Identification under Random Processes *

Rudolf Ahlswede and Vladimir B. Balakirsky

Abstract. After it has been understood how common randomness of sender and receiver's increases identification capacity, we study here the effect of randomness of a more general structure, namely a correlated source, when the sender has access to one marginal-source and the receivers have access to the other marginal source.
The typical phenomena can be and are explained for binary symmetric correlated sources.

1 Introduction

Identification via channels was introduced in [1,2] and can be viewed as a problem of a reliable transmission of one of N messages to N receivers in such a way that each receiver decides whether it is his message, which was sent, or not. If a receiver misses his message then the decoding error of the first kind takes place, and if a receiver accepts a message which was sent for a different receiver, then the decoding error of the second kind takes place. An identification scheme should be constructed in such a way that the decoding error probabilities of the both kinds are small for all receivers.

A key idea of constructing good identification codes is an assignment of the probability distributions (PDs) to the messages instead of fixed codewords. Suppose we a given a discrete memoryless channel W having the input alphabet $\mathcal{X}$ and the output alphabet $\mathcal{Y}$, i.e., the probability to receive a vector $\mathbf{y} = (y^{(1)}, ..., y^{(n)}) \in \mathcal{Y}^n$, when a vector $\mathbf{x} = (x^{(1)}, ..., x^{(n)}) \in \mathcal{X}^n$ was transmitted, is defined as

$$W(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^{n} W(y^{(t)}|x^{(t)}).$$

Definition 1 An $(n, N, \lambda_1, \lambda_2)$ ID code is a collection

$$\{ \ (Q(\cdot|i), \mathcal{D}_i), \ i = 1, ..., N \ \}$$

such that, for all $i, j = 1, ..., N$,

1) $Q(\cdot|i) = \{ \ Q(\mathbf{x}|i), \mathbf{x} \in \mathcal{X}^n \ \}$ is a PD;
2) $\mathcal{D}_i \subset \mathcal{Y}^n$;
3) $\sum_{\mathbf{x}} Q(\mathbf{x}|i) \cdot W(\mathcal{D}_i^c|\mathbf{x}) \leq \lambda_1$;
4) $\sum_{\mathbf{x}} Q(\mathbf{x}|j) \cdot W(\mathcal{D}_i|\mathbf{x}) \leq \lambda_2$ if $i \neq j$.

The input distributions of an ID code can be realized if the sender has N collections of not necessarily distinct codewords. To send the i-th message, he selects one of the codewords of the i-th collection in accordance with the uniform distribution. Each receiver knows his collection and checks if the received vectors is likely to be generated by one of the codewords of this collection or not. The known results show that the number of messages that can be reliably transmitted using an ID code turns out to be doubly exponential in the block-length n [1].

We will consider the system of information transmission given in Fig.1, where the sender and the receivers have an access to a public random process and get correlated binary sequences $\mathbf{x}^*$ and $\mathbf{y}^*$ of length k before the sender generates a codeword, depending of the message i, and transmits it over a memoryless channel W. The probability to receive a pair $(\mathbf{x}^*, \mathbf{y}^*)$ is defined as

$$P(\mathbf{x}^*, \mathbf{y}^*) = 2^{-k} p^{d_H(\mathbf{x}^*, \mathbf{y}^*)} (1-p)^{k - d_H(\mathbf{x}^*, \mathbf{y}^*)},$$

where d_H denotes the Hamming distance and $p \leq 1/2$. Our goal is to evaluate the decoding error probabilities of the first and of the second kind as functions of the number of possible messages, the channel capacity C, the probability p, and the ratio $\rho = k/n$.

A randomization of the transmission can be represented in such a way that the sender observes a vector $\mathbf{z}$ of length l, whose components are independent identically distributed (i.i.d.) binary random variables taking values 0 and 1 with the probability 1/2, and uses this vector as a pointer to a collection corresponding to the message to be sent. Such a procedure can be also realized if the sender uses the vector $\mathbf{x}^* \in \{0, 1\}^k$ for this purpose. The difference of these two possibilities is concluded in the note that in the second case the receivers have a side information on the pointer, which is $\mathbf{y}^* \in \{0, 1\}^k$, and that the length k is assumed to be given, while the length l can be chosen in an arbitrary way.

Let us denote by $R_\rho(C)$ $\frac{1}{n}$times the log log[2] of the number of messages to be identified with arbitrarily small decoding error probabilities (more explicitly, $R_\rho(C)$ is defined in Section 2) and consider the two cases when $p = 0$ and $p = 1/2$.

* The work was supported in part by SFB-343, Universität Bielefeld, Germany

[2] All the logarithms in the paper are to the base 2. Furthermore, we write 2^z as $\exp_2 z$

If $p = 1/2$, then we have a classical identification model where the channel W is the only link, connecting the sender and the receivers. Therefore, we refer to the known results [1,4] and mention that $R_\rho(C) = C$, that the length l should be chosen approximately as nC, and that the observations of a random process are useless (Fig.2).

Next we discuss the case $p = 0$. Here the sender and the receivers have access to a common random experiment. This case arose in the study of identification in the presence of noiseless feedback. (It also plays a fundamental role for secrecy systems [5]). In particular, asymptotically optimal identification scheme with a noiseless feedback consists of two parts [2]. In the first part, the sender and the receivers distribute the result of a random experiment, which is a realization of a noise sequence during transmission of a given or randomly chosen sequence over the channel W. This result becomes common because of feedback. The distribution of a common randomness occupies almost all transmission time and completely determines the asymptotic characteristics of the identification system. Our model is close to the identification with the feedback. The main difference is that the random experiment is public and it occurs somewhere outside the system. Therefore, we cannot control it, but we also do not include its duration into the transmission time. However, using the results of [2] we can show that $R_\rho(C) = C + \rho$ for all $\rho \geq 0$ when $p = 0$ (Fig.2).

The cases considered above show that there is a limit on the length $k = \rho n$ when $p = 1/2$ and that this limit is absent when $p = 0$. In this paper we examine the general case, $p \in [0, 1/2]$, and come to the conclusion that, from a point of view of the direct coding theorems, this limit does not exist for all $p < 1/2$, but the additional quantity in the maximal identification rate, which we gain while observing the process, is limited at a threshold value for all $p > 0$. We give an expression for that value.

During the analysis, we distinguish between two cases when $\rho H(p) < C$ and when $\rho H(p) \geq C$, where $H(p) = -p \log p - (1-p)\log(1-p)$ is the entropy of the noise sequence which makes the observations of the process different at the sender's and at the receiver's sides. In the first case the encoder can improve the behaviour using an additional randomization, and in the second case there is more than enough randomness in these observations.

There was a hint to represent identification via channels as a model for human communication ("in a stormy night one sailor drowns in the ocean..." [1,p.27]). We can also add to this saying that the observation of a process is like listening for a kind of prayer, which allows the sender to predict a way of thinking of the receivers before he transmits the messages to these receivers.

The paper is organized as follows. In Section 2 we give basic notations and ideas, which concern different ways of partitioning the space, because we meet the problem of upper-bounding the expectation of a product of dependent random variables, and the partitions give an opportunity to select the subsets of independent variables. Two identification schemes, with and without additional randomization, are described in Section 3, and corresponding code ensembles are introduced in Section 4. The direct coding theorems for these cases are given in Section 5. Section 6 is devoted to the calculation of the increase in the maximal identification rate, $\Delta R_\infty(C)$, which comes with the infinite observations. Finally, in Section 7, we put our discoveries in a broader perspective. Appendix contains results needed in the analysis.

2 Basic notations and ideas

Let us start with the definitions.

Definition 2 *An $(n, N, \lambda_1, \lambda_2)$ ID code for identification under a random process defined by the distribution*

$$\mathcal{P}^k = \{ \, P(\mathbf{x}^*, \mathbf{y}^*), \ \mathbf{x}^*, \mathbf{y}^* \in \{0,1\}^k \, \}$$

is a collection

$$\{ \, (Q(\cdot|i, \mathbf{x}^*), \mathcal{D}_i(\mathbf{y}^*)), \ \mathbf{x}^*, \mathbf{y}^* \in \{0,1\}^k, \ i = 1, ..., N \, \}$$

such that, for all $\mathbf{x}^*, \mathbf{y}^* \in \{0,1\}^k$ and $i, j = 1, ..., N$,

1) $Q(\cdot|i, \mathbf{x}^*) = \{ \, Q(\mathbf{x}|i, \mathbf{x}^*), \mathbf{x} \in \mathcal{X}^n \, \}$ is a PD;
2) $\mathcal{D}_i(\mathbf{y}^*) \subset \mathcal{Y}^n$;
3) $\sum_{\mathbf{x}^*, \mathbf{y}^*} P(\mathbf{x}^*, \mathbf{y}^*) \cdot \sum_{\mathbf{x}} Q(\mathbf{x}|i, \mathbf{x}^*) \cdot W(\mathcal{D}_i^c(\mathbf{y}^*)|\mathbf{x}) \leq \lambda_1$;
4) $\sum_{\mathbf{x}^*, \mathbf{y}^*} P(\mathbf{x}^*, \mathbf{y}^*) \cdot \sum_{\mathbf{x}} Q(\mathbf{x}|j, \mathbf{x}^*) \cdot W(\mathcal{D}_i(\mathbf{y}^*)|\mathbf{x}) \leq \lambda_2$ if $i \neq j$.

Definition 3 The triple (R, e_1, e_2) will be referred to as a *ρ-achievable ID triple under a random process $\mathcal{P}^k$, $k = \rho n$,* if, for all $\delta > 0$ and $n \geq n(\delta, |\mathcal{X}|, |\mathcal{Y}|)$, an $(n, N, \lambda_1, \lambda_2)$ ID code for identification under the process $\mathcal{P}^k$ with

$$\frac{1}{n} \log\log N \geq R - \delta, \ \ \lambda_1 \leq 2^{-n(e_1 - \delta)}, \ \ \lambda_2 \leq 2^{-n(e_2 - \delta)}$$

exists.

