variants. As a result, for next possible state $S(k+1)$ we get

$$S_{000}(k) \rightarrow \begin{cases} 
S_{011}(k+1) & \text{with probability } q/3, \\
S_{100}(k+1) & \text{with probability } p/3, \\
S_{101}(k+1) & \text{with probability } p/3, \\
S_{010}(k+1) & \text{with probability } q/3, \\
S_{110}(k+1) & \text{with probability } p/3, \\
S_{001}(k+1) & \text{with probability } q/3.
\end{cases} \quad (42)$$

Indeed, at an instant $k$ each message $\theta_i$ has the probability $\pi_i(k) = 1/3$, $i = 1, 2, 3$. Therefore with probability $1/3$ we have $A(k+1) = \theta_1$. Since we assumed $\theta_{\text{true}} = \theta_1$, then with probability $q/3$ we get $S(k+1) = S_{011}(k+1)$ and with probability $p/3$ we get $S(k+1) = S_{100}(k+1)$. Similarly remaining lines of (42) are obtained.

The easiest case is to describe transitions from states, for which the set $A(k+1)$ is defined uniquely, without randomization (i.e., when there is only one most probable
message). Such states are $S_{011}(k), S_{101}(k), S_{110}(k), \ldots$. For those states we get

$$S_{011}(k) \rightarrow \begin{cases} S_{000}(k + 1) \text{ with probability } p, \\ S_{022}(k + 1) \text{ with probability } q, \end{cases}$$

(43)

$$S_{101}(k) \rightarrow \begin{cases} S_{000}(k + 1) \text{ with probability } q, \\ S_{202}(k + 1) \text{ with probability } p, \end{cases}$$

(44)

$$S_{110}(k) \rightarrow \begin{cases} S_{000}(k + 1) \text{ with probability } q, \\ S_{220}(k + 1) \text{ with probability } p. \end{cases}$$

(45)

Similarly, transitions from analogous states $S_{022}(k), S_{202}(k), S_{220}(k), \ldots$ are described. Transitions from remaining states $S_{100}(k), S_{010}(k), S_{001}(k)$ are described similarly to (42):

$$S_{100}(k) \rightarrow \begin{cases} S_{101}(k + 1) \text{ with probability } q/2, \\ S_{210}(k + 1) \text{ with probability } p/2, \\ S_{110}(k + 1) \text{ with probability } q/2, \\ S_{201}(k + 1) \text{ with probability } p/2, \end{cases}$$

(46)

$$S_{010}(k) \rightarrow \begin{cases} S_{021}(k + 1) \text{ with probability } q/2, \\ S_{110}(k + 1) \text{ with probability } p/2, \\ S_{120}(k + 1) \text{ with probability } p/2, \\ S_{011}(k + 1) \text{ with probability } q/2, \end{cases}$$

(47)

$$S_{001}(k) \rightarrow \begin{cases} S_{012}(k + 1) \text{ with probability } q/2, \\ S_{101}(k + 1) \text{ with probability } p/2, \\ S_{102}(k + 1) \text{ with probability } p/2, \\ S_{011}(k + 1) \text{ with probability } q/2. \end{cases}$$

(48)

5 Proof of Theorem 1

By (40), (41) it is sufficient to estimate from below the value $P_0(n) = P\{S_{000}(0) \Rightarrow S_{000}(n)\}$. Clearly,

$$P_0(n) = \sum_{t_n} P\{t_n\},$$

(49)

where the sum is taken over all paths $t_n$ of length $n$ and of the form $S_{000}(0) \Rightarrow S_{000}(n)$.

We call 3-path any path of length 3 and of the form $S_{000}(k) \Rightarrow S_{000}(k + 3)$. We call also 2-path any path of length 2 and of the form $S_{000}(k) \Rightarrow S_{000}(k + 2)$.

First, we limit ourselves in the right-hand side of (49) to paths $t_n$, passing only through the main and basic states (i.e., they do not pass through tentacles). It is simple to see that any such path $t_n$ consists of 3-paths and 2-paths.
There are six 3-paths:

\[ S_{000} \rightarrow S_{100} \rightarrow S_{101} \rightarrow S_{000} \quad \text{with probability } pq^2/6, \]
\[ S_{000} \rightarrow S_{100} \rightarrow S_{110} \rightarrow S_{000} \quad \text{with probability } pq^2/6, \]
\[ S_{000} \rightarrow S_{010} \rightarrow S_{011} \rightarrow S_{000} \quad \text{with probability } pq^2/6, \]
\[ S_{000} \rightarrow S_{010} \rightarrow S_{110} \rightarrow S_{000} \quad \text{with probability } pq^2/6, \]
\[ S_{000} \rightarrow S_{001} \rightarrow S_{101} \rightarrow S_{000} \quad \text{with probability } pq^2/6, \]
\[ S_{000} \rightarrow S_{001} \rightarrow S_{011} \rightarrow S_{000} \quad \text{with probability } pq^2/6. \]

