

# Modelling Biomolecular Structures in Categorical Systems Theory

Georgy Tolokonnikov<sup>1,\*</sup>

<sup>1</sup>Federal Scientific Agroengineering Center VIM, RU-109428, Moscow, Russia

**Abstract.** In the systemic movement there exist numerous approaches to systems, the most profound of which is the theory of functional systems by Anokhin, which remained largely intuitive science until his pioneering works. The basic principles of functional systems are formalized with the help of the convolutional polycategories in the form of categorical systems theory, which embraced the main systemic approaches, including the traditional mathematical theory of systems. Convolutional polycategories can be built using categorical splices that directly model the external and internal parts of systems. For an algebraic biology using the categorical theory of systems in relation to systemic constructions, the main task of which is to predict the properties of organisms from the genome using strict algebraic methods, new categorical methods are proposed that are widely used in categorical systems theory. These methods are based on the theory of categorical splices, with the help of which the behaviour of quantum-mechanical particles is modelled, in particular, within the framework of the proposed representation of molecules, including RNA and DNA, as categorical systems. Thus, new algebraic and categorical methods (associative algebras with identities, PROP, categorical splices) are involved in the analysis of the genome. The paper presents new results on these matters.

## 1 Introduction

Systems theory was born in biology, there are numerous systems approaches, which include the direction of mathematical systems theory, with the concept of a system given by M. Mesarovich, V.M. Matrosov and other scientists (see [1], review in [2]), ergatic systems "man-machine" of engineering psychology Lomov [3] and others. There are approaches initiated in biology and medicine by L. Bertalanffy, P.K. Anokhin, G.N. Kryzhanovskiy [4] and other researchers. The most profound of the approaches is the theory of functional systems by P.K. Anokhin [5], which remained an essentially intuitive science until the author's works (see [6-9] and references therein). The basic principles of functional systems are formalized with the help of (generalizing polycategories of Szabo (1975)) convolutional polycategories in the form of categorical systems theory, which embraced the main system approaches, including the traditional mathematical theory of systems. Convolutional polycategories can be built using categorical splices that directly model the external and internal parts of systems. For the algebraic biology using the categorical theory of systems [10,11], the main task of which is to predict the properties of organisms from the genome using strict algebraic methods, new categorical methods used in categorical systems theory are proposed. These methods are based on the theory of categorical splices, with the help of which the behaviour of biomolecules, including RNA and DNA, as categorical systems is modelled. Thus, new

algebraic and categorical methods are involved in genome analysis. From the point of view of the systemic paradigm, all concepts must be reduced to systems and their aggregates, just as it is implemented in the set-theoretic paradigm according to N. Burbaki, when all concepts are reduced to sets and subsets. The functional systems of P.K. Anokhin do not describe biomolecules as systems. This paper proposes a strictly mathematical description of biomolecules as categorical systems. Unlike the genetic code, the properties of DNA and RNA molecules in algebraic biology have not yet been rigorously studied mathematically. Thus, from the point of view of applications in algebraic biology, the work is a definite step in the study of the properties of DNA and RNA molecules by algebraic methods. The work is also devoted to the study of the properties of categorical splices, in particular, duality. Duality in category theory is known to be a fundamental property. In the theory of categorical splices, there is not one, but several types of duality, including the traditional duality (generated by a change in the direction of the arrows). In particular, the stated duality theorem turns out to be valid, generalizing to gluings and new types of duality in them the fundamental theorem of Hatcher [12] substantiating the application of the duality principle in category theory [13]. The listed results are new and original.

## 2 Categorical splices

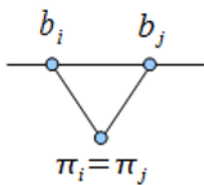
Let's introduce a four-sorted first-order logic language

\* Corresponding author: [admicit@mail.ru](mailto:admicit@mail.ru)

with variables  $p, p', \bar{p}, \bar{p}'$  responsible for splices, convolutions, inner splices and inner convolutions, respectively. Convolutions are represented using: specified letters; functional symbols of co-regions  $C_i, C'_i, \bar{C}_i, \bar{C}'_i$ ; names  $\pi_i, \pi'_i, \bar{\pi}_i, \bar{\pi}'_i$ ; values  $ev_\alpha, ev'_\alpha, \bar{ev}_\alpha, \bar{ev}'_\alpha$ ; predicates  $M_b, M_r, M_t, M_l$ ; using a graphic representation of formulas similar to that introduced by Hetcher [12] when constructing the first-order language of category theory, for example

$$x_i \stackrel{\text{def}}{=} \pi_i(p), b_i \stackrel{\text{def}}{=} C_i(p), x_i = \pi_i(p) \wedge b_i = C_i(p) \wedge$$

