Problems of Information Transmission,
vol. 46, no. 2, pp. 3–23, 2010.

M. V. Burnashev[1], **H. Yamamoto**

ON RELIABILITY FUNCTION OF BSC WITH NOISY FEEDBACK

For information transmission a binary symmetric channel is used. There is also another noisy binary symmetric channel (feedback channel), and the transmitter observes without delay all the outputs of the forward channel via that feedback channel. The transmission of an exponential number of messages (i.e. the transmission rate is positive) is considered. The achievable decoding error exponent for such a combination of channels is investigated. It is shown that if the crossover probability of the feedback channel is less than a certain positive value, then the achievable error exponent is better than the decoding error exponent of the channel without feedback.

§ 1. Introduction and main results

The binary symmetric channel BSC(p) with crossover probability $0 < p < 1/2$ (and $q = 1 - p$) is considered. It is assumed that there is also the feedback BSC(p_1) channel, and the transmitter observes (without delay) all outputs of the forward BSC(p) channel via that noisy feedback channel. No coding is used in the feedback channel (i.e. the receiver simply resends to the transmitter all received outputs). In other words, the feedback channel is "passive" (see Fig. 1).

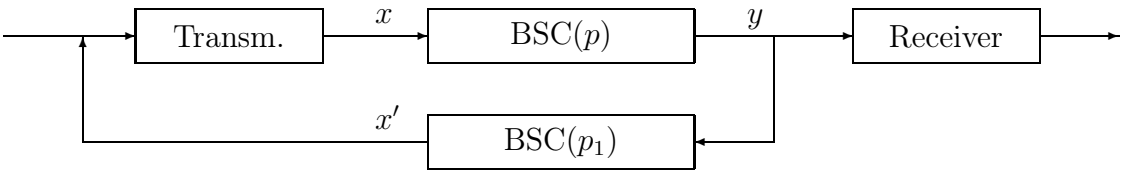

Fig. 1. Channel model

We consider the case when the overall transmission time n and $M = e^{Rn}$ equiprobable messages $\{\theta_1, \ldots, \theta_M\}$ are given. After the moment n, the receiver makes a decision $\hat{\theta}$ on the message transmitted. We are interested in the best possible decoding error exponent (and whether it can exceed the similar exponent of the channel without feedback).

[1]The research described in this publication was made possible in part by the Russian Fund for Fundamental Research (project numbers 06-01-00226 and 09-01-00536).

Such model was considered in [1], where the case of a nonexponential (on n) number M (i.e. $R = 0$) was investigated. In the paper we consider the case $M = e^{Rn}$, $R > 0$, strengthening methods of [1]. The main difference is that since now M is exponential in n, we will need much more accurate investigation of the decoding error probability. Moreover, if M is nonexponential in n, then we know the best code for use during phase I - it is an "almost equidistant" code (i.e. all its codeword distances equal $n/2 + o(n)$). If $R > 0$ then we do not know such best code, and for that reason we choose that code randomly.

Some results for channels with noiseless feedback can be found in [2–12], and in the noisy feedback case – in [13, 14] (see also discussion in [1]).

We show that if the crossover probability p_1 of the feedback channel BSC(p_1) is less then the certain positive value $p_0(p, R)$, then it is possible to improve the best error exponent $E(R, p)$ of BSC(p) without feedback. The transmission method with one "switching" moment, giving such an improvement, is described in § 4. It is similar to the method used in [1].

We will need some definitions and notations. For $L = 1, 2, \ldots$ define the critical rates $R_{\mathrm{crit},1}(p) > R_{\mathrm{crit},2}(p) > \ldots$ [6, 15, 16]

$$R_{\mathrm{crit},L}(p) = \ln 2 - h\left[\frac{p^{1/(L+1)}}{p^{1/(L+1)} + q^{1/(L+1)}}\right], \tag{1}$$

where $h(x) = -x \ln x - (1 - x) \ln(1 - x)$. For $L = 1$ we omit the index L and simply write $R_{\mathrm{crit}}(p) = R_{\mathrm{crit},1}(p)$, $E(R, p) = E(R, p, 1)$, etc.

Define the new critical rate $R_2 = R_2(p)$ as the unique root of the equation [17]

$$\min_{\substack{0 \le \tau \le \alpha \le 1/2 \\ h(\alpha) - h(\tau) = \ln 2 - R_2}} \frac{\alpha(1-\alpha) - \tau(1-\tau)}{1 + 2\sqrt{\tau(1-\tau)}} = \frac{\sqrt{pq}}{1 + 2\sqrt{pq}}\,.$$

Then $0 < R_2(p) < R_{\mathrm{crit}}(p)$, $0 < p < 1/2$.

Denote by $C(p) = \ln 2 - h(p)$ the capacity of the BSC(p), and by $E_{\mathrm{sp}}(R, p)$ the sphere-packing exponent

$$E_{\mathrm{sp}}(R, p) = D\left(\delta_{GV}(R) \,\|\, p\right),$$
$$D(x \,\|\, y) = x \ln \frac{x}{y} + (1 - x) \ln \frac{1 - x}{1 - y},$$

where $\delta_{GV}(R) \le 1/2$ is defined by the relation

$$\ln 2 - R = h(\delta_{GV}(R)).$$

Denote by $E(R, p)$ the best decoding error exponent (the reliability function) of BSC(p) without feedback. For $R_2(p) \le R \le C(p)$, and $R = 0$ the function $E(R, p)$ is known exactly [6, 17]:

$$E(R, p) = E_{\mathrm{r}}(R, p) = \begin{cases} \ln 2 - \ln\left(1 + 2\sqrt{pq}\right) - R, & R_2(p) \le R \le R_{\mathrm{crit}}(p), \\ E_{\mathrm{sp}}(R, p), & R_{\mathrm{crit}}(p) \le R \le C(p), \end{cases}$$
$$E(0, p) = E_{\mathrm{ex}}(0, p) = \frac{1}{4} \ln \frac{1}{4pq}, \tag{2}$$

where $E_{\rm r}(R,p), E_{\rm ex}(R,p)$ – "random coding" bounds [6, 15, 16] (see § 6).

For $0 < R < R_2(p)$ there are known only lower and upper bounds for the function $E(R,p)$. To describe the best known lower bound (the exponent $E_{\rm ex}(R,p)$ of random coding with "expurgation"), introduce the rate $R_{\min}(p)$ (see (43)). Then $0 < R_{\min}(p) < R_2(p) < R_{\rm crit}(p)$, $0 < p < 1/2$, and the best known lower bound [15, 16] has the form

$$E(R,p) \geq E_{\rm ex}(R,p) = \begin{cases} -\delta_{GV}(R) \ln \sqrt{4pq}, & 0 < R \leq R_{\min}(p), \\ \ln 2 - \ln\left(1 + 2\sqrt{pq}\right) - R, & R_{\min}(p) \leq R < R_2(p). \end{cases} \tag{3}$$

Denote by $E(R,p,L)$ the best list size L decoding error exponent of BSC(p) without feedback. It is known that $E(R,p,L) = E_{\rm r}(R,p,L) = E_{\rm sp}(R,p)$, $R_{{\rm crit},L}(p) \leq R < C(p)$ [6, 15, 16] and $E(0,p,L) = E_{\rm ex}(0,p,L)$ [18], where the "random coding" $E_{\rm r}(R,p,L)$ and the "random coding with expurgation" $E_{\rm ex}(R,p,L)$ bounds are described in § 6.

For $0 < R < R_{{\rm crit},L}(p)$ the best known lower bound for $E(R,p,L)$ has the form [15, 16]

$$E(R,p,L) \geq E_{\rm ex}(R,p,L), \quad 0 < R < R_{{\rm crit},L}(p). \tag{4}$$

We also have $E_{\rm ex}(R,p,L) = E_{\rm r}(R,p,L)$, $R_{{\min},L}(p) \leq R \leq R_{{\rm crit},L}(p)$ (see (42)). Denote

$$E_{\rm low}(R,p,L) = \max\{E_{\rm r}(R,p,L), E_{\rm ex}(R,p,L)\}. \tag{5}$$

Denote by $F(R,p)$ the best decoding error exponent of BSC(p) with noiseless feedback. Then

$$\begin{array}{lll} E(R,p) = F(R,p) = E_{\rm sp}(R,p), & R_{\rm crit}(p) \leq R \leq C(p) & [3], \\ E(R,p) \leq F(R,p) \leq E_{\rm sp}(R,p), & 0 < R < R_{\rm crit}(p) & [3], \\ E(0,p) < F(0,p) = -\ln\left(p^{1/3}q^{2/3} + p^{2/3}q^{1/3}\right) & & [5]. \end{array}$$

Denote by $F(R,p,p_1)$ the best decoding error exponent of BSC(p) with the noisy BSC(p_1) feedback channel. Clearly, $E(R,p) \leq F(R,p,p_1) \leq F(R,p)$ for all p, p_1. In particular, $F(R,p,0) = F(R,p)$, $F(R,p,1/2) = E(R,p)$.

Denote by $E_2(p)$ the best error exponent for two codewords over BSC(p) (clearly, it remains the same for the channel with noiseless feedback as well)

$$E_2(p) = \frac{1}{2} \ln \frac{1}{4pq}. \tag{6}$$

Denote by $F_1(R,p,p_1)$ the decoding error exponent of the transmission method described in § 4 (with one switching moment). The inequality $F_1(R,p,p_1) > E(R,p)$ is possible only when $R < R_{\rm crit}(p)$.

To describe the function $p_0(R,p)$ of the critical noise level in the feedback channel, introduce the function

$$t_0(R,p) = \frac{3\left[E_{\rm low}(R,p,2) - E_{\rm low}(R,p)\right]}{\ln(q/p)}, \tag{7}$$

where $E_{\text{low}}(R,p,2)$, $E_{\text{low}}(R,p) = E_{\text{low}}(R,p,1)$ are defined in (5).

