arXiv:1602.02648v2 [cs.IT] 2 May 2016

# Coding in the fork network in the framework of Kolmogorov complexity*

Andrei Romashchenko

October 14, 2018

Abstract

Many statements of the classic information theory (the theory of Shannon's entropy) have natural counterparts in the algorithmic information theory (in the framework of Kolmogorov complexity). In this paper we discuss one simple instance of the parallelism between Shannon's and Kolmogorov's theories: we prove in the setting of Kolmogorov complexity an algorithmic version of Wolf's characterization of admissible rates for the fork network.

## 1 Introduction: the Slepian–Wolf coding scheme

Many remarkable similarities between the probabilistic and algorithmic information theories were studied since the seminal paper of Kolmogorov [1]. In the present article we discuss one particular example of the parallelism between Shannon's and Kolmogorov's frameworks for information theory. We study coding schemes for simple multi-source networks (for so-called fork networks) and show that Wolf's theorem from the classic information theory can be naturally translated in the framework of Kolmogorov complexity.

First of all, we remind the classic Slepian–Wolf theorem and its algorithmic counterpart. A special case of this theorem (its non symmetric version) shows how an auxiliary source $\beta$ can be used for efficient compression of a source $\alpha$:

Theorem 1 ([2]) Let $(\alpha^i, \beta^i)$, $i = 1, 2, \dots$ be a sequence of i.i.d. pairs of random variables jointly distributed on some finite range $A \times B$. Then for every $\varepsilon > 0$ there exist mappings

$$\begin{array}{rcccl} f^n & : & A^n & \to & \{0,1\}^{l(n)}, \\ g^n & : & \{0,1\}^{l(n)} \times B^n & \to & A^n \end{array}$$

such that

$$\mathrm{Prob}[(\alpha^1, \dots, \alpha^n) = g^n(f^n(\alpha^1, \dots, \alpha^n), (\beta^1, \dots, \beta^n))] > 1 - \varepsilon,$$

and $\lim\limits_{n\to\infty} \frac{l(n)}{H(\alpha|\beta)} = 1$.

---

*This article is mostly a translation of the paper A. Romashchenko, A complexity version of the network coding problem. Information Processes (electronic journal) 5 (2005) No. 1, pp. 20–28. (in Russian).

Theorem 1 has a clear intuitive meaning. The sender encodes $n$ randomly chosen values of $\alpha^i$ in the most economic way. The encoding function is denoted $f^n : A^n \to \{0,1\}^{l(n)}$. The receiver has to reconstruct the values of $\alpha^i$ given some additional information (the values of $\beta^i$ correlated with $\alpha^i$). The decoding function is denoted $g^n$. The error probability must be bounded by some $\varepsilon > 0$. The aim is to minimize the length $l(n)$ of the transmitted message.

Shannon's coding theorem claims that we can achieve the error probability $\varepsilon$ with $l(n) = H(\alpha^1 \dots \alpha^n) + o(n)$, even if $\beta^i$ are not used in the decoding. Slepian and Wolf show that given $\beta^i$ we can reduce the length of the message to

$$H(\alpha^1 \dots \alpha^n | \beta^1 \dots \beta^n) + o(n).$$

What makes this theorem notrivial is that the values of $\beta^i$ are available only to the receiver and not the the sender, i.e., $(\beta^1, \dots, \beta^n)$ is an argument of $g^n$ but not of $f^n$.

In the framework of Kolmogorov complexity a counterpart of Theorem 1 was proven by An. A. Muchnik, see [6, 8]:

Theorem 2 For all strings $a, b$ there exists a string $a'$ such that

1. $K(a'|a) = O(\log n)$,
2. $K(a|a', b) = O(\log n)$,
3. $|a'| = K(a|b)$,

where $n = K(a) + K(b)$.

(See also a similar result [4, Theorem 3.11]; an analogous technique was used in [5, 7].) In this theorem the string $a'$ plays the role of a code that allows to "easily" reconstruct $a$ given $b$. Moreover, the code $a'$ can be "easily" computed from $a$. As usual in the theory of Kolmogorov complexity, the words "easily computed" mean that the corresponding conditional Kolmogorov complexity is bounded by $O(\log n)$.

Loosely speaking, Theorem 2 claims that among all almost shortest programs that translate $b$ to $a$ there is one whose complexity conditional on $a$ is negligibly small.