Definition 4 The parameter R will be referred to as a *ρ-achievable ID rate under a random process $\mathcal{P}^k$, $k = \rho n$,* if there exist $e_1, e_2 > 0$ such that (R, e_1, e_2) is a ρ-achievable ID triple under that process. The ρ-achievable ID rate will be denoted by $R_\rho(C)$.

To realize an identification code, we will use an error-correcting code G consisting of

$$m = 2^{nr}$$

codewords $\mathbf{x} \in \mathcal{X}^n$. We suppose that the use of this code to transmit data over a memoryless channel W provides the exponent of the maximal decoding error probability at level $2^{-ne(r)}$ and denote the decoding sets by $\mathcal{D}_G(\mathbf{x}), \mathbf{x} \in G$, i.e.,

$$W(\mathcal{D}_G(\mathbf{x})|\mathbf{x}) \geq 1 - 2^{-ne(r)}, \ \text{ for all } \ \mathbf{x} \in G.$$

We assign the encoding function for the i-th message in such a way that it takes values in the code G. Hence, the transmitted codeword always belongs to G.

Each receiver can construct a list of the vectors $\mathbf{x}^*$ in such a way that, with high probability, the list contains the vector observed by the sender. Therefore, the decoding error probabilities of the first kind are estimated by the sum consisting of the probability that the sender's observation does not belong to the list and the maximal decoding error probability for the code G.

To estimate the decoding error probabilities of the second kind we introduce the code ensembles. These ensembles are different for the cases $\rho H(p) < C$ and $\rho H(p) \geq C$. However, in every case we prove that, for any fixed pair (i,j), $i \neq j$, an upper bound of the following form :

$$Pr\{\ \lambda_j(i) > \lambda\ \} \leq \exp_2\{\ -2^{ne}\ \} \tag{2.1}$$

is valid, where $\lambda_j(i)$ is the decoding error probability of the i-th receiver when the message j was sent, λ and e are constant chosen in a special way, and $Pr\{\ \}$ is the probability in the code ensemble. Then

$$Pr\{\ \lambda_j(i) > \lambda \text{ for at least one } (i,j),\ i \neq j\ \} \leq \sum_{j=1}^{N} \sum_{i \neq j} Pr\{\ \lambda_j(i) > \lambda\ \} < N^2 \cdot \exp_2\{\ -2^{ne}\ \}. \tag{2.2}$$

If $N < \exp_2\{\ 2^{ne}/2\ \}$, then the probability at the left hand side of (2.2) is less than 1 and, with a positive probability, there exists an ID code such that $\lambda_j(i) \leq \lambda$ for all $i \neq j$.

Let us omit a possibility to randomize the transmission on the basis of the vector $\mathbf{z}$, which 'is observed' only by the sender. Then

$$Pr\{\ \lambda_j(i) > \lambda\ \} = Pr\left\{\ 2^{-k} \sum_{\mathbf{y}^*} \eta_{ij}(\mathbf{y}^*) > \lambda \mid F_j(\mathbf{x}^*) \text{ was sent }\right\},$$

where $\eta_{ij}(\mathbf{y}^*)$ is a Boolean random variable, which is equal to 1 if the i-th receiver decides that his message was sent, while the sender had encoded the message j. The encoding process is controlled by the vectors $\mathbf{x}^*$ and, with high probability, the receiver can predict a certain vector if he decides that this vector belongs to the set

$$T^k_{[p]}(\mathbf{y}^*) = \{\ \hat{\mathbf{x}}^* :\ d_H(\hat{\mathbf{x}}^*, \mathbf{y}^*) \in [k(p - \Delta p'),\ k(p + \Delta p)]\ \}, \tag{2.3}$$

where the parameters $\Delta p', \Delta p > 0$ affect the probability that the receiver fails. Since the decoding error probability of the first kind should be small, the receiver has to consider all vectors $\hat{\mathbf{x}}^* \in T^k_{[p]}(\mathbf{y}^*)$ as possible candidates for the sender's observation $\mathbf{x}^*$. If the encoding is realized on the basis of some functions f_i and ψ, i.e., the sender transmits $F_j(\mathbf{x}^*) = f_j(\psi(\mathbf{x}^*))$, then the i-th receiver has to consider $\{f_i(\psi(\hat{\mathbf{x}}^*)), \hat{\mathbf{x}}^* \in T^k_{[p]}(\mathbf{y}^*)\}$ as the list of possible transmitted codewords. These lists can overlap for different $\mathbf{y}^*$ even when $\psi(\mathbf{x}^*) = \mathbf{x}^*$, and $\{\eta_{ij}(\mathbf{y}^*)\}$ is the set of dependent random variables. To prove (2.1) we will use Chernov's upper bound and estimate the generating function of $\sum_{\mathbf{y}^*} \eta_{ij}(\mathbf{y}^*)$. Our estimate is based on selecting the maximal independent subsets of variables $\eta_{ij}(\mathbf{y}^*)$, which is realized using the known constructions of error-correcting codes.

In the further considerations we will use the relation $'\sim'$ and write $a_k \sim b_k$ if $\lim_{k\to\infty} \frac{\log a_k}{k} = \lim_{k\to\infty} \frac{\log b_k}{k}$, where we also assume that the limits exist. Furthermore, we write : $c_k \lesssim b_k$ if $c_k \leq a_k$ and $a_k \sim b_k$. for some a_k.

Given $a, d = 0, ..., k$ and $\mathbf{x}^*, \mathbf{y}^* \in \{0,1\}^k$, we denote

$$\begin{aligned} \tilde{S}^k_a(\mathbf{x}^*) &= \{\ \hat{\mathbf{x}}^* :\ d_H(\hat{\mathbf{x}}^*, \mathbf{x}^*) \leq a\ \}, \\ S^k_d(\mathbf{y}^*) &= \{\ \hat{\mathbf{x}}^* :\ d_H(\hat{\mathbf{x}}^*, \mathbf{y}^*) = d\ \} \end{aligned} \tag{2.4}$$

and

$$c_d = \binom{k}{d},\ \ p_d = p^d(1-p)^{k-d},\ \ C_{[p]} = |T^k_{[p]}(\mathbf{y}^*)|,\ \ P(T^k_{[p]}(\mathbf{y}^*)|\mathbf{y}^*), \tag{2.5}$$

where we have used the fact that the cardinality $|T^k_{[p]}(\mathbf{y}^*)|$ and the probability $P(T^k_{[p]}(\mathbf{y}^*)|\mathbf{y}^*)$ do not depend on $\mathbf{y}^*$. If $d = k\gamma$, $p + \Delta p = \beta < 1/2$, and $\Delta p'$ is assigned in such a way that

$$D(p - \Delta p' \parallel p) = D(\beta \parallel p), \tag{2.6}$$

then, as it is well-known [7],

$$c_d \sim 2^{kH(\gamma)},\ \ C_{[p]} \sim 2^{kH(\beta)},\ \ 1 - P_{[p]} \sim 2 \cdot 2^{-kD(\beta \| p)}. \tag{2.7}$$

Let us introduce the known constructions of error-correcting codes, which will be used in the analysis. Let $\mathcal{B} = \{\ \mathbf{y}_1^*, ..., \mathbf{y}_\mu^*\ \}$, be a binary code consisting of μ codewords of length k. For all $\Delta\mathbf{y}^* \in \{0,1\}^k$, we denote $\Delta\mathbf{y}^* + \mathcal{B} = \{\ \Delta\mathbf{y}^* + \mathbf{y}_1^*, ..., \Delta\mathbf{y}^* + \mathbf{y}_\mu^*\ \}$ and note that, for all $\mathbf{y}^* \in \{0,1\}^k$, the number of occurrences of $\mathbf{y}^*$ in the collection $\{\ \Delta\mathbf{y}^* + \mathcal{B}\ \}$ is equal to

$$\sum_{\Delta\mathbf{y}^* \in \{0,1\}^k} \chi\{\ \mathbf{y}^* \in \{\ \Delta\mathbf{y}^* + \mathcal{B}\ \}\ \} = \mu. \tag{2.8}$$

Given $\mathbf{y}^* \in \mathcal{B}$, let $\mathcal{B}(\mathbf{y}^*) = \{\ \mathbf{x}_1^*, ..., \mathbf{x}_\nu^*\ \}$ be a binary code consisting of ν codewords of length k. Let $\pi = (\pi_1, ..., \pi_k)$ denote a permutation of the components of the vector $(1, ..., k)$ and let $\pi\mathbf{x}^* = (x^{(\pi_1)*}, ..., x^{(\pi_k)*})$ for all $\mathbf{x}^* = (x^{(1)*}, ..., x^{(k)*})$. Furthermore, let $\pi\mathcal{B}(\mathbf{y}^*) = \{\ \pi\mathbf{x}_1^*, ..., \pi\mathbf{x}_\nu^*\ \}$. Then we note that, for all $\mathbf{x}^* \in \{0,1\}^k$ with Hamming weight $w_H(\mathbf{x}^*) = d$, the number of occurrences of $\mathbf{x}^*$ in the collection $\{\ \pi\mathcal{B}(\mathbf{y}^*)\ \}$ is equal to

$$\sum_{\pi \in \Pi} \chi\{\ \mathbf{x}^* \in \pi\mathcal{B}(\mathbf{y}^*)\ \} = \frac{k! \cdot \nu}{c_d}, \tag{2.9}$$

where Π denotes the set consisting of $k!$ possible permutations of the components of $(1, ..., k)$.