The function $t_0(R,p)$ monotonically decreases on R. For a given $R \geq 0$ it first increases on p, and then decreases. Moreover,

$$\max_{R,p} t_0(R,p) = \max_{p} t_0(0,p) \approx t_0(0,0.0124) \approx 0.1322.$$

Introduce the function $p_0 = p_0(R,p) \leq t_0(R,p)$ as the unique root of the equation

$$D\left(t_0(R,p) \,\|\, p_0\right) = 2R. \tag{8}$$

In particular,

$$p_0(0,p) = t_0(0,p) = \frac{3\left[\ln 4 - 3\ln\left(p^{1/3} + q^{1/3}\right)\right]}{4\ln(q/p)}.$$

Define also $t_1 = t_1(R,p_1) \geq p_1$ as the unique root of the equation

$$D\left(t_1 \,\|\, p_1\right) = 2R. \tag{9}$$

The main result of the paper represents

T h e o r e m 1. *If $R < R_{\text{crit}}(p)$ and $p_1 < p_0(R,p)$, then*

$$F_1(R,p,p_1) \geq \max_{0\leq\gamma\leq 1} T(R,p,p_1,\gamma) > \begin{cases} E_{\text{ex}}(R,p), & 0 \leq R \leq R_2(p), \\ E(R,p), & R_2(p) \leq R < R_{\text{crit}}(p), \end{cases} \tag{10}$$

where

$$T = \min\left\{\gamma E_{\text{low}}(R/\gamma,p,2) - \frac{\gamma t_1(R/\gamma,p_1)}{3}\ln\frac{q}{p},\ \gamma E_{\text{low}}(R/\gamma,p) + (1-\gamma)E_2(p)\right\}. \tag{11}$$

In other words, for any $R < R_{\text{crit}}(p)$ and $p_1 < p_0(R,p)$ the function $F_1(R,p,p_1)$ is bigger (i.e. better) than the best known lower bound for the decoding error exponent of BSC(p) without feedback.

Moreover, there exists the positive function $p_2(R,p)$ such that the following result holds.

C o r o l l a r y 1. *If $R < R_{\text{crit}}(p)$ and $p_1 < p_2(R,p)$, then*

$$F_1(R,p,p_1) \geq \max_{0\leq\gamma\leq 1} T(R,p,p_1,\gamma) > E(R,p). \tag{12}$$

This result follows from the proof of the Theorem 2 (see §3) and the fact that the function $T(R,p,p_1,\gamma)$ is continuous on p_1.

Remark 1. We do not try to find the best function $p_0(R,p)$, limiting ourselves to rather simple estimates for it.

On Fig. 2. the plot of the function $p_0(R,p)$ for $p = 0.01$ is given ($R_{\text{crit}} \approx 0.387$). Note that here $p_0(R,p) > p$ for small R.

It is more convenient for us to investigate first the function $F_1(R,p,p_1)$ for $p_1 = 0$, i.e. for the channel with noiseless feedback. Then the next result holds.

T h e o r e m 2. *If* $0 < p < 1/2$, $R < R_{\rm crit}(p)$, *then*

$$F_1(R,p,0) = F_1(R,p) \geq \gamma_0 E_{\rm low}(R/\gamma_0, p, 2) > E(R,p), \tag{13}$$

where $\gamma_0 \in (R/R_{\rm crit}(p), 1)$ *is the largest root of the equation* (20).

Remark 2. If $p_1 \to 0$, then the relations (10), (11) turn into the similar relation (13) for the channel with noiseless feedback (see also remark 6 in §4).

Remark 3. The transmission method described in § 4, reduces the problem to testing of two most probable (at a fixed moment) messages. Such strategy is not optimal even for one switching moment (at least, if p_1 is very small). But it is relatively simple for investigation, and it gives already a reasonable improvement over the channel without feedback.

Remark 4. In the preliminary publication [19, Proposition] it was claimed that $p_0(R,p) = 1/2$ for some range of rates R. In the proof of that result a miscalculation was found.

Below in § 2 informal description of the transmission method is given. In § 3 the transmission method with one switching moment in the case of the channel with noiseless feedback is described and analyzed and the Theorem 2 is proved. In § 4 that method (slightly modified) is investigated for the channel with noisy feedback and the Theorem 1 is proved. In § 5 it is clarified for which p_1 noisy feedback behaves approximately like noiseless. A part of formulas used and some auxiliary results are presented in § 6.

A preliminary (and simplified) paper variant (without detailed proofs) was published in [19].

§ 2. Informal description of the transmission method

We use the transmission method with one fixed switching moment at which the coding function is changed. That method is based on one idea and one useful observation.

Idea. It is based on the inequality which follows from (41)

$$E_{\rm ex}(R,p) < E_{\rm low}(R,p,2), \qquad R < R_{\rm crit}(p). \tag{14}$$

Considering only $R < R_{\rm crit}(p)$ we choose some positive $\gamma < 1$ and partition the total transmission period $[1,n]$ on two phases: $[1,\gamma n]$ (phase I) and $(\gamma n, n]$ (phase II) (at first we may think that γ is rather close to one).

On phase I (i.e. on $[0,\gamma n]$) we use the "best" code of M codewords $\{\boldsymbol{x}_i\}$ of length γn (see below). On that phase the transmitter only observes (via the feedback channel) outputs of the forward channel, but does not change the coding function. We set the value $\gamma = \gamma(R,p)$ such that

$$E_{\rm ex}(R,p) < \gamma E_{\rm low}(R/\gamma, p, 2)\,, \qquad R < R_{\rm crit}(p) \tag{15}$$

(it is always possible due to continuity of the function $\gamma E_{\rm low}(R/\gamma, p, 2)$ on γ and the condition (14)). After phase I (at moment γn) the receiver selects two most probable messages θ_i, θ_j. By the condition (15), the exponent of the probability that the true message $\theta_{\rm true}$ is not among the chosen messages θ_i, θ_j, will be larger (i.e. better) than $E_{\rm ex}(R,p)$. Assume that by some means the transmitter is also able to recover those two most probable messages

θ_i, θ_j (it is certainly so in the noiseless feedback case). Then, on phase II (i.e. on $(\gamma n, n]$) the transmitter only helps the receiver to decide between those two most probable messages θ_i, θ_j, using two opposite codewords of length $(1-\gamma)n$. The error exponent $E_2(p)$ (see (6)) on that phase is better than all other exponents involved. As a result, it gives the overall decoding error exponent better than $E_{\rm ex}(R,p)$.

It remains us to find the way the transmitter will able to recover those two most probable messages θ_i, θ_j. It may seem that it is always possible if the value p_1 is sufficiently small. But it is not true. With high probability (even close to one) the second θ_j and the third θ_k most probable messages will be approximately equiprobable, and then, for any $p_1 > 0$, the transmitter will not be able to rank them correctly (due to noise in the feedback channel).

Observation. Fortunately, in that case (with high probability) the most probable message θ_i will be much more probable than the second most probable message θ_j. In such case the receiver makes a decision immediately after phase I (in favor of the most probable message θ_i), and it ignores all next signals from the transmitter.

The description given is rather intuitive, and it should be checked analytically (which is done below).

§ 3. Channel with noiseless feedback. Proof of Theorem 2

For simplicity, we start with the noiseless feedback case and describe formally the transmission method which (after some modification) will be used for noisy feedback as well. Moreover, in the noisy feedback case we will need some formulas from the noiseless feedback case.

Denote by $F_1(R,p) = F_1(R,p,0)$ the decoding error exponent of the transmission method described below (with one switching moment).

P r o o f o f T h e o r e m 2. We consider $M = e^{Rn}$ messages $\theta_1, \dots, \theta_M$. Using some $\gamma \in [0,1]$ (it will be chosen later), we partition the total transmission time $[1,n]$ on two phases: $[1, \gamma n]$ (phase I) and $(\gamma n, n]$ (phase II). We perform as follows.

1) On phase I (i.e. on $[1,\gamma n]$) we use the "best" code of M codewords $\{\boldsymbol{x}_i\}$ of length γn (see below). On that phase the transmitter only observes (via the feedback channel) outputs of the forward channel, but does not change the coding function.

2) Let $\boldsymbol{x}$ be the transmitted codeword (of length γn) and $\boldsymbol{y}$ be the received (by the receiver) block. After phase I, based on the block $\boldsymbol{y}$, the transmitter selects two messages θ_i, θ_j (codewords $\boldsymbol{x}_i, \boldsymbol{x}_j$) which are the most probable for the receiver, and ignores all the remaining messages $\{\theta_k\}$. If among the selected messages θ_i, θ_j there is the true message $\theta_{\rm true}$, then on phase II (i.e. on $(\gamma n, n]$) the transmitter only helps the receiver to decide between those two most probable messages θ_i, θ_j, using two opposite codewords of length $(1-\gamma)n$. If the true message $\theta_{\rm true}$ is not among those two selected messages, then the transmitter sends an arbitrary block. After moment n the receiver makes a decision in favor of the most probable of those two remaining messages θ_i, θ_j (based on all received on $[1,n]$ signals).

Clearly, a decoding error occurs in the following two cases.

1) After phase I the true message is not among two most probable messages. We denote that probability by P_1.

2) After phase I the true message is among two most probable, but after phase II it is not the most probable. We denote that probability by P_{20}.

Then for the total decoding error probability P_{e} we have

$$P_{\mathrm{e}} \le P_1 + P_{20}\,. \tag{16}$$

On phase I (of length γn) we use a code having small two decoding error probabilities: usual and when decoding with list size $L = 2$. Then there exists a code such that for P_1 we have (see § 6)

$$\frac{1}{n}\ln\frac{1}{P_1} \ge \gamma E_{\mathrm{low}}(R/\gamma, p, 2) + o(1), \qquad n \to \infty\,. \tag{17}$$

Now we evaluate the probability P_{20}. Denote by $d\,(\boldsymbol{x}, \boldsymbol{y})$ the Hamming distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, and $d_{ij} = d\,(\boldsymbol{x}_i, \boldsymbol{x}_j)$. On phase I (of length γn) the distances among codewords are $\{d_{ij}\}$. On phase II (of length $(1-\gamma)n$) the distance between two remaining codewords equals $(1-\gamma)n$. Therefore, the total distance between the true and the concurrent codewords equals $d_{ij} + (1-\gamma)n$. Then there exists a code such that (see derivation in § 6)

$$\frac{1}{n}\ln\frac{1}{P_{20}} \ge \gamma E_{\mathrm{low}}(R/\gamma, p) + (1-\gamma)E_2(p) + o(1). \tag{18}$$

Moreover, there exists a code for which both relations (17) and (18) are fulfilled (see § 6). Then from (16)–(18) we have

$$\frac{1}{n}\ln\frac{1}{P_{\mathrm{e}}} \ge \frac{1}{n}\min\left\{\ln\frac{1}{P_1}, \ln\frac{1}{P_{20}}\right\} - \frac{2}{n} \ge$$
$$\ge \min\left\{\gamma E_{\mathrm{low}}(R/\gamma, p, 2), \gamma E_{\mathrm{low}}(R/\gamma, p) + (1-\gamma)E_2(p)\right\} + o(1)\,,$$

where $E_2(p)$ is defined in (6). Therefore

$$F_1(R, p) \ge \max_{0\le\gamma\le 1} \min\left\{\gamma E_{\mathrm{low}}(R/\gamma, p, 2), \gamma E_{\mathrm{low}}(R/\gamma, p) + (1-\gamma)E_2(p)\right\}, \tag{19}$$

where $E_{\mathrm{low}}(R, p, 2)$ and $E_{\mathrm{low}}(R, p)$ are defined in (5) (see also § 6).