Theorem 1 is optimal in the sense that the ratio $\frac{l(n)}{H(\alpha|\beta)}$ cannot be made less than 1. Similarly, in Theorem 2 under conditions (1) and (2) we have $|a'| \geq K(a|b) - O(\log n)$.

The proofs of both Theorem 1 and Theorem 2 consist in constructing suitable "hash functions"; given the first source of information we compute its fingerprint (a hash value), and then recover the initial value given this fingerprint and another (auxiliary) source of information. However, the technical implementations of this idea in the proofs of Theorem 1 and Theorem 2 are pretty different.

## 2 Fork networks

Theorem 1 can be generalized for a larger class of communication networks. Let us define the admissible rates for the "fork networks".

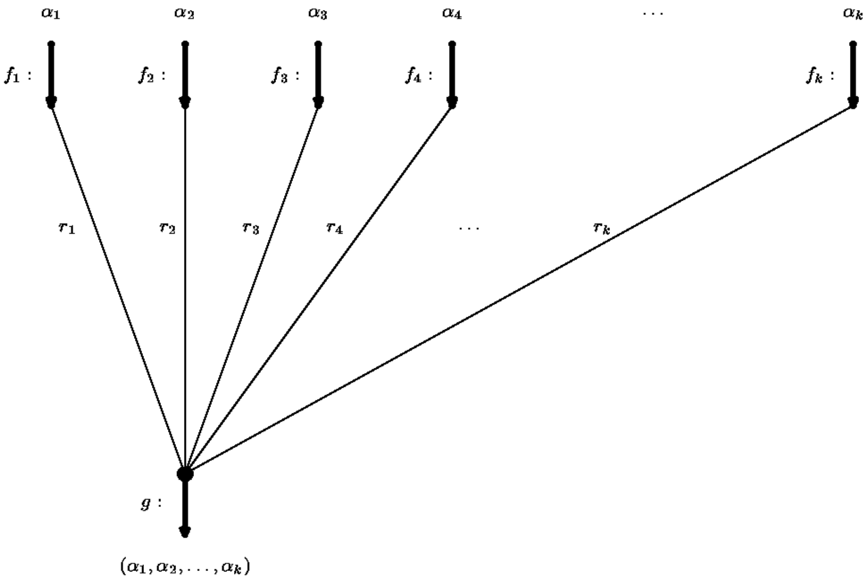

Fig. 1. The fork network with $n$ sources

Definition 1 Let a $k$-dimensional random variable $(\alpha_1, \ldots, \alpha_k)$ be distributed on a finite set $A_1 \times \ldots \times A_k$. Denote by $(\alpha_1^i, \ldots, \alpha_k^i)$ $(i = 1, 2, \ldots)$ a sequence of i.i.d. $k$-dimensional random variables, and let each of them be distributed as $(\alpha_1, \ldots, \alpha_k)$. A tuple of $k$ reals $(r_1, \ldots, r_k)$ is called $\varepsilon$-admissible for the fork network with sources $\alpha_1, \ldots, \alpha_k$, if for every $\delta > 0$ and large enough $n$ there exist functions $f_1^n, \ldots, f_k^n$, $g^n$,

$$\begin{array}{ccccc} f_j^n & : & (A_j)^n & \to & \{0,1\}^{l_j(n)}, \\ g^n & : & \{0,1\}^{l_1(n)+\ldots+l_k(n)} & \to & A_1 \times \ldots \times A_k \end{array}$$

such that $l_j(n) \le (r_j + \delta)n$, and

$$\mathrm{Prob}[g^n(f_1^n(\bar\alpha_1), \ldots, f_k^n(\bar\alpha_k)) = (\bar\alpha_1, \ldots, \bar\alpha_k)] > 1 - \varepsilon,$$

where $\bar\alpha_j$ denotes the $n$-tuple $(\alpha_j^1, \ldots, \alpha_j^n)$ for each $j$.

This definition corresponds to the information transmission in the network shown in Fig. 1. We are given $k$ correlated sources of information, and their distribution is specified by the varables $\alpha_1, \ldots, \alpha_k$. The sources are encoded independently by the block codes $f_j^n$. The definition specifies the lengths of the encoded messages: the senders spend on average $r_j$ bits per each letter of the source $\alpha_j$. The receiver recovers the values of all $k$ sources with the decoding function $g^n$, and the probability of error must be less that $\varepsilon$.

The set of all $\varepsilon$-admissible rates can be characterized in terms of entropies involving the random variables $\alpha_j$.