Therefore

\[ P\{S_{000}(k) \rightarrow S_{000}(k + 3)\} = pq^2. \] (50)

There are three 2-paths:

\[ S_{000} \rightarrow S_{011} \rightarrow S_{000} \quad \text{with probability } qp/3, \]
\[ S_{000} \rightarrow S_{101} \rightarrow S_{000} \quad \text{with probability } qp/3, \] (51)
\[ S_{000} \rightarrow S_{110} \rightarrow S_{000} \quad \text{with probability } qp/3. \]

Therefore

\[ P\{S_{000}(k) \rightarrow S_{000}(k + 2)\} = pq. \] (52)

We estimate the value \( P_0(n) \) from (49), using (50)–(52). Any path \( t_n \), limited to basic states, consists of some number \( n_2 \) of 2-paths and some number \( n_3 \) of 3-paths. Moreover, \( 2n_2 + 3n_3 = n \), \( 0 \leq n_2 \leq n/2 \), and the total number of paths equals to

\[ m = n_2 + n_3 = \frac{n + n_2}{3}. \]

There are \( \binom{m}{n_2} \) ways to distribute \( n_2 \) 2-paths. Remaining \( m - n_2 \) places are occupied by \( n_3 \) 3-paths. Therefore we have (\( z = p/q \))

\[
P_0(n) = \sum_{n_2=0}^{n/2} \binom{(n + n_2)/3}{n_2} (pq)^{n_2} (pq^2)^{n_3} = \sum_{n_2=0}^{n/2} \binom{(n + n_2)/3}{n_2} (pq)^{n_2} (pq^2)^{n-2n_2}/3 = (pq^2)^{n/3} \sum_{n_2=0}^{n/2} \binom{(n + n_2)/3}{n_2} z^{n_2/3}. \] (53)

We estimate from below the sum in the right-hand side of (53). Maximum of the value \( \binom{(n + n_2)/3}{n_2} z^{n_2/3} \) over \( n_2 \) is attained for \( n_2 = a_0 n \), where the value \( a_0 \) will be found below. Then

\[
\sum_{n_2=0}^{n/2} \binom{(n + n_2)/3}{n_2} z^{n_2/3} \geq \left( \frac{n(1 + a_0)/3}{a_0n} \right)^{a_0n/3} \geq \frac{1}{n} \sum_{n_2=0}^{n(1+a_0)/3} \binom{n(1+a_0)/3}{n_2} z^{n_2/3} = \frac{1}{n} (1 + z^{1/3})^{n(1+a_0)/3}. \] (54)
In order to be accurate, we estimate also from above the sum in the right-hand side of (53). We have
\[
\sum_{n_2=0}^{n} \binom{n + n_2}{n_2} z^{n_2/3} \leq n \left( \frac{n(1 + a_0)/3}{a_0 n} \right)^z a_0 n^{1/3} \leq n \sum_{n_2=0}^{n(1+a_0)/3} \binom{n(1 + a_0)/3}{n_2} z^{n_2/3} = n \left( 1 + z^{1/3} \right)^{n(1+a_0)/3}.
\]

As a result, we get from (53) and (54)
\[
P_0(n) \geq \frac{1}{n} \left( pq^n \right)^{n/3} \left( 1 + z^{1/3} \right)^{n(1+a_0)/3}.
\]

We find now the value \( a_0 \) in (54), (55). Since
\[
\ln \left( \frac{n + n_2}{n_2} \right) \approx \frac{n + n_2}{3} h \left( \frac{3n_2}{n + n_2} \right),
\]
then denoting \( n_2 = an, 0 \leq a \leq 1/2 \), introduce the function
\[
f_1(p, a) = (1 + a) h \left( \frac{3a}{1 + a} \right) - a \ln(q/p), \quad 0 \leq a \leq 1/2.
\]

The value \( a_0 \) maximizes the function \( f_1(p, a) \) over \( 0 \leq a \leq 1/2 \). Note that,
\[
(f_1(p, a))' = \ln p (1 + a)(1 - 2a)^2 \quad 27qa^3, \quad (f_1(p, a))'' = 0, \quad (f_1(p, a))'_{a=0} = \infty, \quad (f_1(p, a))'_{a=1/2} = -\infty.
\]