Lemma 1 Let $a = k\alpha,\ d = k\gamma$, and $\tau = k\beta$.

1) There exists a sequence of codes $\{\mathcal{B}_k\}$ such that

$$\mu_k \sim 2^{k(1-H(\gamma))}, \quad |S'_d(\mathcal{B}_k)| \le \varepsilon_k \cdot 2^k, \tag{2.10}$$

where $\varepsilon_k \sim 0$ and

$$\begin{aligned} S'_d(\mathcal{B}_k) &= \{0,1\}^k \setminus \bigcup_{\mathbf{y}^* \in \mathcal{B}_k} S_d(\mathbf{y}^*), \\ S_d(\mathbf{y}^*) &= S_d^k(\mathbf{y}^*) \setminus \bigcup_{\mathbf{y}^{*\prime} \in \mathcal{B}_k \setminus \{\mathbf{y}^*\}} S_d^k(\mathbf{y}^{*\prime}). \end{aligned} \tag{2.11}$$

2) There exists a sequence of codes $\{\tilde{\mathcal{B}}_k\} = \left\{ \bigcup_{\mathbf{y}^* \in \mathcal{B}_k} \bigcup_{\mathbf{x}^* \in \mathcal{B}_k(\mathbf{y}^*)} \{\mathbf{x}^*\} \right\}$ such that $\mathbf{x}^* \in \mathcal{B}_k(\mathbf{y}^*)$ only if $\mathbf{x}^* \in S_d^k(\mathbf{y}^*)$,

$$\mu_k \sim 2^{k(1-H(\alpha*\gamma))}, \quad \nu_k \sim 2^{k(H(\alpha*\gamma)-H(\alpha))}, \quad |\tilde{S}'_a(\tilde{\mathcal{B}}_k)| \le \varepsilon_k \cdot 2^k, \tag{2.12}$$

where $\varepsilon_k \sim 0$ and

$$\begin{aligned} \tilde{S}'_a(\tilde{\mathcal{B}}_k) &= \{0,1\}^k \setminus \bigcup_{\mathbf{x}^* \in \tilde{\mathcal{B}}_k} \tilde{S}_a(\mathbf{x}^*), \\ \tilde{S}_a(\mathbf{x}^*) &= \tilde{S}_a^k(\mathbf{x}^*) \setminus \bigcup_{\mathbf{x}^{*\prime} \in \tilde{\mathcal{B}}_k \setminus \{\mathbf{x}^*\}} \tilde{S}_a^k(\mathbf{x}^{*\prime}). \end{aligned} \tag{2.13}$$

3) There exists a sequence of codes $\{G_k^*\}$ such that $|G_k^*| = m_k^*$,

$$m_k^* \sim 2^{k(1-H(\alpha))}, \quad |\tilde{S}_a(G^*)| \ge (1-\varepsilon_k) \cdot 2^k, \tag{2.14}$$

and

$$(1-\delta'_k) \cdot 2^k \le \sum_{\mathbf{y}^*} \chi\{\ \Psi(\mathbf{y}^*) \le (1+\delta_k) \cdot m^* \cdot 2^{-k(1-H(\alpha*\beta))}\ \}, \tag{2.15}$$

where $\varepsilon_k, \delta_k, \delta'_k \sim 0$ and

$$\tilde{S}_a(G^*) = \bigcup_{\mathbf{c}^* \in G^*} \tilde{S}_a(\mathbf{c}^*), \quad \Psi(\mathbf{y}^*) = \bigcup_{\mathbf{x}^* \in T_{[p]}^k(\mathbf{y}^*)} \{\psi(\mathbf{x}^*)\}. \tag{2.16}$$

3 Identification schemes

Fixed block code $G = \{\mathbf{x}\}$ of length n and rate r, whose decoding sets are given as $\{\mathcal{D}_G(\mathbf{x})\}$, we assign an encoding function under the restriction

$$f_i(\mathbf{x}^*, \mathbf{z}) \in G, \quad \text{for all} \quad (\mathbf{x}^*, \mathbf{z}) \in \{0,1\}^k \times \{0,1\}^l, \ i = 1, ..., N. \tag{3.1}$$

The set

$$\mathcal{D}_i(\mathbf{y}^*) = \bigcup_{\mathbf{x} \in \mathcal{F}_i(\mathbf{y}^*)} \{\mathcal{D}_G(\mathbf{x})\}, \tag{3.2}$$

where

$$\mathcal{F}_i(\mathbf{y}^*) = \bigcup_{\mathbf{x}^* \in T^k_{[p]}(\mathbf{y}^*)} \bigcup_{\mathbf{z}} \{f_i(\mathbf{x}^*, \mathbf{z})\}, \tag{3.3}$$

will be considered as the decoding set of the i-th receiver. Then

$$\lambda(i) = 2^{-l} \sum_{\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}} P(\mathbf{x}^*, \mathbf{y}^*) \sum_{\mathbf{y}} W(\mathbf{y} | f_i(\mathbf{x}^*, \mathbf{z})) \cdot \chi\{ \mathbf{y} \notin \mathcal{D}_i(\mathbf{y}^*) \} \tag{3.4}$$

and

$$\lambda_j(i) = 2^{-l} \sum_{\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}} P(\mathbf{x}^*, \mathbf{y}^*) \sum_{\mathbf{y}} W(\mathbf{y} | f_j(\mathbf{x}^*, \mathbf{z})) \cdot \chi\{ \mathbf{y} \in \mathcal{D}_i(\mathbf{y}^*) \} \tag{3.5}$$

are the decoding error probability of the first kind and the decoding error probability of the second kind provided that the message $j \neq i$ was generated. Hereafter χ denotes the indicator function ($\chi\{S\} = 1$ if the statement S is true, and $\chi\{S\} = 0$ otherwise). If $\rho H(p) \geq C$, then $l = 0$ and we omit dependence on $\mathbf{z}$ in (3.3)-(3.5).

4 Ensembles of ID codes

If $\rho H(p) < C$, we introduce an ensemble of ID codes in such a way that the codewords, assigned to the message i, are selected from a fixed code G in accordance with the uniform distribution. Such a selection is realized independently for all $(\mathbf{x}^*, \mathbf{z}) \in \{0,1\}^k \times \{0,1\}^l$ and $i = 1, ..., N$. We denote the probability in this ensemble by $Pr\{ \}$ and write :

$$Pr\{ f_i(\mathbf{x}^*, \mathbf{z}) = \mathbf{c} \} = \begin{cases} 1/m, & \text{if } \mathbf{c} \in G, \\ 0, & \text{if } \mathbf{c} \notin G. \end{cases}$$

If $\rho H(p) \geq C$, we will use a different code ensemble and denote the probability in this ensemble by $\tilde{Pr}\{ \}$. Let $G^* = \{\mathbf{c}_1^*, ..., \mathbf{c}_{m^*}^*\}$ be a binary code of length k. For all $\mathbf{c}^* \in G^*$, let

$$\tilde{Pr}\{ f_i(\mathbf{c}^*) = \mathbf{c} \} = \begin{cases} 1/m, & \text{if } \mathbf{c} \in G, \\ 0, & \text{if } \mathbf{c} \notin G. \end{cases}$$

To define the encoding for all $\mathbf{x}^* \notin G^*$, we introduce the sets

$$\tilde{S}_a^k(\mathbf{c}^*) = \{ \mathbf{x}^* : d_H(\mathbf{c}^*, \mathbf{x}^*) \leq a \},$$
$$\tilde{S}_a(\mathbf{c}^*) = \tilde{S}_a^k(\mathbf{c}^*) \setminus \bigcup_{\mathbf{c}^{*\prime} \in G^* \setminus \{\mathbf{c}^*\}} \tilde{S}_a^k(\mathbf{c}^{*\prime}),$$

where the integer $a > 0$ will be specified later. Let us introduce the function

$$\psi(\mathbf{x}^*) = \begin{cases} \mathbf{c}^*, & \text{if } \mathbf{x}^* \in \tilde{S}_a(\mathbf{c}^*), \\ \mathbf{c}_0^*, & \text{if } \mathbf{x}^* \notin \tilde{S}_a(\mathbf{c}^*), \end{cases}$$

where $\mathbf{c}_0^* \in G^*$ is some codeword assigned in advance, and define $f_i(\mathbf{x}^*)$ as $f_i(\psi(\mathbf{x}^*))$ for all $\mathbf{x}^* \notin G^*$.

5 Direct coding theorems

5.1 The case $\rho H(p) < C$

Theorem 1 Given $\beta \in (p, 1/2]$ and $\omega > 0$, the triple

$$(R, e_1, e_2) = (\ \rho + \omega - e_2,\ \min\{\ e(r),\ \rho D(\beta \parallel p)\ \},\ \min\{\ e(r),\ r - \rho H(\beta) - \omega\ \}\ ) \tag{5.1}$$

is a ρ-achievable ID triple ($D(\beta \parallel p)$ denotes the I-divergence between the distributions $(\beta, 1-\beta)$ and $(p, 1-p)$, i.e., $D(\beta \parallel p) = \beta \log \frac{\beta}{p} + (1-\beta)\log\frac{1-\beta}{1-p}$).

Corollary

$$R_\rho(C) \geq C + \rho(1 - H(p)), \text{ for all } \rho < \rho_0, \tag{5.2}$$

where $\rho_0 = C/H(p)$.

Proof We may assign $(r, \beta, \omega) = (\ C - \varepsilon_n,\ p + \delta_n,\ C - \rho H(p) - \delta_n'\ )$, where $\varepsilon_n, \delta_n, \delta_n' \to 0$, as $n \to \infty$. Then the exponents of the decoding error probabilities, given in (5.1), are still positive, and the rate R asymptotically coincides with the expression at the right hand side of (5.2). Q.E.D.