Notation: Let $(\alpha_1, \ldots, \alpha_k)$ be a tuple of jointly distributed random variables. In what follows we denote by $\alpha_W$ the tuple of random variables $\alpha_j$ for all $j \in W$, and by $\alpha_{\neg W}$ the tuple of random variables $\alpha_j$ for $j \notin W$. For example, if $k = 5$

and $W = \{1, 2, 5\}$, then $\alpha_W$ denotes $(\alpha_1, \alpha_2, \alpha_5)$ and $\alpha_{\neg W}$ denotes $(\alpha_3, \alpha_4)$. For $W = \emptyset$ we suppose that $\alpha_W$ is a constant (a random variable with zero entropy). In particular, if $W = \{1, \ldots, k\}$ and $\neg W = \emptyset$, then $H(\alpha_W|\alpha_{\neg W}) = H(\alpha_W)$.

Now we can formulate Wolf's theorem that characterizes the set of admissible rates for the fork networks.

Theorem 3 (J.K. Wolf [3]; see also [10]) For every $k$-tuple of jointly distributed random variables $(\alpha_1, \ldots, \alpha_k)$ and for every $\varepsilon > 0$,

(i) [the necessary condition] for every $\varepsilon$-admissible tuple of reals $(r_1, \ldots, r_k)$ and for all $W \subset \{1, \ldots, k\}$ it holds

$$\sum_{j \in W} r_j \geq H(\alpha_W|\alpha_{\neg W}),$$

(ii) [the sufficient condition] if for each set $W \subset \{1, \ldots, k\}$ it holds

$$\sum_{j \in W} r_j > H(\alpha_W|\alpha_{\neg W}),$$

then the tuple of reals $(r_1, \ldots, r_k)$ is $\varepsilon$-admissible for the fork network.

Example 1 For $k = 2$ Theorem 3 implies that a pair $(r_1, r_2)$ is $\varepsilon$-admissible for the network with sources $(\alpha_1, \alpha_2)$, only if

$$\begin{array}{rcl} r_1 & \geq & H(\alpha_1|\alpha_2), \\ r_2 & \geq & H(\alpha_2|\alpha_1), \\ r_1 + r_2 & \geq & H(\alpha_1, \alpha_2). \end{array}$$

Similarly, the conditions

$$\begin{array}{rcl} r_1 & > & H(\alpha_1|\alpha_2), \\ r_2 & > & H(\alpha_2|\alpha_1), \\ r_1 + r_2 & > & H(\alpha_1, \alpha_2). \end{array}$$

are enough to guarantee that a pair $(r_1, r_2)$ is $\varepsilon$-admissible. This special case of Theorem 3 is the general statement of the Slepian–Wolf theorem, [2]. If $r_2 = \infty$, then the only remaining condition is $r_1 \geq H(\alpha_1\ |\alpha_2)$, so we get the statement of Theorem 1 as a special case.

In what follows we prove a counterpart of Theorem 3 for Kolmogorov complexity. Technically, we give a criterion for the following property of a tuple of binary strings $a_1, \ldots, a_n$ (which is a counterpart of the admissibility property from Definition 1 adapted to the Kolmogorov's theory):

> For a tuple of strings $a_1, \ldots, a_k$ and a tuple of integers $r_1, \ldots, r_k$
> there exist strings $a'_1, \ldots, a'_k$, such that
> (1) $|a'_j| \leq r_j,\ j = 1, \ldots, k,$
> (2) $K(a'_j|a_j) \leq C_0 \log n,\ j = 1, \ldots, k,$ (∗)
> (3) $K(a_1, \ldots, a_k|a'_1, \ldots, a'_k) < C_0 \log n,$
> *where* $n = |a_1| + \ldots + |a_k|$.

Theorem 4 (main result)
[The necessary condition]: For all integer $k > 0$ and $C_0 > 0$ there exists a constant $C_1$ such that for all strings $a_1, \ldots, a_k$ (with total length $n = |a_1| + \ldots + |a_k|$) and all integers $r_1, \ldots, r_k$, property (*) holds only if for every non empty set $W \subset \{1, \ldots, k\}$

$$\sum_{j \in W} r_j \geq K(a_W | a_{\neg W}) - C_1 \log n.$$

[The sufficient condition]: For all integer $k > 0$ and $C_2 > 0$ there exists a constant $C_0$ such that for all strings $a_1, \ldots, a_k$ and integers $r_1, \ldots, r_k$ property (*) holds if for every nonempty set $W \subset \{1, \ldots, k\}$

$$\sum_{j \in W} r_j \geq K(a_W | a_{\neg W}) + C_2 \log n.$$

Notation: In this theorem we use the notation $a_W$, which stands for a tuple of all strings $a_j$ for $j \in W$. Similarly, $a_{\neg W}$ stands for a tuple of all strings $a_j$ for $j \notin W$. For the empty $W$ we denote by $a_W$ the empty word. In particular, if $W = \{1, \ldots, k\}$ and $\neg W = \emptyset$, then $K(a_W | a_{\neg W}) = K(a_W) + O(1)$.

