Proceedings of the International Conference on Applied Innovations in IT

Eduard Siemens (editor in chief) et al.

Proceedings of the International Conference on Applied Innovations in IT

Eduard Siemens (editor in chief) et al.

Proceedings of the International Conference on Applied Innovations in IT, 2015

editors: Prof. Dr. Eduard Siemens* (editor in chief),
Dr. Bernd Krause*, Dr. Leonid Mylnikov**
(*Anhalt University of Applied Sciences, ** Perm National Research Polytechnic University)

Content

Practical Aspects of Genetic Algorithms' Implementation in Life Cycle Management of Electrotechnical Equipment

Anton Petrochenkov

Perm National Research Polytechnic University - Electrotechnical Department

Komsomolsky Ave. 29, 614990, Perm, Russia

E-mail: pab@msa.pstu.ru

Abstract—**Some practical aspects of Genetic algorithms' implementation regarding to life cycle management of electrotechnical equipment are considered.**

Keywords: **life cycle, electrotechnical equipment, genetic algorithms.**

I. INTRODUCTION

The task of supporting an optimal conditions of electrotechnical equipment within the specified time is a key task. One of the main stages of the life cycle is the stage of maintenance, which acts as a stage of storage of statistical data, and as a stage of testing and inspection equipment to meet specified performance reliability.

One of the solution techniques, allowing creating effective algorithms for a wide range of tasks, is the usage of a subclass of the directed random search methods – methods of genetic modeling.

Numerous approbation of use and application of genetic methods in scientific and industrial spheres allow saying with confidence that when using the evolutionary methods, realization of the comprehensible well-founded decision search problems is always reached. In overwhelming majority of cases the use of natural analogues gives positive results. This is explained by the fact that the analogue taken from the nature was improved for many years of evolutions and at the present time has the optimum structure [1][2].

The idea is to consider these genetic algorithms regarding to life cycle management of electrotechnical equipment.

Like biological development each technological product develops continuously from a stage of origination (setting up) to a stage of collapse (recycling), going through any environment effects and adapting for them.

Algorithms work with a group of "individuals" – population, each of which represents possible solution of the given problem. Each individual is estimated by a measure of its "suitability" according to how rationally solution of the problem suits to it.

In each generation of the chromosome genetic algorithms (coded solutions) are the result of application of some genetic operations [1]-[3].

II. THE KEY GENETIC RELATIONS REGARDING LIFE CYCLE MANAGEMENT OF ELECTROTECHNICAL EQUIPMENT

Operations must precede the process of algorithms construction [1]-[3]:
- Selection of the initial conditions;
- Criteria selection (criteria functions);
- Analysis and selection of the boundary conditions.

Selection operations of the initial level depend on the type of the current task and goals, which must be reached as a result of algorithm realization.

If it concerns maintenance of the parameters on the certain level and reliability index of the equipment in the course of its operation then as initial conditions, it is reasonable to use data values of parameters and the indexes defined on a development stage in the form, installed by manufacturer and presented in the maintenance documentation (in particular, in maintenance documentation for the specific type of the equipment).

In this case, the main task is conformity «real reliability of the equipment – reliability set as zero conditions» that is support of higher level of reliability [4].

Iinitial conditions – datum values of the equipment parameters (values of reliability including the level of the serviceability, revitalization, mean time between failures, probability of non-failure operation etc.).

Population of candidate solutions P^t [3] :

$$P^t = \{P_1 , P_2 , \ldots , P_{Np}\},$$

where
$t = 0, 1, 2, \ldots$ is the number of generation;
N_p is the population size.
Chromosome consists of genes:

$$P_i = \{g_1 , \ldots, g_L\},$$

where L is the chromosome's length.
Gene is identified by locus (position) and allele a_j:

$$g_j = \{j; a_j\}.$$

It is necessary to introduce criterion of an optimality estimation which will characterize level of considered

solutions under the control of the technical state of the equipment [3]-[6].

Realization of the genetic algorithm assumes the fitness function. This function returns the number, showing how good this chromosome is. The given function will be the criteria for the algorithm forming. In the embodied genetic algorithm it is reasonable at first to define the worse chromosome (having maximum deviations from the target level of parameter) and to measure the index value [3].

The got number is called the bad one with respect to which the quality of the other chromosomes is valued: fitness of the chromosome is calculated as difference between index value, set by the given chromosome and bad index value.

For chromosome P_1 fitness function will be:

$$\omega(P_1) = |n\text{-}m|, \qquad (1)$$

where
n is the current index (feature),
m is the specified index.

Let, for example, have a look some time resolution, limited by the first equipment failure. So as a criteria it is reasonable to use the index value of the mean time between failure of the given type of the electrotechnical equipment. The highest the value of the criterion function, the more optimal is the method of the equipment maintenance, the more optimal is the set of methods, providing this maintenance and their content (the set of the actions, the level and the quality of their realization).

The criteria choice is the main preliminary step when building the algorithm, because only it will define the further algorithm filling, its logical direction and convergence.

One of the important moments is the analysis and the limitation selection of the given system and for the algorithm is its functioning. Limitation can absolutely different indeed:

Economic ones – in the network of the limited founding;

Technological ones – limitation in the instrument base and documentation base, skilled personnel, etc.

Temporal constraints – conducting the actions, etc.

Limitations are the necessary addition, which lets to take into account the particularities of the real technological objects and systems, factors, influencing on its activity. The majority of the real tasks have deterministic character.

Limitations can be also made as some rules and conditions and can be used when selecting and making the possible variants of solutions (heuristics).

In this case, equipment is characterized by particular set of maintenances, with the help of which life cycle following is put into effect [7]. Each application for different methods can be done in different ways. This is the difference in chromosome genes.

This difference will be: in application of various instrument base (for example, in one case optical resources are used at a visual estimation of a state of pendant basic insulators, and in other – there is nothing), taking into account some external factors, in application to various approaches concerning state estimation (up to distance and

points control), and also in application to various methods of results processing.

It is necessary to add that all complex of actions should correspond to the circuit of support of the set technical state of the equipment (and to include not only actions for an estimation of the technical state in the form of monitoring and diagnostics, but also reducing and preventive operations etc.).

Changing of the purposes and the problems solved by system, can lead to revision as maintenances of preliminary operations, and a way of "filling" of chromosomes (for example, concerning economic aspect of a problem, the concepts connected with a given problematic and components can be used).

For the interpretation of genetic concepts, the following definitions are introduced [4].

Reproduction. It includes some elements of standardization, as the developed techniques of service and decision-making. As a result of a reproduction we will receive the element of electrotechnical equipment and a set of the regulated actions by means of which its support and the control of its technical state will be carried out.

Crossover. Refinement of existing techniques of the control and equipment service, on the basis of expert estimations and other sorts of the analysis of existing systems. Framing of concrete solutions on upgrading, support of any solutions of documentation base is necessary. The completion phase is development of the standard of the enterprise regulating and considering all features of a set of actions under the control and support of the set technical state of the equipment.

There are a lot of crossover techniques [3]. Indeed the structure of crossover affects the efficiency of genetic algorithm.

Let's introduce three main crossover techniques, which will later be used in the construction of algorithms for optimal life cycle management system of electrotechnical equipment.

1) *One-point crossover*.

Let's have 2 chromosomes P_1 and P_2 in population P: $P_1 \in P, P_2 \in P$, which will be choose as parents:

$$P_1 = \{g_1, g_2, \ldots, g_L\},$$
$$P_2 = \{g^1, g^2, \ldots, g^L\}.$$

It is necessary to determine the cut-point of crossover k – this point determines the place of two chromosomes, where they should be "cut", i.e. k is the number (or value) of the code of a gene which is performed after incision chromosome:

$$k \in \{1, 2, \ldots, L\text{-}1\}.$$

In this case there will be two new chromosomes P_1' and P_2' formed by permutations of the elements in chromosomes P_1 and P_2 according to the rule:

$$P_1' = \left\{ g_1, g_2, g_k, | g^{k+1}, \ldots, g^L \right\},$$
$$P_2' = \left\{ g^1, g^2, g^k, | g_{k+1}, \ldots, g_L \right\}.$$

Schematically one-point crossover provides convertion two chromosomes and partial exchange of information between them.

2) *N-point crossover.*

Unlike one-point crossover, it is necessary to determine the N cut-points in each of P_1 and P_2 chromosomes. Cut-points divide the chromosome into several blocks, and chromosomes exchange sites located within blocks. E.g., the offspring P_1' is formed from the odd blocks of parent P_1 and even blocks of parent P_2. The offspring P_2' is formed from the odd blocks of parent P_2 and even blocks of parent P_1 respectively:

$$P_1 = \{g_1, \ldots, g_a, | g_{a+1}, \ldots, g_i, | g_{i+1}, \ldots, g_k, | g_{k+1}, \ldots, g_L\}$$
,
$$P_2 = \{g^1, \ldots, g^a, | g^{a+1}, \ldots, g^i, | g^{i+1}, \ldots, g^k, | g^{k+1}, \ldots, g^L\}$$
,
$$P_1' = \{g_1, \ldots, g_a, | g^{a+1}, \ldots, g^i, | g_{i+1}, \ldots, g_k, | g^{k+1}, \ldots, g^L\}$$
,
$$P_2' = \{g^1, \ldots, g^a, | g_{a+1}, \ldots, g_i, | g^{i+1}, \ldots, g^k, | g_{k+1}, \ldots, g_L\}$$
.

However, it is necessary to take into account that a large number of cut points may result in loss of "good" parents properties.

3) *"Cut and splice".*

"Cut and splice" crossover is used by forming the second and further generation.

In this approach, the blocks are analyzed in both chromosomes and partial correspondence between the elements of the first and second parents with forming offsprings is set. In the portable block the duplicate genes (alleles) are replaced by genes (alleles) presented in the locus of original chromosome:

$$P_1 = \{g_1, g_2, \ldots, | g_i^i |, | g_k, \ g_{k+1}, \ldots, g_L\},$$
$$P_2 = \{g^1, g^2, \ldots, \ldots, \ | g^k, | g_i^i \ |, \ldots, g^L\},$$
$$P_1' = \{g_1, g_2, \ldots, | g_i^i |, | g^k, | g_{k+1} |, \ldots, g^L\},$$

where

g_i^i is duplicate gene in chromosomes P_1 and P_2:

$$g_i^i := \{i, a_{k+1}\}.$$

Mutation. In whole, a mutation is the ambiguous phenomenon in most cases calling negative consequences. In this case under mutation we will understand effect of some factors on equipment maintenance. To mutation factors we refer:

- Environment conditions (the external factor, in view of object distribution is important);
- Skills degree of operating staff.

Let's introduce two main mutation techniques, which will later be used in the construction of algorithms for optimal life cycle management system of electrotechnical equipment.

1) *One-point mutation.*

In the chromosome P_1, $P_1 \in P$, it's need to determine (randomly or specifically) the two positions (two loci) (i) and (L-i):

$$P_1 = \{g_1, g_2, \ldots, g_i, \ldots, g_{L-i}, \ldots, g_L\}.$$

Genes corresponding to selected positions, g_i and g_{L-i}, are rearranged. It caused to form a new chromosome:

$$P_1' = \{g_1, g_2, \ldots, g_{L-i}, \ldots, g_i, \ldots, g_L\}.$$

2) *N-point mutation.*

It's necessary to determine the N cut-points in the chromosome P_1, $P_1 \in P$.

After that step by step the genes located to the right of the cut-points are exchange of each other in order of their location. The gene located to the right of the last cut-point goes into position in front of the gene corresponding to the first cut-point:

$$P_1 = \{g_1, \ldots, g_a, | g_{a+1}, \ldots, g_i, | g_{i+1}, \ldots, g_k, | g_{k+1}, \ldots, g_L\}$$
,
$$P_1' = \{g_1, \ldots, g_a, | g_{k+1}, \ldots, g_i, | g_{a+1}, \ldots, g_k, | g_{i+1}, \ldots, g_L\}$$
.

Inversion. Inversion is a mathematical construction that allows the inversion based on parent chromosome (or part thereof) to create a offspring chromosome.

Let's introduce two main inversion techniques, which will later be used in the construction of algorithms for optimal life cycle management system of electrotechnical equipment.

1) *One-point inversion.*

It's necessary to determine the cut-point of inversion k in the chromosome P_1, $P_1 \in P$:

$$k \in \{0, 1, 2, \ldots, L+1\}.$$

Offspring P_1' is formed by inversion of the segment located to the right of cut-point k in the chromosome P_1:

$$P_1 = \{g_1, g_2, g_k, | g_{k+1}, \ldots, g_L\},$$
$$P_1' = \{g_1, g_2, g_k, | g_L, \ldots, g_{k+1}\}.$$

2) *N-point inversion.*

It's necessary to determine the N cut-points in the chromosome P_1, $P_1 \in P$. Then elements formed inside the blocks are inverted:

$$P_1 = \{g_1, \ldots, g_a, | g_{a+1}, \ldots, g_i, | g_{i+1}, \ldots, g_k, | g_{k+1}, \ldots, g_L\}$$

$$P_1' = \{g_1, \ldots, g_a, | g_i, \ldots, g_{a+1}, | g_{i+1}, \ldots, g_k, | g_L, \ldots, g_{k+1}\}$$

Reduction. Reduction is a mathematical construction (based on the analysis of the population) that allows to reduce the size of the population to a predetermined value after one or more generations of genetic algorithm.

The primary purpose of the reduction is eliminating unsuccessful decisions and maintaining population at an optimal level (depending on computer memory).

Reduction carried out the following procedures:

1) the forming a reproduction group from all the solutions generated in the population P^t,

2) the selection of solutions in the following population.

