UDC 621.391
doi:10.15217/issn1684-8853.2018.3.79

RANDOM CODING BOUND FOR CHANNELS WITH MEMORY — DECODING FUNCTION WITH PARTIAL OVERLAPPING

Part 1. Derivation of Main Expression

***A. N. Trofimov*[a]**, *PhD, Tech., Associate Professor, andrei.trofimov@vu.spb.ru*
[a]*Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaia St., 190000, Saint-Petersburg, Russian Federation*

Introduction: *The problem of calculating the random coding exponent in the full range of code rates for finite-state channels is not completely solved and remains relevant. It seems that a good approximation to the optimal random coding exponent can be found by using a mismatched decoding function.* ***Purpose:*** *Deriving a random coding exponent close to the optimal one.* ***Results:*** *A new random coding bound is presented for a wide class of channels, including those for which the complete random coding exponent was not previously derived. The derivation of this bound is based on the use of a mismatched decoding function which depends on two parameters: the length W of the segment of the channel output sequence and the length B of the segment of the channel input sequence. The values of W and B greatly influence the values of the random coding exponent and the complexity of its calculation.*

Keywords *— Random Coding Bound, Finite-State Channel, Mismatched Decoding, Perron — Frobenius Theorem.*

Citation: Trofimov A. N. Random Coding Bound for Channels with Memory — Decoding Function with Partial Overlapping. Part 1. Derivation of Main Expression. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2018, no. 3, pp. 79–88. doi:10.15217/issn1684-8853.2018.3.79

Introduction

According to the random coding theorem [1] there exists a block code of length N and rate R, $R = N^{-1} \log_2 M$, M is the number of code words, such that the error probability of the maximum likelihood (ML) decoding is upper bounded as $P_e < \exp_2(-N(E_r(R) - o(N)))$, where $\exp_2 = 2^x$, and $E_r(R)$ the exponent of the random coding bound. The function $E_r(R) > 0$, if $R < C$, where C is the channel capacity. Another quantity characterizing the channel is the computational cut-off rate R_0. The value R_0 can be found from the function $E_r(R)$ as $R_0 = E_r(0)$. The values of $E_r(R)$ and R_0 are calculated by averaging over a code ensemble and optimizing over a distribution on this ensemble. Generally, the *exact* computation of the function $E_r(R)$ is a difficult problem which can only be solved for some channel models. However, sometimes it is possible to obtain a bound for the function $E_r(R)$ for the decoding algorithm used a decoding function ψ which differs from the ML decoding function (mismatched decoding; see, for example, [2–6] and references wherein). In this case it is possible to obtain near-exact characteristics of the transmission reliability. This approach is used in this paper.

For simplicity, we assume that the distribution on the code ensemble corresponds to the *independent uniformly distributed* (i.u.d.) code symbols. In this case we have the function $E_r^*(R; \psi) \le E_r(R)$ and quantities $R_0^*(\psi) \le R_0$, $C^*(\psi) \le C$ (the asterisk in the superscript hereafter means that the code symbols are chosen as i. u. d. random variables).

We assume that the channel model is given by the well known finite-state model [1]. For this model we present the derivation of the random coding exponent for the decoding function which is in the form of the product of the *a posteriori* probabilities (APP) of segments of the channel input sequence relative to the overlapped segments of the channel output sequence. One of the main challenges of this study was the choice of a suitable *decoding function* enabling a good final result. With the usage of such a decoding function the problem can be reduced to the evaluation of the logarithm of a bilinear form defined by a power of a nonnegative matrix. Then we used a known technique based on the usage of the Perron — Frobenius theorem to obtain the final result. Using this approach a new suboptimal random coding bound which is applicable to a wide class of the channel with memory has been obtained. The discrete-time model with intersymbol interference, additive noise and fixed inputs is one of the important examples of such channels [7].

Under these or similar conditions the *maximum achievable information rate* for the finite-state channel model has been estimated in [8–10] using a simulation-based algorithm. Bounds on the computational cut-off rate R_0^* for the finite-state channel models were studied, particularly, in [11–16]. Some of the very first random coding bounds for channels with memory were published in [17, 18]. In [18] random coding bounds for the discrete additive finite-state channel were obtained for mismatched decoding different from the ML decoding function, and that publication gave a basic idea for this paper. The main known results for the random coding bounds for discrete-time channels with time-invariant intersymbol interference (ISI) consist in evaluation the value R_0^* for *ML decoding* and *i. u. d. code symbols*. In [11], this task was solved for a particular case of the ISI channel with the length of interference equal to 1. More general cases of ISI were studied in [12–16]. Therefore, one may say that for channels with ISI it is known how to calculate the value of R_0 (the technique based on the Perron — Frobenius theorem) and the value of *maximum achievable information rate* C^* (simulation-based algorithm [9]). Taking these values into account let us introduce function $\tilde{E}_r^*(R)$, as

$$\tilde{E}_r^*(R) = \begin{cases} R_0^* - R, \ 0 \le R \le R_0^*, \\ 0, \ R \ge C^*. \end{cases} \quad (1)$$

Note that the function $\tilde{E}_r^*(R)$ in is not defined for rates in the range $R_0^* < R < C^*$. Let us denote by $E_r^*(R)$ the random coding exponent obtained for the discrete-time ISI channel, ML decoding, and a code ensemble with i. u. d. code symbols. Then, for the values of $R \in [0, R_0^*] \cup \{C^*\}$ the inequality $E_r^*(R) \ge \tilde{E}_r^*(R)$ is valid. Clearly, $E_r^*(R) = \tilde{E}_r^*(R)$ in the interval $0 \le R \le R_{cr}$, where R_{cr} is the *critical rate*, and in the point $R = C^*$. In this paper we introduce a decoding function ψ with partial overlapping which depends on two integer parameters W and B, $W \ge B \ge 0$. With the use of this function we obtain the random coding exponent $E_r^*(R; \psi)$ which can be a good approximation of the function $E_r^*(R)$ if $W, B \to \infty$ even for the those values of R for which the function $\tilde{E}_r^*(R)$ is not defined.

This paper is the first part of a general study, consisting of two parts, published separately. In the second part of the work we intend to present a number of examples and their discussion.

Notation and Basic Equations

Let $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$ be the transition probability of the discrete-time channel; for the continuous-output channel it is instead a probability density function (p.d.f.); $\mathbf{x} \in X^N$, where X be a discrete input channel alphabet and $q_x = |X| < \infty$; $\mathbf{y} \in Y^N$, where Y is the channel output alphabet and N is the length of a block code. For the continuous channel output $Y = \mathbb{R}$. The set Y may also represent a quantized version of the continuous channel outputs, i. e. $|Y| = q$. In this study consider this case.

The notation $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$ will mainly be used when vectors $\mathbf{x}$ and $\mathbf{y}$ have the equal lengths, e. g., N. For subvectors, or segments of vectors, $\mathbf{x}$ and $\mathbf{y}$ the notation x and y is used. The difference between them is noted due to the use of ordinary and *sans serif* font. This notation is context-dependent; in particular, the length of x and/or y can differ in the various contexts.

To indicate a segment of an arbitrary vector $\mathbf{z}$ we use the notation $\mathbf{z}_a^b = (z^{(\max(1,a))}, z^{(\max(1,a)+1)}, \ldots, z^{(\min(b,L))})$, where L is length of the vector $\mathbf{z}$.

Let $p_{\mathbf{x}}(\mathbf{x})$ be a distribution on the code ensemble, where $\mathbf{x} = (x^{(1)}, x^{(2)}, \ldots, x^{(N)})$, and $p_x(x^{(n)})$ be a one-dimensional distribution giving the distribution of a single code symbol, $n = 1, 2, \ldots, N$.

We assume the decoding rule is given by $\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \psi(\mathbf{y}; \mathbf{x})$, where $\psi(\mathbf{y}; \mathbf{x})$ is a real-valued positive decoding function, and the maximization is performed over all code words.