Note that the function $\gamma E_{\mathrm{low}}(R/\gamma, p, 2)$ from the right-hand side of (19) monotonically increases in γ. On the contrary, the function $S(\gamma, R, p) = \gamma E_{\mathrm{low}}(R/\gamma, p) + (1-\gamma)E_2(p)$ monotonically decreases in γ. Indeed, denoting $r = R/\gamma$ and omitting p, we have $S'_\gamma(\gamma, R) = E_{\mathrm{low}}(r) - rE'_{\mathrm{low}}(r) - E_2$ and $S''_{\gamma r}(\gamma, R) = -rE''_{\mathrm{low}}(r) < 0$. Therefore maximum over R, γ of the value $S'_\gamma(\gamma, R)$ is attained when $r \to 0$. Since $rE'_{\mathrm{low}}(r) \to 0$, $r \to 0$, then we get $\max\limits_{R,\gamma} S'_\gamma(\gamma, R) = E_{\mathrm{low}}(0) - E_2 < 0$.

We consider only the case $R < R_{\mathrm{crit}}(p)$, i.e. when $E_{\mathrm{low}}(R, p, 2) > E_{\mathrm{low}}(R, p)$. For such R the best is to set $\gamma = \gamma_0$ such that $P_1 = P_{20}$, i.e.

$$\gamma_0 E_{\mathrm{low}}(R/\gamma_0, p, 2) = \gamma_0 E_{\mathrm{low}}(R/\gamma_0, p) + (1-\gamma_0)E_2(p)\,. \tag{20}$$

Both sides of (20) are continuous functions in γ_0. The left-hand side of (20) monotonically increases in γ_0, and the right-hand one monotonically decreases in γ_0. With $\gamma_0 = 1$ the left-hand side is greater than its right-hand side, which equals $E_{\rm low}(R,p)$. On the contrary, for $\gamma_0 = R/R_{\rm crit}$ the right-hand side is greater than the left-hand side. Then there exists the unique $\gamma_0 \in (R/R_{\rm crit}, 1)$ satisfying (20). Therefore we get

$$F_1(R,p) \geq \gamma_0 E_{\rm low}(R/\gamma_0, p) + (1-\gamma_0)E_2(p) > E_{\rm low}(R,p). \tag{21}$$

We show that, in fact, $F_1(R,p)$ satisfies the stronger inequality (13), although we know exactly only part of the function $E(R,p), 0 < R < R_{\rm crit}(p)$ (see (2)). If we connect the points $E(0,p)$ and $E(R_{\rm crit},p)$ by the piece of the straight line, then due to the "straight-line bound" [20], for $0 \leq R \leq R_{\rm crit}$ the function $E(R,p)$ does not exceed that straight line. Therefore, if $0 < R < R_{\rm crit}(p)$ and $0 < p < 1/2$ then the inequality holds

$$E(R,p) < E(0,p) - \frac{[E(0,p) - E(R_{\rm crit}(p),p)]R}{R_{\rm crit}(p)}\,.$$

Now, to establish the formula (13), it is sufficient to check that for such p, R the following strict inequality is valid

$$\gamma_0 E_{\rm ex}(R/\gamma_0, p, 2) > E(0,p) - \frac{[E(0,p) - E(R_{\rm crit}(p),p)]R}{R_{\rm crit}(p)}\,. \tag{22}$$

For that purpose it is convenient to introduce the parameter $u = R/\gamma_0$, $u \in (0, R_{\rm crit})$. Then we get the parametric representation for $\gamma_0 = \gamma_0(u,p)$ and $R = R(u,p)$:

$$\gamma_0 = \frac{E_2(p)}{E_2(p) + E_{\rm ex}(u,p,2) - E_{\rm ex}(u,p)}\,, \qquad R = u\gamma_0.$$

Then combining analytical and numerical methods, it is not difficult to check validity of the inequality (22). It concludes proof of the Theorem 2. ▲.

In Fig. 3 the plots of the functions $F_1(R,p)$ and $E_{\rm ex}(R,p)$ for $p = 0.01$ ($R_{\rm crit} \approx 0.387$) are shown.

To compare the functions $F_1(R,p)$ and $E(R,p)$ consider

E x a m p l e 1. Let $p = (1-\varepsilon)/2$, $\varepsilon \to 0$. Then

$$C(p) = \frac{\epsilon^2}{2} + O(\varepsilon^4), \quad R_{\rm crit}(p) = \frac{C(p)[1+o(1)]}{4}, \quad R_{\rm min2}(p) \leq R_{\rm min}(p) = O(C^2).$$

Therefore when $p \to 1/2$ the expurgation bound, essentially, is not applicable and we get the known results [15]

$$E(R,p)[1+o(1)] = \begin{cases} C/2 - R, & 0 \leq R \leq C/4, \\ (\sqrt{C} - \sqrt{R})^2, & C/4 \leq R \leq C, \end{cases}$$

and

$$E(R,p,2)[1+o(1)] \geq E_r(R,p,2) = \begin{cases} 2C/3 - 2R, & 0 \leq R \leq C/9, \\ (\sqrt{C} - \sqrt{R})^2, & C/9 \leq R \leq C. \end{cases}$$

From those formulas and (7) we also have

$$4\varepsilon t_0(R,p)[1+o(1)] = \begin{cases} C-6R, & 0 \le R \le C/9, \\ 3(\sqrt{C}-2\sqrt{R})^2, & C/9 \le R \le C/4, \\ 0, & C/4 \le R \le C. \end{cases} \tag{23}$$

Consider the equation (20). For $R < R_{\rm crit}(p) = C(p)[1+o(1)]/4$, there are possible two cases: $R/\gamma_0 \le C/9$ and $C/9 \le R/\gamma_0 < C/4$.

1) Let $R/\gamma_0 \le C/9$. Then from (20) we get

$$\gamma_0 = \frac{6(R+C)}{7C}\,, \qquad F_1(R,p) = \frac{4C-10R}{7}\,, \qquad R \le \frac{2C}{19}\,,$$

and

$$\frac{F_1(R,p)}{E(R,p)} = \frac{8}{7} - \frac{4R}{7(C-2R)}\,, \qquad R \le \frac{2C}{19}\,.$$

The ratio $F_1(R,p)/E(R,p)$ monotonically decreases from $8/7$ (for $R=0$) down to $16/15$ (for $R = 2C/19$).

2) Let $C/9 \le R/\gamma_0 < C/4$. Then we get

$$\sqrt{\gamma_0} = \frac{2\sqrt{R}+\sqrt{6C-8R}}{3\sqrt{C}}\,, \qquad \frac{2C}{19} \le R < \frac{C}{4}\,,$$

and

$$F_1(R,p) = \frac{1}{9}\left[6C-7R-2\sqrt{2R(3C-4R)}\right].$$

The ratio $F_1(R,p)/E(R,p)$ monotonically decreases from $16/15$ (for $R = 2C/19$) down to 1 (for $R = C/4$).

It is natural to expect that similar results will also hold in the case of the noisy feedback channel $\mathrm{BSC}(p_1)$, if p_1 is sufficiently small.

§ 4. Channel with noisy feedback. Proof of Theorem 1

In the noisy feedback case we will still use the transmission method with one switching moment. But if we try to use exactly the same method as in the noiseless feedback case, we will face with the following problem. After phase I, the transmitter should find the two most probable (for the receiver) codewords $\boldsymbol{x}^1, \boldsymbol{x}^2$. But with relatively high probability, the second and the third ranked codewords $\boldsymbol{x}^2$ and $\boldsymbol{x}^3$ will be approximately equiprobable, and therefore it will be difficult to the transmitter to rank them correctly (due to noise in the feedback). Fortunately, in that case (with high probability) the most probable codeword $\boldsymbol{x}^1$ will be much more probable than $\boldsymbol{x}^2$, and then (again with high probability) $\boldsymbol{x}^1$ is the true codeword. We use this observation as follows: if posterior probabilities of the second $\boldsymbol{x}^2$ and the third $\boldsymbol{x}^3$ ranked codewords are not very different, the receiver makes a decision immediately after phase I (in favor of the most probable codeword $\boldsymbol{x}^1$), and it ignores all next signals from the transmitter on phase II.

As a result, we use the following transmission and decoding method.

Transmission. We set a number $0 < \gamma < 1$. On phase I, of length $m = \gamma n$, we use a "good" code (it is explained below). Let $\boldsymbol{x}_{\text{true}}$ be the transmitted codeword of length m, $\boldsymbol{y}$ be the received (by the receiver) block, and $\boldsymbol{x}'$ be the received (by the transmitter) block. The transmitter selects one more codeword $\boldsymbol{x}_i \neq \boldsymbol{x}_{\text{true}}$, closest to $\boldsymbol{x}'$. For example, the codeword $\boldsymbol{x}_1 \neq \boldsymbol{x}_{\text{true}}$ is chosen, if $d(\boldsymbol{x}_1, \boldsymbol{x}') = \min\limits_{\boldsymbol{x}_i \neq \boldsymbol{x}_{\text{true}}} d(\boldsymbol{x}_i, \boldsymbol{x}')$. As a result, the transmitter builds a list of two messages: the true one θ_{true} and another message $\theta_i \neq \theta_{\text{true}}$, which looks most probable among remaining ones.

A "good" code in use of length m should have the following properties:
1) Its decoding error probability P_e satisfies the inequality $P_e \leq e^{-E_{\text{low}}(R,p)m}$;
2) Its list size $L = 2$ decoding error probability $P_e(2)$ satisfies similar inequality $P_e(2) \leq e^{-E_{\text{low}}(R,p,2)m}$;
3) The relations (18) and (26) hold for it.

Existence of such code is shown in § 6, slightly modifying standard Gallager's arguments for expurgation bound [15, 16].

On phase II (i.e. on $(\gamma n, n]$) the transmitter uses the two opposite codewords of length $n - m = (1-\gamma)n$ (for example, consisting of all zeros and all ones), in order to help the receiver to decide between the true message θ_{true} and another most probable message $\theta_i \neq \theta_{\text{true}}$.