Size of new population N^{t+1} is defined as:

$$N^{t+1} = N^t + N_{OC} + N_{OM} + N_{OI},$$

where

N^t is the size of population on step before (t), N_{OC} is the

number of offsprings produced by crossover,
N_{OM} is the number of offsprings produced by mutation,
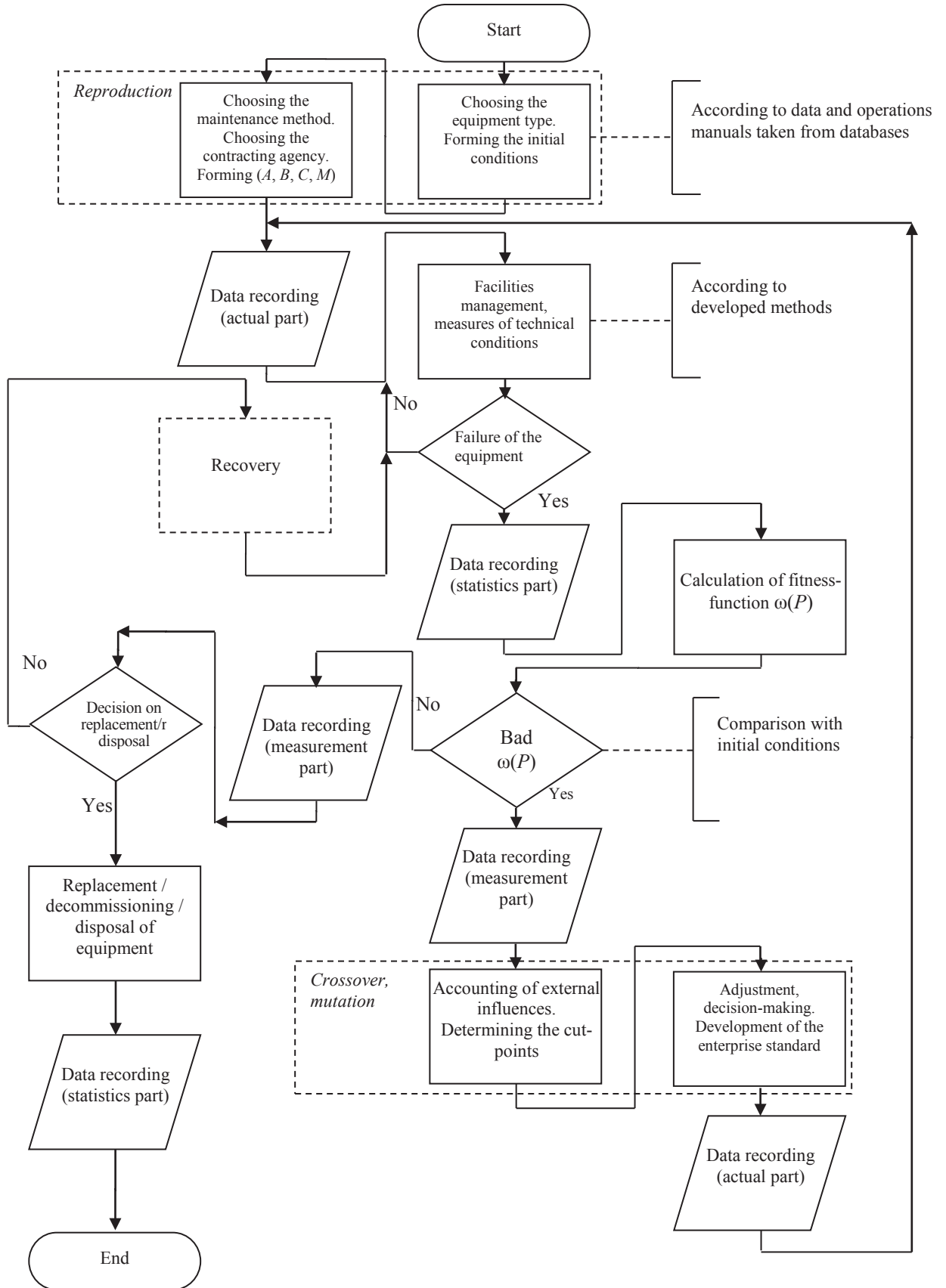N_{OI} is the number of offsprings produced by inversion.

Fig. 1. Generic flowchart of life cycle management system of electrotechnical equipment with the use of genetic algorithms.

III. Implementation Of Genetic Algorithms In The Task Of Supporting An Optimal Conditions Of Electrotechnical Equipment

Generic flowchart of life cycle management system of electrotechnical equipment with the use of genetic algorithms is showed in the Fig. 1. From the figure it can be seen that the performance of each group of measures to support the optimal conditions affects the function of the life cycle equipment as a whole [8]-[11].

The reliability plan also includes a detailed description of reliability tools and criteria for critical items list [7]. As the function of reliability, the function of the life cycle equipment $R'(t)$ should provide maximum effort to specified conditions, i.e., the effective maintenance model should ensure the maintenance of the set of indicators at the appropriate level:

$$R'(t) = R_0 - (A(p) + B(p) + C(p) + M(p) + ...) \leq 0,$$

where R_0 is the specified reliability, $A(p)$, $B(p)$, $C(p)$, $M(p)$ are the "parts" of reliability, provides a variety of measures [9][10][12]. $A = \{a_1, a_2, a_3\}$ is a variety of information-obtain-actions,
a_1 – monitoring by «hectic rush», a_2 – monitoring, based on the visual examination of the equipment, a_3 – monitoring with the use of infrared and ultraviolet control,
$B = \{b_1, b_2, b_3\}$ is a variety of rules of estimation of technical conditions, b_1 – analysis based on risk model, b_2 – analysis based on probabilistic assessment of failures, b_3 – analysis based on expert judgment,
$C = \{c_1, c_2, c_3\}$ is a variety of maintenance and repair activities, c_1 – maintenance by «hectic rush», c_2 – preventative maintenance, c_3 – maintenance within a system of service within a given technical condition,
$M = \{m_1, m_2, m_3\}$ is a variety of management methods.

Let's introduce the next basic heuristics:

$$H_1 = \{a_1 c_1\},$$
$$H_2 = \{a_2 c_2\},$$
$$H_3 = \{a_3 c_3\}.$$

Each group of measures provides a certain level of selected indicators, the value of which varies from 1 to 10. The initial conditions are based on the available basic parameters of reliability of different types of electrotechnical equipment.

Let's assume that each of the groups of measures has the same weight in forming the reliability in current time. Thus the maximum value $(A(p) + B(p) + C(p) + M(p) + ...)$ is equal to 40.

In case of absolute terms (e.g. time between failure t_{max}) it can be used the calibration scale groups of measures of reliability indices of a particular type of equipment:

- t_{max} corresponds to 40 relative units,

- calculating the calibration factor $k_\kappa = \dfrac{t_{max}}{40}$,

- the computational procedures returns to the operator the value of the function of the life cycle equipment in absolute terms $(a_n + b_n + c_n + m_n) \cdot k_\kappa$.

E.g., to suspension insulators PS-70 with $t_{max} = 360$ months, the calibration factor is $k_K = 9$ and the equation of the function of the life cycle equipment takes the form:

$$a_n + b_n + c_n + m_n \geq 40, \tag{2}$$

where
$n = 1...3$ is the name of heuristic.

The first generation of chromosomes is formed basing on the evaluation of a particular type of electrotechnical equipment in four different power systems (Table 1).

TABLE I
GENERATION OF NEW POPULATION

Chromosome	(a, b, c, m)
P_1	(5, 5, 8, 7)
P_2	(2, 5, 4, 3)
P_3	(5, 4, 4, 3)
P_4	(8, 4, 7, 7)

To calculate the coefficients of fitness $\omega(P_i)$, it's necessary to substitute each solution in (2) and find the deviation from the formula (1) (Table 2).

TABLE II
CALCULATION OF THE FITNESS-FUNCTION

Chromosome	$\omega(P)$
P_1	40-25=15
P_2	40-14=26
P_3	40-16=24
P_4	40-26=14

The higher fitness (less the value of $\omega_i(P)$), the greater the chance of being selected as a parent chromosome (Table 3).

TABLE III
SELECTION OF PARENT CHROMOSOMES

1st parent chromosome	2nd parent chromosome
P_4	P_1
P_4	P_3
P_4	P_2
P_1	P_3
P_1	P_2
P_3	P_2

Let's use the crossover for generation the offspring chromosome (Table 4). The result is a second generation of chromosomes with their usual fitness (Table 5).

TABLE IV
CROSSOVER

1st parent chromosome	2nd parent chromosome	Offspring chromosome
(8| 4, 7, 7)	(5| 5, 8, 7)	(8, 5, 8, 7)
(8| 4, 7, 7)	(5| 4, 4, 3)	(8, 4, 4, 3)
(8| 4, 7, 7)	(2| 5, 4, 3)	(8, 5, 4, 3)
(5| 5, 8, 7)	(5| 4, 4, 3)	(5, 4, 4, 3)
(5| 5, 8, 7)	(2| 5, 4, 3)	(5, 5, 4, 3)
(5| 4, 4, 3)	(2| 5, 4, 3)	(5, 5, 4, 3)

TABLE V
OFFSPING'S FITNESS

Offspring chromosome	$\omega(P)$
(8, 5, 8, 7)	40-28=12
(8, 4, 4, 3)	40-19=21
(8, 5, 4, 3)	40-20=20
(5, 4, 4, 3)	40-16=24
(5, 5, 4, 3)	40-17=23
(5, 5, 4, 3)	40-17=23

Continuing in this way, one chromosome eventually reaches fitness, equal to 0 or close to it.

IV. CONCLUSION

It is also necessary to take into account additional "financial" criterion, calculating the additional factors of fitness. Thus, the choice of the optimal solution will be a synthesis of solutions for two functions: the function of "reliability" and function of "costs" [3][13].

The difficulty and the goodness of action realizations in technical state of the equipment increases, the number of the operations required for the algorithm realization increases. But in spite of the apparent bulking, algorithms are simple in their accomplishment because of their coherency.

Electronic educational resources were developed using an educational process for training students with the specializations "Electrical Power Supply," "Automation of Technological Processes and Production," and "Automated Management of Product Life Cycle" of Perm National Research Polytechnic University.

Works on this direction are conducted within the Russian Foundation for Basic Research Grant of Russia No 14-07-96000 "Development of an intellectual decision support system to ensure of energy facilities trouble-free operation".

REFERENCES

[1] L. A. Gladkov, V. V. Kurejchik, V. M. Kurejchik, "Genetic algorithms", Under the editorship of V.M. Kurejchik. – M: FIZMATLIT, 2006. (rus)

[2] D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley, 1989.

[3] H. Aytug , M. Khouja and F. E. Vergara, "Use of genetic algorithms to solve production and operations management problems: A review", International Journal of Production Research, 2003, 41:17, pp.3955-4009, doi: 10.1080/00207540310001626319

[4] A. Petrochenkov, "Regarding to Implementation of Genetic Algorithms in Life Cycle Management of Electrotechnical Equipment", Proc. of the Second International Conference on Applied Innovations in IT, E. Siemens (editor in chief) et al. Kothen, Anhalt University of Applied Sciences, 2014. – pp. 79-83. doi: 10.13142/kt10002.13.

[5] D. Garg, K. Kumar and Meenu, "Availability Optimization for Screw Plant Based on Genetic Algorithm," International Journal of Engineering Science and Technology, vol. 2, No. 4, pp 658-668, 2010.

[6] S.P. Sharma, Y.Vishwakarma, "Availability optimization of refining system of Sugar Industry by Markov process and Genetic Algorithm", Proc. of the 2014 International Conference on Reliability, Optimization and Information Technology, art. no. 6798290, pp. 29-33. doi: 10.1109/ICROIT.2014.6798290

[7] E.V. Cota, L. Gullo, R. Mujal, "Applying Design for Reliability to increase reliability confidence", Proc. of Annual Reliability and Maintainability Symposium, 2014, art. no. 6798454, doi: 10.1109/RAMS.2014.6798454

[8] A. Petrochenkov, "Methodical Bases of the Integrated Electrotechnical Complexes Life Cycle Logistic Support", Proc. of the First International Conference on Applied Innovations in IT, E. Siemens (editor in chief) et al. Dessau, Anhalt University of Applied Sciences, 2013. – P.7-11. doi: 10.13142/kt10001.02.

[9] A. B. Petrochenkov, "Regarding Life-Cycle Management of Electrotechnical Complexes in Oil Production", Russian Electrical Engineering, 2012, vol. 83, No.11, pp.621-627. doi: 10.3103/S1068371212110090.

[10] A. B. Petrochenkov, "On the Problem of Development of Models of Processing Operations Performed during Repair of Electrical Engineering Complex Components", Russian Electrical Engineering, 2013, Vol. 84, No. 11, pp. 613–616. doi: 10.3103/S1068371213110096.

[11] A. B. Petrochenkov, S. V. Bochkarev, A. V. Romodin, D. K. Eltyshev, "The Planning Operation Process of Electrotechnical Equipment Using the Markov Process", Russian Electrical Engineering, 2011, Vol. 82, No.11., pp.592-595. doi: 10.3103/S1068371211110113.

[12] N.A.J. Hastings, "Physical Asset Management", 2015. doi: 10.1007/978-3-319-14777-2_3

[13] A. B. Petrochenkov, A. V. Romodin, "Energy-optimizer complex", Russian Electrical Engineering, 2010, vol. 81, no. 6, pp. 323-327. doi: 10.3103/S106837121006009X.