Example 2 For $k = 2$ this theorem gives the necessary and sufficient conditions

$$\begin{array}{rcl} r_1 & \geq & K(a_1|a_2) + O(\log n), \\ r_2 & \geq & K(a_2|a_1) + O(\log n), \\ r_1 + r_2 & \geq & K(a_1, a_2) + O(\log n). \end{array}$$

For $k = 2, 3$ Theorem 4 was proven in [9]. In the present paper we prove this theorem for all integer $k > 0$.

The standard proof of Theorem 3 (see [10]) cannot be translated in the language of Kolmogorov complexity. The crucial point is that the proof in [10] employs the principle of time sharing, which does not apply in the framework of Kolmogorov complexity. We prove Theorem 4 using the following version of Muchnik's theorem (which is somewhat stronger than Theorem 2):

Theorem 5 ([8]) For every integer $k > 0$ there exists a number $C = C(k)$ with the following property. Let $x_0, x_1, \ldots, x_k$ be binary strings, $n = |x_1| + \ldots + |x_k|$, and $r$ be a number less than $|x_0|$. Then there exists a string $y$ such that

(1) $|y| = r$,

(2) $K(y|x_0) \leq C \log n$,

and $K(y|x_j) \geq \min\{K(x_0|x_j), r\} - C \log n$ for every $j = 1, \ldots, k$.

Informally, Theorem 5 claims that we can extract (with only a logarithmic advice) from a string $x_0$ a fingerprint of length $r$ that looks maximally "random" given each of the strings $x_j$ as a condition.

Remark 1: Since $K(y|x_0) = O(\log n)$, for every $j$ we have

$$K(y|x_j) \leq K(x_0|x_j) + O(\log n).$$

Remark 2: If $r > K(x_0|x_j)$, then $x_0$ can be completely reconstructed given $y$ and $x_j$ (and some logarithmic advice), i.e.,

$$K(x_0|y, x_j) = O(\log n).$$

Indeed, in this case we have $K(y|x_j) = K(x_0|x_j) + O(\log n)$, so

$$\begin{array}{rcl} K(x_0|y, x_j) & = & K(x_0, y|x_j) - K(y|x_j) + O(\log n) \\ & = & K(x_0|x_j) - K(y|x_j) + O(\log n) \\ & = & O(\log n). \end{array}$$

Remark 3: Theorem 5 implies Theorem 2. Indeed, let us apply Theorem 5 for $k = 1$, with $x_0 = a$, $x_1 = b$, $r_1 = K(a|b)$. We obtain a string $y$ such that

1. $|y| = r_1$,
2. $K(y|a) = O(\log n)$,
3. $K(y|b) \geq K(a|b) - O(\log n)$.

These conditions imply that $K(a|y, b) = O(\log n)$. Thus, we may let $a' = y$. $\square$

## 3 Proof of theorem 4

For the sake of brevity, we use the following asymptotic notation:

$$\begin{array}{rcl} F(n) \leq_n G(n) & \leftrightharpoons & F(n) \leq G(n) + O(\log n), \\ F(n) \geq_n G(n) & \leftrightharpoons & G(n) \leq F(n) + O(\log n), \\ F(n) =_n G(n) & \leftrightharpoons & F(n) = G(n) + O(\log n). \end{array}$$

Proof of the necessity condition: Let $W \subset \{1, \ldots, k\}$ be any set of indices, and $\neg W = \{1, \ldots, k\} \setminus W$. By the condition of the theorem, all strings $a_i$ have logarithmic complexity conditional on the tuple $\langle a'_1, \ldots, a'_k \rangle$. It follows that $K(a_W|a'_W, a_{\neg W}) =_n 0$. Hence, complexity of the tuple $a'_W$ cannot be less than the conditional complexity $K(a_W|a_{\neg W})$. On the other hand, Kolmogorov complexity of $a'_W$ is not greater than $r_i$ for each $i \in W$, and we are done. More formally this argument can be presented as a chain of inequalities:

$$\begin{array}{rcl} K(a_W|a_{\neg W}) & \leq_n & K(a'_W) + K(a_W|a'_W, a_{\neg W}) \leq_n \\ & \leq_n & K(a'_W) + K(a_W|a'_W, a'_{\neg W}) \leq_n \\ & \leq_n & K(a'_W) \leq_n \sum\limits_{j \in W} r_j. \end{array}$$