This transmission method is a slight modification of the method used in [1]. It gives the same decoding error probability exponent, but it is simpler for analysis. If the true message θ_{true} is not among the two most probable messages for the receiver, then there will always be the decoding error. A slight modification of the transmission method from [1] used here helps in the case when the true message θ_{true} is among the two most probable messages for the receiver, but it is not such one for the transmitter.

Decoding. We set a number $t > 0$. Arrange the Hamming distances $\{d(\boldsymbol{x}_i, \boldsymbol{y}),\ i = 1, \ldots, M\}$ after phase I in the increasing order, denoting

$$d^{(1)} = \min_i d(\boldsymbol{x}_i, \boldsymbol{y}) \leq d^{(2)} \leq \ldots \leq d^{(M)} = \max_i d(\boldsymbol{x}_i, \boldsymbol{y}),$$

(in case of tie we use any order). Let also $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^M$ be the corresponding ranking of codewords after phase I, i.e $\boldsymbol{x}^1$ is the closest to $\boldsymbol{y}$ codeword, etc. Two cases are possible.

C a s e 1. If $d^{(3)} \leq d^{(2)} + t\gamma n$, then the receiver makes the decoding immediately after phase I (in favor of the closest to $\boldsymbol{y}$ codeword $\boldsymbol{x}^1$). Although the transmitter will still send some signals on phase II, the receiver has already made its decision.

C a s e 2. If $d^{(3)} > d^{(2)} + t\gamma n$, then after phase I the receiver selects two most probable messages θ_i, θ_j, and after transmission on phase II (i.e. after moment n) makes a decision between those two remaining messages θ_i, θ_j in favor of more probable of them (based on all received on $[0, n]$ signals).

In the case 2 the transmitter and the receiver will perform in coordination, if the lists of two messages build by each of them coincide. Remind that the receiver's list always contains the true message. Of course, those lists may be different (and then there will be the decoding error), but probability of such event should be sufficiently small (which will be secured below).

Remarks 5. a) In the case of noiseless feedback (i.e. when $p_1 = 0$) the strategy described reduces to the strategy from § 3 if we set $t = 0$.
b) The strategy described can be improved by introducing an additional parameter $\tau \geq 0$, such that if $d^{(2)} \geq d^{(1)} + \tau\gamma n$ then the receiver also makes the decoding immediately after phase I (in favor of the closest to $\boldsymbol{y}$ codeword $\boldsymbol{x}^1$). But introduction of such parameter leads to too bulky formulas.

To evaluate the decoding error probability $P_{\rm e}$, denote by P_1 and P_2 the decoding error probabilities in the case 1 (i.e. after the moment γn), and in the case 2 (i.e. after the moment n), respectively. Then for $P_{\rm e}$ we have

$$P_{\rm e} \leq P_1 + P_2\,. \tag{24}$$

We evaluate the probabilities P_1, P_2 in the right-hand side of (24). Denoting $d_i = d(\boldsymbol{x}_i, \boldsymbol{y})$, $i = 1, \ldots, M$, for P_1 we have

$$P_1 \leq M^{-1} \sum_{k=1}^{M} \mathbf{P}(d_k \neq d^{(1)}; d_k \geq d^{(3)} - t\gamma n | \boldsymbol{x}_k). \tag{25}$$

We show that there exists a code such that for P_1 we have ($n \to \infty$)

$$\frac{1}{n} \ln \frac{1}{P_1} \geq \gamma E_{\rm low}(R/\gamma, p, 2) - \frac{t\gamma}{3} \ln \frac{q}{p} + o(1)\,. \tag{26}$$

Indeed, using the inequality $\left(\sum a_i\right)^{1/\rho} \leq \sum a_i^{1/\rho}$, $\rho \geq 1$, we have

$$\begin{gathered}
\mathbf{P}^{1/\rho}\left(d_k \neq d^{(1)}; d_k \geq d^{(3)} - t\gamma n | \boldsymbol{x}_k\right) \leq \\
\leq 2^{1/\rho} \mathbf{P}^{1/\rho}\left(d_k = d^{(2)} \geq d^{(3)} - t\gamma n | \boldsymbol{x}_k\right) + 2^{1/\rho} \mathbf{P}^{1/\rho}\left(d_k \geq d^{(3)} | \boldsymbol{x}_k\right) \leq \\
\leq 2^{1+1/\rho} \left(\frac{q}{p}\right)^{t\gamma n/(3\rho)} \sum_{m_1, m_2} \left[\sum_{\boldsymbol{y}} \left[P\left(\boldsymbol{y}|\boldsymbol{x}_k\right) P\left(\boldsymbol{y}|\boldsymbol{x}_{m_1}\right) P\left(\boldsymbol{y}|\boldsymbol{x}_{m_2}\right)\right]^{1/3}\right]^{1/\rho},
\end{gathered}$$

and then

$$\left[E\mathbf{P}^{1/\rho}\left(d_k \neq d^{(1)}; d_k \geq d^{(3)} - t\gamma n | \boldsymbol{x}_k\right)\right]^{\rho/n} \leq 2^{(1+\rho)/n} \left(\frac{q}{p}\right)^{t\gamma/3} e^{-\gamma E_{\rm ex}(R/\gamma, p, 2)}.$$

A similar inequality holds with $E_{\rm r}(R/\gamma, p, 2)$ instead of $E_{\rm ex}(R/\gamma, p, 2)$. Therefore using the definition of $E_{\rm low}(R/\gamma, p, 2)$ (see (5)), we get the formula (26).

For the value P_2 we have

$$P_2 \leq P_{20} + P_{2n}\,, \tag{27}$$

where P_{20} is the decoding error probability in the case 2 for the channel with noiseless feedback, and P_{2n} is the probability that the most probable codeword (excluding the true

codeword $\boldsymbol{x}_{\text{true}}$) for the receiver is not such one for the transmitter (moreover, the true codeword is among two most probable codewords for the receiver).

For the value P_{20} the formula (18) remains valid.

It remains us to evaluate P_{2n}. For that purpose consider the ensemble of codes $\mathcal{C}$ in which each codeword is selected independently with the probability 2^{-m} among all possible binary vectors of length m. We are interested in the value $\mathbf{E}_{\mathcal{C}} P_{2n}^{1/\rho}(\mathcal{C})$, $\rho \geq 1$, where expectation is taken over randomly chosen codes $\mathcal{C}$. Clearly,

$$\mathbf{P}\left(\boldsymbol{y}|\boldsymbol{x}_{\text{true}}\right) = q^m \binom{m}{d} \left(\frac{p}{q}\right)^d, \qquad d = d\left(\boldsymbol{x}_{\text{true}}, \boldsymbol{y}\right).$$

For given blocks $\boldsymbol{x}_{\text{true}}$ and $\boldsymbol{y}$ all $(M-1)$ remaining codewords are independently and equiprobably distributed among all 2^m binary vectors of length m. The vector $\boldsymbol{y}$ is transmited over the feedback channel BSC(p_1) and the transmitter receives the vector $\boldsymbol{x}'$.

Without loss of generality we assume that $\boldsymbol{x}_{\text{true}} = \boldsymbol{x}_M$. For the received block $\boldsymbol{y}$ we arrange all remaining codewords $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}$ as $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^{M-1}$, in increasing by their distance $d\left(\boldsymbol{x}^i, \boldsymbol{y}\right)$ order, i.e. $d\left(\boldsymbol{x}^1, \boldsymbol{y}\right)$ is the minimal distance, etc. In the case 2 it is necessary to have $d(\boldsymbol{x}^i, \boldsymbol{y}) - d(\boldsymbol{x}^1, \boldsymbol{y}) \geq tm$, $i = 2, \ldots, M-1$ (otherwise, the case 1 occurs). Moreover, we may assume that the distance $d(\boldsymbol{x}^1, \boldsymbol{y})$ satisfies the condition ($m \to \infty$)

$$d(\boldsymbol{x}^1, \boldsymbol{y})/m \leq \delta_{GV}(R/\gamma) - t + o(1), \qquad R > 0, \tag{28}$$

which is equivalent to the inequality

$$h\left\{d(\boldsymbol{x}^1, \boldsymbol{y})/m + t\right\} \leq \ln 2 - R/\gamma\,, \qquad d(\boldsymbol{x}^1, \boldsymbol{y})/m + t < 1/2\,.$$

Indeed, blocks $\boldsymbol{y}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}$ are distributed independently and equiprobably among all 2^m binary vectors of length m. For $u \geq 0$ introduce the random event

$$\mathcal{A}(u) = \{d(\boldsymbol{x}^1, \boldsymbol{y}) > (u-t)m;\ d(\boldsymbol{x}^2, \boldsymbol{y}) - d(\boldsymbol{x}^1, \boldsymbol{y}) \geq tm\}.$$

Then

$$\mathbf{P}\{\mathcal{A}(u)\} \leq (M-1)\mathbf{P}\left\{d(\boldsymbol{x}_1, \boldsymbol{y}) > (u-t)m\right\} \prod_{i=2}^{M-1} \mathbf{P}\left\{d(\boldsymbol{x}_i, \boldsymbol{y}) > um\right\} =$$

$$= (M-1)\mathbf{P}\{w(\boldsymbol{x}_1) > (u-t)m\}\mathbf{P}^{M-2}\{w(\boldsymbol{x}_2) > um\} \leq$$

$$\leq (M-1)\left[1 - \mathbf{P}\left\{w(\boldsymbol{x}_2) \leq um\right\}\right]^{M-2} \leq (M-1)\exp\left\{-(M-2)P\left\{w(\boldsymbol{x}) \leq um\right\}\right\},$$

where the inequality $(1-a)^b \leq e^{-ab}$, $b \geq 0$ was used. Note that

$$P\left\{w(\boldsymbol{x}) \leq um\right\} \geq 2^{-m} \binom{m}{um} \geq \frac{1}{(m+1)} 2^{-m} e^{mh(u)},$$

since [21, формула (12.40)] for any $0 \leq k \leq n$ the inequalities hold

$$\frac{1}{n+1} 2^{nh(k/n)} \leq \binom{n}{k} \leq 2^{nh(k/n)}.$$

Therefore

$$\mathbf{P}\{\mathcal{A}(u)\} \le \exp\left\{\frac{Rm}{\gamma} - \frac{(M-2)}{M(m+1)}\, e^{[R/\gamma + h(u) - \ln 2]m}\right\}.$$

We set u such that $[R/\gamma + h(u) - \ln 2]m \ge 4 \ln m$. Then for sufficiently large m we have $\mathbf{P}\{\mathcal{A}(u)\} \le e^{-m^2}$, and we may neglect the event of such small probability. Therefore the inequality (28) holds.