Using standard techniques [1] one can obtain the random coding bound $P_e < \exp_2(-NE_r(N, R, \psi))$, where P_e is the block error probability; $E_r(N, R, \psi)$ is the generalized exponent of the random coding bound defined as

$$E_r(N, R, \psi) = \max_{p_{\mathbf{x}}} \max_{1 \ge \rho \ge 0} \max_{\lambda > 0} (E_0(N, \psi, \rho, \lambda, p_{\mathbf{x}}) - \rho R),$$

where λ, ρ are the optimization parameters; R is the code rate, and (see also [5])

$$E_0(N, \psi, \rho, \lambda, p_{\mathbf{x}}) = = -\frac{1}{N} \log \sum_{\mathbf{y}} \sum_{\mathbf{x}} p_{\mathbf{x}}(\mathbf{x}) p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) \psi(\mathbf{y}; \mathbf{x})^{-\lambda\rho} \times \times \left(\sum_{\mathbf{x}'} p_{\mathbf{x}}(\mathbf{x}') \psi(\mathbf{y}; \mathbf{x}')^{\lambda} \right)^{\rho}. \quad (2)$$

Hereafter $\log(\cdot)$ denotes the binary logarithm. The expression is an approximation to the exponent of random coding for a channel with a continuous output and the accuracy of this approximation increases with increasing number of quantization levels q.

Assigning $\psi(\mathbf{y}; \mathbf{x}) = p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$ corresponds to ML decoding; in this case the optimal value of the parameter λ is equal to $\lambda = 1/(1 + \rho)$, and we get classical expression for the random coding exponent [1]

$$E_r(N, R, p_{\mathbf{y}|\mathbf{x}}) = E_r(N, R) = = \max_{p_{\mathbf{x}}} \max_{1 \ge \rho \ge 0} \left(E_0(N, \rho, p_{\mathbf{x}}) - \rho R \right),$$

where

$$E_0(N, \rho, p_{\mathbf{x}}) = -\frac{1}{N} \log \sum_{\mathbf{y}} \left(\sum_{\mathbf{x}} p_{\mathbf{x}}(\mathbf{x}) p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}.$$

As $N \to \infty$, there is an *asymptotic* generalized random coding bound $P_e < \exp_2(-N(E_r(R;\ \psi) - o(N)))$, where

$$E_r(R;\ \psi) = \max_{p_{\mathbf{x}}} \max_{1 \ge \rho \ge 0} \left(\max_{\lambda > 0} (E_0(\psi,\ \rho,\ \lambda,\ p_{\mathbf{x}}) - \rho R \right),$$

and

$$E_0(\psi,\ \rho,\ \lambda,\ p_{\mathbf{x}}) = \lim_{N \to \infty} E_0(N,\ \psi,\ \rho,\ \lambda,\ p_{\mathbf{x}}).$$

If $\psi(\mathbf{y};\ \mathbf{x}) = p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$, then again the optimal value of the parameter λ is $\lambda = 1/(1+\rho)$ and we get asymptotic expression for the asymptotic random coding exponent for ML decoding [1]

$$E_r(R, p_{\mathbf{y}|\mathbf{x}}) = E_r(R) = \max_{p_{\mathbf{x}}} \max_{1 \ge \rho \ge 0} (E_0(\rho, p_{\mathbf{x}}) - \rho R),$$

where $E_0(\rho,\ p_{\mathbf{x}}) = \lim_{N \to \infty} E_0(N,\ \rho,\ p_{\mathbf{x}})$. It is known [1] that $E_r(R) > 0$, if $R < C$, where C is the channel capacity, which can be found from the function $E_0(\rho,\ p_{\mathbf{x}})$ as

$$C = \max_{p_{\mathbf{x}}} \left. \frac{\partial E_0(\rho, p_{\mathbf{x}})}{\partial \rho} \right|_{\rho=0}.$$

We assume that the code ensemble distribution is given as $p_{\mathbf{x}}(\mathbf{x}) = \prod_{n=1}^{N} p_x(x^{(n)})$, $\mathbf{x} \in X^N$, and $p_x(x) = 1/q_x$, $x \in X$, i. e. it corresponds to the i. u. d. channel input. This assumption leads to *loss of optimality* but simplifies further consideration. Under these assumptions one can derive the suboptimal exponent of the random coding bound in *asymptotic form*

$$E_r^*(R;\ \psi) = \max_{1 \ge \rho \ge 0} \left(\max_{\lambda > 0} E_0^*(\psi,\ \rho,\ \lambda) - \rho R \right), \tag{3}$$

where

$$E_0^*(\psi,\ \rho,\ \lambda) = (1+\rho)\log q_x - \\ - \lim_{N \to \infty} \frac{1}{N} \log \sum_{\mathbf{y}} \sum_{\mathbf{x}} p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\,|\,\mathbf{x}) \psi(\mathbf{y};\ \mathbf{x})^{-\lambda\rho} \times \\ \times \left(\sum_{\mathbf{x}'} \psi(\mathbf{y};\ \mathbf{x}')^{\lambda} \right)^{\rho}. \tag{4}$$

By analogy with the channel capacity C let us define the lower bound on maximum achievable code rate $C^*(\psi)$ as

$$C^*(\psi) = \left. \frac{\partial \max_{\lambda > 0} E_0^*(\psi,\ \rho,\ \lambda)}{\partial \rho} \right|_{\rho=0} \le C. \tag{5}$$

The value $R_0^*(\psi) = \max_{\lambda > 0} E_0^*(\psi,\ 1,\ \lambda)$ gives a bound on the cut-off rate $R_0 = \max_{p_{\mathbf{x}}} E_0(1,\ p_{\mathbf{x}})$; evidently the inequalities $R_0^*(\psi) \le R_0^* \le R_0$ are valid.

Similarly we can write the equations for function $E_r^*(N,\ R)$ for the ML decoding as

$$E_r^*(N,\ R) = \max_{1 \ge \rho \ge 0} \left(E_0^*(N,\ \rho) - \rho R \right), \tag{6}$$

where

$$E_0^*(N,\ \rho) = (1+\rho)\log q_x - \\ - \frac{1}{N} \log \sum_{\mathbf{y}} \left(\sum_{\mathbf{x}} p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\,|\,\mathbf{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}, \tag{7}$$

and the asymptotic random coding exponent as $E_r^*(R) = \lim_{N \to \infty} E_r^*(N,\ R)$ with bound on the maximum achievable information rate

$$R_{\max}^*(N) = \partial E_0^*(N,\ \rho) / \partial \rho \big|_{\rho=0},\ \ R_{\max}^*(N) \le C.$$

Channel Model and Decoding Function

Let the channel transition probabilities be given as $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\,|\,\mathbf{x}) = \left(\sum_{\mathbf{s}} p_{\mathbf{yxs}}(\mathbf{y},\ \mathbf{x},\ \mathbf{s}) \right) / p_{\mathbf{x}}(\mathbf{x})$, where $\mathbf{s} = (s^{(0)}, s^{(1)}, \ldots, s^{(n)}, \ldots)$ is the sequence of the channel states, $s^{(n)} \in S$, S is a set of the channel states and $|S| < \infty$, $p_{\mathbf{yxs}}(\mathbf{y},\ \mathbf{x},\ \mathbf{s})$ is the simultaneous probability of the vectors $\mathbf{y}$, $\mathbf{x}$ and $\mathbf{s}$. Note that

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\,|\,\mathbf{x}) = \frac{\sum_{\mathbf{s}} p_{\mathbf{yxs}}(\mathbf{y},\ \mathbf{x}, \mathbf{s})}{p_{\mathbf{x}}(\mathbf{x})} = \\ = \frac{\sum_{\mathbf{s}} p_{\mathbf{y}|\mathbf{xs}}(\mathbf{y}\,|\,\mathbf{x},\ \mathbf{s}) p_{\mathbf{xs}}(\mathbf{x},\ \mathbf{s})}{p_{\mathbf{x}}(\mathbf{x})} = \sum_{\mathbf{s}} p_{\mathbf{y}|\mathbf{xs}}(\mathbf{y}\,|\,\mathbf{x},\ \mathbf{s}) p_{\mathbf{s}|\mathbf{x}}(\mathbf{s}\,|\,\mathbf{x}),$$

where $p_{\mathbf{y}|\mathbf{xs}}(\mathbf{y}|\mathbf{x},\ \mathbf{s})$ is the conditional probability of the channel output for the fixed vectors $\mathbf{x}$ and $\mathbf{s}$, and $p_{\mathbf{s}|\mathbf{x}}(\mathbf{x}|\mathbf{s})$ is a conditional probability of the channel states for the given input vector $\mathbf{x}$.