Regarding Information Systems Dependability Analysis

Ivan Luzyanin, Anton Petrochenkov

Perm National Research Polytechnic University - Electrotechnical Department

Komsomolsky Ave. 29, 614990, Perm, Russia

E-mail: {lis, pab}@msa.pstu.ru

Abstract—**The article presents classification of information systems by different parameters. Factors influencing information systems dependability are also presented. The article describes the strategy of information systems dependability analysis and methods of its increase. The example of analysis of real information system is considered to show how to implement the strategy.**

Keywords: **information system, dependability, maintainability, reparability, integrated logistics support, structural and functional analysis.**

I. INTRODUCTION

Nowadays information systems perform different tasks almost in all fields of human activity. Their responsibility increase rapidly for the last time. For that reason, providing the information systems dependability becomes more and more important.

The aims of research are the information systems dependability estimation criteria determination and dependability increasing methods development.

We consider the information system as a hardware system. The software of this system is considered as the component, which provides working of the system, and delivers a human-machine interface. The special attention is paid to hardware dependability.

In this report, we will consider information systems as a whole and highlight their features. We will also describe the generic algorithm of information systems dependability analysis and give the methods of dependability increase.

As a main dependability definition, we use the definition given in the standards IEC 60050-191:1990 [1] and IEC 60300-3-1:2003 [2].

In this term, availability performance is the ability of an item to be in a state to perform a required function under given conditions at a given instant of time or over a given time interval, assuming that the required external resources are provided.

Reliability performance is the ability of an item to perform a required function under given conditions for a given time interval.

Maintainability performance is the ability of an item under given conditions of use, to be retired in, or restored to, a state in which it can perform a required function when maintenance is performed under given conditions and using stated procedures and resources.

Maintenance support performance is the ability of a maintenance organization, under given conditions, to provide upon demand, the resources required to maintain an item, under a given maintenance policy.

The given conditions in these definitions are related to the item itself and to the conditions under which the item is used and maintained.

II. OBJECT ANALYSIS

The generic structure of the information system can be presented as a three-level model (Fig. 1).
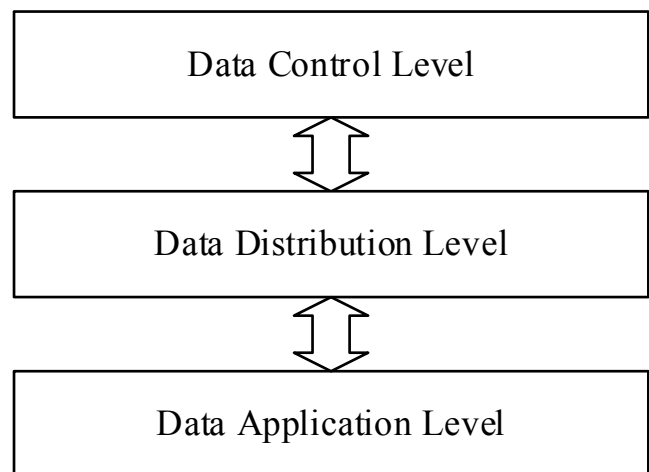

Fig. 1. Information systems generic structure

Upper level is responsible for working of IT-infrastructure. The centralized data processing, general-system services delivering, data storage and shared resources control are performed here. This level also provides data integrity and safety.

Middle level provides safe data transmission, data integrity control during the transmission and data routing.

Lower level provides client workstations and peripherals working.

Every level has a specific structure and performs different tasks; therefore, it is necessary to study working of these levels separately.

Moreover, every real implementation of the information system has specific parameters. Every specific parameter needs additional studying.

It is impossible to describe all variants of information systems implementation. So that, let's try to classify these

implementations according to different parameters. After that one can present the most common specifics and regularities of real information systems in each class.

The information systems can be classified by following parameters:

a) On geographical distribution:
- Distributed systems;
- Centralized systems.
b) On performed tasks:
- Corporate Networks;
- Data processing centers.
- Information systems of network providers.
c) On application field:
- Industrial information systems;
- Administrative Information systems;
- Information systems included both industrial and administrative levels.

Distributed information systems are located in several buildings. The system consists of several segments. This type of systems usually has central segment placed in the main building and a number of local segments, which are able to work independently from each other. Central segment equipment delivers general services such as connection between segments, providing access to the Internet and others. If the central segment equipment fails, other segments are not able to use general services but are still able to perform local operations. The most critical levels in these systems are central data control level and central data distribution level. Its dependability must be higher than dependability of other ones. The failures of central data application level may not cause general functions failures and do not affect working of other segments but in this case central layer administrators may lose control of the system. Therefore, its dependability has to be higher as well.

The structure of centralized information system can be presented as one local segment of distributed information system.

The purpose of corporate networks is to support users working. Hence, the tasks of data application level determine the structure and functions of these systems. This layer is usually the largest one in this system.

The data processing centers are high-performance systems, which perform complex operations with the large amount of data. The tasks of data control level determine its structure and functions.

The data distribution level can be conventionally divided into internal (which provides communication between data center servers) and external (which connects the data center and external users) levels.

In these centers, it may be no data application level for example if the data center performs the cloud computing.

The main task of the information systems of network providers is to provide data transmission. There is no data application level in this type of information systems. Its structure and functions are defined by the data distribution level, and the task of data control level is to distribute information between clients according to their requests.

The industrial information systems are usually real-time systems. They provide working of automated control systems. These systems deliver the human-machine interface and allow controlling real state of the technological process. They are critical systems. Their failures may cause massive destruction or any more critical aftermaths. So that, these systems must have high dependability and long lifetime.

The administrative information systems are not real-time systems. These systems support office working and their failures are not so critical as industrial information systems failures. The faults of them may cause economic loses but usually not affect human health and life and not cause massive destruction. Therefore, the dependability of these systems depends on aftermath scale caused by their failures.

III. INFORMATION SYSTEMS DEPENDABILITY FACTORS ANALYZING

After information systems analyzing let's study the factors influencing information systems dependability.

These factors are divided into internal and external ones.

1) The internal factors are:
- Elements dependability;
- Amount of elements;
- Elements joining method.
2) The external factors are:
- Environmental conditions;
- Usage strategy (including usage time and intensity, user's qualification, work organization);
- System structure;
- Software used.

The internal factors affect working of information system equipment and the external factors affect working of information system as a whole.

The analysis of studies in the field of microelectronics and computing equipment dependability [3]-[7] showed that there are quite different methods of eliminating every concrete factor. The analysis of different examples of real information systems implementations [8][9] provides distinguish three basic groups of arrangements used to eliminate influence of described factors:

1) Technological arrangements that increase the system dependability using the maintenance and repair policy.

2) Organization arrangements, that allow increasing dependability by changing usage strategy (e.g. by optimizing usage time and intensity or employing more qualified specialists).

3) Arrangements on structure optimization, that increase dependability due to changing of system structure.

On this step, the problem of arrangement choosing is faced. It is also necessary to have to develop the concrete methods of implementing these arrangements in real information systems.

To solving these problems the strategy of information systems dependability analysis and methods of its increase are developed. If data are to be reliable, the instrumentation used has to be reliable too [10].

The algorithm has iterative structure. The amount of iterations depends on amount of functions and priorities. This algorithm is the implementation of the integrated logistics support (ILS) methodic for information systems.

The first part of this algorithm corresponds to the Failure Modes, Effects and Criticality Analysis (FMECA) procedure [11]. (Fig. 2).

Using the accumulated statistics corresponding to the separate junctions of a system, each type of the system component can be related to a definite law of failure intensity, assuming the latter to be a constant value at minimum statistics. The influence of external unfavorable factors on the system in this case cannot be determined; thus, the distribution law and its parameters cannot be correctly obtained either. Use of the guide data corresponding to the analogs of the system components may be a way out of this situation. Knowledge of the dependability component does not provide information concerning the dependability of the system as a whole. For analyzing the system one should know its structure and inner functional connections. In structural design, the dependability of a structural component is evaluated with respect to one or more failure modes. There are a lot of methods for constructing failure models of complex systems [12]. One of the approaches for models under incomplete information based on the set of canonical analytical expressions for computing imprecise structural reliability has been obtained by Lev V.Utkin, and Igor O.Kozine [13].

The second part implements the reliability centered maintenance including condition based monitoring procedure [11]. The latter task entails selection of the servicing system, development of control and diagnostics systems, processing and analysis the data for obtaining information on performance quality, carrying out various stages of maintenance according to technical and economy criteria, and increasing the quality of reconstruction [14]-[17].

The third part defines the profitability of arrangements to be determined [2][5][14]. The arrangements planning and implementation are also performed in this part.

One can use this algorithm in case of a small enterprise budget and having problems with working of IT-infrastructure. It may be problems with a low data transmission time in the network or frequent failures of the network equipment. So, this is the most general algorithm.

There is the task of implementing this algorithm for solving real systems problems. For this purpose, it is necessary to describe formally all procedures of this algorithm as applied to information systems.

It is also need to develop concrete realization methods of these algorithm positions, which allow solving real problems.

Let's describe only the structural and functional analysis methods for information systems. Let's consider these methods on the example of the Microprocessor Units of Automation chair infrastructure in the Perm National Research Polytechnic University. The hardware structure of this system is shown on Fig. 3.
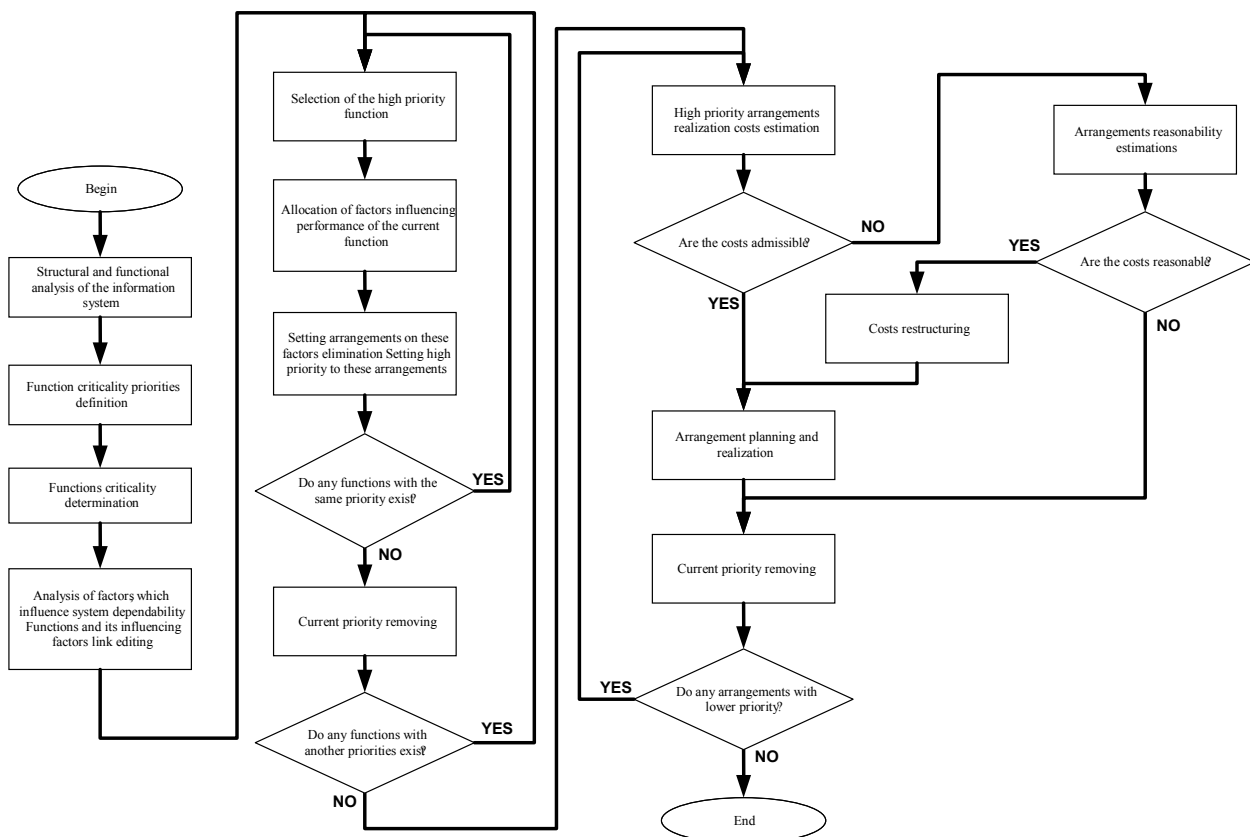

Fig. 2. The strategy of information systems analysis and their dependability increasing methods choosing
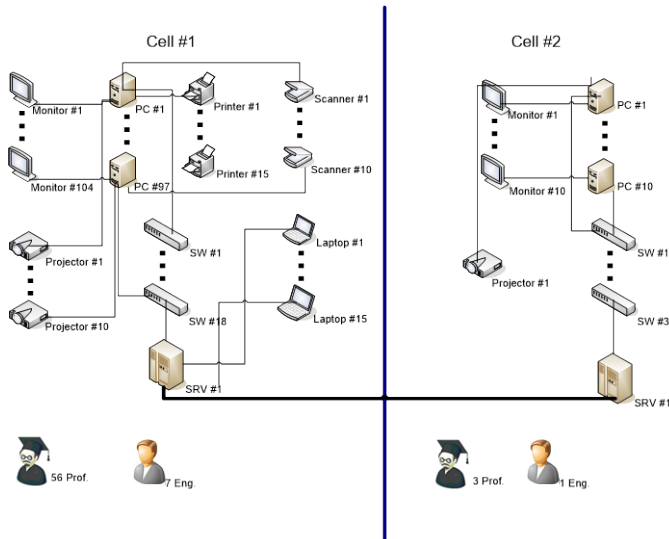
Fig. 3. The IT-infrastructure of the Microprocessor Units of Automation chair in Perm National Research Polytechnic University

The system consists of two segments. The main segment supports working environment for fifty six professors. Seven engineers provide technical support, maintenance and repair of the system. The second segment supports working environment for three professors and maintained by one engineer.