The inequality (5.2) provides a straight line given in Fig.3.

Proof of Theorem 1

Let us fix $\Delta p = \beta - p > 0$ and assign $\Delta p' > 0$ satisfying (2.6). Then using (3.1)-(3.5) and (2.5)-(2.7) we write :

$$\begin{aligned}
\lambda(i) &\leq \sum_{\mathbf{y}^*} \sum_{\mathbf{x}^* \notin T^k_{[p]}(\mathbf{y}^*)} P(\mathbf{x}^*, \mathbf{y}^*) + 2^{-l} \sum_{\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}} P(\mathbf{x}^*, \mathbf{y}^*) \sum_{\mathbf{y} \notin \mathcal{D}_G(f_i(\mathbf{x}^*, \mathbf{z}))} W(\mathbf{y} | f_i(\mathbf{x}^*, \mathbf{z})) \\
&\lesssim 2 \cdot 2^{-\rho n D(\beta \| p)} + 2^{-ne(r)}, \\
\lambda_j(i) &= 2^{-l} \sum_{\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}} P(\mathbf{x}^*, \mathbf{y}^*) \sum_{\mathbf{y} \notin \mathcal{D}_G(f_j(\mathbf{x}^*, \mathbf{z}))} W(\mathbf{y} | f_j(\mathbf{x}^*, \mathbf{z})) \\
&+ 2^{-l} \sum_{\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}} P(\mathbf{x}^*, \mathbf{y}^*) \sum_{\mathbf{y} \in \mathcal{D}_G(f_j(\mathbf{x}^*, \mathbf{z}))} W(\mathbf{y} | f_j(\mathbf{x}^*, \mathbf{z})) \cdot \chi\{\ \mathbf{y} \in \mathcal{D}_i(\mathbf{y}^*)\ \} \\
&\leq 2^{-ne(r)} + 2^{-l} \sum_{\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}} P(\mathbf{y}^*) P(\mathbf{x}^* | \mathbf{y}^*) \sum_{\mathbf{y} \in \mathcal{D}_G(f_j(\mathbf{x}^*, \mathbf{z}))} W(\mathbf{y} | f_j(\mathbf{x}^*, \mathbf{z})) \cdot \chi\{\ \mathbf{y} \in \mathcal{D}_i(\mathbf{y}^*)\ \} \\
&= 2^{-ne(r)} + 2^{-(k+l)} \sum_{d = k(p - \Delta p')}^{k\beta} p_d \sum_{\mathbf{y}^*, \mathbf{z}} \eta_{ij}^{(d)}(\mathbf{y}^*, \mathbf{z}),
\end{aligned} \tag{5.3}$$

where

$$\eta_{ij}^{(d)}(\mathbf{y}^*, \mathbf{z}) = \sum_{\mathbf{x}^* \in S^k_d(\mathbf{y}^*)} \chi\{\ f_j(\mathbf{x}^*, \mathbf{z}) \in \mathcal{F}_i(\mathbf{y}^*)\ \} \tag{5.4}$$

and the set $S^k_d(\mathbf{y}^*)$ is defined in (2.4). Therefore, e_1 is an asymptotically achievable exponent of the decoding error probability of the first kind. To estimate the decoding error probability of the second kind we use Chernov's inequality and write

$$\begin{aligned}
\log Pr \left\{ \sum_{\mathbf{y}^*, \mathbf{z}} \eta_{ij}^{(d)}(\mathbf{y}^*, \mathbf{z}) > 2^{k+l} c_d \lambda \right\} &\leq -s \cdot 2^{k+l} c_d \lambda + \log G^{(d)}(s), \\
G^{(d)}(s) = E\left[\prod_{\mathbf{y}^*, \mathbf{z}} 2^{s \cdot \eta_{ij}^{(d)}(\mathbf{y}^*, \mathbf{z})} \right],
\end{aligned} \tag{5.5}$$

where $s > 0$ and $E[\,]$ denotes the expectation in the code ensemble.

The following result is proved in the Appendix.

Lemma 2 If there exists a code $\mathcal{B}$ of length k consisting of μ codewords and satisfying (2.10) with $\varepsilon_k = \varepsilon$, then

$$\log G^{(d)}(s) \leq 2^{k+l} \cdot (\ (1-\varepsilon) \cdot \log g(s2^k/\mu) + s\varepsilon 2^k/\mu\ ), \tag{5.6}$$

where

$$g(s) = 1 - \lambda_0 + \lambda_0 \cdot 2^s, \quad \lambda_0 = C_{[p]} \cdot 2^l / m. \tag{5.7}$$

We apply Lemma 2 to every component of a sequence $\{\mathcal{B}_k\}$ with the properties guaranteed by Lemma 1. Then using (2.7), (2.10), and (5.5)-(5.7) we conclude that

$$\log G^{(d)}(s) \lesssim 2^{k+l} \cdot \log g\left(s \cdot 2^{kH(\gamma)}\right), \tag{5.8}$$

$$\log Pr\left\{\sum_{\mathbf{y}^*,\mathbf{z}} \eta_{ij}^{(d)}(\mathbf{y}^*,\mathbf{z}) > 2^{k+l} c_d \lambda\right\} \lesssim 2^{k+l}\left(-s \cdot 2^{kH(\gamma)}\lambda + \log g\left(s \cdot 2^{kH(\gamma)}\right)\right),$$

$$\lambda_0 \sim 2^{-n(r-\rho H(\beta)-\omega)},$$

where $\omega = l/n$.
Let $\lambda > \lambda_0$. Then we can set $s = 2^{-kH(\gamma)} \cdot \log \frac{\lambda}{1-\lambda}\frac{1-\lambda_0}{\lambda_0} > 0$ at the right hand side of (5.8) and obtain the estimate

$$\log Pr\left\{\sum_{\mathbf{y}^*,\mathbf{z}} \eta_{ij}^{(d)}(\mathbf{y}^*,\mathbf{z}) > 2^{k+l} c_d \lambda\right\} \lesssim -2^{k+l} \cdot D(\lambda \,\|\, \lambda_0),$$

which does not depend on d. Therefore,

$$Pr\left\{\sum_{d=k(p-\Delta p')}^{k\beta} p_d \sum_{\mathbf{y}^*,\mathbf{z}} \eta_{ij}^{(d)}(\mathbf{y}^*,\mathbf{z}) > P_{[p]} \cdot \lambda\right\} = Pr\left\{\sum_{d=k(p-\Delta p')}^{k\beta} p_d \left(\sum_{\mathbf{y}^*,\mathbf{z}} \eta_{ij}^{(d)}(\mathbf{y}^*,\mathbf{z}) - c_d\lambda\right) > 0\right\} \tag{5.9}$$

$$\leq \sum_{d=\sigma}^{\tau} Pr\left\{\sum_{\mathbf{y}^*,\mathbf{z}} \eta_{ij}^{(d)}(\mathbf{y}^*,\mathbf{z}) > c_d\lambda\right\} \lesssim \tau \exp_2\left\{-2^{k+l} \cdot D(\lambda \,\|\, \lambda_0)\right\}.$$

If $\lambda, \lambda_0 \to 0$, as $n \to \infty$, then $D(\lambda \,\|\, \lambda_0) \sim \lambda$ and (5.9) leads to the inequality :

$$Pr\left\{\sum_{d=k(p-\Delta p')}^{k\beta} p_d \sum_{\mathbf{y}^*,\mathbf{z}} \eta_{ij}^{(d)}(\mathbf{y}^*,\mathbf{z}) > P_{[p]} \cdot \lambda\right\} \lesssim \exp_2\left\{-2^{k+l} \cdot \lambda\right\}. \tag{5.10}$$

Let us assign $\lambda = P_{[p]}^{-1} \cdot (2^{-ne_2} + 2^{-ne(r)})$, where e_2 is given in (5.1). Then we can refer to the considerations of Section 2 (see (2.1),(2.2)) and using (5.8), (5.10) conclude that the rate R and the exponent e_2, given in (5.1), are asymptotically achievable. Q.E.D.

5.2 The case $\rho H(p) \geq C$

Theorem 2 For all $\beta \in (p, 1/2]$ and $\alpha > 0$, the triple

$$(R, e_1, e_2) = (\ \rho(1-H(\alpha)) - e_2,\ \min\{\ e(r),\ \rho D(\beta \,\|\, p)\ \},\ \min\{\ e(r),\ r - \rho(H(\alpha * \beta) - H(\alpha))\ \}\ )$$

is a ρ-achievable ID triple ($\alpha * p$ denotes the convolution of the distributions $(\alpha, 1-\alpha)$ and $(p, 1-p)$, i.e., $\alpha * p = \alpha(1-p) + (1-\alpha)p$).
Corollary

$$R_\rho(C) \geq \rho(1-H(\alpha)) = C + \rho(1 - H(\alpha * p)), \text{ for all } \rho \geq \rho_0, \tag{5.11}$$

where $\rho_0 = C/H(p)$ and the parameter α is chosen in such a way that

$$C = \rho(H(\alpha * p) - H(\alpha)). \tag{5.12}$$

Proof We assign r and β in the same way as in the proof of the corollary from Theorem 1 and complete the proof. Q.E.D.
The inequality (5.11) provides a convex up curve given in Fig.3.
Proof of Theorem 2 The same considerations as in the proof of Theorem 1 lead to the inequalities :

$$\lambda(i) \leq 2 \cdot 2^{-\rho n D(\beta\|p)} + 2^{-ne(r)},$$

$$\lambda_j(i) \leq 2^{-ne(r)} + 2^{-k} \sum_{d=k(p-\Delta p')}^{k\beta} p_d \sum_{\mathbf{y}^*} \eta_{ij}^{(d)}(\mathbf{y}^*),$$

where the variables $\eta_{ij}^{(d)}(\mathbf{y}^*)$ are defined similarly to (5.4), i.e.,

$$\eta_{ij}^{(d)}(\mathbf{y}^*) = \sum_{\mathbf{x}^* \in S_d^k(\mathbf{y}^*)} \chi\{\ f_j(\mathbf{x}^*) \in \mathcal{F}_i(\mathbf{y}^*)\ \}. \tag{5.13}$$

Furthermore, for a given constant $\tilde{\lambda} > 0$,

$$\log \tilde{Pr} \left\{ \sum_{\mathbf{y}^*} \eta_{ij}^{(d)}(\mathbf{y}^*) > 2^k c_d \tilde{\lambda} \right\} \leq -s \cdot 2^k c_d \tilde{\lambda} + \log \tilde{G}^{(d)}(s), \tag{5.14}$$
$$\tilde{G}^{(d)}(s) = \tilde{E} \left[\prod_{\mathbf{y}^*} 2^{s \cdot \eta_{ij}^{(d)}(\mathbf{y}^*)} \right],$$

where $s > 0$ and $\tilde{E}[\,]$ denotes the expectation in the code ensemble.