Let us assume that the channel is a *probabilistic finite-state machine*, i. e.

$$p_{\mathbf{y}|\mathbf{xs}}(\mathbf{y}\,|\,\mathbf{x},\ \mathbf{s}) = \prod_{n=1}^{N} p_{y|xs}(y^{(n)}\,|\,x^{(n)},\ s^{(n-1)});$$

$$p_{\mathbf{s}|\mathbf{x}}(\mathbf{s}\,|\,\mathbf{x}) = p_s(s^{(0)}) \prod_{n=1}^{N} p_{s|xs}(s^{(n)}\,|\,x^{(n)},\ s^{(n-1)}),$$

where $p_s(\cdot)$ is an unconditional (stationary) distribution on the set of channel states; $p_{s|xs}(s^{(n)}|x^{(n)},\ s^{(n-1)})$ is the conditional channel state transition probability. In addition, we assume that the input channel symbol $x^{(n)}$ and the current channel state $s^{(n-1)}$ are independent. Such a model is quite general enough and has been widely used (see, for example, [8, 9] and references wherein). An explanatory illustration is given in Fig. 1 [9].

Some particular cases of this model are:

1. State transitions not depending on the input symbol (*channel with freely evolving states*) [9], i. e. $p_{s|xs}(s^{(n)}| x^{(n)}\ s^{(n-1)}) = p_{s|s}(s^{(n)}|s^{(n-1)})$. The Gilbert — Elliot [19, 20] channel is a specific example of such a model.

2. Deterministic state transitions (*deterministic finite state machine)*. In this case the state transition is given as

$$p_{s|xs}(s^{(n)} \mid x^{(n)},\ s^{(n-1)}) = \begin{cases} 1, & \text{if } s^{(n-1)} \xrightarrow{x^{(n)}} s^{(n)}; \\ 0, & \text{otherwise.} \end{cases}$$

The ISI channel is an example of a channel in this class.

The probabilities $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x})$ can be represented as

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x}) = \sum_{s^{(0)}} \cdots \sum_{s^{(N-1)}} \sum_{s^{(N)}} p_s(s^{(0)}) \times$$
$$\times \left(\prod_{n=1}^{N} p_{y|xs}(y^{(n)} \mid x^{(n)},\ s^{(n-1)}) \, p_{s|xs}(s^{(n)} \mid x^{(n)},\ s^{(n-1)}) \right),$$

or in the following form of *matrix product* proved similarly to the derivation of [1, eq. (5.9.39)]:

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x}) = \mathbf{p}_s \left(\prod_{n=1}^{N} \mathbf{P}(y^{(n)} \mid x^{(n)}) \right) \mathbf{1}^{\mathrm{T}}, \qquad (8)$$

where

$$\mathbf{P}(y \mid x) = [p_{y|xs}(y \mid x,\ s)\, p_{s|xs}(s' \mid x,\ s)] \qquad (9)$$

is a matrix of size $|S| \times |S|$; $\mathbf{p}_s = [p_s(1), \ldots, p_s(|S|)]$ is the vector of the unconditional state probabilities at $n = 0$, and $\mathbf{1} = (1, \ldots, 1)$ is vector of 1's of dimensions $1 \times |S|$.

Let us choose the decoding function as the product of APP of input segments of length $2B + 1$ for fixed output segments of length $2W + 1$, $W \geq B \geq 0$ are integer parameters,

$$\psi(\mathbf{y};\ \mathbf{x}) = \prod_{n=0}^{N(B)-1} \Pr\left[\mathbf{x}_{n(2B+1)-B}^{n(2B+1)+B} \mid \mathbf{y}_{n(2B+1)-W}^{n(2B+1)+W} \right],$$

and

$$N(B) = \lceil (N - B) / (2B + 1) \rceil. \qquad (10)$$

The value $N(B)$, defined in (10), gives the number of the code block segments, or subblocks, of the length $2B + 1$. As will be seen in the following, for such a decoding function, it is possible to obtain a good final result for the suboptimal random coding exponent.

Let us denote

$$k(n) = n(2B + 1) + 1. \qquad (11)$$

This value gives the position of the central element of the n^{th} code block segment of length $2B + 1$. For equiprobable segments $\mathbf{x}_{k(n)-B}^{k(n)+B}$ we can write the expression for the decoding function $\psi(\mathbf{y};\ \mathbf{x})$ in another equivalent form

$$\psi(\mathbf{y};\ \mathbf{x}) = \prod_{n=0}^{N(B)-1} p_{\mathsf{y}|\mathsf{x}}\left(\mathbf{y}_{k(n)-W}^{k(n)+W} \mid \mathbf{x}_{k(n)-B}^{k(n)+B} \right), \qquad (12)$$

where $p_{\mathsf{y}|\mathsf{x}}(\cdot|\cdot)$ is the conditional probability for segments of *different*, in general, lengths $2W + 1$ and $2B + 1$. Note, that in general $p_{\mathsf{y}|\mathsf{x}}(\cdot|\cdot) \neq p_{\mathbf{y}|\mathbf{x}}(\cdot|\cdot)$. The difference in these values is emphasized by their index notation typed in Roman bold font and *sans serif* bold font respectively.

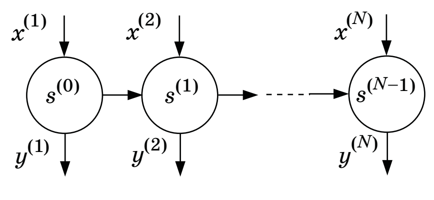
The segments $\mathbf{x}_{k(n)-B}^{k(n)+B}$ of length $2B + 1$ of the input vector $\mathbf{x}$ *do not overlap*, but the segments $\mathbf{y}_{k(n)-W}^{k(n)+W}$ of length $2W + 1$ *overlap* on a segment of length $2(W - B)$. The illustration for $B = 2$, $W = 4$, $2B + 1 = 5$, $2W + 1 = 9$ is shown in Fig. 2.

Let us note some specific cases of the decoding function (12).

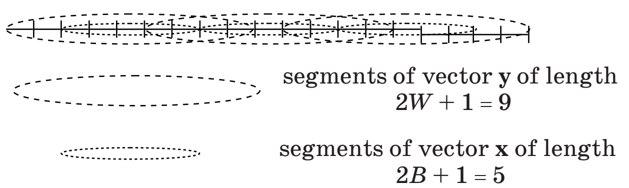
Case 1. If $W = B = 0$, then $N(B) = N$, and the decoding function is

$$\psi(\mathbf{y};\ \mathbf{x}) = \prod_{n=0}^{N-1} p_{\mathsf{y}|\mathsf{x}}\left(\mathbf{y}_{n+1}^{n+1} \mid \mathbf{x}_{n+1}^{n+1} \right) = \prod_{n=0}^{N-1} p_{y|x}\left(y^{(n)} \mid x^{(n)} \right),$$

i. e. it is matched with the memoryless channel.