The detailed structure of this information system is shown on Fig. 4.

The PCs and servers are detailed into their internal replaceable modules but network equipment and peripherals are detailed into concrete devices due to repairing them in service centers.
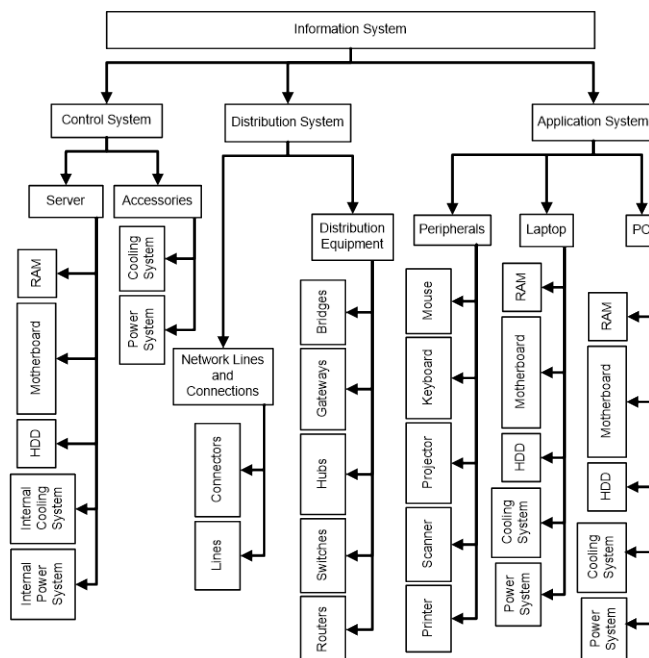

Fig. 4. Information system detailed structure

The functional structure of the information systems is determined by its software. The basic software used in the chair's information systemis shown on the Fig. 5.
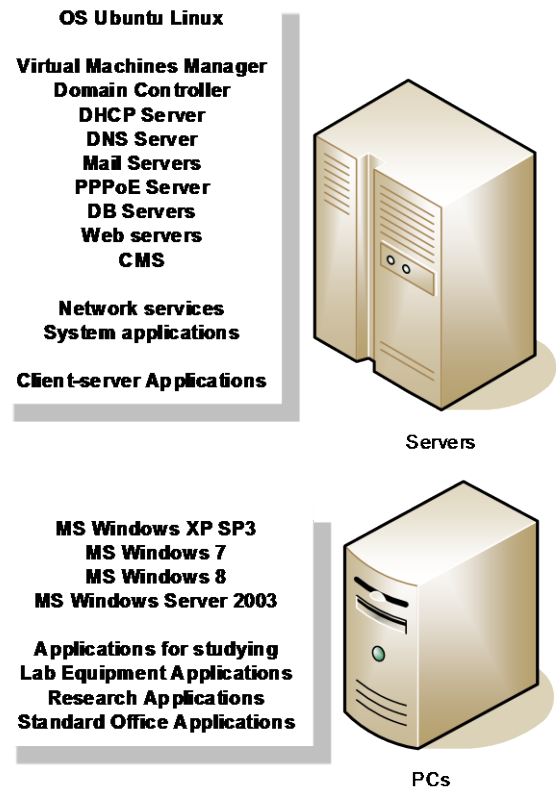

Fig. 5. Software used on the chair's information system

On the described system, there are organization problems on all levels. For example: there are no technical descriptions and schemes of this system, there are no concrete usage rules, which makes impossible to support this system. There also structural problems (the network structure do not provide stable interconnection between equipment). There are some technological problems because of some equipment is very old.

IV. CONCLUSION

This example illustrated the implementation of the first part of the algorithm described previously. In future, we plan to develop implementation methods for all steps of the algorithm. The experimental analysis of our chair information system using this algorithm will be carried out. It is also necessary to generalize the described strategy on all information systems classes. The task is to find concrete enterprises with different types of information systems to conduct more experiments and test this strategy in real conditions.

Works on this direction are conducted within the Russian Foundation for Basic Research Grant of Russia No 14-07-96000 "Development of an intellectual decision support system to ensure of energy facilities trouble-free operation".

REFERENCES

[1] IEC 60050-191:1990 International Electrotechnical Vocabulary. Chapter 191: Dependability and Quality of Service.
[2] IEC 60300-3-1:2003 "Dependability management - Part 3-1: Application guide - Analysis techniques for dependability - Guide on methodology".
[3] R. Longbottom "Computer System Reliability", New York: Wiley-Blackwell, 1980.
[4] P. Gill, N. Jain, N. Nagappan, "Understanding Network Failures in Data Centers: Measurement, Analysis and Implementations" Microsoft Research, Redmond WA.

[5] B. Guenter, N. Jain, C. Williams, "Managing Cost, Performance, and Reliability Tradeoffs for Energy-Aware Server Provisioning", Microsoft Research, Redmond WA.

[6] J. Gray, "Why Do Computers Stop and What Can Be Done About It?", 1985.

[7] S.C. Malik, "Reliability modeling of a computer system with preventive maintenance and priority subject to maximum operation and repair times", International Journal of System Assurance Engineering and Management, vol. 4, no. 1, pp. 94-100, Mar. 2013.

[8] R. Reinertsen, "Residual life of technical systems; diagnosis, prediction and life extension", Reliability Engineering and System Safety, 1996, 54 (1), pp. 23-34. doi: 10.1016/S0951-8320(96)00092-0.

[9] A.K.S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance", Mechanical Systems and Signal Processing, 2006, 20 (7), pp. 1483-1510. doi: 10.1016/j.ymssp.2005.09.012.

[10] M. Bagajewicz, "A review of techniques for instrumentation design and upgrade in process plants", Canadian Journal of Chemical Engineering, 2002, 80 (1), pp. 3-16.

[11] JSP 886 The Defence Logistic Support Chain Manual vol. 7 Integrated Logistics Support

[12] A. B. Petrochenkov, E. M. Solodkii, "On the Methods for Constructing Failure Models of Complex Systems", Russian Electrical Engineering, 2011, Vol. 82, No.11., pp.623-627. doi: 10.3103/S1068371211110125.

[13] Lev V. Utkin, and Igor O. Kozine, "Stress-strength reliability models under incomplete information@, International Journal of General Systems, 2002, 31:6, 549-568, DOI: 10.1080/0308107021000061885

[14] N.A.J. Hastings, "Physical Asset Management", 2015. doi: 10.1007/978-3-319-14777-2_3

[15] A. B. Petrochenkov, S. V. Bochkarev, A. V. Romodin, D. K. Eltyshev, "The Planning Operation Process of Electrotechnical Equipment Using the Markov Process", Russian Electrical Engineering, 2011, Vol. 82, No.11., pp.592-595. doi: 10.3103/S1068371211110113.

[16] A. B. Petrochenkov, "On the Problem of Development of Models of Processing Operations Performed during Repair of Electrical Engineering Complex Components", Russian Electrical Engineering, 2013, Vol. 84, No. 11, pp. 613–616. doi: 10.3103/S1068371213110096.

[17] E.V. Cota, L. Gullo, R. Mujal, "Applying Design for Reliability to increase reliability confidence", Proc. of Annual Reliability and Maintainability Symposium, 2014, art. no. 6798454, doi: 10.1109/RAMS.2014.6798454

Conceptual Foundations of Modelling of Innovative Production Projects

Leonid Mylnikov

Perm National Research Polytechnic University - Electrotechnical Department
Komsomolsky Ave. 29, 614990, Perm, Russia
E-mail: leonid@pstu.ru

Abstract—**At the moment there is a lack of methodological approaches to formalization of management of innovative projects relating to production systems, as well as to adaptation and practical use of the existing approaches. This article is about one potential approach to the management of innovative projects, which makes the building of innovative process models possible based on objective approach. It outlines the frameworks for the building of innovative project models, and describes the method of transition from conceptual modelling to innovative project management. In this case, the model alone and together with parameters used for evaluation of the project may be unique and depends on the special features of the project, preferences of decision-making person, and production and economic system in which it is to be implemented. Unlike existing approaches, this concept does not place any restrictions on types of models and makes it possible to take into account the specificities of economic and production systems. Principles embodied in the model allow its usage as a basis for simulation model to be used in one of specialized simulation systems, as well as for information system providing information support of decision-making process in production and economic systems both newly developed by the company (enterprise) and designed on the basis of available information systems that interact through the exchange of data. In addition, this article shows that the development of conceptual foundations of innovative project management in the economic and production systems is inseparable from the development of the theory of industrial control systems, and their comprehensive study may be reduced to a set of elements represented as certain algorithms, models and evaluations. Thus, the study of innovative process may be conducted in both directions: from general to particular, and vice versa.**

Keywords: **innovative project, management, decision-making, algorithm, analysis, concept, search for optimal solution innovative project, management, decision-making, algorithm, analysis, concept, search for optimal solution.**

I. Introduction

The existence of fresh results that can be used in the products has become a resource which gives strategic advantages. However, it is essential to use such resource properly.

In large production systems time and budget for implementing any changes are strictly limited. Each of these projects requires key competencies some of which a company may not have. In such a situation, execution of the project gives the company new competencies and skills, as well as new impetus for its development.

Effect of innovations may take the form of improved consumer properties and higher level of standardization and automation of production, which in turn leads to reduced costs.

High rates of economic growth become an additional factor under a free market economy. According to data from the Federal State Statistics Service in the period from 2009 to 2013, the product output in the Russian Federation (in million rubles) grew by 181,92%; and the output of innovative products grew by 350% which demonstrates economic expediency of innovative products.

System manufacturing innovative products had to be in a process of transformation manifested as the need to increase the number of product modifications, relationship links with other enterprises, and proprietary sub-systems as compared to traditional products. This makes higher demands for quick administrative decisions and high-quality innovation management, therefore the quality of management, as a rule, becomes a critical factor in the success of projects aimed at development and manufacture of innovative products and technologies, and addressing the problem of efficient management of innovative projects becomes more and more important.

Theory and practice of the management were developed through addressing local problems because it was hard to manage innovation as a single system [15] which created a lack of methodological approaches to formalization [formal characterization] of the management of innovative projects as a single system.

Because of this, while noting all benefits of holistic consideration of the problem of innovative project management, modern publications do not describe any models of innovative projects as complicated systems, which allows no effective solution of the issues of management, expertise and rationale. Such issues are often viewed as separate not interconnected challenges rather than a single process [12][15].

There are some special approaches to certain types of innovations. They depend largely on the sector of economy and sphere they are related to (organizational innovations, innovations in the field of service, technological innovations, innovations applicable to the products which affect their consumer properties, etc.).

Methods of solving management tasks and decision making used in product innovative projects are based on analysis of the project properties. Scientific and technical literature relating to innovative programs and processes contains description of several stages and functions of

manager at each stage. A common feature is that all his decisions must be informed and reasoned [11].

The rationale needs the use of model. The way and method of modelling must solve the problem of increasing the efficiency of administrative decisions.

A model allows the manager to rule out the prospect that only some of available methods were used during project execution, determine the dynamics of its progress and set its parameters reasoning from the current and desirable dynamics.

II. MODELLING OF INNOVATIVE PROJECTS

Development of methods of mathematic formalization of innovative project management has been reflected in several approaches.

In the context of this approach volumetric production planning and procurement (acquisition) planning are the most studied. The principal research in this area focuses on support of administrative decisions in nonlinear-cost or combined-structure systems, and most studies focuses on cost minimization and demand forecasting. Only little research deals with training and promotion of decision-making skills with due regard for requirements of production systems [10]. Important problem arising from the above challenges is concurrent consideration of challenges of pricing (price formation), volumetric production planning and procurement planning. Solving of this problem raises the issues of market selection and internal structural organization of economic and production systems, and is defined "Wagner-Whitin's problem" [9].

Formalization of concurrent consideration of challenges (tasks) relating to market selection and volumetric production planning was first mentioned in [1], and continued to be a challenge. It has been shown that this is a NP-full task, and can be tackled only when certain factors are determined[1] [10].

The perception of single innovations as constants is a very rough assumption. In 1995 Peppal suggested the use of game theory [8] in description of duplicating and improving innovations so that to take account of interaction and reciprocal influence between projects. This approach has generated new developments related to the issues of change management. Its promotion in the field of innovation management is currently connected with agent-based modelling and the use of forecasts. In early 2000s this approach was furthered by object-oriented modelling, and, when applied to economic and production systems, is known as multi-agent systems [4] taking into account such factors as independence, exposure to external shocks (influences), flexibility, pro-activities, and availability of intellectual control [3][7]. The most complex issue is to coordinate interaction between different parts of economic and production system [5].

Another approach is management model oriented at such event as innovation diffusion (Eric von Hippel) [1], namely, adoption and study of best practices of similar product manufacturing used in various production systems in an effort to find common features. This approach helps to uncover common features and to provide guidance based on the same type of products and similar production systems.

Approach to innovation management in production systems is often based on some pattern of events that make up the process concerned (Jordan's modelling, Hein's logical modelling), through it does not take into account the relationship between different sub-systems and their interference.

Present-day development of methods of statistics-based management is establishment of databases of actual status of process, event and object. This approach presupposes that all facts within the framework of innovative process paradigm are true, and involves the use of interdependent and agreed databases, their relationship and processing rules [13]. This method may serve as a good base for designing information infrastructures in systems with well-established processes, but is not efficient for creating information infrastructures in dynamic systems, that is, introduction of innovations (novelties) entails a lot of changes in production systems.