All the peculiarities of the analysis are concentrated in a new upper bound on $\tilde{G}^{(d)}(s)$. The following result is proved in the Appendix.

Lemma 3 If there exists a code $\tilde{\mathcal{B}}$ of length k consisting of $\mu\nu$ codewords and satisfying (2.12) with $\varepsilon_k = \varepsilon$ and a code G^* of length k consisting of m^* codewords and satisfying (2.14),(2.15) with $(\varepsilon_k, \delta_k, \delta'_k) = (\varepsilon, \delta, \delta')$, then

$$\log \tilde{G}^{(d)}(s) \leq \mu\nu \cdot \left(\, (1 - \varepsilon - \delta') \cdot \log \tilde{g}(s 2^k c_d / (\mu\nu)) + s(\varepsilon + \delta') 2^k c_d / (\mu\nu) \, \right), \tag{5.15}$$

where

$$\tilde{g}(s) = 1 - \tilde{\lambda}_0 + \tilde{\lambda}_0 \cdot 2^s, \;\; \tilde{\lambda}_0 = (1 + \delta) \cdot m^* \cdot 2^{-k(1 - H(\alpha * \beta))}/m. \tag{5.16}$$

We apply Lemma 2 to every component of a sequence $\{\tilde{\mathcal{B}}_k\}$ satisfying Lemma 1. Then using (2.5),(2.12),(2.14), and (5.14)-(5.16) we conclude that

$$\log \tilde{G}^{(d)}(s) \lesssim 2^{k(1 - H(\alpha))} \cdot \log g \left(s 2^{k(H(\gamma) + H(\alpha))} \right),$$
$$\log \tilde{Pr} \left\{ \sum_{\mathbf{y}^*, \mathbf{z}} \eta_{ij}^{(d)}(\mathbf{y}^*) > 2^k c_d \tilde{\lambda} \right\} \lesssim -s 2^{k(1 + H(\gamma))} \tilde{\lambda} + 2^{k(1 - H(\alpha))} \cdot \log \tilde{g} \left(s 2^{k(H(\gamma) + H(\alpha))} \right),$$
$$\tilde{\lambda}_0 \lesssim 2^{-n(r - \rho(H(\alpha * \beta) - H(\alpha)))}.$$

Let $\tilde{\lambda} > \tilde{\lambda}_0$. Then we can set $s = 2^{-k(H(\alpha) + H(\gamma))} \cdot \log \frac{\tilde{\lambda}}{1 - \tilde{\lambda}} \frac{1 - \tilde{\lambda}_0}{\tilde{\lambda}_0} > 0$ and

$$\log Pr \left\{ \sum_{\mathbf{y}^*, \mathbf{z}} \eta_{ij}^{(d)}(\mathbf{y}^*, \mathbf{z}) > 2^k c_d \tilde{\lambda} \right\} \lesssim -2^{k(1 - H(\alpha))} \cdot D(\tilde{\lambda} \parallel \tilde{\lambda}_0).$$

We set $\tilde{\lambda}, \tilde{\lambda}_0 \to 0$, as $n \to \infty$, note that $D(\tilde{\lambda} \parallel \tilde{\lambda}_0) \sim \tilde{\lambda}$, and repeat the same considerations as in the proof of Theorem 1. Q.E.D.

6 Calculation of the function $\Delta R_\infty(C)$

Using (5.11) we conclude that the increase in the identification rate, which we gain because of the observations of the process, is measured as $\rho(1 - H(\alpha * p))$, where the parameter α is defined by the equation (5.12). Let us consider the function

$$\Delta R_\infty(C) = \lim_{\rho \to \infty} \rho \cdot (1 - H(\alpha * p)),$$

which measures this quantity conditioned on the infinite length of the vectors $\mathbf{x}^*, \mathbf{y}^*$. This function can be calculated exactly.

Theorem 3

$$\Delta R_\infty(C) = \frac{(1 - 2p)^2}{1 - (1 - 2p)^2} \cdot C. \tag{6.1}$$

Proof The parameter α satisfying (5.12) is a function of ρ, and we may write :

$$\alpha = \frac{1}{2} - \varepsilon_\rho, \;\; \alpha * p = \frac{1}{2} - (1 - 2p) \cdot \varepsilon_\rho,$$

where

$$\varepsilon_\rho \to 0, \;\; \text{as} \;\; \rho \to \infty. \tag{6.2}$$

Since, for all $\varepsilon \in (0, 1/2)$,

$$\log \left(\frac{1}{2} - \varepsilon \right) = -1 - \frac{1}{\ln 2} \cdot \sum_{i \geq 1} \frac{(2\varepsilon)^i}{i}, \;\; \log \left(\frac{1}{2} + \varepsilon \right) = -1 - \frac{1}{\ln 2} \cdot \sum_{i \geq 1} (-1)^i \cdot \frac{(2\varepsilon)^i}{i},$$

we have

$$H(\alpha) = 1 - \frac{1}{\ln 2} \cdot \sum_{k \geq 1} (2\varepsilon_\rho)^{2k} \cdot \left(\frac{1}{2k-1} - \frac{1}{2k} \right),$$

$$H(\alpha * p) = 1 - \frac{1}{\ln 2} \cdot \sum_{k \geq 1} (1-2p)^{2k} \cdot (2\varepsilon_\rho)^{2k} \cdot \left(\frac{1}{2k-1} - \frac{1}{2k} \right)$$

and

$$\Delta R_\infty(C) = \lim_{\rho \to \infty} \frac{\rho}{\ln 2} \cdot \sum_{k \geq 1} (1-2p)^{2k} \cdot (2\varepsilon_\rho)^{2k} \cdot \left(\frac{1}{2k-1} - \frac{1}{2k} \right),$$

$$C = \frac{\rho}{\ln 2} \cdot \sum_{k \geq 1} (1-(1-2p)^{2k}) \cdot (2\varepsilon_\rho)^{2k} \cdot \left(\frac{1}{2k-1} - \frac{1}{2k} \right).$$

Therefore,

$$\Delta R_\infty(C) = \lim_{\rho \to \infty} \frac{2(1-2p)^2 + \delta_\rho^{(R)}}{2(1-(1-2p)^2) + \delta_\rho^{(C)}} \cdot C, \tag{6.3}$$

where

$$\delta_\rho^{(R)} = \sum_{k \geq 2} (1-2p)^{2k} \cdot (2\varepsilon_\rho)^{2k-2} \cdot \left(\frac{1}{2k-1} - \frac{1}{2k} \right),$$

$$\delta_\rho^{(C)} = \sum_{k \geq 2} (1-(1-2p)^{2k}) \cdot (2\varepsilon_\rho)^{2k-2} \cdot \left(\frac{1}{2k-1} - \frac{1}{2k} \right).$$

Using (6.2) we conclude that

$$\delta_\rho^{(R)}, \delta_\rho^{(C)} \to 0, \quad \text{as} \quad \rho \to \infty, \tag{6.4}$$

and (6.1) follows from (6.3),(6.4). Q.E.D.

7 Discussion

We look now at our coding theorems in a somewhat more general context.
Let now $(X^*, Y^*) = ((X^{(1)}, \ldots, X^{(k)}), (Y^{(1)}, \ldots, Y^{(k)}))$ stand for a general i.i.d. correlated source and W for any channel of capacity C.
Observing Y^* a receiver has "list-knowledge" of "list-rate" $H(X|Y)$ about X^*. If now $H(X|Y) < C$ then, using bining and the channel to transmit the color (in the bining), list size can be reduced to 1. Thus we have established X^* as common random experiment between sender and receivers. Its rate is $H(X)$. It is obtained at a prize of channel rate, namely $H(X|Y)$. The remaining channel rate $C - H(X|Y)$ can be used to establish additional common randomness. Thus we have achieved in total common randomness of rate

$$H(X) + (C - H(X|Y)) = I(X \wedge Y) + C. \tag{7.1}$$

In the case $C < H(X|Y)$ X^* can be replaced by a function $f(X^*)$ with $H(f(X^*)|Y) < C$. Such functions can be produced by partitions in terms of auxiliary $RV's$ U, a standard technique in multi-user information theory. The resulting rate of common randomness is

$$R(C) = \max\{I(U \wedge X) : I(U \wedge X) - I(U \wedge Y) \leq C, U \multimap X \multimap Y, |U| \leq |X|\} \tag{7.2}$$

The equations (7.1) and (7.2) correspond to theorems 1 and 2. By the method of [2] the expressions give achievable second order identification rates. There is another remarkable connection. In case $C = 0$ formula (7.2) gives for correlated sources the common information in the sense of [8]. This was shown in [9]. Common randomness extends in this situation the notion of common information to the case where a channel of capacity C helps to extract a common part.
By standard techniques it can actually be shown that (7.2) gives the optimal value (see [7]). Detailed definitions and proofs can be found in a subsequent paper [11], where also several other new models are introduced and analyzed. It has been argued in [9] that the concept of common information of [8] is likely the most meaningful among such notions.
Now it is naturally incorporated in the theory of common randomness and thus has for instance operational significance in the theory of identification and cryptography.
Notice also that it ranges between 0 and the mutual information, which it always achieves, if C is large enough. We conclude with a general problem: Under which conditions can the equality of the optimal rate of common randomness and the identification capacity be proved ? Cases with complete feedback are settled in [10].