Therefore, \( a_0(p) \) is the unique root of the equation
\[
27qa^3 - p(1 + a)(1 - 2a)^2 = 0 = (27 - 31p)a^3 + 3pa - p.
\]

For that root we have \[\text{[3] Ch. 1.8-3}\]
\[
a_0(p) = \left[ \frac{p}{2(27 - 31p)} \right]^{1/3} \left\{ 1 + \sqrt[1/3]{\frac{27(1 - p)}{27 - 31p}} \right\}^{1/3} + \left\{ 1 - \sqrt[1/3]{\frac{27(1 - p)}{27 - 31p}} \right\}^{1/3}.
\]

For small \( p \) we have \( 3a_0(p) \approx p^{1/3} \). Since \( a_0 < 2 \), the estimate (55) yields to the upper bound (55)-(56) for \( P_0(3, n, p) \). But the estimate (55) shows that when investigating the value \( P_0(n) \), we may not limit ourselves only to basic states, but should take into account also states on tentacles.
We strengthen the estimate (55), taking also into account states on tentacles. We call by 2-loop any path of length 2 with the same starting and final states (not necessarily states \( S_{000} \)). Besides 2-paths from (51), other examples of 2-loops are also

\[
\begin{align*}
S_{011} &\rightarrow S_{022} \rightarrow S_{011} \quad \text{with probability } qp, \\
S_{100} &\rightarrow S_{201} \rightarrow S_{100} \quad \text{with probability } qp/2, \\
S_{100} &\rightarrow S_{210} \rightarrow S_{100} \quad \text{with probability } qp/2, \\
S_{101} &\rightarrow S_{202} \rightarrow S_{101} \quad \text{with probability } qp, \ldots.
\end{align*}
\]

Such 2-loops go out to tentacles.

We consider paths \( t_n \), consisting of some number \( n_3 \) of 3-paths and some number \( k_2 \) of 2-loops. Assume that we distributed \( n_3 \) 3-paths on \([1, n]\). After that we insert \( k_2 \) 2-loops in any different \( k_2 \) instants on \([1, n]\). If such 2-loop hits on the initial state of a 3-path, then that 3-path is simply moved to the right on two steps. If such 2-loop hits an internal state of a 3-path, then the part of that 3-path is moved to the right on two steps, in order to imbed that 2-loop. Similarly, 2-loops can be inserted into other 2-loops.

Since it is necessary to have \( n = 3n_3 + 2k_2 \), then

\[
P_0(n) \geq \sum_{k_2=0}^{n/2} \binom{n}{k_2} (pq)^{k_2} (pq^2)^{n_3} = \sum_{k_2=0}^{n/2} \binom{n}{k_2} (pq)^{k_2} (pq^2)^{(n-2k_2)/3} = (pq^2)^{n/3} \sum_{k_2=0}^{n/2} \binom{n}{k_2} z^{k_2/3}. \tag{56}
\]

Note that,

\[
\binom{n}{k_2} z^{k_2/3} + \binom{n}{n-k_2} z^{(n-k_2)/3} \leq 2 \binom{n}{k_2} z^{k_2/3}, \quad k_2 \leq n/2, \quad z < 1.
\]

Then (56) can be continued as follows:

\[
P_0(n) \geq \frac{1}{2} (pq^2)^{n/3} \sum_{k_2=0}^{n} \binom{n}{k_2} z^{k_2/3} = \frac{1}{2} (pq^2)^{n/3} (1 + z^{1/3})^n. \tag{57}
\]

Therefore from (57), (40) and (41) we get

\[
P_e(n) \geq \frac{2}{3} P_0(n) \geq \frac{1}{3} (pq^2)^{n/3} (1 + z^{1/3})^n. \tag{58}
\]

From (58) it follow (8) and Theorem 1 (formulas (11), (12)). Δ

**APPENDIX**

**Proof of equation (14).** Consider \( n \)-simplex code \((x_1, x_2, x_3)\), where \( x_1 \) first has \( n/3 \) ones and then \( 2n/3 \) zeros, \( x_2 \) first has \( n/3 \) zeros, then \( n/3 \) ones and then \( n/3 \) zeros, and \( x_3 \) first has \( 2n/3 \) zeros and then \( n/3 \) ones. Then \( w(x_1) = w(x_2) = w(x_3) = n/3 \) and
\(d_{12} = d_{13} = d_{23} = 2n/3\). Let an output \(y\) has \(u_1n/3\) ones on the first \(n/3\) positions, \(u_2n/3\) ones on next \(n/3\) positions and \(u_3n/3\) ones on last \(n/3\) positions. Then

\[
d(x_1, y)/n = (1 - u_1 + u_2 + u_3)/3, \]
\[
d(x_2, y)/n = (1 + u_1 - u_2 + u_3)/3, \]
\[
d(x_3, y)/n = (1 + u_1 + u_2 - u_3)/3.
\]