Proof of the sufficiency condition: We prove the theorem by induction on $k$. To make the inductive step work, we need to reformulate the theorem and make it somewhat stronger:

Inductive claim: For every integer $k > 0$ and for all $C_2 > 0$ there exists a number $C_0$ with the following property. Let $a_1, \ldots, a_k, b$ be binary strings and $r_1, \ldots, r_k$ be integers, denote $n = |a_1| + \ldots + |a_k| + |b|$. Assume that for every nonempty $W \subset \{1, \ldots, k\}$ it holds $\sum\limits_{j \in W} r_j \geq K(a_W|a_{\neg W}, b) + C_2 \log n$. Then it follows that there exist binary strings $a'_1, \ldots, a'_k$ such that

1. $|a'_j| \leq r_j$, $j = 1, \ldots, k$,
2. $K(a'_j|a_j) \leq C_0 \log n$, $j = 1, \ldots, k$,
3. $K(a_1, \ldots, a_k | a'_1, \ldots, a'_k, b) < C_0 \log n$.

The difference between this claim and (*) is a new parameter $b$.

For $k = 1$ the inductive claim follows immediately from Theorem 2. Let us perform the inductive step. Fix some binary strings $a_1, \ldots, a_k, a_{k+1}, b$. From Theorem 5 it follows that there exists a string $a'_{k+1}$ such that

- $|a'_{k+1}| \leq r_{k+1}$,
- $K(a'_{k+1}|a_{k+1}) \leq C \log n$,

and for every nonempty $W \subset \{1, \ldots, k\}$

$$(**) \qquad K(a'_{k+1}|a_W, b) \geq \min\{K(a_{k+1}|a_W, b), r_{k+1}\} - C \log n$$

(the value of $C$ depends on $k$ but not on $a_i$ and $b$).

We are going to use the inductive hypothesis with the tuple of strings

$$a_1, \ldots, a_k, b' := \langle a'_{k+1}, b \rangle$$

and the tuple of integers $r_1, \ldots r_k$. To this end we should verify that the inductive claim is applicable to these strings, i.e., we need to prove the following lemma.

Lemma 1 There exists a $C'_2 = C'_2(k, C_2)$ such that for every non-empty $V \subset \{1, \ldots, k\}$ and its complement $\neg V = \{1, \ldots, k\} \setminus V$ it holds

$$\sum_{j \in V} r_j \geq K(a_V | a_{\neg V}, b') + C'_2 \log n.$$

Proof: We consider separately two cases.

Case 1: Assume that $r_{k+1} \geq K(a_{k+1}|a_{\neg V}, b)$. From Remark 2 we know that

$$K(a_{k+1}|a'_{k+1}, a_{\neg V}, b) =_n 0.$$

Hence,

$$K(a_V|a_{\neg V}, a'_{k+1}, b) \leq_n K(a_V|a_{\neg V}, a_{k+1}, b) \leq_n \sum_{j \in V} r_j$$

(the last inequality is a part of the condition of the inductive claim).

Case 2: Now we assume that $r_{k+1} < K(a_{k+1}|a_{\neg V}, b)$. We are given the condition

$$r_{k+1} + \sum_{j \in V} r_j \geq_n K(a_{k+1}, a_V | a_{\neg V}, b).$$

Using the Kolmogorov–Levin theorem, we can reformulate this inequality as

$$(***) \qquad \begin{aligned} r_{k+1} + \textstyle\sum_{j \in V} r_j \quad \geq_n \quad & K(a'_{k+1}|a_{\neg V}, b) + K(a_{k+1}|a'_{k+1}, a_{\neg V}, b) \\ & + K(a_V, |a_{k+1}, a_{\neg V}, b). \end{aligned}$$