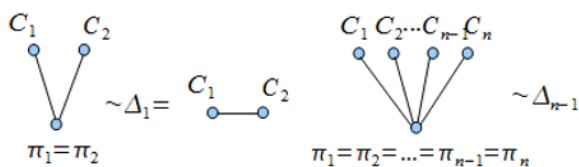
$$\wedge x_j = \pi_j(p) \wedge b_j = C_j(p) \wedge x_i = x_j$$



The properties of predicates and other letters of the signature are determined by axioms. Formulas  $x_i = \pi_i(p), b_i = C_i(p)$  will be denoted by  $\pi_i, C_i$ . The convolution  $S$  is defined by the ratio, the symbol  $ev$  has the opposite and the opposite convolution is determined by the  $S = M_b ev M_l$ . The complete splices can be represented as a conjunction formula  $a \wedge b' \wedge \bar{c} \wedge \bar{d}'$ , divided into four subformulas in accordance with the sort of variables, located in the 2x2 quadrants of the table

$a$	$b$
$c$	$d$

Equivalences between co-regions, between names that correspond to the equal sign, are depicted by connecting the right and left letters inequality with dashed lines. The predicates  $M_t, M_r, M_l, M_b$  are determined by the following formula transformations.  $M_b$  moves to the lower quadrants from the upper quadrants only those pairs  $C_k \pi_k$  that have co-area connections,  $M_t$  moves to the upper quadrants from the lower quadrants only those pairs  $C_k \pi_k$  that have co-area connections,  $M_l$  moves vertically and inside a pair from right quadrants to left ones,  $M_r$  moves connected pairs from left quadrants to rights.



The function symbol  $ev$  translates  $\pi_i, C_i$  to  $ev(\pi_i), ev(C_i)$  and places the = sign in the appropriate places. If you do not impose restrictions on the choice of splices and convolutions, then we speak of free general splices. Special categorical splices are constructed from a fixed set of convolutions and a fixed set of splices, while the original splices are called generators, and all the rest are obtained by all possible applications of these convolutions to splices. The splices model the set of affine simplexes (Fig. 1). For clarity, consider the following formula as an example.

$$C_k \wedge \pi_k \wedge C_i \wedge \pi_i \wedge C_m \wedge \pi_m \wedge C_n \wedge \pi_n \wedge C'_j \wedge \pi'_j$$

$$\wedge C'_w \wedge \pi'_w \wedge C'_v \wedge \pi'_v \wedge \pi_k$$

$$= \pi_i \wedge \pi_m = \pi_n \wedge \pi'_j = \pi'_v \wedge C_i$$

$$= C'_j \wedge C_m = C_v$$

We will omit the conjunction sign, replace equalities with a line connecting the right and left sides of the equalities. The table for the formula is

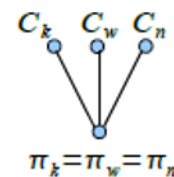
$C_k \pi_k C_i \pi_i C_m \pi_m C_n \pi_n$	$C'_j \pi'_j C'_w \pi'_w C'_v \pi'_v$
$\emptyset$	$\emptyset$

For convolution, we have a table

$C_k \pi_k C_n \pi_n$	$C'_w \pi'_w$
$C_i \pi_i C_m \pi_m$	$C'_j \pi'_j C'_v \pi'_v$

Applying,  $M_l$  we get

$C_k \pi_k C_n \pi_n$	$C'_w \pi'_w$	$\emptyset$
$C_i \pi_i C_m \pi_m$	$C'_j \pi'_j C'_v \pi'_v$	$\emptyset$



A complete restoration of the original splice is possible as follows. We apply a transfer  $M_r$  to the final table in relation to  $C_w \pi_w \bar{C}_j \bar{\pi}_j \bar{C}_v \bar{\pi}_v$ , we will get a table that coincides with the previous final table. Then we

**Fig. 1.** Modeling simplexes by splices

apply  $M_t$  to it, which gives the original table and, thereby, the original splice.