8 Acknowledgment

The authors are grateful to Imre Csiszár for interesting discussions.

Appendix

1 Proof of Lemma 2

With (2.8) write

$$\sum_{\mathbf{y}^*} \eta_{ij}^{(d)}(\mathbf{y}^*) = \sum_{\Delta\mathbf{y}^*} \sum_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \eta_{ij}^{(d)}(\mathbf{y}^*)/\mu$$

and with Hölder's inequality

$$G^{(d)}(s) = E\left[\prod_{\mathbf{y}^*,\mathbf{z}} 2^{s\cdot\eta_{ij}^{(d)}(\mathbf{y}^*,\mathbf{z})}\right] = \prod_{\mathbf{z}} E\left[\prod_{\mathbf{y}^*} 2^{s\cdot\eta_{ij}^{(d)}(\mathbf{y}^*,\mathbf{z})}\right] = \prod_{\mathbf{z}} E\left[\prod_{\Delta\mathbf{y}^*} \prod_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} 2^{s\cdot\eta_{ij}^{(d)}(\mathbf{y}^*,\mathbf{z})/\mu}\right] \tag{A.1}$$

$$\leq \prod_{\mathbf{z}} \prod_{\Delta\mathbf{y}^*} E^{2^{-k}}\left[\prod_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} 2^{s\cdot\eta_{ij}^{(d)}(\mathbf{y}^*,\mathbf{z})\cdot 2^k/\mu}\right].$$

Let $\eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*,\mathbf{z}) = \chi\{\ f_j(\mathbf{x}^*,\mathbf{z}) \in \mathcal{F}_i(\mathbf{y}^*)\ \}$. Then

$$\sum_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \eta_{ij}^{(d)}(\mathbf{y}^*,\mathbf{z}) = \sum_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \sum_{\mathbf{x}^* \notin S_d(\mathbf{y}^*)} \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*,\mathbf{z}) + \sum_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \sum_{\mathbf{x}^* \in S_d(\mathbf{y}^*)} \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*,\mathbf{z}) \tag{A.2}$$

$$\leq |S'_d(\mathcal{B})| + \sum_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \sum_{\mathbf{x}^* \in S_d(\mathbf{y}^*)} \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*,\mathbf{z})$$

$$\leq \varepsilon 2^k + \sum_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \sum_{\mathbf{x}^* \in S_d(\mathbf{y}^*)} \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*,\mathbf{z})$$

by (2.10) and (2.11). Therefore, (A.1), (A.2) lead to the inequality :

$$\log G^{(d)}(s) \leq 2^{k+l}\cdot s\varepsilon 2^k/\mu + \log \prod_{\mathbf{z}} \prod_{\Delta\mathbf{y}^*} E^{2^{-k}}\left[\prod_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \prod_{\mathbf{x}^* \in S_d(\mathbf{y}^*)} 2^{s\cdot\eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*,\mathbf{z})\cdot 2^k/\mu}\right]. \tag{A.3}$$

Let us represent the expectation $E[\ ]$ as a concatenation of the expectation $E_j[\ ]$ taken on all assignments of the codewords for the message j and the expectation $E_i[\ ]$ taken on all assignments of the codewords for the message i, i.e., $E\,[\,(*)\,] = E_i\,[E_j\,[\,(*)\,]]$. The codewords, which are used to encode the j-th message, are independent random vectors for different pairs $(\mathbf{x}^*,\mathbf{y}^*)$, where $\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}$ and $\mathbf{x}^* \in S_d(\mathbf{y}^*)$. Therefore,

$$E\left[\prod_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \prod_{\mathbf{x}^* \in S_d(\mathbf{y}^*)} 2^{s\cdot\eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*,\mathbf{z})\cdot 2^k/\mu}\right] = E_i\left[E_j\left[\prod_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \prod_{\mathbf{x}^* \in S_d(\mathbf{y}^*)} 2^{s\cdot\eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*,\mathbf{z})\cdot 2^k/\mu}\right]\right]$$

$$= E_i\left[\prod_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \prod_{\mathbf{x}^* \in S_d(\mathbf{y}^*)} E_j\left[2^{s\cdot\eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*,\mathbf{z})\cdot 2^k/\mu}\right]\right]$$

$$= E_i\left[\prod_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \prod_{\mathbf{x}^* \in S_d(\mathbf{y}^*)} \left(1 - \frac{|\mathcal{F}_i(\mathbf{y}^*)|}{m} + \frac{|\mathcal{F}_i(\mathbf{y}^*)|}{m}\cdot 2^{s2^k/\mu}\right)\right]$$

$$\leq \prod_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \prod_{\mathbf{x}^* \in S_d(\mathbf{y}^*)} \left(1 - \frac{C_{[p]}2^l}{m} + \frac{C_{[p]}2^l}{m}\cdot 2^{s2^k/\mu}\right),$$

where the inequality $|\mathcal{F}_i(\mathbf{y}^*)| \leq C_{[p]}2^l$ was used. Hence,

$$\log E\left[\prod_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} \prod_{\mathbf{x}^* \in S_d(\mathbf{y}^*)} 2^{s\cdot\eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*,\mathbf{z})\cdot 2^k/\mu}\right] \leq \sum_{\mathbf{y}^* \in \Delta\mathbf{y}^* + \mathcal{B}} |S_d(\mathbf{y}^*)|\cdot \log g(s2^k/\mu) \leq (1-\varepsilon)2^k\cdot \log g(s2^k/\mu), \tag{A.4}$$

where we used the fact that the code $\Delta\mathbf{y}^* + \mathcal{B}$ has the same properties as $\mathcal{B}$ for all $\Delta\mathbf{y}^*$. Combining (A.3) and (A.4) we complete the proof. Q.E.D.

2 Proof of Lemma 3

The first step of the proof is the same as in the proof of Lemma 2, i.e., we write the inequality :

$$\tilde{G}^{(d)}(s) \leq \prod_{\Delta \mathbf{y}^*} \tilde{E}^{2^{-k}} \left[\prod_{\mathbf{y}^* \in \Delta \mathbf{y}^* + \mathcal{B}} 2^{s \cdot \eta_{ij}^{(d)}(\mathbf{y}^*) \cdot 2^k / \mu} \right].$$

Let us denote $\eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*) = \chi\{\ f_j(\mathbf{x}^*) \in \mathcal{F}_i(\mathbf{y}^*)\ \}$. Then using (5.13) and (2.9) we have

$$\eta_{ij}^{(d)}(\mathbf{y}^*) = \sum_{\mathbf{x}^* \in S_d^k(\mathbf{y}^*)} \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*) = \sum_{\pi \in \Pi} \sum_{\mathbf{x}^* \in \pi \mathcal{B}(\mathbf{y}^*)} \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*) \cdot c_d/(k! \cdot \nu).$$

Hence,

$$\begin{aligned}\tilde{E}\left[\prod_{\mathbf{y}^* \in \Delta \mathbf{y}^* + \mathcal{B}} 2^{s \cdot \eta_{ij}^{(d)}(\mathbf{y}^*) \cdot 2^k / \mu} \right] &= \tilde{E}\left[\prod_{\mathbf{y}^* \in \Delta \mathbf{y}^* + \mathcal{B}} \prod_{\pi \in \Pi} \prod_{\mathbf{x}^* \in \pi \mathcal{B}(\mathbf{y}^*)} 2^{s \cdot \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*) \cdot 2^k c_d / (\mu \nu k!)} \right] \\ &\leq \prod_{\pi \in \Pi} \tilde{E}^{1/k!} \left[\prod_{\mathbf{y}^* \in \Delta \mathbf{y}^* + \mathcal{B}} \prod_{\mathbf{x}^* \in \pi \mathcal{B}(\mathbf{y}^*)} 2^{s \cdot \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*) \cdot 2^k c_d / (\mu \nu)} \right].\end{aligned}$$

Given $\Delta \mathbf{y}^* \in \{0,1\}^k$ and $\pi \in \Pi$, let

$$\begin{aligned}\mathcal{L}(\pi, \Delta \mathbf{y}^*) = \{\ (\mathbf{x}^*, \mathbf{y}^*)\ :\ &\mathbf{y}^* \in \Delta \mathbf{y}^* + \mathcal{B},\ \mathbf{x}^* \in \pi \mathcal{B}(\mathbf{y}^*),\ \psi(\mathbf{x}^*) \notin \bigcup_{\mathbf{x}^{*\prime} \in \tilde{\mathcal{B}}' \setminus \{\mathbf{x}^*\}} \{\ \psi(\mathbf{x}^{*\prime})\ \}, \\ &|\Psi(\mathbf{y}^*)| \leq (1+\delta) \cdot m^* \cdot 2^{-k(1-H(\alpha * \beta))}\ \}\end{aligned}$$

and

$$\mathcal{L}'(\pi, \Delta \mathbf{y}^*) = \{\ (\mathbf{x}^*, \mathbf{y}^*)\ :\ \mathbf{y}^* \in \Delta \mathbf{y}^* + \mathcal{B},\ \mathbf{x}^* \in \pi \mathcal{B}(\mathbf{y}^*)\ \ \} \setminus \mathcal{L}(\pi, \Delta \mathbf{y}^*),$$

where

$$\tilde{\mathcal{B}}' = \bigcup_{\mathbf{y}^* \in \Delta \mathbf{y}^* + \mathcal{B}} \bigcup_{\mathbf{x}^* \in \pi \mathcal{B}(\mathbf{y}^*)} \{\mathbf{x}^*\}.$$