Since \(d(x_1, y) = d(x_2, y) = d(x_3, y)\), then we get \(u_1 = u_2 = u_3\) and

\[
p(y^n|x_1) = p^{d(x_1, y)} q^{n-d(x_1, y)} = q^n z^{d(x_1, y)} = q^n z^{(1+u)n/3}, \quad z = p/q < 1.
\]

Therefore

\[
P \left\{ p(y^n|x_1) \approx p(y^n|x_2) \approx p(y^n|x_3) \right\} \sim \max_{0 \leq u \leq 1} \left\{ \begin{array}{l} n \frac{n}{u n} p^{(1+u)n/3} q^{(2-u)n/3} \\ \end{array} \right\}
\]
\[
\sim q^n \max_{0 \leq u \leq 1} \left\{ \begin{array}{l} n \frac{n}{u n} z^{(1+u)n/3} \\ \end{array} \right\},
\]

and

\[
\frac{1}{n} \max_{0 \leq u \leq 1} \ln P \{ p(y^n|x_1) \} = \ln q + \max_{0 \leq u \leq 1} g(u), \quad (59)
\]

where

\[
g(u) = h(u) + (1 + u) \ln(z^{1/3}), \quad g'(u) = \ln \frac{1 - u}{u} + \ln(z^{1/3}), \quad g''(u) < 0.
\]

For the maximizing \(u_0\) we get

\[
u_0 = \frac{1}{1 + z^{-1/3}} = \frac{p^{1/3}}{p^{1/3} + q^{1/3}},
\]

and after simple algebra

\[
\ln q + g(u_0) = \ln \left( p^{1/3} q^{2/3} + p^{2/3} q^{1/3} \right). \quad (60)
\]

From (59) and (60) formulas (14) and (15) follow. \(\triangle\)

The author would like to thank Bassalygo L.A. and Kabatianski G.A. for useful discussions and constructive critical remarks, which improved the paper.

References

[1] Berlekamp, E.R., Block Coding with Noiseless Feedback, *PhD Thesis*, MIT, Cambridge, USA, 1964. Available at [http://hdl.handle.net/1721.1/14783](http://hdl.handle.net/1721.1/14783)
[2] Zigangirov, K.Sh., Upper Bounds for the Error Probability for Channels with Feedback, *Probl. Peredachi Inf.*, 1970, vol. 6, no. 2, pp. 87–92 [Probl. Inf. Transm. (Engl. Transl.), 1970, vol. 6, no. 2, pp. 159–163]. 
http://mi.mathnet.ru/eng/ppi1740

[3] Burnashev, M.V., On the Reliability Function of a Binary Symmetrical Channel with Feedback, *Probl. Peredachi Inf.*, 1988, vol. 24, no. 1, pp. 3–10 [Probl. Inf. Transm. (Engl. Transl.), 1988, vol. 24, no. 1, pp. 1–7]. 
http://mi.mathnet.ru/eng/ppi681

[4] Zigangirov, K.Sh., Optimum Zero Rate Transmission through Binary Symmetric Channel with Feedback, *Probl. Control Inform. Theory*, 1978, vol. 7, no. 3, pp. 183–198.

[5] Shannon, C.E., Probability of Error for Optimal Codes in a Gaussian Channel, *Bell Syst. Tech. J.*, 1959, vol. 38, no. 3, pp. 611–656. 
https://doi.org/10.1002/j.1538-7305.1959.tb03905.x

[6] Gallager, R.G., *Information Theory and Reliable Communication*, New York: Wiley, 1968. Translated under the title *Teoriya informatsii i nadezhnaya svyaz’,* Moscow: Sov. Radio, 1974.

[7] Pinsker, M.S., The Probability of Error in Block Transmission in a Memoryless Gaussian Channel with Feedback, *Probl. Peredachi Inf.*, 1968, vol. 4, no. 4, pp. 3–19 [Probl. Inf. Transm. (Engl. Transl.), 1968, vol. 4, no. 4, pp. 1–14].
http://mi.mathnet.ru/eng/ppi1868

[8] Korn, G.A. and Korn, T.M., *Mathematical Handbook for Scientists and Engineers*, New York: McGraw-Hill, 1968. Translated under the title *Spravochnik po matematike dlya nauchnykh rabotnikov i inzhenerov*, Moscow: Nauka, 1974.