For a given splice, it is possible to construct a new  $\Delta'$ -dual splice by transferring the contents in the right parts of the splice table to the left and at the same time vice versa (splice-convolution duality). Similarly, you can construct a new  $\bar{\Delta}$ -dual splice by transferring the content in the upper parts of the splice table to the lower parts and at the same time vice versa (duality "outside-inside"). If the names in the splice formula are replaced by the corresponding co-domains and simultaneously vice versa, then again the splice is obtained,  $\Delta_\pi^c$ -dual of the original splice, in which the equalities of the co-domains turn into the equal names and vice versa (the duality "name-co-domain").

### 3 Higher categorical k-splices and general systems

Let there be k-splices with sets of variables of four sorts, functional symbols, and predicates. It is not difficult to construct a unified formal theory of higher categorical k-splices. Let's go to a table with 2k columns, in the first k columns, in order, insert  $a_i, c_i, i = 1, \dots, k$ , in the second k columns, insert  $b_i, d_i, i = 1, \dots, k$ , as indicated below

$a_1$	...	$a_k$	$b_1$	...	$b_k$
$c_1$	...	$c_k$	$d_1$	...	$d_k$

The above types of duality ( $\Delta', \bar{\Delta}, \Delta_\pi^c$ ) are naturally carried over to the theory of k-splices. Here a form of duality  $\hat{\Delta}$ , already known from category theory, arises, associated with the replacement of the direction of the arrows. It takes the form of interchange by places in the formula of a general k-splices or any other formula of the first-order language of the formal theory of k-splices of one type of splices to another. There can be several such replacements in k-splices.

Let formulas be given in the language of categorical splices. The transformations of formulas, respectively,  $\Delta', \bar{\Delta}, \Delta_\pi^c, \hat{\Delta}$ -types defined above are called their transformation  $\Delta', \bar{\Delta}, \Delta_\pi^c, \hat{\Delta}$ -duality. The fundamental role played by duality in the categories is well known. It is logically substantiated in [12]. For the listed dualities in the theory of categorical k-splices, an analogue of the theorem proved in [12] holds.

**Theorem.** Let in the formal theory of categorical k-splices be given a provable well-constructed formula A. Then dual - formulas are also provable in this theory.

We considered systems in convolutional polycategories and in 1-splices [6-9], having higher k-splices, we can naturally define k-systems or higher systems on them. At the same time, k-systems are used

for modeling neurons with connections other than spike connections.

In the last decade, research has been expanding on the non-spike connections of neurons (see, for example, [14] and references therein). Traditional neural networks in physiological terminology realize only the level of reflexes, which, as taught by modern physiology and the theory of functional systems, is not enough for the implementation of cognitive functions. A possible way out is to complicate the models of neurons [15] and to model new types of connections between neurons [14]. More formally, we can say that the main idea we propose with regard to neural networks is the transition from associative compositional convolutional polycategories with corona-type convolutions, which simulate traditional artificial neural networks of arbitrary topology [7,9], to networks based on higher categorical splices, modeling also other types of neuronal connections besides the usual communication carried out by spikes.

### 4 Higher Convolutional polycategories and their gluing representation

In convolutional polycategories, poly-arrows have inputs and outputs. Their modeling is possible using 2-splices. Let's look at an example of traditional categories. Two arrows, when the beginning of one arrow coincides with the end of the other arrow, and the convolution (corresponding to the composition of the arrows), corresponds to the formula (between the letters the conjunction sign is omitted)

$$C_1^{(1)} \pi_1^{(1)} C_2^{(1)} \pi_2^{(1)} C_3^{(1)'} \pi_3^{(1)'} C_1^{(2)} \pi_1^{(2)} C_2^{(2)} \pi_2^{(2)} C_3^{(2)'} \pi_3^{(2)'}$$

$$(\pi_1^{(1)} = \pi_1^{(2)}) (\pi_2^{(1)} = \pi_2^{(2)}) (\pi_3^{(1)'} = \pi_3^{(2)'})$$