■ ***Fig. 1.*** Finite-state channel transitions



■ ***Fig. 2.*** Positions of the subblocks in product (12)

Case 2. If $B = 0$ and $W \geq 1$, then $N(B) = N$, and the decoding function is

$$\psi(\mathbf{y};\ \mathbf{x}) = \prod_{n=0}^{N-1} p_{\mathsf{y}|x}\left(\mathbf{y}_{n+1-W}^{n+1+W} \mid x^{(n)}\right) = \prod_{n=0}^{N-1} p_{\mathsf{y}|x}\left(\mathbf{y}_{n-W}^{n+W} \mid x^{(n)}\right),$$

and such decoding is equivalent to APP symbol decoding in a window of the length $2W + 1$, and it is similar to windowed version of the Bahl — Cocke — Jelinek — Raviv (BCJR) algorithm [21], (see e. g. [22]).

Case 3. If $W = B = N$, then $N(B) = 1$, and the decoding function is

$$\psi(\mathbf{y};\ \mathbf{x}) = p_{\mathsf{y}|\mathsf{x}}\left(\mathbf{y}_{1-N}^{1+N} \mid \mathbf{x}_{1-N}^{1+N}\right) = p_{\mathsf{y}|\mathsf{x}}\left(\mathbf{y}_1^N \mid \mathbf{x}_1^N\right) = p_{\mathsf{y}|\mathsf{x}}(\mathbf{y} \mid \mathbf{x}),$$

and corresponds to the ML decoding.

It can be shown (see Appendix A) that the probabilities $p_{\mathsf{y}|\mathsf{x}}(\mathsf{y}|\mathsf{x})$ on the right-hand side of can be found for any $\mathsf{y} \in Y^{2W+1}$ and $\mathsf{x} \in X^{2B+1}$ as

$$p_{\mathsf{y}|\mathsf{x}}(\mathsf{y} \mid \mathsf{x}) = q_x^{-2(W-B)}\mathbf{p}_s\left(\prod_{l=1}^{W-B}\mathbf{P}(\mathsf{y}^{(l)})\right)\times \left(\prod_{l=W-B+1}^{W+B+1}\mathbf{P}(\mathsf{y}^{(l)} \mid \mathsf{x}^{(l)})\right)\left(\prod_{l=W+B+2}^{2W+1}\mathbf{P}(\mathsf{y}^{(l)})\right)\mathbf{1}^{\mathrm{T}}, \quad (13)$$

where $\mathsf{y}^{(l)}$, $\mathsf{x}^{(l)}$ are components of the vectors y and x respectively; $\mathbf{P}(y|x)$ is $|S| \times |S|$ *matrix* defined by equation (9), and $\mathbf{P}(y)$ is *matrix* defined as the sum

$$\mathbf{P}(y) = \sum_{x \in X}\mathbf{P}(y \mid x). \quad (14)$$

Suboptimal Random Coding Exponent

Consider the sums in the right-hand part of the equation (4). It can be shown (see Appendix B) that

$$\sum_{\mathbf{x}}\psi(\mathbf{y};\ \mathbf{x})^{\lambda} = a\left(\mathbf{y}_1^{k(n_1-1)+W}\right)\times \left(\prod_{n=n_1}^{n_2} D_1\left(\mathbf{y}_{k(n)-W}^{k(n)+W};\ \lambda\right)\right) b\left(\mathbf{y}_{k(n_2+1)-W}^{N}\right),$$

where

$$D_1(\mathsf{y};\ \lambda) = \sum_{\mathsf{x} \in X^{2B+1}} p_{\mathsf{y}|\mathsf{x}}(\mathsf{y} \mid \mathsf{x})^{\lambda},\ \ \mathsf{y} \in Y^{2W+1},$$

n_1 is the least positive integer such that $k(n_1) - W \geq 1$, and n_2 is the greatest integer such that, $k(n_2) + W \leq N$, or according to the definition (11)

$$n_1 = \lceil W / (2B+1)\rceil,\quad n_2 = \lfloor (N - W - 1) / (2B+1)\rfloor. \quad (15)$$

We denote by $a(\cdot)$ and $b(\cdot)$ the non-essential (not affecting the exponent of random coding) positive factors. It is also shown in Appendix B that

$$\sum_{\mathbf{x}} p_{\mathsf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x})\psi(\mathbf{y};\ \mathbf{x})^{-\lambda\rho} = \mathbf{p}_s\mathbf{A}\left(\mathbf{y}_1^{k(n_1-1)+W}\right)\times \left(\prod_{n=n_1}^{n_2}\mathbf{D}_2(\mathbf{y}_{k(n)-W}^{k(n)+W};\ \lambda\rho)\right)\mathbf{B}\left(\mathbf{y}_{k(n_2+1)-W}^{N}\right)\mathbf{1}^{\mathrm{T}},$$

where $\mathbf{D}_2(\mathsf{y};\ \lambda\rho)$ is $|S| \times |S|$ matrix defined as

$$\mathbf{D}_2(\mathsf{y};\ \lambda\rho) = \sum_{\mathsf{x} \in X^{2B+1}}\mathbf{P}_{\mathsf{y}|\mathsf{x}}\left(\mathsf{y}_{W-B+1}^{W+B+1} \mid \mathsf{x}\right)p_{\mathsf{y}|\mathsf{x}}(\mathsf{y} \mid \mathsf{x})^{-\lambda\rho}, \quad (16)$$

$\mathsf{y} \in Y^{2W+1}$; $\mathbf{A}(\cdot)$ и $\mathbf{B}(\cdot)$ are inessential nonnegative matrix multipliers of dimension $|S| \times |S|$, and

$$\mathbf{P}_{\mathsf{y}|\mathsf{x}}(\mathsf{y} \mid \mathsf{x}) = \prod_{l=1}^{2B+1}\mathbf{P}(\mathsf{y}^{(l)} \mid \mathsf{x}^{(l)}),\ \mathsf{y} \in Y^{2B+1},\ \mathsf{x} \in X^{2B+1}, \quad (17)$$

where the matrices $\mathbf{P}(\cdot|\cdot)$ are defined by the equation (9). Note, that $\mathsf{y}_{W-B+1}^{W+B+1}$ in the right-hand part of (16) is the middle part of the vector y having length $2B + 1$.

Using these notations we get from (4) that

$$E_0^*(\psi,\ \rho,\ \lambda) = (1+\rho)\log q_x - \lim_{N\to\infty}\frac{1}{N}\log \mathbf{p}_s\sum_{\mathbf{y}}\left(\mathbf{U}\left(\mathbf{y}_1^{k(n_1-1)+W}\right)\times \left(\prod_{n=n_1}^{n_2}\mathbf{D}\left(\mathbf{y}_{k(n)-W}^{k(n)+W};\ \lambda,\ \rho\right)\right)\mathbf{V}\left(\mathbf{y}_{k(n_2+1)-W}^{N}\right)\right)\mathbf{1}^{\mathrm{T}}, \quad (18)$$

where

$$\mathbf{U}(\mathsf{y}) = a(\mathsf{y})^{\rho}\mathbf{A}(\mathsf{y}),\ \ \mathsf{y} \in Y^{k(n_1-1)+W};$$

$$\mathbf{V}(\mathsf{y}) = \mathbf{B}(\mathsf{y})b(\mathsf{y})^{\rho},\ \ \mathsf{y} \in Y^{N-k(n_2+1)+W+1}; \quad (19)$$

$$\mathbf{D}(\mathsf{y};\ \lambda,\ \rho) = D_1(\mathsf{y};\ \lambda)^{\rho}\mathbf{D}_2(\mathsf{y};\ \lambda\rho),\ \ \mathsf{y} \in Y^{2W+1}. \quad (20)$$