Assuming that $\boldsymbol{x}_{\text{true}} = \boldsymbol{x}_M$, For given $\boldsymbol{y}, \boldsymbol{x}', \boldsymbol{x}_M$ and randomly (equiprobably) chosen $\boldsymbol{x}_1, \boldsymbol{x}_2$ introduce the set

$$\mathcal{F}(\boldsymbol{y}, \boldsymbol{x}', \boldsymbol{x}_M) = \left\{\boldsymbol{x}_1, \boldsymbol{x}_2 :\ d(\boldsymbol{x}_1, \boldsymbol{y}) \le \delta_{GV}(R/\gamma)m - tm,\ d(\boldsymbol{x}_1, \boldsymbol{x}') \ge d(\boldsymbol{x}_2, \boldsymbol{x}')\ \right\}.$$

We are interested in the values $P_3 = \mathbf{P}\left\{\mathcal{F}(\boldsymbol{y}, \boldsymbol{x}', \boldsymbol{x}_M) \middle| \boldsymbol{y}, \boldsymbol{x}', \boldsymbol{x}_M\right\}$ and $\mathbf{E}_{\boldsymbol{y}, \boldsymbol{x}', \boldsymbol{x}_M} P_3^s$, $s \ge 0$.

Remark 6. In the definition of the set $\mathcal{F}(\boldsymbol{y}, \boldsymbol{x}', \boldsymbol{x}_M)$ we might include additional constraints: $d(\boldsymbol{x}_2, \boldsymbol{y}) \ge \delta_{GV}(R/\gamma)m$; $d(\boldsymbol{x}_2, \boldsymbol{y}) \ge d(\boldsymbol{x}_M, \boldsymbol{y})$. But it seems that they do not improve the exponent of P_3.

Note that if $d(\boldsymbol{y}, \boldsymbol{y}') \le tm$ then $P_{2n} = P_3 = 0$. Moreover, if $p_1 < t$ then

$$P_{2n} \le P\left\{d(\boldsymbol{y}, \boldsymbol{x}') \ge tm\right\} \le e^{-mD(t\|p_1)}. \tag{29}$$

If $d(\boldsymbol{y}, \boldsymbol{x}') > tm$, then for any nonnegative α, φ

$$\begin{aligned} P_3 &\le \mathbf{E}_{\boldsymbol{x}_1, \boldsymbol{x}_2}\left\{e^{\alpha[(\delta - t)m - d(\boldsymbol{x}_1, \boldsymbol{y})] + \varphi[d(\boldsymbol{x}_1, \boldsymbol{x}') - d(\boldsymbol{x}_2, \boldsymbol{x}')]} \middle| \boldsymbol{y}, \boldsymbol{x}', \boldsymbol{x}_M\right\} = \\ &= e^{\alpha(\delta - t)m} \mathbf{E}_{\boldsymbol{x}_1, \boldsymbol{x}_2}\left\{e^{-\alpha d(\boldsymbol{x}_1, \boldsymbol{y}) + \varphi[d(\boldsymbol{x}_1, \boldsymbol{x}') - d(\boldsymbol{x}_2, \boldsymbol{x}')]} \middle| \boldsymbol{y}, \boldsymbol{x}', \boldsymbol{x}_M\right\}. \end{aligned}$$

For any a, b and equiprobable $\boldsymbol{x}$

$$\mathbf{E}_{\boldsymbol{x}}\left[e^{ad(\boldsymbol{x}, \boldsymbol{y}) + bd(\boldsymbol{x}, \boldsymbol{x}')} \middle| \boldsymbol{y}, \boldsymbol{x}'\right] = 2^{-m}\left(1 + e^{a+b}\right)^m \left(\frac{e^a + e^b}{1 + e^{a+b}}\right)^{d(\boldsymbol{y}, \boldsymbol{x}')}.$$

Then when $d(\boldsymbol{y}, \boldsymbol{x}') > tm$, we have

$$P_3 \le 2^{-2m} e^{\alpha(\delta - t)m}\left(1 + e^{\varphi - \alpha}\right)^m \left(1 + e^{-\varphi}\right)^m \left[\frac{e^{-\alpha} + e^{\varphi}}{1 + e^{\varphi - \alpha}}\right]^{d(\boldsymbol{y}, \boldsymbol{x}')}.$$

Since $\mathbf{E} b^{d(\boldsymbol{y}, \boldsymbol{x}')} = (q_1 + p_1 b)^m$, then

$$\begin{aligned} \left\{\mathbf{E}\left[b^{d(\boldsymbol{y}, \boldsymbol{x}')}; d(\boldsymbol{y}, \boldsymbol{x}') > tm\right]\right\}^{1/m} &\le \left\{\min_{\mu \ge 0} \mathbf{E} b^{d(\boldsymbol{y}, \boldsymbol{x}') + \mu[d(\boldsymbol{y}, \boldsymbol{x}') - tm]}\right\}^{1/m} = \\ &= \min_{\mu \ge 0}\left\{b^{-\mu t}(q_1 + p_1 b^{1+\mu})\right\}. \end{aligned}$$

Note that ($b \ge 1$)

$$\begin{gathered} \min_{\mu \ge 0}\left\{b^{-\mu t}\left(z_1 + b^{1+\mu}\right)\right\} = e^{f_4(b, t, p_1)}, \\ f_4(b, t, p_1) = \begin{cases} h(t) + (1-t)\ln z_1 + t \ln b, & \ln(tz_1/(1-t)) \ge \ln b, \\ \ln(z_1 + b), & \ln(tz_1/(1-t)) \le \ln b, \end{cases} \end{gathered} \tag{30}$$

where minimum is attained when

$$\mu = \mu_0 = \left[\frac{\ln(tz_1/(1-t))}{\ln b} - 1\right]_+ .$$

Therefore for $b_1 \geq 1$ we have

$$\left(\mathbf{E}_{\boldsymbol{y},\boldsymbol{x}',\boldsymbol{x}_M} P_3^s\right)^{1/m} \leq 2^{-2s} e^{\alpha(\delta-t)s} e^{-(\alpha+\varphi)s + f_4(b_1^s,t,p_1)} \left(e^{\alpha}+e^{\varphi}\right)^s \left(e^{\varphi}+1\right)^s ,$$

where

$$b_1 = \frac{1+e^{\varphi+\alpha}}{e^{\alpha}+e^{\varphi}}.$$

We should minimize that expression over nonnegative α, φ. We have

$$\mathbf{E} e^{ad(\boldsymbol{x}_M,\boldsymbol{y})} = \left(q+pe^a\right)^m , \qquad \mathbf{E} b^{d(\boldsymbol{y},\boldsymbol{x}')} = (q_1+p_1 b)^m .$$

Denote

$$z = \frac{q}{p}, \qquad z_1 = \frac{q_1}{p_1}, \tag{31}$$

and note that

$$b_1 - 1 \sim \left(e^{\varphi} - e^{-\varphi}\right)\left(1-e^{-\alpha}\right) \geq 0 .$$

Then

$$\left(\mathbf{E}_{\boldsymbol{y},\boldsymbol{x}',\boldsymbol{x}_M} P_3^s\right)^{1/m} \leq 2^{-2s} p_1 e^{-[\alpha(1-\delta+t)+\varphi]s + f_4(b_1^s,t,p_1)} \left(e^{\alpha}+e^{\varphi}\right)^s \left(e^{\varphi}+1\right)^s . \tag{32}$$

We apply the random coding with expurgation method, using the inequality $\left(\sum a_i\right)^{1/\rho} \leq \sum a_i^{1/\rho}$, $\rho \geq 1$. We have

$$\mathbf{E}_{\mathcal{C}} P_{2n}^{1/\rho}(\mathcal{C}) \leq M^2 \mathbf{E}_{\boldsymbol{y},\boldsymbol{x}',\boldsymbol{x}_M} \mathbf{P}^{1/\rho}\left\{\mathcal{F}(\boldsymbol{y},\boldsymbol{x}',\boldsymbol{x}_M)\middle|\boldsymbol{y},\boldsymbol{x}',\boldsymbol{x}_M\right\} = M^2 E_{\boldsymbol{y},\boldsymbol{x}',\boldsymbol{x}_M} P_3^{1/\rho}$$

and then from (32) we get ($\rho = 1/s \geq 1$)

$$\left[\mathbf{E}_{\mathcal{C}} P_{2n}^{1/\rho}(\mathcal{C})\right]^{\rho/m} \leq e^{2R\rho/\gamma} 2^{-2} p_1^{\rho} e^{\rho f_4(b_1^s,t,p_1) - \alpha(1-\delta+t) - \varphi} \left(e^{\alpha}+e^{\varphi}\right)\left(e^{\varphi}+1\right).$$

To avoid bulky formulas, we choose the parameters such that the inequality holds (see (30))

$$\rho \ln(tz_1/(1-t)) \geq \ln b_1 . \tag{33}$$

Then

$$\left[\mathbf{E}_{\mathcal{C}} P_{2n}^{1/\rho}(\mathcal{C})\right]^{\rho/m} \leq 2^{-2} e^{G\rho+F_2}, \qquad b_1 = \frac{1+cd}{c+d},$$
$$G = 2R/\gamma + h(t) + \ln\left[p_1^t q_1^{1-t}\right] = 2R/\gamma - D\left(t\|p_1\right),$$
$$F_2 = -(1-\delta+t)\ln d - \ln c + t\ln(1+dc) + \ln(1+c) + (1-t)\ln(d+c),$$

and we should minimize F_2 over $c, d \geq 1$.

Note that F_2 does not depend on ρ. If $G < 0$ then the best is $\rho \to \infty$. Since $\Big[\mathbf{E}_{\mathcal{C}} P_{2n}^{1/\rho}(\mathcal{C})\Big]^{\rho/m} \to 0$, $\rho \to \infty$, we may assume that $P_{2n} = 0$. If $G \geq 0$ then the best is $\rho = 1$ (and then it is better to use simply the random coding method). In both cases we need the condition (33) be satisfied.