There is a great number of practically developed methods and techniques of formalizing individual solutions that can be used in different special cases irrespective of management levels and types of innovation.

By knowing the peculiarities of innovative project to be executed, and the relevant economic and production system (its identification), one can reduce the modelling of its management to application of a number of standard or author's models and methods to each component of innovative project. This approach is justified by the high level of some challenges. So it would be reasonable to give due consideration for the decisive results which can be received from solving local tasks, especially in connection with the fact that evolution of technical and economic systems becomes more complicated both in number of elements and quality of their relations, and hence structuring of models allows changes of this type to be easily incorporated [17].

The challenge of choosing right methods and models to be used can be addressed by a decision-making person. To simplify the process, existing and known methods can be presented in a convenient form based on classification characteristics of innovative projects (refer to Fig. 1).

The application of this scheme is only to choose the classification characteristics[2] as shown in Fig. 1 [2]. In this case, to present methods and models means to indentify their location by reference to the chosen characteristics.

The next challenge is to combine them in an effort to make general assessments at decision-making points. Thus, it is essential to establish relationship between methods when they are combined.

On the one hand, interrelation of methods will depend on interrelation of sub-systems, their elements and tasks they help to address (structural scheme of economic and production system, sequence of management tasks to be solved, etc.). On the other hand, successive (application of some methods to determine internal parameters of others) or parallel (general assessments) application of some methods can be used to determine coefficients of others.

---

[1] For instance, Jean Tirole has successfully addressed the issue of management in sector markets.

[2] Only characteristics having common angles (Fig. 2) may be chosen.

If there are known structural interrelations a model of innovative project (innovative process) introduction can be presented as a matrix [16]:

$$[M] = \left([A][K]^T \begin{bmatrix} M_1 \\ \cdots \\ M_l \end{bmatrix} \right),$$

where

$[A]$ - incidence matrix (demonstrates interrelations between sub-systems, methods and project parameters);

$[M]$ - vector of summative assessments for each methods employed;

$[K]$ - diagonal matrix of corrective coefficients;

M_i – assessment of method employed;
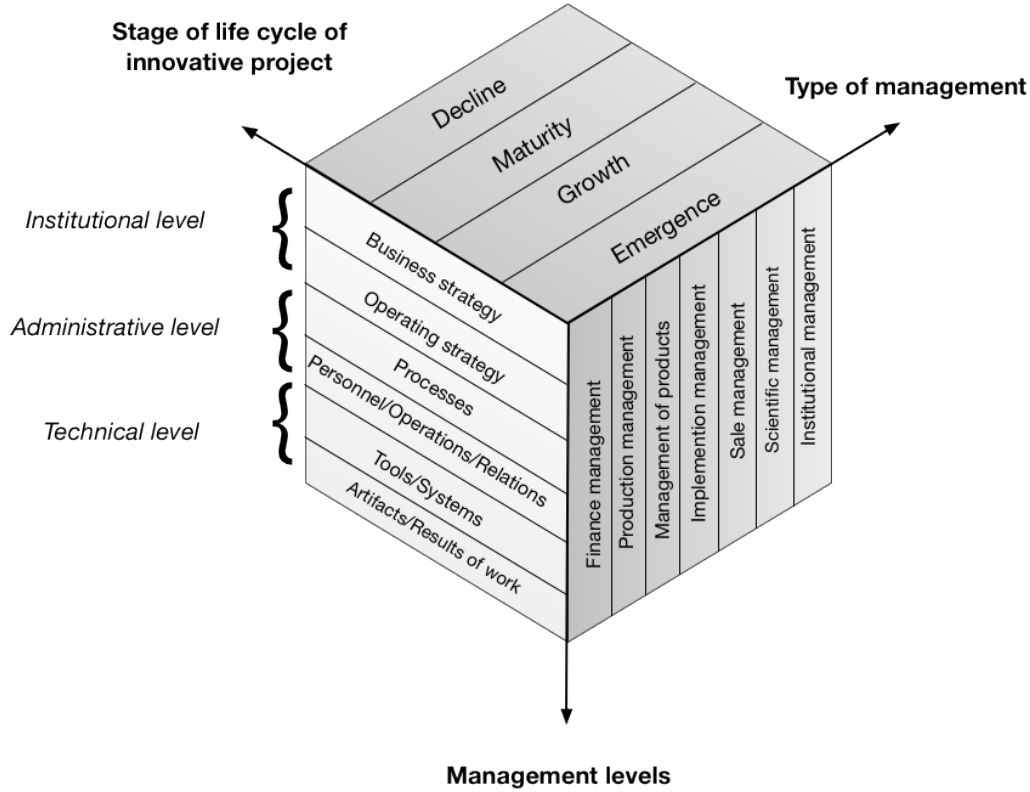
l - number of combined methods employed.

Fig. 1. The Way of Combining Management Presentations to Choose the Right Methods and Techniques Suitable for Assessment of Outcome of Administrative Influences in Execution of Project

For the purpose of addressing economic challenges, results enabling to take optimal or admissible decisions are particularly important. Thus, there is the need for generic criterion which can be obtained by adding or deducting assessments made with the methods described herein [16].

In other cases, method interrelation can also be represented through adding, deducting, multiplying and dividing assessments obtained with these methods.

Thus, implementation of this approach reduces the modelling to a number of operations with mathematical descriptions of methods, techniques and models relating to certain tasks.

In this regard, it could be concluded that implementation of innovations and manufacturing of innovative products have "nucleus" - "generic production function" in the form of non-linear multi-factor dependence between output figures and vector of generic production factors.

III. FUNCTIONAL APPROACH TO ADMINISTRATIVE DECISION MAKING IN THE EXECUTION OF INNOVATIVE PROJECT

The modelling theory outlines a few tasks that can be addressed with the use of models depending on the unknown. If we look at a model of innovative project implementation, we can highlight two objectives:

1) to establish internal setpoint of project parameters aimed at ensuring the achievement of given targets (goals) of the projects;

2) to determine production parameter values (settings) at the decision-making points.

If we look at examples of formalized objectives, then in the first case production function will dictate general criterial function as follows (refer to more detailed information [16]):

$$\sum_{j=1}^{l_1} u_j k_j (h_j - M_j)^2 + \sum_{j=l_1+1}^{l_2} u_j k_j M_j - \sum_{j=l_2+1}^{l} u_j k_j M_j \to min$$

where

h_j - desirable value;

k_j - elements of vector of corrective coefficients $[K]$ which, in turn, is calculated by the above formula:

$$k_j = \frac{k_{\text{в}j} - k_{\text{н}j}}{k},$$

where

$k_{\text{в}j}$ - upper edge of the range of parameter variations in j-method;

$k_{\text{н}j}$ - lower edge of the range of parameter variations in j-method);

u_j - assessment of importance of chosen methods (can be made with the use of both expert approach and assessment methods);

l_1, l_2 - boundaries of groups depending on the criterial function ($l \geq l_2 \geq l_1$),

M_j - summative assessment of each method employed.

Search for optimal solution may restrict the targets of methods employed m_{ij} (components of summative assessments of methods M_j):

$$m_{ij} \in G_{ij}; \; j = \overline{1, l}; \; i = \overline{1, n_l},$$

where

G_{ij} – multitude of alternative parameter values (settings) for j – method;

i - parameter.

This objective can be reduced to a class of discrete multi-parameter optimization tasks with certain restrictions, and achieved by using Bellman-Ford algorithm which is adapted method of dynamic programming for graphs.

As an example of second-type formalization, we will consider a volumetric scheduling task based on discrete volumetric production targets.

$$\sum_i \sum_h K_{ih} \sum_t (C_h(t) x_h(t) + C_h(t) x_h(t)) \to max$$

$$\sum_i \sum_t R_{ij} x_i(t) \leq P_j(t), \; j = \overline{1, M}$$

$$\sum_i \sum_t S_{ki} x_i(t) \leq T_k(t), \; k = \overline{1, K}$$

$$\sum_i \sum_t \alpha_i^q x_i(t) \leq G^q(t), \; k = \overline{1, Q}$$

$$\sum_h \sum_t R_{hj} x_h(t) \leq P_j(t), \; j = \overline{1, M}$$

$$\sum_h \sum_t S_{kh} x_h(t) \leq T_k(t), \; k = \overline{1, K}$$

$$\sum_h \sum_t \alpha_h^q x_h(t) \leq G^q(t), \; k = \overline{1, Q}$$

where

$x_i, i = \overline{1, N}$ – vector unknown;

i type, $C_i, \; i = \overline{1, N}$ - net revenue from production of i-goods (estimated);

$R_{ij}, \; j = \overline{1, M}, \; i = \overline{1, M}$ - required capacity of each type of equipment per finished product;

$P_j, \; j = \overline{1, M}$ - total capacity resources for each type of equipment in terms of average production rate of all equipment of that type;

$S_{ki}, \; k = \overline{1, K}, \; i = \overline{1, N}$ - required volume of key materials per finished product;

$T_k, \; k = \overline{1, K}$ - volume of available key materials based on inventory in stock and procurement plan;

$$\alpha_i^q = \begin{cases} 1 - \text{if } i - \text{goods belong to } q \text{ category} \\ 0 - \text{if } i - \text{goods does not belong } q \text{ category} \end{cases}$$

$G^q, \; q = \overline{1, Q}$ - restriction on sales market (marketing area) (estimated);

K_{ih} - coefficient of correspondence between i and h goods, which shows economic feasibility of joint manufacturing of goods or group of goods within one production (as evaluated by one of methods of Slope One group on the basis of sales statistic data).

The use of forecasts in second-type tasks, even if they are very adequate, may result in variations (especially if there are crisis phenomena in the economy). If we include accidental variations in our model, we will get Markovsky process, which requires further research on the basis repeated modelling and statistic data – refer to Fig 2.

Thus, the designed models and production functions can be considered, as a set of "black boxes" of a sort, each with the relevant functional description [18]. This approach was first proposed by an American economist J.B. Clark.

In addition to the above models which take into account production functions, there are many models of certain units (production department, warehouse, etc.); models of types of activities (production, sales); and models of personnel, product, finance and other models.
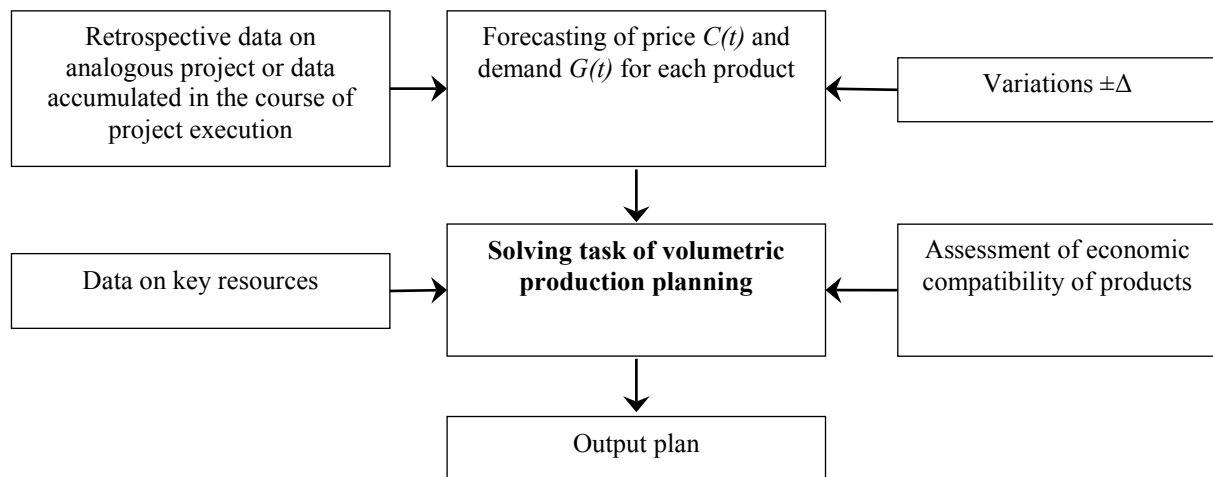
Fig. 2. Hierarchy of Selection of school based on analytical hierarchy process [6].

IV. CONCLUSION

Based on the above, we can conclude that at present there is no uniform concept of project management that would allow us to perform focused search for administrative decisions instead of their analysis and sorting out. Development of conceptual business tends to focus either on separate management sub-systems (such as enterprise management information system [14]) or on separate sub-tasks (such as management of warehouse, sales, deliveries, etc. [18]).

Thus, development of the conceptual foundations of the theory of innovative project management is as topic as ever. It is precisely the understanding of conceptual foundations that facilitates the use of methods and approaches in this field of study as a tool of practical task solving.

It is shown, that development of conceptual foundations of management of product innovative projects is inseparable from development of theory of production system management. Their comprehensive study can be reduced to a number of elements embodied with specific algorithms, models and assessments, which makes it possible not only to take administrative decisions, but to design information system to manage economic and production systems as well.

REFERENCES

[1] F. Gault, E von Hippel, "The Prevalence of User Innovation and free Innovation Transfers," Implications for Statistical Indicators and Innovation Policy.MIT Sloan School of Management Working Paper #4722-09, January, Cambridge, MA, 2009.

[2] M. Gepp, M. Amberg, T. Schaeffler, S. Horn, J. Vollmar, "Framework and rationale for economic considerations in industrial plant business," Systems Conference (SysCon), IEEE International, 2013.- pp 14 – 19.

[3] V. Dignum, "An overview of agents in knowledge management," Heidelberg: Springer. 2006.

[4] N. Jennings, "On agent-based software engineering," Artificial Intelligence, 117 (2), 2000, pp. 277-296.