In general (see Appendix C), the sum over $\mathbf{y}$ on the right-hand side of (16) can be written in the form

$$\sum_{\mathbf{y}}\prod_{n=n_1}^{n_2}\mathbf{D}\left(\mathbf{y}_{k(n)-W}^{k(n)+W};\ \lambda,\ \rho\right) = \mathbf{F}\,\mathbf{K}(\lambda,\ \rho)^{n_2-n_1+1}\,\mathbf{G}, \quad (21)$$

where

$$\mathbf{K}(\lambda,\ \rho)=[\mathbf{K}_{ij}(\lambda,\ \rho)]=$$
$$=\begin{bmatrix} \mathbf{K}_{11}(\lambda,\ \rho) & \dots & \mathbf{K}_{1\,q^{2(W-B)}}(\lambda,\ \rho) \\ \dots & \dots & \dots \\ \mathbf{K}_{q^{2(W-B)}\,1}(\lambda,\rho) & \dots & \mathbf{K}_{q^{2(W-B)}\,q^{2(W-B)}}(\lambda,\rho) \end{bmatrix} \quad (22)$$

is a square block matrix of order $|S|q^{2(W-B)}$ built of he blocks $\mathbf{K}_{ij}(\lambda,\ \rho)$ of dimension $|S|\times|S|$ defined as

$$\mathbf{K}_{ij}(\lambda,\ \rho)=\begin{cases} \mathbf{D}(\mathbf{y};\ \lambda,\ \rho), & W\ge 2B+1; \\ \sum\limits_{\mathbf{y}_{2(W-B)+1}^{2B+1}} \mathbf{D}(\mathbf{y};\ \lambda,\ \rho), & W<2B+1 \end{cases} \quad (23)$$

for $i, j=1, ..., q^{2(W-B)}$. The matrix $\mathbf{D}(\mathbf{y};\ \lambda,\ \rho)$ in equations (21) and (23) is defined in (20). The correspondence of the indices i, j and the vector $\mathbf{y}$ in the expression (23) is given as $i\leftrightarrow \mathbf{y}_1^{2(W-B)}$ and $j\leftrightarrow \mathbf{y}_{2B+2}^{2W+1}$. In other words, the components of the vectors $\mathbf{y}_1^{2(W-B)}$ and $j\leftrightarrow \mathbf{y}_{2B+2}^{2W+1}$ are the digits in q-ary representation of the indices i and j respectively.

The matrices $\mathbf{F}$ and $\mathbf{G}$ on the right-hand side of (21) are nonnegative matrix multipliers of dimension $|S|\times|S|q^{2(W-B)}$ and $|S|q^{2(W-B)}\times|S|$.

With these definitions equation (18) can be rewritten as follows

$$E_0^*(\psi,\ \rho,\ \lambda)=(1+\rho)\log q_x -$$
$$-\lim_{N\to\infty}\frac{1}{N}\log\left(\mathbf{f}\,\mathbf{K}(\lambda,\ \rho)^{n_2-n_1+1}\mathbf{g}\right), \quad (24)$$

where n_1, n_2 are defined in equations (15) and $\mathbf{f}=\mathbf{p}_s\mathbf{G};\ \ \mathbf{g}=\mathbf{G}\mathbf{1}^{\mathrm{T}}$ are inessential nonnegative vectors. In what follows we use the following assertion.

Corollary from the Perron — Frobenius theorem [1, 23]. Let $\mathbf{A}$ be a nonnegative irreducible square matrix, $\mathbf{a}$ and $\mathbf{b}$ be nonnegative vectors of the corresponding dimensions, then

$$\lim_{N\to\infty} N^{-1}\log \mathbf{a}\mathbf{A}^N\mathbf{b}=\log r(\mathbf{A}),$$

where $r(\mathbf{A})$ is maximum eigenvalue (spectral radius) of the matrix $\mathbf{A}$.

Using this corollary and the definitions in (15) we get from the equation (24) that, if the matrix $\mathbf{K}(\lambda,\ \rho)$ is irreducible, then

$$E_0^*(\psi,\ \rho,\ \lambda)=(1+\rho)\log q_x -$$
$$-\ (2B+1)^{-1}\log r(\mathbf{K}(\lambda,\ \rho)), \quad (25)$$

where $r(\mathbf{K}(\lambda,\ \rho))$ is maximum eigenvalue (spectral radius) of the matrix $\mathbf{K}(\lambda,\ \rho)$. The similar approach has been used many times in early publications such as [1, 17, 18], and later, for example, in [12–16].

Let us consider the conditions for matrix $\mathbf{K}(\lambda,\ \rho)$, defined in (22), to be irreducible. Obviously for the matrix $\mathbf{K}(\lambda,\ \rho)$ to be irreducible it is sufficient that each of its blocks $\mathbf{K}_{ij}(\lambda,\ \rho)$ is irreducible. Matrices $\mathbf{K}_{ij}(\lambda,\ \rho)$ defined by the equation (23) are linear combinations of the matrices $\mathbf{P}_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$ [see (16) and (17)]. Hence, for irreducibility of the matrix $\mathbf{K}_{ij}(\lambda,\ \rho)$ it is sufficient that the matrices in (17) are irreducible. This condition is satisfied if the matrices (9) are irreducible. Irreducibility of the matrices (9) means that any channel state is reachable from any other state over a finite number of steps when receiving independent, equally distributed symbols to the channel input. Below we assume that the matrices (9) in are irreducible for any x and y.

After substitution (25) into (3) we have the following theorem.

Theorem. Let channel be specified by conditional probabilities (8), where the matrices (9) are irreducible, and let the decoding function ψ be given by equation (12) with integer parameters W and B, where $W\ge B\ge 0$. Then the achievable random coding exponent $E_r^*(R;\ \psi)$ for the code ensemble with i. u. d. code symbols is

$$E_r^*(R;\psi)=\max_{0\le\rho\le 1}\left(E_0^*(\psi,\ \rho)-\rho R\right),$$

where $\quad E_0^*(\psi,\ \rho)=\max\limits_{\lambda>0}E_0^*(\psi,\ \rho,\ \lambda)=(1+\rho)\log q_x -$

$-\ (2B+1)^{-1}\log\left(\min\limits_{\lambda>0} r\left(\mathbf{K}(\lambda,\ \rho)\right)\right)$, and $r(\mathbf{K}(\lambda,\ \rho))$ is the maximum eigenvalue (spectral radius) of matrix $\mathbf{K}(\lambda,\ \rho)$, given in equation (22).

Conclusion

In this paper representing the first part of the general study we built a random coding bound applicable to a wide class of channels with memory defined as probabilistic finite-state machine. This class of the models describes many transmission channels important for theory and practice. Among them, we can highlight channel models with intersymbol interference, which are widely used for description of data transmission and recording systems. To obtain the main result, an approach used mismatched decoding function is applied. The choice of decoding function was the main problem of this study. In the second part of the work we will give examples of calculating the exponent of random coding for several models of channels with memory, their discussion and comparison of these results with known ones.

Appendix A

Let us consider the computation of the probabilities $p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}_{k-W}^{k+W} \,|\, \mathbf{x}_{k-B}^{k+B}\right)$ for some k. The segment $\mathbf{x}_{k-W}^{k+W}$ can be represented as a concatenation $\mathbf{x}_{k-W}^{k+W} = \left(\mathbf{x}_{k-W}^{k-B-1},\, \mathbf{x}_{k-B}^{k+B},\, \mathbf{x}_{k+B+1}^{k+W}\right)$. Therefore,

$$p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}_{k-W}^{k+W} \,|\, \mathbf{x}_{k-B}^{k+B}\right) = \sum_{\mathbf{x}_{k-W}^{k-B-1}} \sum_{\mathbf{x}_{k+B+1}^{k+W}} p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}_{k-W}^{k+W} \,|\, \mathbf{x}_{k-W}^{k+W}\right) \times$$
$$\times p_{\mathbf{x}}\left(\mathbf{x}_{k-W}^{k-B-1}\right) p_{\mathbf{x}}\left(\mathbf{x}_{k+B+1}^{k+W}\right).$$

For i. u. d. components of the vectors $\mathbf{x}_{k-W}^{k-B-1}$ and $\mathbf{x}_{k+B+1}^{k+W}$ the following equation is valid: $p_{\mathbf{x}}\left(\mathbf{x}_{k-W}^{k-B-1}\right) = p_{\mathbf{x}}\left(\mathbf{x}_{k+B+1}^{k+W}\right) = q_x^{-(W-B)}$. Hence,

$$p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}_{k-W}^{k+W} \,|\, \mathbf{x}_{k-B}^{k+B}\right) =$$
$$= q_x^{-2(W-B)} \sum_{\mathbf{x}_{k-W}^{k-B-1}} \sum_{\mathbf{x}_{k+B+1}^{k+W}} p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}_{k-W}^{k+W} \,|\, \mathbf{x}_{k-W}^{k+W}\right). \quad \text{(A1)}$$