If $\rho \to \infty$ then the inequality (33) is equivalent to the condition $tz_1/(1-t) > 1$, i.e. $p_1 < t$. We set $t > p_1$ such that $2R/\gamma - D(t \,\|\, p_1) < 0$. Then $G < 0$, $P_{2n} = 0$, and from (26), (18) we get

$$F_1(R, p, p_1) \geq \max_{\gamma, t > p_1} \min \left\{ \gamma E_{\text{low}}(R/\gamma, p, 2) - \frac{t\gamma}{3} \ln \frac{q}{p}, \gamma E_{\text{low}}(R/\gamma, p) + (1-\gamma) E_2(p) \right\}. \quad (34)$$

Using $t = t_1(R, p_1) \geq p_1$ (see (9)) we get from (34)

$$\begin{aligned} F_1(R, p, p_1) \geq \max_{\gamma} \min \Big\{ & \gamma E_{\text{low}}(R/\gamma, p, 2) - \frac{\gamma t_1(R/\gamma, p_1)}{3} \ln \frac{q}{p}, \\ & \gamma E_{\text{low}}(R/\gamma, p) + (1-\gamma) E_2(p) \Big\}, \end{aligned} \quad (35)$$

from which the formulas (10), (11) and the Theorem 1 follow. ▲

Remark 7. Note that if $p_1 \to 0$, then $t_1 \to 0$ and the relation (35) transfers to the similar relation (19) for the channel with noiseless feedback.

To find the function $p_0(R, p)$ of the critical noise level in the feedback channel we set $\gamma \to 1$. Then $p_0 = p_0(R, p)$ is defined by the system of equations

$$\begin{aligned} E_{\text{low}}(R, p, 2) - \frac{t}{3} \ln \frac{q}{p} &= E_{\text{low}}(R, p), \\ D(t \,\|\, p_0) &= 2R. \end{aligned}$$

In other words, $t_0(R, p)$ and $p_0(R, p) \leq t_0(R, p)$ are defined by the formulas (7) and (8), respectively.

§ 5. When noisy feedback behaves like noiseless ?

How small should be p_1 in order to have the error exponent $F_1(R, p, p_1)$ close to the similar exponent $F_1(R, p)$ for noiseless feedback ? More exactly, when for a given $\alpha \in (0, 1)$ the inequality holds $F_1(R, p, p_1) - E(R, p) \leq (1-\alpha)[F_1(R, p) - E(R, p)]$?

We give a simple estimate for such p_1, considering only the case $R = 0$. For the optimal $\gamma = \gamma_0$ from (10), (11) we have ($E_2(p) = 2E(0, p)$)

$$\gamma_0 = \frac{2E(0, p)}{E(0, p, 2) + E(0, p) - p_1 \ln(q/p)/3}$$

and then

$$\begin{aligned} F_1(0, p, p_1) &= \frac{2E(0, p)[E(0, p, 2) - p_1 \ln(q/p)/3]}{E(0, p, 2) + E(0, p) - p_1 \ln(q/p)/3}, \\ F_1(0, p, p_1) - E(0, p) &= \frac{E(0, p)[E(0, p, 2) - E(0, p) - p_1 \ln(q/p)/3]}{E(0, p, 2) + E(0, p) - p_1 \ln(q/p)/3}. \end{aligned}$$

Now in order to have

$$F_1(0,p,p_1) - E(0,p) \geq (1-\alpha)[F_1(0,p) - E(0,p)],$$

it is sufficient to have

$$p_1 \leq \frac{3\alpha\,[E^2(0,p,2) - E^2(0,p)]}{[\alpha E(0,p,2) + (2-\alpha)E(0,p)]\ln(q/p)}.$$

Since $E(0,p,2) \geq E(0,p)$, without much loss, we may replace the last inequality by a stronger one:

$$p_1 \leq \frac{3\alpha\,[E(0,p,2) - E(0,p)]}{\ln(q/p)} = p_{11}(p,\alpha).$$

On Fig. 4 the plot of the function $p_{11}(p, 0.1)$ is given.

E x a m p l e 2. Consider the case $p = (1-\varepsilon)/2$, $\varepsilon \to 0$. Then $C(p) \approx \epsilon^2/2$ and $E(0,p,2) \approx 2C/3$, $E(0,p) \approx C/2$. As a result, we get

$$p_{11}(p,\alpha) = \frac{\alpha(1-2p)[1+o(1)]}{8}, \qquad p \to 1/2.$$

In other words, if the forward BSC(p) is very bad, then in order to improve its error exponent we need a very good feedback channel BSC(p_1).

§ 6. Auxiliary formulas and results

Lower bounds for the decoding error exponents. All formulas below are derived following Gallager's technique [15, 16].

1) *Random coding* bounds:

$$E(R,p,L) \geq E_{\rm r}(R,p,L), \qquad R \geq 0. \tag{36}$$

Moreover ($R_{{\rm crit},L}(p)$ определено в (1)),

$$E(R,p,L) = E_{\rm r}(R,p,L) = E_{\rm sp}(R,p), \qquad R_{{\rm crit},L}(p) \leq R \leq C(p), \tag{37}$$

and for $R \leq R_{{\rm crit},L}(p)$ we have

$$E(R,p,L) \geq E_{\rm r}(R,p,L) = L(\ln 2 - R) - (1+L)\ln\left[p^{1/(1+L)} + q^{1/(1+L)}\right]. \tag{38}$$

Since $R_{{\rm crit},L}(p) \to 0$, $L \to \infty$, then $E(R,p,L) \to E_{\rm sp}(R,p)$, $L \to \infty$ for any $R \geq 0$.

2) *Random coding with expurgation bound*:

$$E(R,p,L) \geq E_{\rm ex}(R,p,L) = \max_{\rho \geq 1}\left\{-\rho L R - \rho \ln f(p,L,\rho)\right\}, \qquad R \geq 0, \tag{39}$$

where

$$f(p,L,\rho)=2^{-(L+1)}\left\{2+\sum_{i=1}^{L}\binom{L+1}{i}a_i^{1/\rho}\right\},$$
$$a_i=p\left(\frac{q}{p}\right)^{i/(L+1)}+q\left(\frac{p}{q}\right)^{i/(L+1)}.$$

The bound (39) improves the random coding bound (38) for $0\le R<R_{\min,L}(p)$ (see (42), but it does not give $E_{\rm sp}(R,p)$. Note also that

$$f(p,L,\rho)=\mathbf{E}\sum_{m,m_1,\ldots,m_L}\left[\sum_{\boldsymbol{y}}\left[\mathbf{P}\left(\boldsymbol{y}|\boldsymbol{x}_m\right)\mathbf{P}\left(\boldsymbol{y}|\boldsymbol{x}_{m_1}\right)\ldots\mathbf{P}\left(\boldsymbol{y}|\boldsymbol{x}_{m_L}\right)\right]^{1/(L+1)}\right]^{1/\rho}, \tag{40}$$

where all components of each codeword $\boldsymbol{x}_i$ are chosen independently and equiprobably from 0 and 1.

In particular,

$$E_{\rm ex}(R,p)=E_{\rm ex}(R,p,1)=\max_{\rho\ge1}\left\{\rho\ln2-\rho R-\rho\ln\left[1+(2\sqrt{pq})^{1/\rho}\right]\right\},$$
$$E_{\rm ex}(R,p,2)=\max_{\rho\ge1}\left\{\rho\ln4-2\rho R-\rho\ln\left[1+3\left(p^{1/3}q^{2/3}+p^{2/3}q^{1/3}\right)^{1/\rho}\right]\right\}.$$

The functions $E(R,p,L)$, $E_{\rm r}(R,p,L)$ and $E_{\rm ex}(R,p,L)$ does not decreases on L. In particular,

$$E_{\rm ex}(R,p)<E_{\rm ex}(R,p,2),\qquad R<R_{\rm crit}(p). \tag{41}$$

In order to get a more convenient representation for the functions $E_{\rm ex}(R,p)$ and $E_{\rm ex}(R,p,L)$, introduce rates

$$R_{\min,L}(p)=\ln2-\frac{(L+1)}{L}\ln\left[p^{1/(L+1)}+q^{1/(L+1)}\right]-\frac{\sum_{i=1}^{L}\binom{L+1}{i}a_i\ln a_i}{2L\left[p^{1/(L+1)}+q^{1/(L+1)}\right]^{L+1}}. \tag{42}$$

The function $R_{\min,L}(p)$ monotonically decreases on L and $R_{\min,L}(p)<R_{{\rm crit},L}(p)$, if $L\ge1$ and $0<p<1/2$. In particular,

$$\begin{aligned}R_{\min}(p)=R_{\min,1}(p)&=\ln2-h\left(\frac{2\sqrt{pq}}{1+2\sqrt{pq}}\right),\\ R_{\min,2}(p)&=\ln2-\frac{1}{2}\left[\ln(1+3a_1)-\frac{3a_1\ln a_1}{1+3a_1}\right],\qquad a_1=p^{1/3}q^{2/3}+p^{2/3}q^{1/3}.\end{aligned} \tag{43}$$

We also have $R_{\min,2}(p)<R_{\min,1}(p)<R_{\rm crit}(p)$, $0<p<1/2$.

Now

$$\begin{gathered}E_{\rm ex}(R,p,L)<E_{\rm r}(R,p,L)=E_{\rm sp}(R,p)\qquad R>R_{{\rm crit},L}(p),\\ E_{\rm ex}(R,p,L)=E_{\rm r}(R,p,L),\qquad R_{\min,L}(p)\le R\le R_{{\rm crit},L}(p),\\ E_{\rm ex}(R,p,L)>E_{\rm r}(R,p,L),\qquad 0\le R<R_{\min,L}(p).\end{gathered}$$

Moreover,

$$E_{\rm ex}(R,p)=\frac{\delta_{GV}(R)}{2}\ln\frac{1}{4pq}\,,\qquad 0\le R\le R_{\min}(p). \tag{44}$$

Note also that $0\le R\le R_{\min}(p)$ corresponds to the case $\delta_{GV}(R)\ge(2\sqrt{pq})/(1+2\sqrt{pq})$.

If $L=2$, the

$$E_{\rm ex}(R,p,2)=-v\ln a_1\,,\qquad 0\le R\le R_{\min,2}(p),$$

where a_1 is defined in в (43), and v is the unique root of the equation

$$\ln 4-h(v)-v\ln 3=2R\,,\qquad 0\le v<\frac{3}{4}\,.$$

In particular,

$$\begin{aligned}E_{\rm ex}(0,p)=E(0,p)&=\frac{1}{2}\ln\frac{1}{2\sqrt{pq}}\,,\\ E_{\rm ex}(0,p,2)=E(0,p,2)&=-\frac{3}{4}\ln\left(p^{1/3}q^{2/3}+p^{2/3}q^{1/3}\right),\end{aligned} \tag{45}$$

(the second relation is established in [18]).