[5] T. Kaihara, S. Fujii, "Game theoretic enterprise management in industrial collaborative networks with multi-agent systems," International journal of production research. 46 (1). 2008. pp. 1297-1313.

[6] M. Lang, S. Kammerer, M. Amberg, "Projektportfoliomanagement in der IT: Priorisierung, Investition, Steuerung," Düsseldorf: Symposion, 2012. p. 442.

[7] J. H. Lee, C. O. Kim, "Multi-agent systems applications in manufacturing systems and supply chain management: a review paper, "International Journal of Production research, 46(1). 2008, pp.233-265.

[8] L. Pepall, "Imitative Competition and Product Innovation in a Dupoly Model," Economica. №64, 1995. pp. 265-269.

[9] H. M. Wagner, T. M. Whitin, "Dynamic version of the economic lot size model," Management Science. 5. 1958. pp. 89-96.

[10] W. V. den Heuvel, E. Kundakcioglu, J. Geunes, E. Romeijn, T. C. Sharkey, A. P. M. Wagelmans, "Integrated market selection and production planning: complexity and solution approaches," Mathematical Programming, Volume 134, Issue 2, 2012, pp. 395-424.

[11] P. Young, "The evolution of conventions," Econometrica № 61, 1993. pp. 57-84.

[12] V. Vinokur, A. V. Trusov, "Intelectual properties as basis of innovation activities," Perm: PSTU Publ., 2004. – 271 p. (In Russian)

[13] T. Gavrilova, V. Khoroshevskiy, "Knowledge bases of intelligent systems,". SPb: Piter Publ., 2000. – 382 p. (In Russian)

[14] Y. Zelenkov, "Methodology of strategic management of inforamtion systerm development for huge industrial enterprise in modern conditions," Dr. tech. sci. autor. diss. Chelybinsk. 2013.-42 p. (In Russian)

[15] L. Mylnikov, "Microeconomic problems of innovation project management," Management problems, 2011. №3, pp. 2-11. (In Russian)

[16] L. Mylnikov, "Management of innovation projects based on an integral criteria," Management of big systems. Issue. Moscow: IPU RAS Publ., 2010, pp. 128-151. (In Russian)

[17] L. Mylnikov, A. Trusov, N. Khoroshev, "Review of concepts for information management of innovation projects," Information resources of Russia, 2010. №3, pp.34-39. (In Russian)

[18] V. Shiriaev, E. Shiriaev, "Management of business process," – Moscow: Finansy I Kredit Publ.; INFRA-M Publ., 2009. 464 p. (In Russian).

City Outdoor Lighting System Control.
Problems and Prospects

Andrei Bachurin

Perm National Research Polytechnic University – Electrotechnical Department

Komsomolsky Ave, 29, 614990, Perm, Russia

E-mail: bachurin@ellips.ru

Abstract—**This review describes current problems, trends and prospects of the quasi-automatic light control system. The lighting systems of million cities.**

Keywords: **control system, lighting system, management problems, prospects for control systems, control in lighting trends.**

I. OBJECT OF STUDY

a. Cities

The objects of study are the largest cities in Russia: Ekaterinburg, Kazan, Perm. These cities are administrative, industrial and cultural centers of large areas and they have developed network of outdoor lighting (Table 1).

TABLE I
MAIN CHARACTERISTICS OF CITIES

City	Square, sq.km.	Official number of residents	Official number of light points
Perm	800	1026000	29000
Ekaterinburg	468	1428000	41000
Kazan	614	1206000	65000

Cities are characterized by a large area, considerable extent in latitude and large difference in weather conditions at one and the same time (Fig. 1).
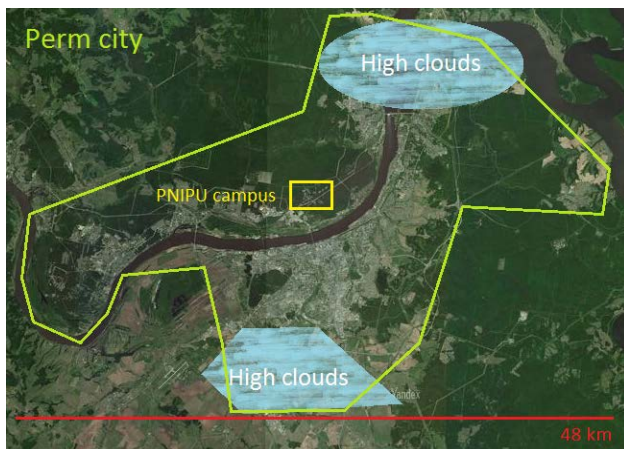

Fig. 1. Perm city from satellite photo.

Lighting systems of these cities are characterized by the presence of all types of lamps (Fig. 5), different types of lighting control systems, illuminated objects of all possible types (Table 2) and modes (Table 3) of illumination, a single

lighting control center, a small number of service personnel, high demands on energy-efficient lighting.

b. Classification of illuminated objects

Street lighting can be divided by illuminated objects into 3 main categories and some subcategories (Table 2), each of them can be characterized by the following common features:

- Illumination is regulated by standards (STD);
- Illumination of the object is normalized (NORM);
- Illumination of the object can be changed depending on external conditions (VAR NORM);
- There is a standardized schedule of the object illumination (SHED);
- Schedule of the object illumination can be changed depending on external conditions (VAR SHED).

TABLE II
CATEGORIES AND OBJECTS

Category	Object	STD	NORM	VAR NORM	SHED	VAR SHED
Street lighting	Highways	*	*	*	*	*
	Streets	*	*	*	*	*
	Car parking	-	-	*	*	*
	Tunnels	*	*	-	*	-
	Road signs	-	-	-	-	*
	Road markings	-	-	-	-	*
	Pedestrian crossings	*	*	-	-	*
	Crossroads	*	*	-	-	*
Pedestrian lighting	Pedestrian areas	*	*	*	*	*
	Yards	*	*	*	*	*
	Pedestrian paths	*	*	*	*	*
	Athletic facilities	*	*	-	-	*
Decorative lighting	Parks and gardens	-	-	*	*	*
	Architectural lighting	-	-	*	*	*
	Outdoor advertising	-	-	*	*	*
	House numbers	-	-	*	*	*

It should be noted that the existing standards for most objects are very flexible and lack of base building nouns and other standards concerning lighting of there is a certain objects [5][6].

Lamps of different categories are located near each other and have to work synchronously. Ideally, the lamps must insure each other in case of failure.

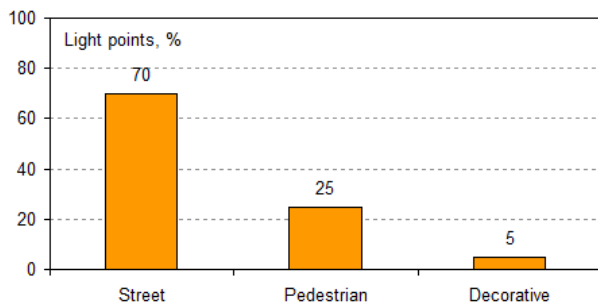The approximate ratio of the objects in city lighting categories is shown on Figure 2.

Fig. 2. The ratio of the urban lighting categories.

Objects (categories) can be grouped by lighting modes (Table 3) [5][6].

TABLE III
TYPICAL LIGHTING MODES OF OBJECTS

Objects	Lighting mode
Highways, Streets, Pedestrian areas	The lights are switched on and off according to the schedule with adjustment for the current value of illumination, at night by low traffic density, the luminous flux can be reduced by 30% while decreasing traffic to 33% and by 50% while decreasing traffic to 20%. Depending on the object, there are different rates of light and illumination should not be reduced at least 4 lux.
Road markings, Pedestrian crossings, Crossroads	Lighting should switch on and off simultaneously with the appropriate lighting of streets, but the change of the luminous flux at night is not allowed.
Tunnels, Road signs	Lighting should work around the clock without changing the luminous flux.
Pedestrian paths, Car parking	Lighting should switch on selectively (some light points), if it is necessary, according to presence and light sensors data.
Parks and gardens, Yards	Lighting should switch on and off simultaneously with the street lighting, at night the luminous flux should be reduced to a minimum, but the switching on of some light points at full power is allowed on the basis of presence, sensor data.
Architectural lighting, Advertising, House numbers	Lighting should switch on and off together with street lighting and work without changing the luminous flux.

Let's consider the classification of illumination objects by the example of PNIPU campus (Fig. 3):
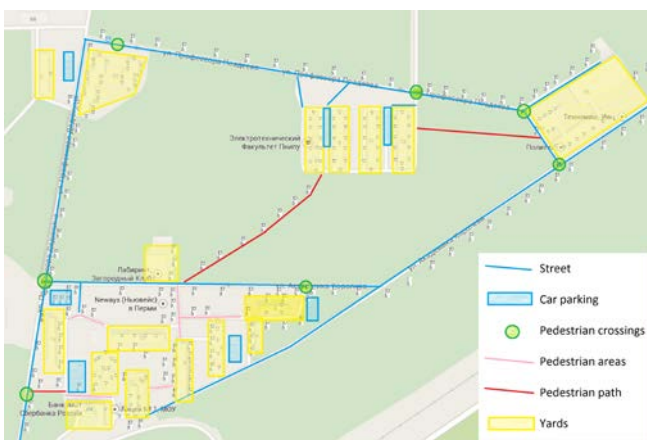

Fig. 3. Outdoor lighting of PNIPU campus.

We can see that control object is very diverse and complicated. Its automatic control system (ACS) must have more than one level of control and some not-trivial algorithms of control for effective energy usage [2].

As shown on fig. 4, energy consumption by city lighting systems makes up 25% from all city energy consumption.
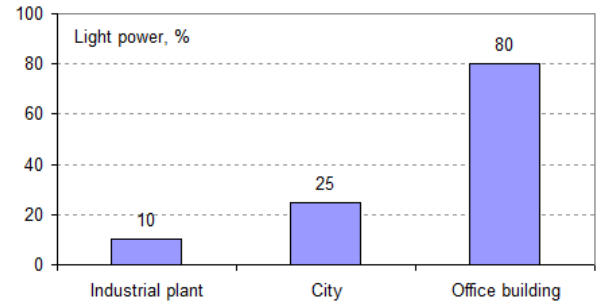

Fig. 4. Lighting share in energy consumption of object.

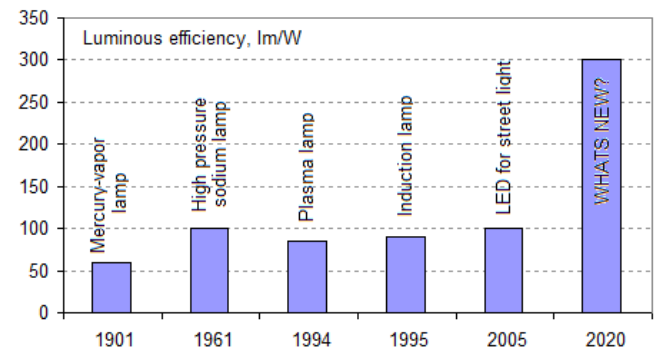*c. Historical, current and forecast data on the main indicators*

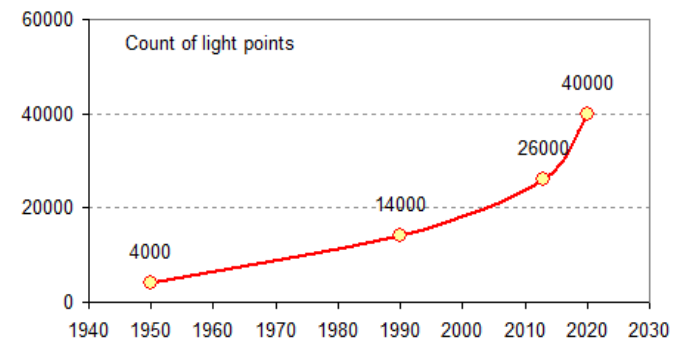Fig. 5. Types of lamps used in outdoor lighting.

Fig. 6. Count of light points in different years in Perm city.

Number of light points used in outdoor lighting of the city increases with growth rate. The reasons for this - increase of light sources light output, growth of city territory, growth of illuminated city area, growing need for different types of lighting.
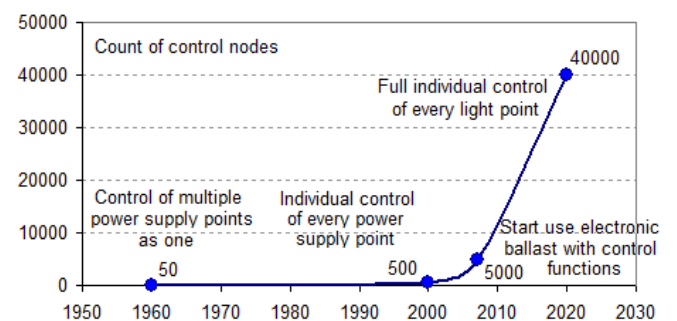

Fig. 7. Growing number of control nodes in the lighting system.

The number of levels in the ACS and the number of managed nodes grow exponentially, it is connected with the need for management and energy efficiency.
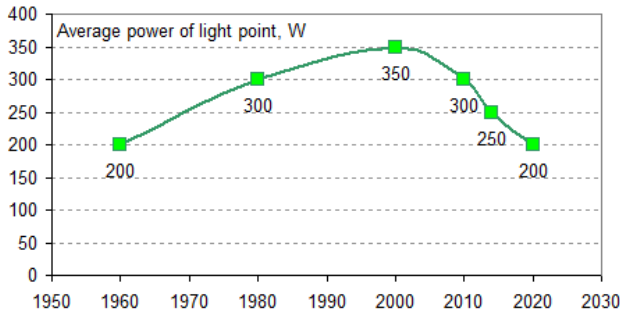

Fig. 8. Average power of light point by years.