Then we get from (**)

$$K(a'_{k+1}|a_{\neg V}, b) =_n r_{k+1}.$$

So (***) rewrites to

$$r_{k+1} + \sum_{j \in V} r_j \geq_n r_{k+1} + K(a_{k+1}|a'_{k+1}, a_{\neg V}, b) + K(a_V, |a_{k+1}, a_{\neg V}, b),$$

which implies

$$\sum_{j \in V} r_j \geq_n K(a_V, |a_{k+1}, a_{\neg V}, b),$$

and we are done. □

With Lemma 1 we can apply the inductive hypothesis. We obtain some strings $a'_1, \dots, a'_k$ such that

1. $|a'_j| \leq r_j$, $j = 1, \dots, k$,
2. $K(a'_j|a_j) \leq C'_0 \log n$, $j = 1, \dots, k$,
3. $K(a_1, \dots, a_k|a'_1, \dots, a'_k, a'_{k+1}, b) \leq C'_0 \log n$

for some $C'_0 = C'_0(k, C_2)$. It remains to show that

$$K(a_{k+1}|a'_1, \dots, a'_k, a'_{k+1}, b) \leq C_0 \log n$$

for some $C_0 \geq C'_0$ (which may depend on $k$ and $C_2$). To this end, it is enough to prove $K(a_{k+1}|a_1, \dots, a_k, a'_{k+1}, b) =_n 0$. For the sake of brevity, we use the asymptotic notation:

$$\begin{aligned}
&K(a_{k+1}|a_1, \dots, a_k, a'_{k+1}, b) =_n \\
&=_n K(a_{k+1}, a'_{k+1}|a_1, \dots, a_k, b) - K(a'_{k+1}|a_1, \dots, a_k, b) \\
&=_n K(a_{k+1}|a_1, \dots, a_k, b) - K(a'_{k+1}|a_1, \dots, a_k, b) \\
&=_n K(a_{k+1}|a_1, \dots, a_k, b) - \min\{K(a_{k+1}|a_1, \dots, a_k, b), r_{k+1}\} =_n 0.
\end{aligned}$$

□

## 4 Conclusion

It seems natural to ask whether a version of Theorem 4 holds for resource bounded versions of Kolmgorov complexity, e.g., for programs running in polynomial time or polynomial space. Recently M. Zimand proved a variant of Thereom 4 where the encoding procedures $a'_j = Enc_j(a_j)$, $j = 1, \dots, k$ can be performed by probabilistic polynomial time algorithms, see [11]. It seems unlikely that the optimal lengths of "codewords" $a'_j$ and polynomial time encoding could be combined also with polynomial time decoding $(a'_1 \dots a'_k) \mapsto (a_1 \dots a_k)$.

Acknowledgments. The author is grateful to Marius Zimand, who pointed out an error in the first version of this paper.

# References

[1] Kolmogorov A.N., Three approaches to the quantitative definition of information, Problems of information transmission, 1(1), 1–7, 1965.

[2] Slepian D., Wolf J.K., Noiseless coding of correlated information sources. IEEE Transactions on Information Theory, 19, 471–480, 1973.

[3] Wolf J.K., Data reduction for multiple correlated sources. In: Proc. of the Fifth Colloquium on Microwave Communication. Budapest, 287–295, 1974.

[4] Bennett C.H., Gács P., Li M., Vitányi P. M., Zurek W. H. Information distance. IEEE Transactions on Information Theory, 44(4), 1998,1407–1423.

[5] Fortnow L., Laplante S. Nearly optimal language compression using extractors. In Proc. STACS, 1998. 84–93.

[6] Muchnik An.A., Semenov A.L. Multi-conditional Descriptions and codes in Kolmogorov complexity. Electronic Collocuium on Computational Complexity (ECCC), 7(15), 2000.

[7] Buhrman H., Fortnow L., Laplante S. Resource-bounded Kolmogorov complexity revisited. SIAM Journal on Computing. 31(3), 887–905, 2001.

[8] Muchnik, A.A. Conditional complexity and codes. Theoretical Computer Science, 271(1), 97–109, 2002.

[9] Izmailova A.A., Information transmission in the fork network with bounded channel capacities. Master thesis. Moscow State University, 2004. In Russian. (Измайлова А.А., Передача сообщений в вилочной сети с ограниченными пропускными способностями каналов. Дипломная работа. Москва, МГУ им. Ломоносова. 2004.)

[10] Csiszar I., Körner J., Information theory: coding theorems for discrete memoryless systems, Cambridge University Press, 2011.

[11] Zimand, M. Kolmogorov complexity version of Slepian-Wolf coding. arXiv:1511.03602 (2015).