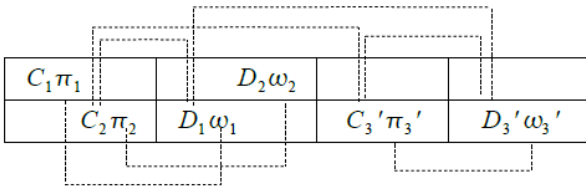
$$(C_1^{(2)} = C_2^{(1)}) (C_2^{(1)} = C_3^{(1)'})$$

$$(C_1^{(2)} = C_3^{(2)'}) (C_3^{(1)'} = C_3^{(2)'})$$

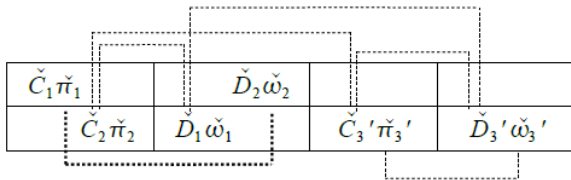
The second pairs  $C^{(2)}, \pi^{(2)}$  will be denoted by  $D, \omega$ . We have a table

$C_1 \pi_1 C_2 \pi_2$	$D_1 \omega_1 D_2 \omega_2$	$C_3' \pi_3'$	$D_3' \omega_3'$
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

Apply  $M_b$  from convolution  $S = M_t \text{ev} M_b$ , we get



Now we apply  $ev$ ,  $ev C_i = \check{C}_i$ ,  $ev \pi_i = \check{\pi}_i$ ,  $ev D_i = \check{D}_i$  which changes the parameters available in the models, with equal parameters equates (we connect with an equal sign, in the diagram it is highlighted in bold dotted line) the names. For each model (category)  $ev$  must be explicitly set. We get



Convolutions of several arrows for a category in a Set form a nerve of this category, the nerve construction is transferred to convolutional polycategories and categorical splices.

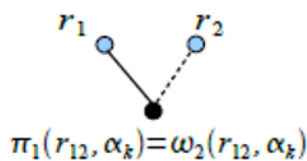
We will explicitly take into account the parameters of the general case using an easily generalized example of a simplicial category, which is one of the main examples of categories used, for example, in algebraic topology. Its objects are finite ordinals  $r_j \in \bar{m} = \{0, 1, \dots, m-1\}$ , arrows are of non-decreasing functions between ordinals. We construct arrows from elementary splices

$$C_i \pi_i = r_i \pi_i(r_i, r_i, id(i, i)), D_j \omega_j = r_j \pi_j(r_j, r_j, id(j, j))$$

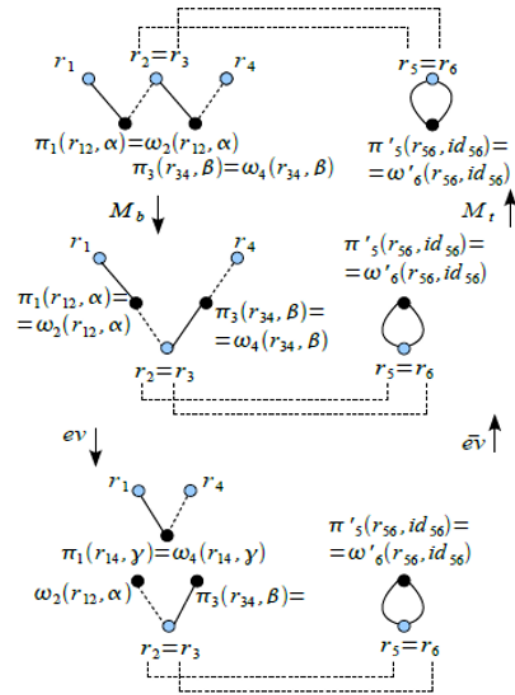
as follows

$$r_i \pi_i(r_i, r_j, \alpha_{ij}) r_j \omega_j(r_i, r_j, \alpha_{ij}) (\pi_i(r_i, r_j, \alpha_{ij}) = \omega_j(r_i, r_j, \alpha_{ij})), \alpha_{ij} \in Hom(r_i, r_j).$$

For clarity, the lines between the names and objects will be dashed  $r\omega$ , the arrow will take a graphical form



The indices  $i, j$  are omitted from the mapping  $\alpha_k \in Hom(r_i, r_j)$  (they are restored, if necessary, from the indices of the objects,  $k$  numbers the arrows from  $om(r_i, r_j), (r_i, r_j) = r_{ij}$ . We conduct the convolution  $\gamma = \alpha^\circ \beta$



We got the result of the convolution along with the "history" from which the original splice is restored. To restore, we apply the convolution with  $\bar{ev}$  and  $M_t$  to the last scheme (as indicated in the diagram on the right).