Today and in the future the average power of one light point will decrease, thanks to the introduction of modern types of lamps with high luminous efficiency (HPS, LED, plasma, induction) and further improvement of the light sources [1].
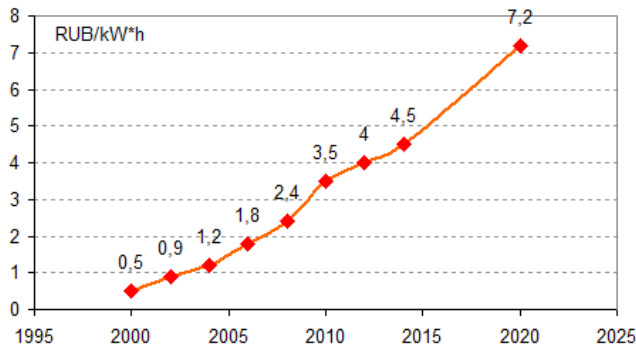

Fig. 9. Cost of electricity in Moscow by years.

The cost of electricity in Russia is growing steadily and this trend will continue in the near term.
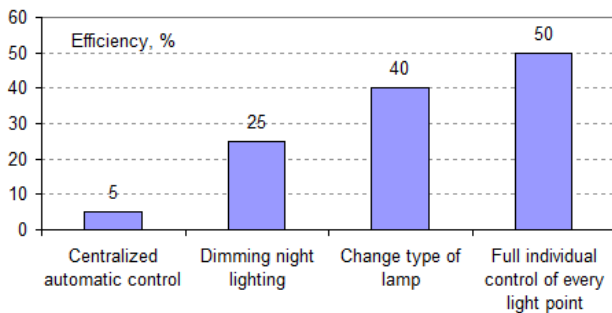

Fig. 10. Measures to improve lighting energy efficiency.

At the moment, not all MVL lamps are replaced by HPS or LED and a very small number of lamps equipped with electronic ballast. There is practically no individual control of light points.

As we can see and calculate from fig. 10, in the near future the efficiency of lighting can be increased up to 30-80% from current level.

II. CURRENT TRENDS OF LIGHTING CONTROL SYSTEMS

a. Current state and trends of lighting control.

The use of light points with individual control is the basis for the construction of "smart" outdoor lighting system,

when the operation mode of each light point changes depending on the outdoor light, traffic density or pedestrians and others.

In the area of light points is the trend in the implementation of individual lighting control systems (electronic ballasts), which have their own controller, provide the switching on / off actions of illumination and control the level of lighting. Entry level electronic ballasts can change level of lighting by two steps: full level (100%) and low level (50% or 70% or other predefined value), changing initiated by internal timer. Middle level electronic ballasts can change level of lighting by three steps by analogy. Osram named this algorithm – AstroDIM (Fig. 11) [8].
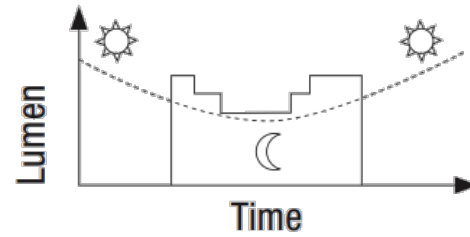
Fig. 11. AstroDIM algorithm by Osram [8].
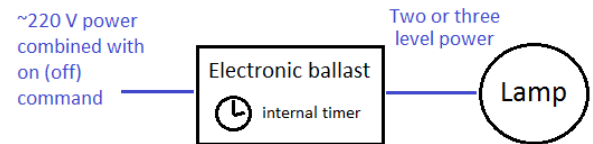
Fig. 12. Middle level electronic ballast connections.

Advanced level electronic ballasts can change level of lighting variously (Fig. 13) and equipped with network interfaces (RS485, DALI, 0-10V, etc.), allowing to integrate them in the local control system and gradually increasing their share, also their can connect some sensors (Illumination, presence).

Fig. 13. Variously changing level of lighting [8].

Fig. 14. Advanced level electronic ballast connections.

As we can see on Fig. 10, entry level ballasts (Fig. 12) can save only 25% of electric energy, but use of advanced level ballasts (Fig. 14) allow to save up to 50% of energy. Also we can see that entry level ballasts work autonomous and need only ~220V connection, but advanced level ballasts must have advanced connection to control network (for synchronous group work).
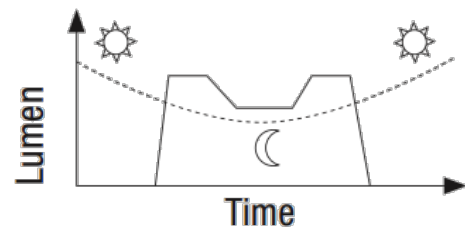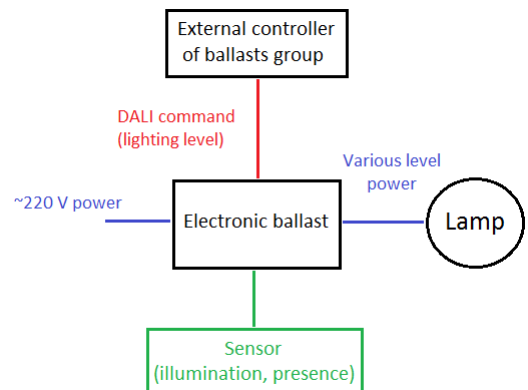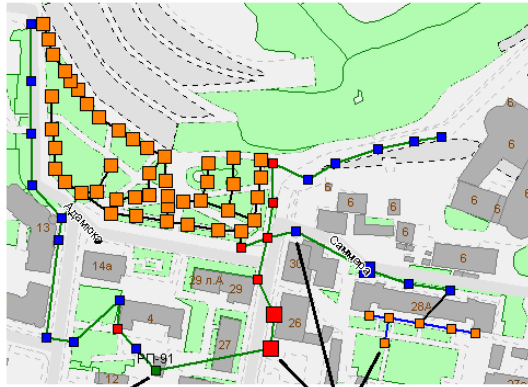

Fig. 15. Sample part of real lighting net from Kazan city.

b. Current trends of data transfer in lightning systems.

Today lighting system consists of some local energy supply centers (SC) and groups of light points which are connected to SC via power cable (Fig. 15).

External controller of ballasts group typically placed in supply center and we must connect all light points from this SC to controller via control network. Several variants are possible (Fig. 16, 17, 18) [1]-[4][7][8].
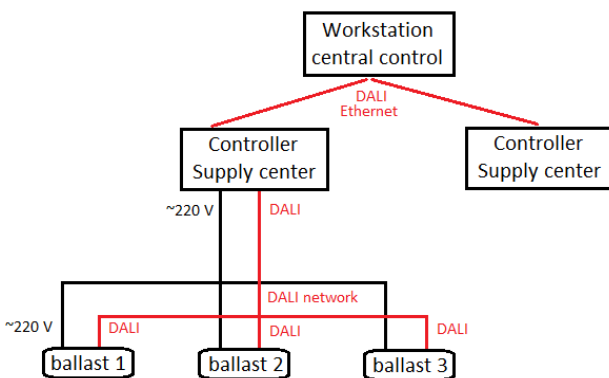

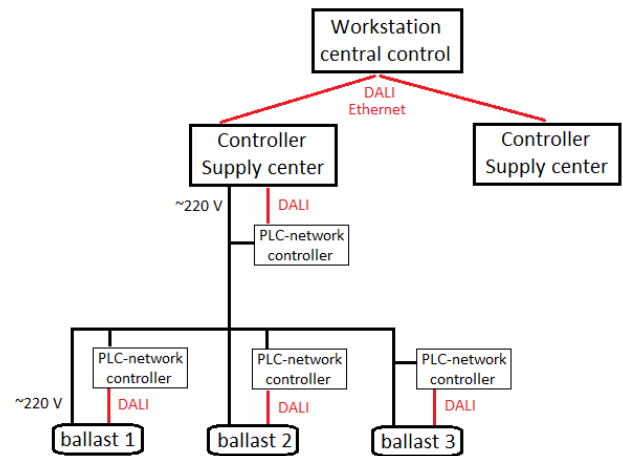Fig. 16. Electronic ballast connections via DALI-network [8].

Fig. 17. Electronic ballast connections via PLC-network.
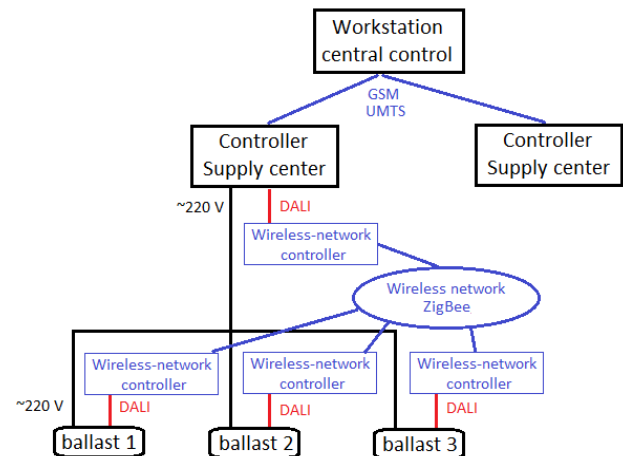
Fig. 18. Electronic ballast connections via ZigBee-network [4].

The surroundings, the methods and data transfer protocols evolve. Over long distances (supply center – central control) die slow and expensive ones - modem connection over a telephone pair, radio, SMS, Ethernet. And rapidly gain popularity those that become cheaper and expand geographically - GSM, UMTS. At close distances (light point – supply center) wire ones - PLC, DALI, RS485 are superseded by cheaper wireless ZigBee and WiFi.

In the field of wireless communication we can see a rapid growth of the so-called M2M devices (smart sensors that are connected to the Internet)/ That's due to the reduced cost of electronic components providing connection to networks (WiFi, GSM, UMTS, ZigBee), increasing network coverage and range of devices, reduction of tariffs for connection to the grid.

In the future these trends will intensify and that will lead to a serious displacement other data channels by cellular networks and it will be possible to connect to the Internet practically "each pen".

Also it is likely that advanced electronic ballast will be separated into more simple electronic ballast and controller light point [7]. Simple ballast will be more unified and cost effective. Controller light point will take following functions: sensor connection, link with central control workstation, link with other near light points, control of electronic ballast via simple unified interface (0-10V) by typical algorithms (Fig. 19).

Fig. 19. Structure of future lighting control system.

c. Current corporate trends.

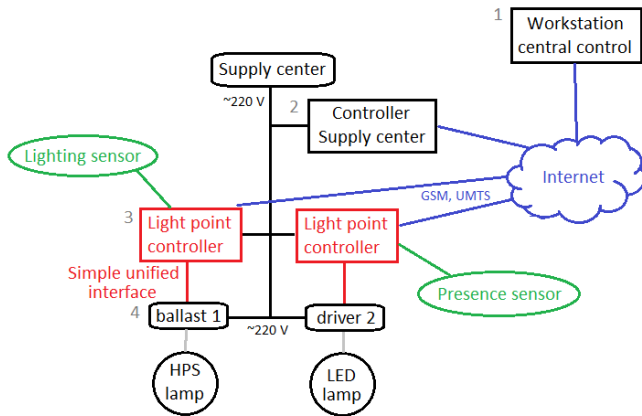In the area of corporate management continues unabated trend towards outsourcing usage. It is gradual applies to the production functions, including power supply, telecommunication services and data transmission.

We can suppose that in the field of lighting will increase the number of energy service contracts, increase the number of light points of light transmitted to the specialized service organizations. Energy service organization will be integrated and their number will be reduced and the load for a specific electrician will increase.

Giants of lighting industry (General Electric, Philips, Osram) for several years develop and implement projects of "smart" light systems for the city [8]-[10]. But the cost of such systems is rather great for mass distribution, and also the used technology (Fig. 12-18), design technology and equipment setup technology – are already out of date today. All that do not allow you to create quickly and inexpensively large lighting systems, that are simple and inexpensive to operation.

III. CURRENT PROBLEMS

Here is the list of problems facing the large systems of outdoor lighting:

- A large number of light points and its constant growth;
- A large area where light points are placed;
- A wide variety of illuminated objects types, and as a consequence - wide variety of local lighting systems control modes;
- Low level of lamp types unification (MVL, HPS, LED), and as a consequence - low level of unification of electronic ballasts types and "drivers" (various ballasts and LED-drivers);
- The lack of widespread low cost universal interface for controlling electronic ballasts and "drivers" (the predominance of proprietary technologies in control systems);
- Low distribution of tools for the diagnosis of lighting lines and light points;
- Low level of management culture and operation of lighting networks, which dictate the need for a complex but unified control algorithms for typical lighting objects;

- Almost complete absence of individual control of light points which can give up to 50% additional energy savings.

IV. PROJECTED SOLUTIONS

Try to imagine a system of outdoor lighting in the future.

Maintenance and management of the most city' outdoor lighting installations will be administered by a one service organization.

All lighting systems will be combined into a single network management using public, common, reliable communication channels provided by the operators.

All light points and power supply centers will be energized and on-line connected all the time.

Each light point will have its own individual fully functional control system that can operate completely autonomously and interact with neighboring light points, the controller of power supply center and general control center [9][10].