Then

$$\begin{aligned}\sum_{\mathbf{y}^* \in \Delta \mathbf{y}^* + \mathcal{B}} \sum_{\mathbf{x}^* \in \pi \mathcal{B}(\mathbf{y}^*)} \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*) &= \sum_{(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{L}'(\pi, \Delta \mathbf{y}^*)} \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*) + \sum_{(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{L}(\pi, \Delta \mathbf{y}^*)} \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*) \\ &\leq |\mathcal{L}'(\pi, \Delta \mathbf{y}^*)| + \sum_{(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{L}(\pi, \Delta \mathbf{y}^*)} \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*).\end{aligned} \tag{A.5}$$

Since

$$|\mathcal{L}'(\pi, \Delta \mathbf{y}^*)| = \sum_{\mathbf{y}^* \in \Delta \mathbf{y}^* + \mathcal{B}} |\ \{\pi \mathcal{B}(\mathbf{y}^*)\}\ | - |\mathcal{L}(\pi, \Delta \mathbf{y}^*)| == \mu\nu - |\mathcal{L}(\pi, \Delta \mathbf{y}^*)|, \tag{A.6}$$

using (A.5) and (A.6) we obtain

$$\log \tilde{G}^{(d)}(s) \leq s \cdot \frac{2^k c_d}{\mu\nu} \cdot (\mu\nu - \overline{\mathcal{L}}) + \log \prod_{\Delta \mathbf{y}^*} \prod_{\pi \in \Pi} \tilde{E}^{2^{-k}/k!} \left[\prod_{(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{L}(\pi, \Delta \mathbf{y}^*)} 2^{s \cdot \eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*) \cdot 2^k c_d / (\mu\nu)} \right], \tag{A.7}$$

where

$$\overline{\mathcal{L}} = \frac{1}{2^k \cdot k!} \cdot \sum_{\pi \in \Pi} \sum_{\Delta \mathbf{y}^*} |\mathcal{L}(\pi, \Delta \mathbf{y}^*)|. \tag{A.8}$$

Similarly to the proof of Lemma 2, we represent the expectation $\tilde{E}[\ ]$ as a concatenation of the expectation $\tilde{E}_j[\ ]$ taken on all assignments of the codewords for the message j and the expectation $\tilde{E}_i[\ ]$ taken on all assignments of the codewords for the message i, i.e., $\tilde{E}\left[\,(*)\,\right] = \tilde{E}_i\left[\tilde{E}_j\left[\,(*)\,\right]\right]$. The codewords, which are used to encode the i-th message, are independent random vectors for different pairs $(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{L}(\pi, \Delta\mathbf{y}^*)$. Therefore,

$$\begin{aligned} \tilde{E}\left[\prod_{(\mathbf{x}^*,\mathbf{y}^*)\in\mathcal{L}(\pi,\Delta\mathbf{y}^*)} 2^{s\cdot\eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*)\cdot 2^k c_d/(\mu\nu)}\right] &= \tilde{E}_i\left[\tilde{E}_j\left[\prod_{(\mathbf{x}^*,\mathbf{y}^*)\in\mathcal{L}(\pi,\Delta\mathbf{y}^*)} 2^{s\cdot\eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*)\cdot 2^k c_d/(\mu\nu)}\right]\right] \\ &= \tilde{E}_i\left[\prod_{(\mathbf{x}^*,\mathbf{y}^*)\in\mathcal{L}(\pi,\Delta\mathbf{y}^*)} \tilde{E}_j\left[\,2^{s\cdot\eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*)\cdot 2^k c_d/(\mu\nu)}\,\right]\right] \\ &= \tilde{E}_i\left[\prod_{(\mathbf{x}^*,\mathbf{y}^*)\in\mathcal{L}(\pi,\Delta\mathbf{y}^*)} \left(1 - \frac{|\mathcal{F}_i(\mathbf{y}^*)|}{m} + \frac{|\mathcal{F}_i(\mathbf{y}^*)|}{m}\cdot 2^{s2^k c_d/(\mu\nu)}\right)\right]. \end{aligned} \tag{A.9}$$

Then

$$\log\tilde{E}\left[\prod_{(\mathbf{x}^*,\mathbf{y}^*)\in\mathcal{L}(\pi,\Delta\mathbf{y}^*)} 2^{s\cdot\eta_{ij}^{(d)}(\mathbf{x}^*|\mathbf{y}^*)\cdot 2^k c_d/(\mu\nu)}\right] \le |\mathcal{L}(\pi,\Delta\mathbf{y}^*)|\cdot\log\tilde{g}(s2^k c_d/(\mu\nu)) \tag{A.10}$$

and (A.7)-(A.10) lead to the inequality

$$\log\tilde{G}^{(d)}(s) \le s\cdot\frac{2^k c_d}{\mu\nu}\cdot(\mu\nu - \overline{\mathcal{L}}) + \overline{\mathcal{L}}\cdot\log\tilde{g}(s2^k c_d/(\mu\nu)). \tag{A.11}$$

Let us prove that

$$\overline{\mathcal{L}} \ge \mu\nu\cdot(1-\varepsilon-\delta'), \tag{A.12}$$

by representing the set $\mathcal{L}(\pi, \Delta\mathbf{y}^*)$ as

$$\mathcal{L}(\pi,\Delta\mathbf{y}^*) = \mathcal{L}_1(\pi,\Delta\mathbf{y}^*)\bigcap\mathcal{L}_2(\pi,\Delta\mathbf{y}^*),$$

where

$$\begin{aligned} \mathcal{L}_1(\pi,\Delta\mathbf{y}^*) &= \{\,(\mathbf{x}^*,\mathbf{y}^*)\ :\ \mathbf{y}^*\in\Delta\mathbf{y}^*+\mathcal{B},\ \mathbf{x}^*\in\pi\mathcal{B}(\mathbf{y}^*),\ \psi(\mathbf{x}^*)\notin\bigcup_{\mathbf{x}^{*\prime}\in\tilde{\mathcal{B}}'\setminus\{\mathbf{x}^*\}}\{\,\psi(\mathbf{x}^{*\prime})\,\}\,\}, \\ \mathcal{L}_2(\pi,\Delta\mathbf{y}^*) &= \{\,(\mathbf{x}^*,\mathbf{y}^*)\ :\ \mathbf{y}^*\in\Delta\mathbf{y}^*+\mathcal{B},\ \mathbf{x}^*\in\pi\mathcal{B}(\mathbf{y}^*),\ |\Psi(\mathbf{y}^*)|\le(1+\delta)\cdot m^*\cdot 2^{-k(1-H(\alpha*\beta))}\,\}. \end{aligned}$$

Using (2.15),(2.16) we note that

$$\begin{aligned} \sum_{\pi\in\Pi}\sum_{\Delta\mathbf{y}^*}|\mathcal{L}_1(\pi,\Delta\mathbf{y}^*)| &\ge \sum_{\mathbf{x}^*,\mathbf{y}^*}\sum_{\pi\in\Pi}\sum_{\Delta\mathbf{y}^*}\chi\{\,(\mathbf{x}^*,\mathbf{y}^*)\in\mathcal{L}_1(\pi,\Delta\mathbf{y}^*)\,\}\cdot\chi\{\,\mathbf{x}^*\in\tilde{S}_a(G^*)\,\} \\ &= \sum_{\mathbf{y}^*,\Delta\mathbf{y}^*}\chi\{\,\mathbf{y}^*\in\Delta\mathbf{y}^*+\mathcal{B}\,\}\sum_{\mathbf{x}^*\in S_d^k(\mathbf{y}^*)}\chi\{\,\mathbf{x}^*\in\tilde{S}_a(G^*)\,\}\sum_{\pi\in\Pi}\chi\{\,\mathbf{x}^*\in\pi\mathcal{B}\,\} \\ &= \mu\cdot\frac{k!\nu}{c_d}\cdot\sum_{\mathbf{y}^*}\sum_{\mathbf{x}^*\in S_d^k(\mathbf{y}^*)}\chi\{\,\mathbf{x}^*\in\tilde{S}_a(G^*)\,\} \\ &= \mu\cdot\frac{k!\nu}{c_d}\cdot\sum_{\mathbf{x}^*\in\tilde{S}_a(G^*)}\sum_{\mathbf{y}^*}\chi\{\,\mathbf{y}^*\ :\ \mathbf{x}^*\in S_d^k(\mathbf{y}^*)\,\} \\ &= \mu\cdot\frac{k!\nu}{c_d}\cdot|\tilde{S}_a(G^*)|\cdot c_d \\ &\ge 2^k k!\mu\nu(1-\varepsilon) \end{aligned} \tag{A.13}$$

and

$$\begin{aligned} \sum_{\pi\in\Pi}\sum_{\Delta\mathbf{y}^*}|\mathcal{L}_2(\pi,\Delta\mathbf{y}^*)| &= \sum_{\mathbf{y}^*,\Delta\mathbf{y}^*}\chi\{\,\mathbf{y}^*\in\Delta\mathbf{y}^*+\mathcal{B}\,\}\cdot\chi\{\,\Psi(\mathbf{y}^*)\le(1+\delta_k)\cdot m^*\cdot 2^{-k(1-H(\alpha*\beta))}\,\} \\ &\quad\cdot\sum_{\mathbf{x}^*\in S_d^k(\mathbf{y}^*)}\sum_{\pi\in\Pi}\chi\{\,\mathbf{x}^*\in\pi\mathcal{B}\,\} \\ &= \mu c_d\cdot\frac{k!\nu}{c_d}\cdot\sum_{\mathbf{y}^*}\chi\{\,\Psi(\mathbf{y}^*)\le(1+\delta_k)\cdot m^*\cdot 2^{-k(1-H(\alpha*\beta))}\,\} \\ &\ge 2^k k!\mu\nu(1-\delta'). \end{aligned} \tag{A.14}$$

Furthermore,

$$\sum_{\mathbf{x}^*,\mathbf{y}^*}\sum_{\pi\in\Pi}\sum_{\Delta\mathbf{y}^*}\chi\{\,\mathbf{y}^*\in\Delta\mathbf{y}^*+\mathcal{B}\,\}\cdot\chi\{\,\mathbf{x}^*\in\pi\mathcal{B})\,\} = 2^k k!\mu\nu. \tag{A.15}$$

Combining (A.13)-(A.15) we obtain (A.12) and using this inequality in (A.11) complete the proof. Q.E.D.