Then, using the equation (8), we have

$$p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}_{k-W}^{k+W} \,|\, \mathbf{x}_{k-W}^{k+W}\right) = \mathbf{p}_s \left(\prod_{l=k-W}^{k+W} \mathbf{P}(y^{(l)} \,|\, x^{(l)}) \right) \mathbf{1}^{\mathrm{T}} =$$
$$= \mathbf{p}_s \left(\prod_{l=k-W}^{k-B-1} \mathbf{P}(y^{(l)} \,|\, x^{(l)}) \right) \left(\prod_{l=k-B}^{k+B} \mathbf{P}(y^{(l)} \,|\, x^{(l)}) \right) \times$$
$$\times \left(\prod_{l=k+B+1}^{k+W} \mathbf{P}(y^{(l)} \,|\, x^{(l)}) \right) \mathbf{1}^{\mathrm{T}}$$

and further

$$\sum_{\mathbf{x}_{k-W}^{k-B-1}} \sum_{\mathbf{x}_{k+B+1}^{k-W}} p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}_{k-W}^{k+W} \,|\, \mathbf{x}_{k-W}^{k+W}\right) =$$
$$= \mathbf{p}_s \left(\prod_{l=k-W}^{k-B-1} \mathbf{P}(y^{(l)}) \right) \left(\prod_{l=k-B}^{k+B} \mathbf{P}(y^{(l)} \,|\, x^{(l)}) \right) \times$$
$$\times \left(\prod_{l=k+B+1}^{k+W} \mathbf{P}(y^{(l)}) \right) \mathbf{1}^{\mathrm{T}},$$

where $\mathbf{P}(y) = \sum_{x \in X} \mathbf{P}(y \,|\, x)$. Then using equation (A1) and notation (14) we obtain the expression (13) for $k = W + 1$.

Appendix B

To calculate the sum $\sum_{\mathbf{x}} \psi(\mathbf{y};\, \mathbf{x})^{\lambda}$, let us represent the vector $\mathbf{x}$ as a sequence of $N(B)$ subvectors of length $2B + 1$ except perhaps the first and last segments,

$$\mathbf{x} = \left(\mathbf{x}_{k(0)-B}^{k(0)+B},\, \mathbf{x}_{k(1)-B}^{k(1)+B},\, \ldots,\right.$$
$$\left.\mathbf{x}_{k(n)-B}^{k(n)+B},\, \ldots,\, \mathbf{x}_{k(N(B)-1)-B}^{k(N(B)-1)+B}\right), \quad \text{(B1)}$$

where $k(n)$ is defined in equation (11). Then from the equation (12), it follows that

$$\psi(\mathbf{y};\, \mathbf{x})^{\lambda} = a_1^{\lambda} \left(\prod_{n=n_1}^{n_2} p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}_{k(n)-W}^{k(n)+W} \,|\, \mathbf{x}_{k(n)-B}^{k(n)+B}\right)^{\lambda} \right) b_1^{\lambda}, \quad \text{(B2)}$$

where n_1 is the smallest integer such that $k(n_1) - W \geq 1$, n_2 is the greatest integer such that $k(n_2) + W \leq N$, and a_1, b_1 are positive multipliers depending on initial and final segments of the vectors $\mathbf{x}$ and $\mathbf{y}$.

After a summation over i. u. d. and disjoint segments of the vector $\mathbf{x}$, we have

$$\sum_{\mathbf{x}} \psi(\mathbf{y};\, \mathbf{x})^{\lambda} = a(\lambda) \left(\prod_{n=n_1}^{n_2} D_1\left(\mathbf{y}_{k(n)-W}^{k(n)+W};\, \lambda\right) \right) b(\lambda), \quad \text{(B3)}$$

where

$$D_1(\mathbf{y};\, \lambda) = \sum_{\mathbf{x} \in X^{2B+1}} p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \,|\, \mathbf{x})^{\lambda}, \quad \mathbf{y} \in Y^{2W+1},$$

and $a(\lambda)$, $b(\lambda)$ are positive multipliers obtained after summing the quantities a_1^{λ} and b_1^{λ} over the initial and final segments of the vector $\mathbf{x}$.

Let us now derive an expression for the sum $\sum_{\mathbf{x}} p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \,|\, \mathbf{x}) \psi(\mathbf{y};\, \mathbf{x})^{-\lambda\rho}$. In this case, we also represent the vector $\mathbf{x}$ as a sequence of subvectors of length $2B + 1$ except the first and last ones, which may have a different length [see equation (B1)]. Using equation (B1) one can rewrite the expression

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \,|\, \mathbf{x}) =$$
$$= \mathbf{p}_s \mathbf{A}_1 \left(\prod_{n=n_1}^{n_2} \mathbf{P}_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}_{k(n)-B}^{k(n)+B} \,|\, \mathbf{x}_{k(n)-B}^{k(n)+B}\right) \right) \mathbf{B}_1 \mathbf{1}^{\mathrm{T}}, \quad \text{(B4)}$$

where, as before, n_1 is the least integer such that $k(n_1) - W \geq 1$, and n_2 is the greatest integer such that $k(n_2) + W \leq N$; $\mathbf{A}_1$ и $\mathbf{B}_1$ nonnegative matrix multipliers of dimension $|S| \times |S|$, corresponding to the first and last segments of the vector $\mathbf{x}$, and

$$\mathbf{P}_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \,|\, \mathbf{x}) = \prod_{l=1}^{2B+1} \mathbf{P}(\mathrm{y}^{(l)} \,|\, \mathrm{x}^{(l)}),\ \mathbf{y} \in Y^{2B+1},\ \mathbf{x} \in X^{2B+1}, \quad \text{(B5)}$$

where the matrices $\mathbf{P}(\mathsf{y}^{(l)}|\mathsf{x}^{(l)})$ in the right-hand part of (B5) is defined by equation (9).

Using expressions (B4) and (B2), we can write

$$p_{\mathsf{y}|\mathsf{x}}(\mathbf{y}\,|\,\mathbf{x})\psi(\mathbf{y};\ \mathbf{x})^{-\lambda\rho} =$$

$$= \mathbf{p}_s a_1^{-\lambda\rho}\mathbf{A}_1\left(\prod_{n=n_1}^{n_2}\mathbf{P}_{\mathsf{y}|\mathsf{x}}\left(\mathbf{y}_{k(n)-B}^{k(n)+B}\,|\,\mathbf{x}_{k(n)-B}^{k(n)+B}\right)\times\right.$$

$$\left.\times p_{\mathsf{y}|\mathsf{x}}\left(\mathbf{y}_{k(n)-W}^{k(n)+W}\,|\,\mathbf{x}_{k(n)-B}^{k(n)+B}\right)^{-\lambda\rho}\right)b_1^{-\lambda\rho}\mathbf{B}_1\mathbf{1}^{\mathrm{T}}.$$