### 5 Categorical models of chemical bond

The adiabatic approximation in quantum chemistry consists in a separate consideration of the motions of nuclei and electrons. It can be assumed that chemical bonds are determined by the density of the distribution of the electron charge in the electron cloud, specified by the wave function of the molecule, which ensures the equilibrium position of the nuclei (the Gelman-Feynman theorem). An electron cloud in different molecules can have the same substructures, each of these substructures can correspond to one or another type of chemical bond. After some refinements, the structural chemical fomrules become equivalent to the corresponding splices and their convolutions (covalent strokes are associated with splices co-regions). For example, methane Fig. 2. is obtained by simple convolutions.

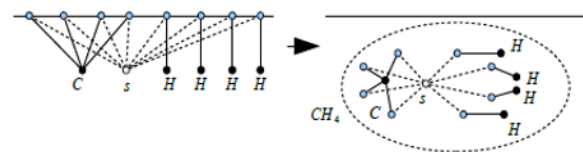
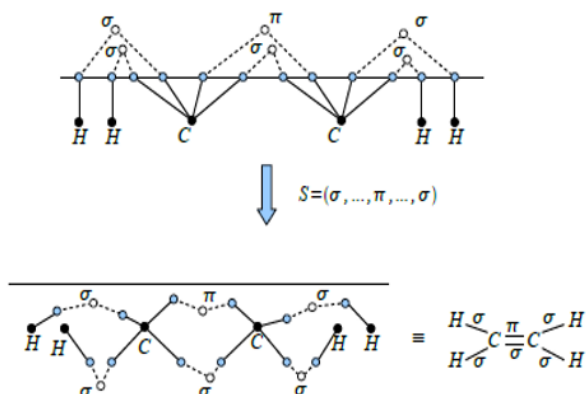


Fig. 2. Convolutions for methane

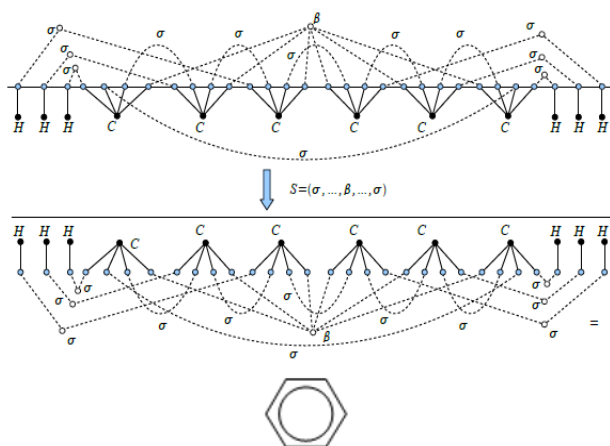
Alkane models are represented by associative algebras with identities, the multiplication operation is ternary. Hydrocarbons with double and triple bonds are

modeled taking into account different types of convolutions for bonds (Fig.3)



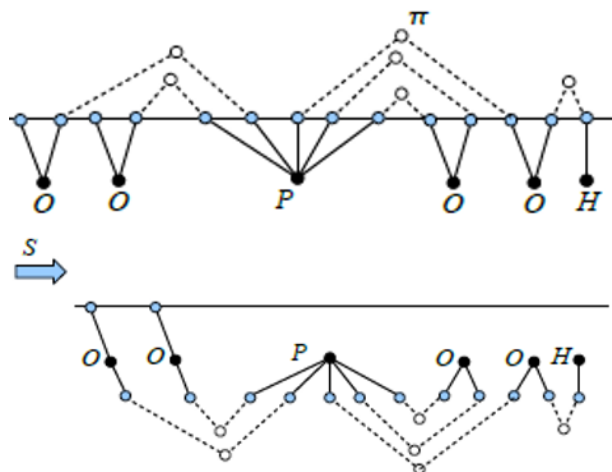
**Fig. 3.** Convolutions for ethylene

For the aromatic bond, a convolution with six (for benzene) and a large number of H atoms (for other aromatic hydrocarbons) is selected.