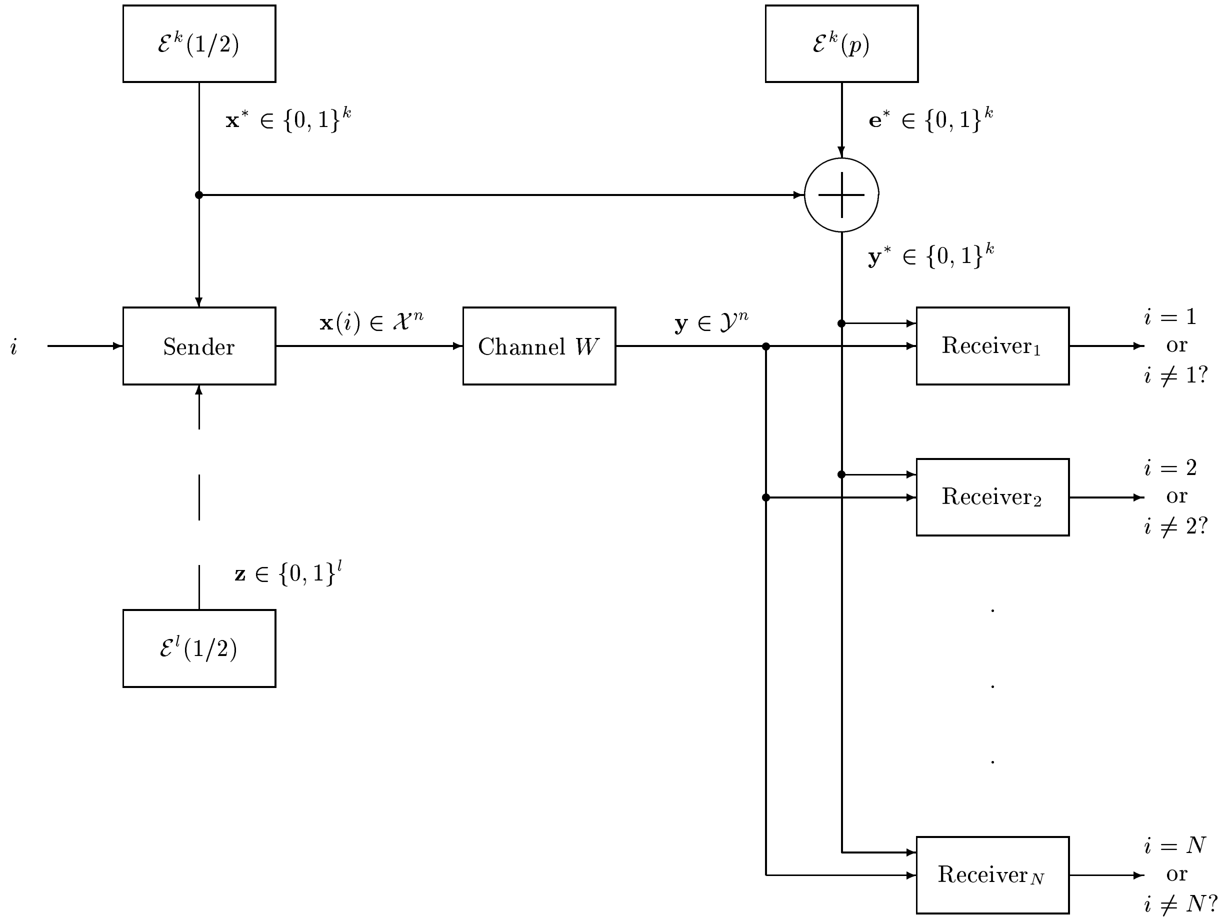

Fig. 1. A model of identification over a channel W under a random process generated by a binary symmetric source $(\mathcal{E}^k(1/2), \mathcal{E}^k(p))$.

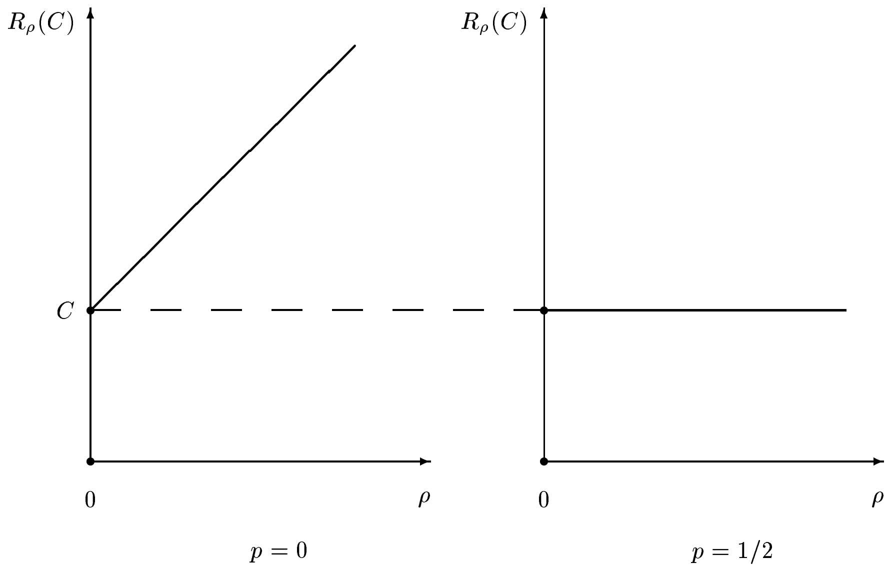

Fig. 2. The ρ-achievable ID rate $R_\rho(C)$ as a function of ρ for $p = 0$ and $p = 1/2$.

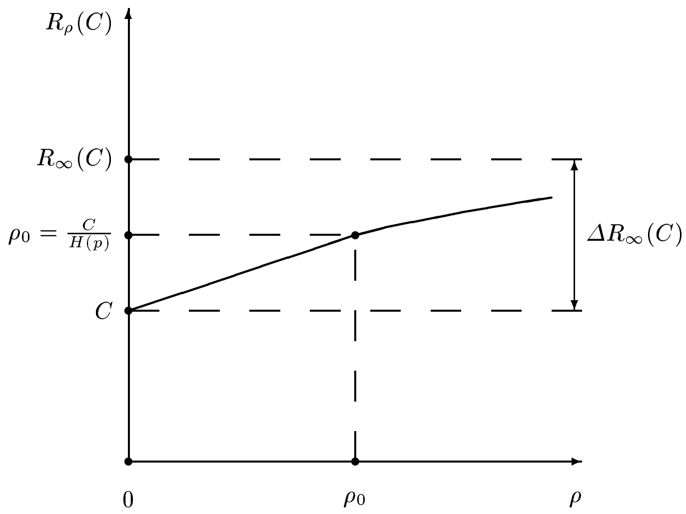

Fig. 3. A lower bound on the ρ-achievable ID rate $R_\rho(C)$ as a function of ρ for $p \in (0, 1/2)$.

References

[1] R.Ahlswede and G.Dueck, "Identification via channels," *IEEE Trans.Inform.Theory,* vol.35, 15-29, 1989.

[2] R.Ahlswede and G.Dueck, "Identification in the presence of feedback - A discovery of new capacity formulas," *IEEE Trans.Inform.Theory,* vol.35, 30-36, 1989.

[3] T.S.Han and S.Verdu, "New results in the theory of identification via channels," *IEEE Trans.Inform.Theory,* vol.38, 14-25, 1992.

[4] T.S.Han and S.Verdu, "Approximation theory of output statistics," *IEEE Trans.Inform.Theory,* vol.39, 752-772, 1993.

[5] R.Ahlswede and I.Csiszár, "Common randomness in information theory and cryptography - Part 1: Secret sharing," *IEEE Trans.Inform.Theory,* vol.39, 1121-1131, 1993.

[6] R.Ahlswede and Z.Zhang, "New directions in the theory of identification via channels,", *IEEE Trans. Inform. Theory*, vol.41, 1040-1050, 1995.

[7] I.Csiszár and J.Körner, *Information Theory : Coding Theorems for Discrete Memoryless Systems.* New York : Academic, 1981.

[8] P. Gács and J. Körner, "Common information in far less than mutual information", *Problems of Control and Information Theory*, vol.2, 149-162, 1973.

[9] R. Ahlswede and J. Körner, "On common information and related characteristics of correlated information sources", SFB *Diskrete Strukturen in der Mathematik*, Ergänzungsreihe 95-003, Bielefeld 1995.

[10] R. Ahlswede and B. Verboven, "On identification via multi-way channels with feedback", *IEEE Trans. Inform. Theory*, vol.37, 1519-1526, 1991.

[11] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography - Part 2", in preparation.