Summing over i. u. d. and disjoint components of the vector $\mathbf{x}$, we get

$$\sum_{\mathbf{x}} p_{\mathsf{y}|\mathsf{x}}(\mathbf{y}\,|\,\mathbf{x})\psi(\mathbf{y};\ \mathbf{x})^{-\lambda\rho} =$$

$$= \mathbf{p}_s\mathbf{A}\left(\prod_{n=n_1}^{n_2}\sum_{\mathsf{x}\in X^{2B+1}}\mathbf{P}_{\mathsf{y}|\mathsf{x}}\left(\mathbf{y}_{k(n)-B}^{k(n)+B}\,|\,\mathsf{x}\right)\times\right.$$

$$\left.\times p_{\mathsf{y}|\mathsf{x}}\left(\mathbf{y}_{k(n)-W}^{k(n)+W}\,|\,\mathsf{x}\right)^{-\lambda\rho}\right)\mathbf{B}\mathbf{1}^{\mathrm{T}}, \qquad \text{(B6)}$$

where $\mathbf{A}$ и $\mathbf{B}$ are inessential matrix multipliers obtained after summation $a_1^{-\lambda\rho}\mathbf{A}_1$ and $b_1^{-\lambda\rho}\mathbf{B}_1$ over the initial and final segments of the vector $\mathbf{x}$ respectively. And finally it follows from (B6) that

$$\sum_{\mathbf{x}} p_{\mathsf{y}|\mathsf{x}}(\mathbf{y}\,|\,\mathbf{x})\psi(\mathbf{y};\ \mathbf{x})^{-\lambda\rho} =$$

$$= \mathbf{p}_s\mathbf{A}\left(\prod_{n=n_1}^{n_2}\mathbf{D}_2\left(\mathbf{y}_{k(n)-W}^{k(n)+W};\ \lambda\rho\right)\right)\mathbf{B}\mathbf{1}^{\mathrm{T}}, \qquad \text{(B7)}$$

where $\mathbf{D}_2(\mathsf{y};\ \lambda\rho)$ is matrix $|S|\times|S|$, defined as

$$\mathbf{D}_2(\mathsf{y};\ \lambda\rho) = \sum_{\mathsf{x}\in X^{2B+1}}\mathbf{P}_{\mathsf{y}|\mathsf{x}}\left(\mathsf{y}_{W-B+1}^{W+B+1}\,|\,\mathsf{x}\right)\times$$

$$\times p_{\mathsf{y}|\mathsf{x}}(\mathsf{y}\,|\,\mathsf{x})^{-\lambda\rho},\ \mathsf{y}\in Y^{2W+1}, \qquad \text{(B8)}$$

and matrix $\mathbf{P}_{\mathsf{y}|\mathsf{x}}(\mathsf{y}\,|\,\mathsf{x})$ is given by equation (B5).

Appendix C

Consider two adjacent terms in the product on the left-hand side of (21)

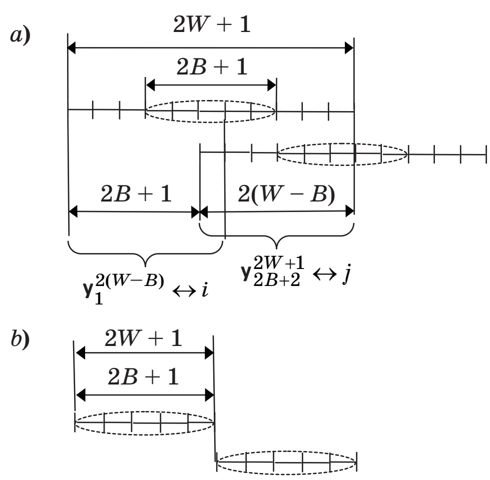
$$\prod_{n=n_1}^{n_2}\mathbf{D}\left(\mathbf{y}_{k(n)-W}^{k(n)+W};\ \lambda,\ \rho\right) = \mathbf{D}\left(\mathbf{y}_{k(n_1)-W}^{k(n_1)+W};\ \lambda,\ \rho\right)\times$$

$$\times\mathbf{D}\left(\mathbf{y}_{k(n_1+1)-W}^{k(n_1+1)+W};\ \lambda,\ \rho\right)\ldots\mathbf{D}\left(\mathbf{y}_{k(n_2)-W}^{k(n_2)+W};\ \lambda,\ \rho\right).$$

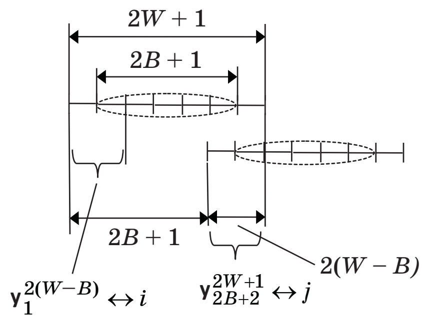
They depend on two adjacent blocks $\mathbf{y}_{k(n)-W}^{k(n)+W}$ and $\mathbf{y}_{k(n+1)-W}^{k(n+1)+W}$. The indices of the first block are $n(2B+1)+1-W$, ..., $n(2B+1)+1+W$, and indices of the second block are $(n+1)(2B+1)+1-W$, ..., $(n+1)(2B+1)+1+W$, i. e. the second set of indices is shifted right to $2B+1$ positions. An intersection of the position numbers exists on an interval of length $(2W+1)-(2B+1) = 2(W-B)$. If $W > B$, then this intersection is not empty; an illustration is shown in Fig. C1, *a*. Since $W \geq B$, the case $W = B$ is also possible. In this case there is no intersection and the summing in over vector $\mathbf{y}$ is reduced to summing separate factors over segments $\mathbf{y}_{k(n)-W}^{k(n)+W}$. This case is simpler to analyze and hence omitted. The illustration is given in Fig. C1, *b*.

Consider the case that $W > B$. Let us introduce indices $i, j = 1, 2, ..., q^{2(W-B)}$ and and establish a one-to-one correspondence between indices i, j and vectors $\mathsf{y}\in Y^{2W+1}$ as $i \leftrightarrow \mathsf{y}_1^{2(W-B)}$ и $j \leftrightarrow \mathsf{y}_{2B+2}^{2W+1}$. Then we consider two cases: 1) the vectors $\mathsf{y}_1^{2(W-B)}$ and y_{2B+2}^{2W+1} have common elements (see Fig. C1, *a*), and 2) the vectors $\mathsf{y}_1^{2(W-B)}$ and y_{2B+2}^{2W+1} have no common elements (see Fig. C2).

The first case takes place, if $2W+1 < 2\times\times 2(W-B)$, or if $W \geq 2B+1$, because W is integer.



■ ***Fig. C1.*** Layout of adjacent segments: *a* — intersection of segments, $W > B$; *b* — disjoint segments, $W = B$



■ ***Fig. C2.*** Intersection of adjacent segments, $2W < 4B+1$

The second case takes place if $W < 2B + 1$. Since in the second case two blocks $\mathbf{y}_1^{2(W-B)}$ and $\mathbf{y}_{2B+2}^{2W+1}$ have no common elements, it is possible to sum independently over elements with indices from $2(W - B) + 1$ to $2B + 1$. Therefore, we can define the $|S| \times |S|$ matrices $\mathbf{K}_{ij}(\lambda, \rho)$ as

$$\mathbf{K}_{ij}(\lambda, \rho) = \begin{cases} \mathbf{D}(\mathbf{y}; \lambda, \rho), W \geq 2B + 1; \\ \sum\limits_{\mathbf{y}_{2(W-B)+1}^{2B+1}} \mathbf{D}(\mathbf{y}; \lambda, \rho), W < 2B+1, \end{cases} \quad \text{(C1)}$$

where $\mathbf{y} \in Y^{2W+1}$, $i \leftrightarrow \mathbf{y}_1^{2(W-B)}$ and $j \leftrightarrow \mathbf{y}_{2B+2}^{2W+1}$.

Also define a block matrix $\mathbf{K}(\lambda, \rho) = [\mathbf{K}_{ij}(\lambda, \rho)]$ of order $|S|q^{2(W-B)}$. Then the sum over $\mathbf{y}$ on right-hand side of (21) can be written as

$$\sum_{\mathbf{y}} \left(\mathbf{U}\left(\mathbf{y}_1^{k(n_1-1)+W}\right) \prod_{n=n_1}^{n_2} \mathbf{D}\left(\mathbf{y}_{k(n)-W}^{k(n)+W}; \lambda, \rho\right) \times \right.$$
$$\left. \times \mathbf{V}\left(\mathbf{y}_{k(n_2+1)-W}^{N}\right)\right) =$$
$$= \mathbf{F} \prod_{n=n_1}^{n} \mathbf{K}(\lambda, \rho)\, \mathbf{G} = \mathbf{F}\, \mathbf{K}(\lambda, \rho)^{n_2-n_1+1}\, \mathbf{G}, \quad \text{(C2)}$$

where matrices $\mathbf{U}(\cdot)$ and $\mathbf{V}(\cdot)$ are defined by equations (19), $\mathbf{F}$ is a block matrix of size $|S| \times |S|q^{2(W-B)}$, and $\mathbf{G}$ is a block matrix of size $|S|\, q^{2(W-B)} \times |S|$.