Existence of code with given properties. We are interested in a code $\mathcal{C}$ such that each its codeword has certain properties $\mathcal{A}_1,\mathcal{A}_2,\ldots$. For that purpose we use the following result which is a natural modification of the cute Lemma 5.7 from [16].

Assume that we choose randomly (in arbitrary way) a code $\mathcal{C}$ with M' codewords $\boldsymbol{x}_m$, and for each $\boldsymbol{x}_m$, $m=1,\ldots,M'$ we have

$$\mathbf{P}_{\text{over codes}}\{\boldsymbol{x}_m \text{ does not have property } \mathcal{A}\}\le 1/2\,. \tag{46}$$

L e m m a. *If the condition* (46) *is satisfied then there exists a code in the ensemble of codes with* $M'=2M-1$ *codewords for which, at least, for* M *its codewords the property* $\mathcal{A}$ *is fulfilled.*

P r o o f remains the same as in [16, Lemma 5.7] (it is the changing of the summation order in the corresponding double sum). ▲

If there are, say, four properties $\mathcal{A}_1,\ldots,\mathcal{A}_4$, then assume that for each $\boldsymbol{x}_m$, $m=1,\ldots,M'$, we have

$$\mathbf{P}_{\text{over codes}}\{\boldsymbol{x}_m \text{ does not have property } \mathcal{A}_i\}\le 1/8\,,\qquad i=1,\ldots,4\,. \tag{47}$$

C o r o l l a r y 2. *If the condition* (47) *is satisfied, then there exists a code in the ensemble of codes with* $M'=2M-1$ *codewords for which, at least, for* M *its codewords all four properties* $\mathcal{A}_i$, $i=1,\ldots,4$ *are fulfilled.*

In our case the property $\mathcal{A}_1$ means that the codeword $\boldsymbol{x}_m$ has small decoding error probability; $\mathcal{A}_2$ means that $\boldsymbol{x}_m$ has small list size $L=2$ decoding error probability; $\mathcal{A}_3,\mathcal{A}_4$ mean that for the codeword $\boldsymbol{x}_m$ the relations (18) and (26), respectively, hold.

Proof of the formula (18). Consider a code $\mathcal{C}$ with M codewords $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_M$ of length $n+k$. Each codeword $\boldsymbol{x}_i$ has the form $\boldsymbol{x}_i=(\boldsymbol{x}_i',\boldsymbol{x}_i'')$, where $\boldsymbol{x}_i'$ has length n and $\boldsymbol{x}_i''$ has length

k. We suppose that the parts $\{\boldsymbol{x}''_i\}$ are given, while the parts $\{\boldsymbol{x}'_i\}$ are chosen randomly (in some way). We also assume that

$$\min_{i\neq j} d\left(\boldsymbol{x}''_i, \boldsymbol{x}''_j\right) = \delta k\,. \tag{48}$$

Using maximum likelihood decoding, denote by $P_{e,m}$ the conditional decoding error probability provided the codeword $\boldsymbol{x}_m$ was transmitted. An output block $\boldsymbol{y}$ has the form $\boldsymbol{y} = (\boldsymbol{y}', \boldsymbol{y}'')$, where $\boldsymbol{y}', \boldsymbol{y}''$ have length n and k, respectively. Then $\mathbf{P}\left(\boldsymbol{y}|\boldsymbol{x}_m\right) = \mathbf{P}\left(\boldsymbol{y}'|\boldsymbol{x}'_m\right)\mathbf{P}\left(\boldsymbol{y}''|\boldsymbol{x}''_m\right)$. Using the inequality $\left(\sum a_i\right)^s \leq \sum a_i^s$, $0 \leq s \leq 1$, and the formula

$$\sum_{\boldsymbol{y}'} \sqrt{\mathbf{P}\left(\boldsymbol{y}'|\boldsymbol{x}'_m\right)\mathbf{P}\left(\boldsymbol{y}'|\boldsymbol{x}'_{m'}\right)} = (4pq)^{d\left(\boldsymbol{x}'_m, \boldsymbol{x}'_{m'}\right)/2}\,,$$

we have

$$\begin{aligned} P_{e,m}^s &\leq \sum_{m'\neq m}\left[\sum_{\boldsymbol{y}} \sqrt{\mathbf{P}\left(\boldsymbol{y}|\boldsymbol{x}_m\right)\mathbf{P}\left(\boldsymbol{y}|\boldsymbol{x}_{m'}\right)}\right]^s = \\ &= \sum_{m'\neq m}\left[\sum_{\boldsymbol{y}'} \sqrt{\mathbf{P}\left(\boldsymbol{y}'|\boldsymbol{x}'_m\right)\mathbf{P}\left(\boldsymbol{y}'|\boldsymbol{x}'_{m'}\right)}\right]^s \left[\sum_{\boldsymbol{y}''} \sqrt{\mathbf{P}\left(\boldsymbol{y}''|\boldsymbol{x}''_m\right)\mathbf{P}\left(\boldsymbol{y}''\boldsymbol{x}''_{m'}\right)}\right]^s \leq \\ &\leq \sum_{m'\neq m}\left[\sum_{\boldsymbol{y}'} \sqrt{\mathbf{P}\left(\boldsymbol{y}'|\boldsymbol{x}'_m\right)\mathbf{P}\left(\boldsymbol{y}'|\boldsymbol{x}'_{m'}\right)}\right]^s \left[\max_{m_1\neq m_2} (2\sqrt{pq})^{d\left(\boldsymbol{x}''_{m_1}, \boldsymbol{x}''_{m_2}\right)}\right]^s = \\ &= (2\sqrt{pq})^{\delta sk} \sum_{m'\neq m}\left[\sum_{\boldsymbol{y}'} \sqrt{\mathbf{P}\left(\boldsymbol{y}'|\boldsymbol{x}'_m\right)\mathbf{P}\left(\boldsymbol{y}'|\boldsymbol{x}'_{m'}\right)}\right]^s = (2\sqrt{pq})^{\delta sk} \sum_{m'\neq m} (4pq)^{sd\left(\boldsymbol{x}'_m, \boldsymbol{x}'_{m'}\right)/2}\,. \end{aligned} \tag{49}$$

Consider an ensemble of codes in which each codeword $\boldsymbol{x}'_m$ is selected independently with the probability 2^{-n} among all possible binary vectors of length n. Since

$$\mathbf{E} z^{d\left(\boldsymbol{x}'_m, \boldsymbol{x}'_{m'}\right)} = \mathbf{E} z^{w\left(\boldsymbol{x}'_m\right)} = \left(\frac{1+z}{2}\right)^n\,,$$

we get

$$\left(\mathbf{E} P_{e,m}^s\right)^{1/s} \leq (2\sqrt{pq})^{\delta k}\left\{e^R 2^{-1}\left[1 + (2\sqrt{pq})^s\right]\right\}^{n/s}.$$

Further derivation follows Theorem 5.7.1 from [16]. As a result, defining $\rho = 1/s$, $\rho \geq 1$, we get that there exists a code with M codewords such that for any $m = 1, \ldots, M$ we have

$$\frac{1}{n}\ln\frac{1}{P_{e,m}} \geq \frac{\delta k}{n}\ln\frac{1}{2\sqrt{pq}} + \max_{\rho\geq 1}\left\{\rho\ln 2 - \rho R - \rho\ln\left[1 + (2\sqrt{pq})^{1/\rho}\right]\right\}.$$

From that relation the formula (18) follows. ▲

The authors wish to thank the University of Tokyo for supporting this joint research.

REFERENCES

1. *Burnashev M.V., Yamamoto H.* On zero-rate error exponent for BSC with noisy feedback // Problems of Inform. Transm. 2008. V. 44, № 3. P. 33–49.

2. *Shannon C.E.* The Zero Error Capacity of a Noisy Channel // IRE Trans. Inform. Theory. 1956. V. 2. № 3. P. 8–19.

3. *Dobrushin R.L.* Asymptotic bounds on error probability for message transmission in a memoryless channel with feedback // Probl. Kibern. No. 8. M.: Fizmatgiz, 1962. P. 161–168.

4. *Horstein M.* Sequential Decoding Using Noiseless Feedback // IEEE Trans. Inform. Theory. 1963. V. 9. № 3. P. 136–143.

5. *Berlekamp E.R.*, Block Coding with Noiseless Feedback. Ph.D. Thesis. MIT, Dept. Electrical Enginering, 1964.

6. *Elias P.* Coding for Noisy Channels // IRE Conv. Rec. 1955. V. 4. P. 37–46. Reprinted in Key Papers in the Development of Information Theory. New York: IEEE Press, 1974. P. 102–111.

7. *Burnashev M.V.* Data transmission over a discrete channel with feedback: Random transmission time // Problems of Inform. Transm. 1976. V. 12, № 4. P. 10–30.

8. *Burnashev M.V.* On a Reliability Function of Binary Symmetric Channel with Feedback // Problems of Inform. Transm. 1988. V. 24, № 1. P. 3–10.

9. *Pinsker M.S.* The probability of error in block transmission in a memoryless Gaussian channel with feedback // Problems of Inform. Transm. 1968. V. 4, № 4. P. 3–19.

10. *Schalkwijk J.P.M., Kailath T.* A Coding Scheme for Additive Noise Channels with Feedback - I: No Bandwidth Constraint // IEEE Trans. Inform. Theory. 1966. V. 12. № 2. P. 172–182.

11. *Tchamkerten A., Telatar E.* Variable Length Coding over an Unknown Channel // IEEE Trans. Inform. Theory. 2006. V. 52. № 5. P. 2126–2145.

12. *Yamamoto H., Itoh R.* Asymptotic Performance of a Modified Schalkwijk–Barron Scheme for Channels with Noiseless Feedback // IEEE Trans. Inform. Theory. 1979. V. 25. № 6. P. 729–733.

13. *Draper S.C., Sahai A.* Noisy Feedback Improves Communication Reliability // Proc. IEEE Int. Sympos. on Information Theory. Seattle, USA. July 9–14, 2006, P. 69–73.

14. *Kim Y.-H., Lapidoth A., Weissman T.* The Gaussian Channel with Noisy Feedback // Proc. IEEE Int. Sympos. on Information Theory. Nice, France. June 24–29, 2007. P. 1416–1420.