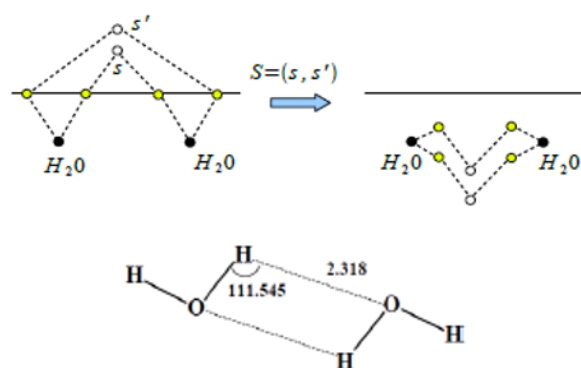
**Fig. 4.** Modeling a benzene molecule by splices

Cyclic hydrocarbons are modeled naturally. The properties of the algebras, PROPs and splices that appeared in this way, studied by strictly mathematical methods, give the properties of biomolecules. The most important biomolecules for algebraic biology, RNA and DNA, are also modeled by splices. It is not difficult to write out splices corresponding to nucleotides and nucleosides. Further construction of RNA and DNA molecules from them is reduced to algebraic convolutions. We emphasize the strictly mathematical nature of these models and the operations in them, which equips algebraic biology with a number of categorical algebraic research methods. The DNA molecule uses a phosphodiester bond. The polynucleotide chain is built through the formation of phosphodiester bonds between the third and fifth carbons of sugars (3C' and 5C'). Phosphodiester bond is modeled by the following splices



**Fig. 5.** Modeling a phosphodiester bond by splices

It is possible to simulate hydrogen bonds in DNA and RNA that are responsible for the secondary structure by adding new types of elements (marked with yellow circles) into splices co-regions and convolutions between them. We also note another option for taking into account the differences in bonds using higher splices, when certain types of bonds are modeled by the corresponding splices included in the higher splices, which completely model the molecule. Let us illustrate the formation of water dimers and their models



**Fig. 6.** Modeling of water dimer isomers by splices

The geometry of molecules, in particular the angles and positions of atomic nuclei for dimers, can also be modeled by adding parameters to the corresponding convolutions. The disassembled types of communication and their models are sufficient to represent the entire DNA (or RNA) molecule, including various levels of its structure. The examples given make it possible to construct formulas and their graphic designations for DNA and RNA, due to their cumbersomeness, they are not presented here.

## 6 Conclusion

One of the main tasks of algebraic biology is to predict the properties of organisms from the genome using

rigorous algebraic methods. The paper proposes new categorical methods used in categorical systems theory. These methods are based on the theory of categorical splices, the elements of which are presented in the work. With their help, the behavior of quantum-mechanical objects, in particular, RNA and DNA biomolecules, as categorical systems, is modeled. New algebraic and categorical methods (associative algebras with identities, PROP, categorical splices) are involved in the analysis of the genome. The presented results are new and original.

## References

1. Yi Lin, *General systems theory: A Mathematical Approach* (Kluwer Academic Publishers New York, 2002)
2. S.N. Vasiliev, A.K. Zherlov, E.A. Fedosov, B.E. Fedunov, *Intelligent control of dynamic systems* (Moscow, Phys.-math. literature, 2000)
3. B.F. Lomov, *Man and technology* (M., Soviet radio, 1966)
4. G.N. Kryzhanovsky, *Fundamentals of General Pathophysiology* (Moscow, Medical Information Agency, 2011)
5. P.K. Anokhin, *Fundamental questions of the general theory of functional systems, Principles of systemic organization of functions* (M., Science, 1973)
6. G.K. Tolokonnikov, *Biomachsystems. Theory and applications*, **1**, 31-213 (2016)
7. G.K. Tolokonnikov, *Biomachsystems*, **2**(1), 127-174 (2018)
8. G.K. Tolokonnikov, *Biomachsystems*, **2**(4), 41-144 (2018)
9. G.K. Tolokonnikov, *Advances in Intelligent Systems and Computing*, **938**, 259-267 (2020)
10. S. Petoukhov, *Matrix genetics, algebra of genetic code, noise immunity* (M., RHD, 2008)
11. G.K. Tolokonnikov, S.V. Petoukhov. *Advances in Intelligent Systems and Computing*, **1126**, 55-64 (2020)
12. W. Hatcher, *The logical foundations of mathematics* (Perg.Pr., 1982)
13. S. Mac Lane, *Categories for a working mathematician* (M., 2004)
14. O. Kuznetsov, *AI and decision making*, **2**, 3-20 (2018)
15. A.A. Zhdanov, *Autonomous artificial intelligence* (M., BINOM, 2012)