The matrices $\mathbf{F}$ and $\mathbf{G}$ on the right-hand side of (C2) do not affect the asymptotic expression for the random coding exponent, so the description of their structure is not given due to space savings.

References

1. **Gallager R.** *Information Theory and Reliable Communication*. New York, John Wiley & Sons, 1968. 588 p.
2. **Ganti A., Lapidoth A., Telatar I. E.** Mismatched Decoding Revisited: General Alphabets, Channels with Memory, and the Wide-Band Limit. *IEEE Transactions on Information Theory*, 2000, vol. 46, no. 7, pp. 2315–2328.
3. **Lapidoth A., Narayan P.** Reliable Communication under Channel Uncertainty. *IEEE Transactions on Information Theory*, 1998, vol. 44, no. 7, pp. 2148–2177.
4. **Merhav N., Kaplan G., Lapidoth A., Shamai (Shitz) S.** On Information Rate for Mismatched Decoders. *IEEE Transactions on Information Theory*, 1994, vol. 40, no. 10, pp. 1953–1967.
5. **Kaplan G., Shamai (Shitz) S.** Information Rates and Error Exponents of Compound Channels with Application to Antipodal Signaling in a Fading Environment. *AEU International Journal of Electronics and Communications*, 1993, vol. 47, no. 4, pp. 228–230.
6. **Scarlett J., Martinez A., Guillén I Fàbregas A.** Mismatched Decoding: Error Exponents, Second-Order Rates and Saddlepoint Approximations. *IEEE Transactions on Information Theory*, 2014, vol. 60, no. 5, pp. 2647–2666.
7. **Forney Jr. G. D.** Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference. *IEEE Transactions on Information Theory*, 1972, vol. IT-18, no. 3, pp. 363–378.
8. **Kavcic A., Xiao Ma, Mitzenmacher M.** Binary Intersymbol Interference Channels: Gallager Codes, Density Evolution, and Code Performance Bounds. *IEEE Transactions on Information Theory*, 2003, vol. 47, no. 7, pp. 1636–1652.
9. **Arnold D. M., Loeliger H.-A., Vontobel P. O., Kavcic A., Zeng W.** Simulation-based Computation of Information Rates for Channels with Memory. *IEEE Transactions on Information Theory*, 2006, vol. 52, no. 8, pp. 3498–3508.
10. **Rusek F., Fertonani D.** Bounds on the Information Rate of Intersymbol Interference Channels based on Mismatched Receivers. *IEEE Transactions on Information Theory*, 2012, vol. 58, no. 3, pp. 1470–1482.
11. **Sasano H., Kasahara M., Namekawa T.** Evaluation of the Exponent Function E(R) for Channels with Intersymbol Interference. *Electronics and Communications in Japan*, 1982, vol. 65-A, no. 8, pp. 28–37.
12. **Biglieri E.** The Computational Cutoff Rate of Channel Having Memory. *IEEE Transactions on Information Theory*, 1981, vol. 27, pp. 352–357.
13. **Raghavan S. A., Wolf J. K., Milstein L. B.** On the Cutoff Rate of a Discrete Memoryless Channel with (d, k)-constrained Input Sequences. *IEEE Journal on Selected Areas in Communications*, 1992, vol. 10, no. 1, pp. 233–241.
14. **Raghavan S., Kaplan G.** Optimum Soft Decision Demodulation for ISI Channels. *IEEE Transactions on Communications*, 1993, vol. 41, no. 1, pp. 83–89.
15. **Shamai (Shitz) S., Raghavan S.** On the Generalized Symmetric Cutoff Rate for Finite-State Channels. *IEEE Transactions on Information Theory*, 1995, vol. 41, no. 9, pp. 1333–1346.
16. **Trofimov A., Chan Keong Sann.** Complexity-Performance Trade-off for Intersymbol Interference Channels — Random Coding Analysis. *IEEE Transactions on Magnetics*, 2010, vol. 46, no. 4, pp. 1077–1091.
17. **Egarmin V. K.** Lower and Upper Bounds on Decoding Error Probability for Discrete Channels. *Problemy peredachi informatsii*, 1969, vol. 5, no. 1, pp. 23–39 (In Russian).
18. **Poltyrev G. Sh., Shekhunova N. A.** Decoding in Discrete Channels with Memory. *Voprosy kibernetiki*, 1977, vol. 34, pp. 130–158 (In Russian).
19. **Gilbert E. N.** Capacity of a Burst-Noise Channel. *Bell System Technical Journal*, 1960, vol. 39, no. 5, pp. 1253–1265.

20. **Elliott E. O.** Estimates of Error Rates for Codes on Burst-Noise Channels. *Bell System Technical Journal*, 1963, vol. 42, no. 5, pp. 1977–1997.
21. **Bahl L. R., Cocke J., Jelinek F., Raviv J.** Optimal Decoding of Linear Codes for Minimum Symbol Error Rate. *IEEE Transactions on Information Theory*, 1974, vol. 20, no. 2, pp. 284–287.
22. **Benedetto S., Divsalar D., Montorsi G., Pollara F.** Soft-output Decoding Algorithms for Continuous Decoding of Parallel Concatenated Convolutional Codes. *Proc. IEEE Intern. Conf. on Communication (ICC 96)*, 1996, vol. 1, pp. 112–117.
23. **Gantmacher F. R.** *Teoriia matrits* [Theory of Matrices]. Moscow, Nauka Publ., 1988. 552 p. (In Russian).

УДК 621.391
doi:10.15217/issn1684-8853.2018.3.79

Граница случайного кодирования для каналов с памятью — декодирующая функция с частичным перекрытием

Часть 1: Вывод основного выражения

Трофимов А. Н.[a], канд. техн. наук, доцент, andrei.trofimov@vu.spb.ru
[a]Санкт-Петербургский государственный университет аэрокосмического приборостроения, Б. Морская ул., 67, Санкт-Петербург, 190000, РФ

Введение: задача вычисления экспоненты случайного кодирования во всем диапазоне скоростей кода для каналов с конечным числом состояний не решена полностью и остается актуальной. Представляется, что хорошее приближение к оптимальной экспоненте случайного кодирования может быть найдено при использовании несогласованной декодирующей функции. **Цель:** построить экспоненту случайного кодирования, близкую к оптимальной. **Результаты:** представлена новая граница случайного кодирования, применимая для широкого класса каналов, в том числе для тех, для которых полная экспонента случайного кодирования ранее не была построена. Вывод этой границы основан на использовании несогласованной декодирующей функции, которая зависит от двух параметров: длины сегмента выходной последовательности канала W и длины сегмента последовательности на входе канала B. Величины W и B в существенной степени влияют на значения экспоненты случайного кодирования и на сложность ее вычисления.

Ключевые слова — граница случайного кодирования, канал с конечным числом состояний, несогласованное декодирование, теорема Перрона — Фробениуса.

Цитирование: Trofimov A. N. Random Coding Bound for Channels With Memory – Decoding Function with Partial Overlapping. Part 1: Derivation of Main Expression // Информационно-управляющие системы. 2018. № 3. С. 79–88. doi:10.15217/issn1684-8853.2018.3.79
Citation: Trofimov A. N. Random Coding Bound for Channels With Memory – Decoding Function with Partial Overlapping. Part 1. Derivation of Main Expression. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2018, no. 3, pp. 79–88. doi:10.15217/issn1684-8853.2018.3.79