15. *Gallager R.G.* A Simple Derivation of the Coding Theorem and some Applications // IEEE Trans. Inform. Theory. 1965. V. 11. P. 3–18.

16. *Gallager R.G.* Information theory and reliable communication. Wiley, NY, 1968.

17. *Burnashev M.V.* Code spectrum and reliability function: binary symmetric channel – II // Problems of Inform. Transm. (in press).

18. *Blinovsky V.M.* Error probability exponent of list decoding at low rates // Problems of Inform. Transm. 2001. V. 37, № 4. P. 277–287.

19. *Burnashev M.V., Yamamoto H.* Noisy Feedback Improves the BSC Reliability Function // Proc. IEEE Int. Sympos. on Information Theory. Seoul, Korea. June 28–July 3, 2009. P. 1501–1505.

20. *Shannon C.E., Gallager R.G.. Berlekamp E.R.* Lower bounds to error probability for codes on discrete memoryless channels // Information and Control. 1967. V. 10, Part I, P. 65–103; Part II, P. 522–552.

21. *Cover T.M., Thomas J.A.* Elements of Information Theory. New York: Wiley. 1991.

Burnashev Marat Valievich
Institute for Information Transmission Problems RAS
burn@iitp.ru
Yamamoto Hirosuke
The University of Tokyo, Japan
hirosuke@ieee.org

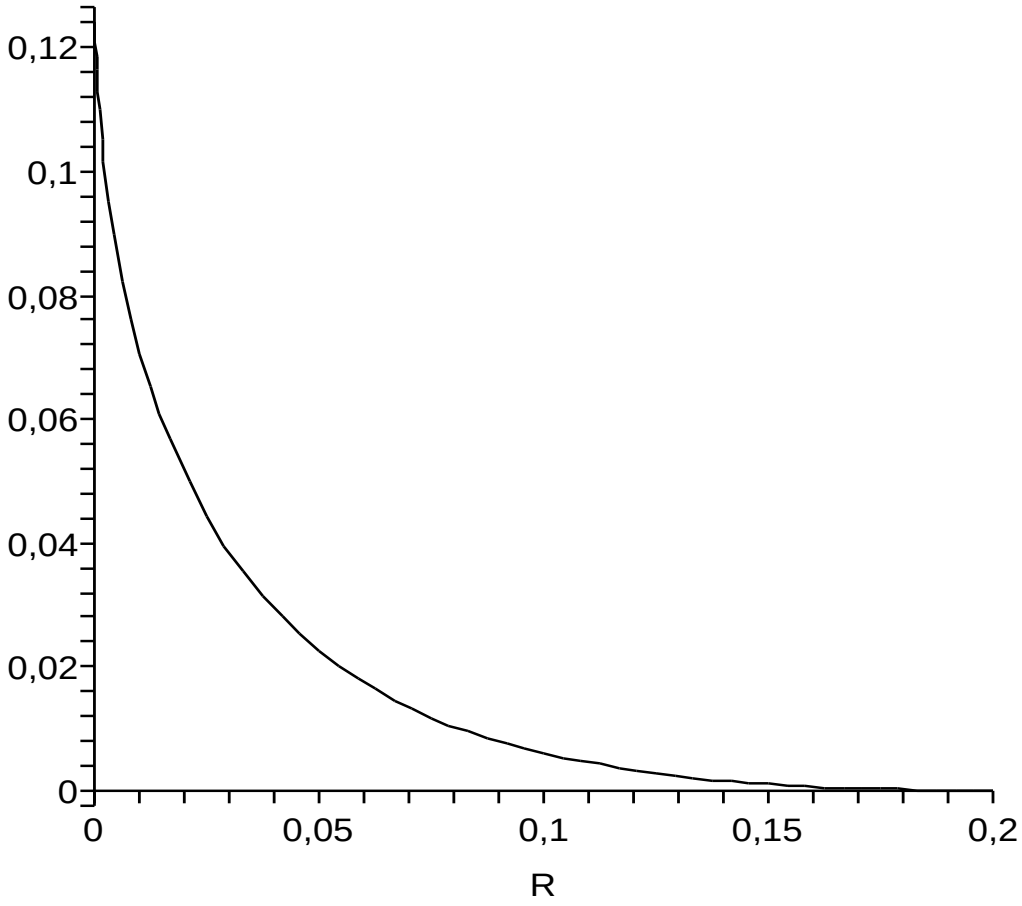

Fig. 2. The plot of the function $p_1(R, 0.01)$ ($R_{\text{crit}} \approx 0.387$).

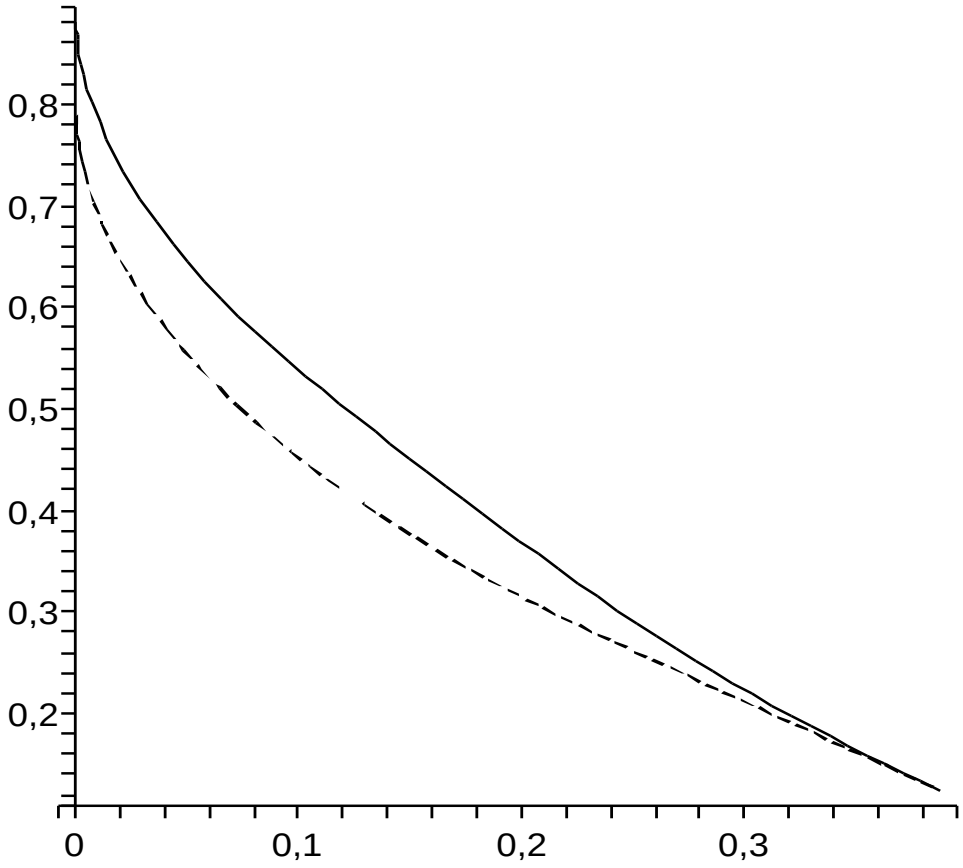

Fig. 3. The plots of the functions $F_1(R,p)$ and $E_{\rm ex}(R,p)$ for $p=0.01$ ($R_{\rm crit}\approx 0.387$).

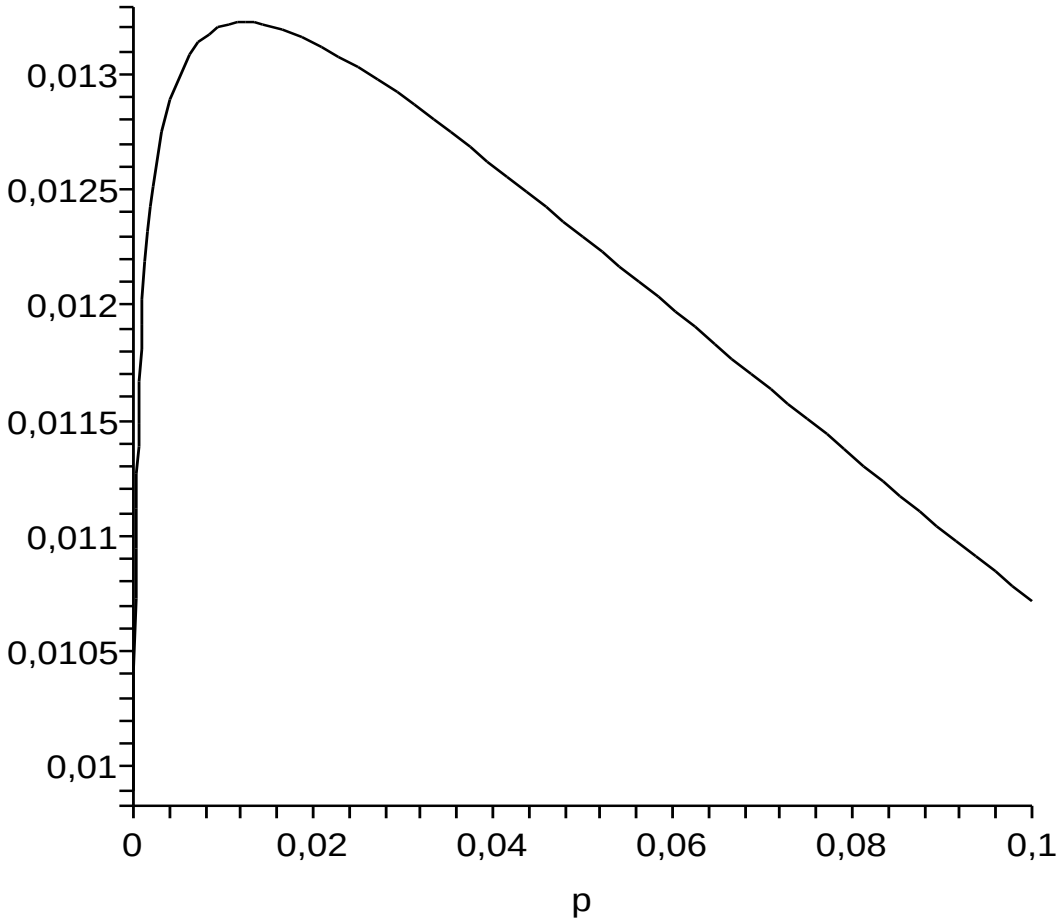

Fig. 4. The plot of the function $p_{11}(p, 0.1)$