Future innovate changes in the lighting control system:

- From electronic ballast will be picked up functions of light point control (working with sensors, network protocol support, link with central control workstation, link with other near light points, control of light point via simple unified interface by typical algorithms) and they will be given to light point controller;
- Electronic ballast control will be organized according to a simple low-cost and reliable unified interface;
- From SC controller will be picked up functions of light points group control (switch on/off, light level adjusting), because individual control functions will be given to light point controller and functions of group control will be given to central control workstation;
- All light points all the time will be energized and on-line connected to central control workstation (not uncontrol, not undefined status of light points);
- The light points controller will have all the features which are necessary for high-grade lighting control in stand-alone mode (parameter monitoring mains and control object, light control point on the schedule, external commands, depending on the values of external sensors, and the state of the neighboring light points using standard algorithms);
- The use of standardized common communication channels;
- The use of standardized common control algorithms for typical lighting objects;
- Group work of light points (the ability to combine any point of light in any group) using communication via a higher level;
- Implementation of possible function of a single light point by an individual algorithm (one of the types, but with individual settings).

To realize all these innovations its required the creation of a new control system on the upper level, that is based on new principles of light points individual control. It is our future work.

REFERENCES

[1] "Handbook of Illumination", Edited by J. B. Eisenberg. - M: Energoatomizdat, 1995. (rus)

[2] Feasibility efficient installation automatic control unit outdoor lighting, Karelin A.N., Industrial ACS and controllers, 2008, № 1, pp 62-66.

[3] Decentralized wireless control system outdoor lighting, Obraztsov S.A., Panfilov D.I., Light & engineering, 2012, № 1, pp 32-36.

[4] DE102010049121 (A1) "Method for lighting e.g. road, involves switching on the lamp on detection of movement of person, and sending message to neighboring lamps through communication unit"

[5] GOST R 54305-2011 "Roads public road. Horizontal illuminance from the artificial lighting".

[6] SP 52.13330.2011 "Natural and artificial lighting - Code of Practice. The updated edition of SNIP 23-05-95"

[7] Teliko Intelligent street lighting. [Online]. Available: http://www.teliko.com

[8] Osram products. [Online]. Available: http://www.osram.com

[9] GE's LightGrid Big Data System. [Online]. Available: http://www.scientificcomputing.com/news/2013/04/lights-data-action-ge's-lightgrid-big-data-system-cut-la's-electricity-bill

[10] Philips IntelligentCity. [Online]. Available: http://www.lighting.philips.com/main/systems/public/intelligentcity.html

Long-term Forecasting of Frame Alignment Losses for Circuit Emulation Implementation

Yuriy Babich

Odessa National A.S. Popov Academy of Telecommunications

1 Kovalska Str., Odessa, Ukraine

E:mail: y.babich@onat.edu.ua

Abstract⸺**The article provides a method for long-term forecast of frame alignment losses based on the bit-error rate monitoring for structure-agnostic circuit emulation service over Ethernet in a mobile backhaul network. The developed method with corresponding algorithm allows to detect instants of probable frame alignment losses in a long term perspective in order to give engineering personnel extra time to take some measures aimed at losses prevention. Moreover, long-term forecast of frame alignment losses allows to make a decision about the volume of TDM data encapsulated into a circuit emulation frame in order to increase utilization of the emulated circuit.**

The developed long-term forecast method formalized with the corresponding algorithm is recognized as cognitive and can act as a part of network predictive monitoring system.

Keywords: **frame alignment loss, long-term forecast, predictive monitoring.**

I. Introduction

Convergent development of modern information communication networks makes actual the task of monitoring of the states and time/probabilistic characteristics of equipment. Monitoring is the main way to infer a specific state and/or performance characteristics of a maintenance entity in order to diagnose disturbances, faults and degradations [1], which allows engineering personnel to take the necessary measures to restore normal functioning of the entity.

This article proposes a method of the long-term forecasting of the frame alignment loss probability P_{AL} as the new approach to time/probabilistic characteristics monitoring. The method is based on monitoring the bit error rate ε in a mobile backhaul network that uses Circuit Emulation Service over Ethernet (CESoETH). The developed method performs forecasting of long-term change of the P_{AL} value in order to detect in advance a time instant when the P_{AL} value may exceed its threshold value denoted as P_{ALT}. Frame alignment is supposed to be lost in the time instant when the P_{AL} value exceeds the P_{ALT} threshold value.

II. Analysis of Previous Researches and Publications

ITU Recommendations [1–4] define the various kinds of monitoring including telemonitoring, operational monitoring, state monitoring, quality monitoring, on-demand monitoring, performance monitoring, and proactive monitoring. All these kinds of monitoring immediately react to the faults and degradations after their occurrence. However, there is a definition of the completely new kind of monitoring called the predictive monitoring available in [5]. Predictive monitoring allows to predict emergency situations, which gives engineering personnel extra time to take the necessary actions to prevent the emergency situation or if it is unavoidable to ease the consequences.

Predictive monitoring is now commonly used in industry, healthcare, transport and logistics, environmental protection, and business [6–12]. However, possibilities of using the predictive monitoring for information communication networks are studied insufficiently.

The article [13] proposes a method for short-term forecasting of frame alignment losses in mobile backhauls using CESoETH defined in Technical Specifications MEF 3 [14] and MEF 8 [15]. CESoETH supports both structure-agnostic and structure-aware modes of operation. Unlike the structure-aware mode, an interworking function (IWF) of equipment operating in the structure-agnostic mode encapsulates incoming TDM bit stream into Ethernet frames "as is" without any structural conversions and then frames travel throughout an asynchronous part of a network.

The method for short-term forecasting of frame alignment losses, as a part of a predictive monitoring implementation, predicts the P_{AL} value over a prediction interval L = 1 using bit error rate ε monitoring in a mobile backhaul network. Using the short-term P_{AL} forecast it is possible to increase utilization of an emulated channel by increasing the volume of encapsulated data.

The article [16] describes the $P_{AL} = f(\varepsilon, n_C)$ dependence, which can be obtained for the given bit error rate ε of a TDM stream and a known number n_C of TDM frames encapsulated into an Ethernet frame.

The Recommendation [17] defines Errored Second Ratio − ESR value suitable for the synchronous part of the object under research.

The work [18] contains mathematical tools necessary to calculate bit error rate in order to simulate changes of ε within the wireless part of the research object.

The work [19] brings mathematical tools including polynomial and exponential moving average extrapolations, the Foster-Stuart method along with the method of sequential differences, which are used as a basis of the

developed method for long-term forecasting of frame alignment losses using bit error rate ε monitoring in a mobile backhaul network with CESoETH implemented.

III. TASK STATEMENT

The method of predicative monitoring of frame alignment losses is applied for frame aligners (FA) residing in a base transceiver station (BTS) and in a mobile switching center (MSC) of the research object (Fig. 1). The research object includes synchronous and asynchronous segments. Performance degradation in each segment may affect probability of frame alignment loss.
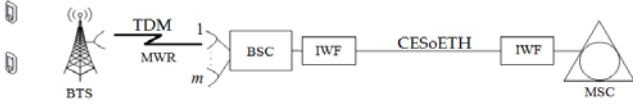

Fig. 1. Functional architecture of the research object

TDM bit stream with the bit rate of 2 Mbit/s is transferred with the help of the microwave radio (MWR) between the BTS and a base station controller (BSC). Between the BSC and MSC synchronous TDM data is carried through asynchronous segment by means of CESoETH. This article considers the case of structure-agnostic CESoETH operation mode because, compared to the structure-aware mode, the research object is exposed to higher risks of frame alignment losses [16].

The purpose of this work is to develop the method of long-term forecasting of the frame alignment losses based on monitoring the bit error rate ε in a mobile backhaul network using Circuit Emulation Service over Ethernet (CESoETH). Unlike the method for short-term forecasting of frame alignment losses [13] the method of long-term forecasting of the frame alignment losses uses bigger prediction interval that incorporates the t_{cr} time instant when a frame alignment loss is supposed to happen. This can be used for making a decision about the volume of TDM data encapsulated into an Ethernet frame.

IV. THE METHOD OF LONG-TERM FORECASTING OF FRAME ALIGNMENT LOSSES BASED ON BIT ERROR RATE MONITORING

The method of long-term forecasting of the frame alignment losses based on bit error rate ε monitoring implies calculation of the n values of the frame alignment loss probability P_{AL} performed according to [13] as:

$$P_{AL}(\mathrm{i}) = \left[1 - \left(1 - \varepsilon_i \right)^a \cdot \left(1 - \beta \cdot n_C \right) \right]^3, \qquad (1)$$

where
a – quantity of bits in the frame alignment signal,
$\beta = ESR \cdot 0,175 \cdot \Lambda / f_0$ – coefficient defining a type of TDM bit stream (f_0 – bit rate of the TDM bit stream; Λ – length of a TDM frame),
n_C – quantity of TDM frames encapsulated into an Ethernet frame,
ε – bit error rate that varies over time due to disturbances over the microwave radio path.

The threshold value $P_{ALT}(n_C)$ of the frame alignment loss probability is obtained for the known n_C value by

substituting $\varepsilon_i = 10^{-3}$ for the corresponding variable of the (1). The value of $\varepsilon_i = 10^{-3}$ is used because ITU-T Recommendation G.706 [20] states that with this value it is almost impossible to distinguish whether Cyclic Redundancy Check errors are caused by the false frame alignment or by transmission bit errors. For the threshold value of the frame alignment loss probability the (1) is rewritten as:

$$P_{ALT}(n_C) = \left[1 - 0,993 \cdot \left(1 - \beta \cdot n_C \right) \right]^3. \qquad (2)$$

In the case of real MWR equipment, the values of bit error rate are obtained through corresponding monitoring but for the research object simulation can be calculated with the following expression [18] considering the Binary Phase Shift Keying implemented in the microwave radio path:

$$\varepsilon_i = \frac{1}{2} \left(1 - erf \left(\sqrt{ \frac{2 \cdot \frac{10^{0,1 \cdot p_c}}{R}}{4,002 \cdot 10^{-21} \cdot k} } \middle/ \sqrt{2} \right) \right), \qquad (3)$$

where
p_c – signal power level at an input of MWR receiver,
R – bit rate of MWR,
k – noise coefficient of the MWR receiver.

The series of $\{P_{AL}(i)\}$ values may have significant fluctuations in long-term perspective. These fluctuations can be made smoother with the help of an exponential moving average obtained as follows [19]:

$$Q(i) = Q(i-1) + \alpha \cdot \left(P_{AL}(i) - Q(i-1) \right), \qquad (4)$$

where
α – coefficient that characterizes a weight of current observation and belongs to the range of $0 < \alpha \le 1$.

It is necessary to detect an increasing or decreasing trend in the series of $\{P_{AL}(i)\}$ values before performing extrapolation. The trend detection is carried out with the help of the Foster-Stuart method [19]. This method associates every $P_{AL}(i)$ value with two variables u_i and l_i. If a current $P_{AL}(i)$ is the biggest one in a series then $u_i = 1$ while $l_i = 1$ if the current $P_{AL}(i)$ is the smallest one in the series. Otherwise $u_i = l_i = 0$. Trend in the series of $\{P_{AL}(i)\}$ values can be proved by testing the following inequality:

$$\frac{\sum_{i=1}^{n} u_i - l_i}{\sigma_2} > t_\alpha; \qquad (5)$$

where
σ_2 – mean-square error of the $\sum_{i=1}^{n} u_i - l_i$ value,
t_α – value of Student t-statistics.

If the inequality (5) is fair then trend in the series of $\{P_{AL}(i)\}$ values is proven with the certain confidence probability.

If the quantity of available P_{AL} values is insufficient for the polynomial extrapolation or there is no trend found in the P_{AL} values then the exponential moving average extrapolation is used to detect the possibility of a frame alignment loss. The possibility of a frame alignment loss is detected with the help of the following inequality:

$$Q(n) + t_\alpha \cdot s \cdot \sqrt{1 + \frac{\alpha}{2-\alpha}} > P_{ALT}(n_C), \qquad (6)$$

where

s − mean square deviation from frame alignment loss probability series obtained considering f degrees of freedom.

If a trend in the series of $\{P_{AL}(i)\}$ has been proven and the quantity of available P_{AL} values is sufficient ($n \geq n_{min}$; where $n_{min} = f(L, \lambda)$ according to [19]) for the polynomial extrapolation then the order λ of a polynomial is determined with the help of the method of sequential differences.

The next step is to use the least squares method [19] to obtain the statistical estimations $a_j \quad j = \overline{0, \lambda}$ of polynomial coefficients. The general form of an extrapolation polynomial is $\tilde{P}_{AL}(i) = a_0 + a_1 \cdot i + a_2 \cdot i^2 + \cdots + a_\lambda \cdot i^\lambda$.

The purpose of long-term forecasting of frame alignment losses is to determine a time instant in a long-term perspective when a frame alignment loss may occur. This can be done by finding a minimum value of i that makes the following expression fair.

$$P_{ALT}(n_C) \leq a_0 + a_1 \cdot i + a_2 \cdot i^2 + \cdots + a_\lambda \cdot i^\lambda + t_\alpha \cdot s. \qquad (7)$$

The minimum value of i that fits expression (7) refers to the time instant t_{cr} of possible frame alignment loss.

It should be noted that simulation involves time shift from event to event where the new event refers to obtaining a new ε_i value.

Fig. 2 depicts the developed method for long-term forecasting of frame alignment losses using bit error rate monitoring in a mobile backhaul network with CESoETH implemented.

Let us take a look at the developed method operation example considering the following initial conditions: 1) p_c changes from −123 dBW to −124.67 dBW; 2) $k = 9$, 3) $R = 2$ Mbit/s; 4) $n_C = 2$; 5) BPSK modulation is used in MWR path; 6) the value of ε_i is updated every second considering Nokia FlexiHopper [21] MWR; 7) trend in the series of $\{P_{AL}(i)\}$ is proven with the confidence probability of 0.95; 8) quantity of values in the series of $\{P_{AL}(i)\}$ is $n = 23$; 9) simulation lasts for 70 seconds; 10) $\alpha = 1$ because there is no need to apply smoothening.
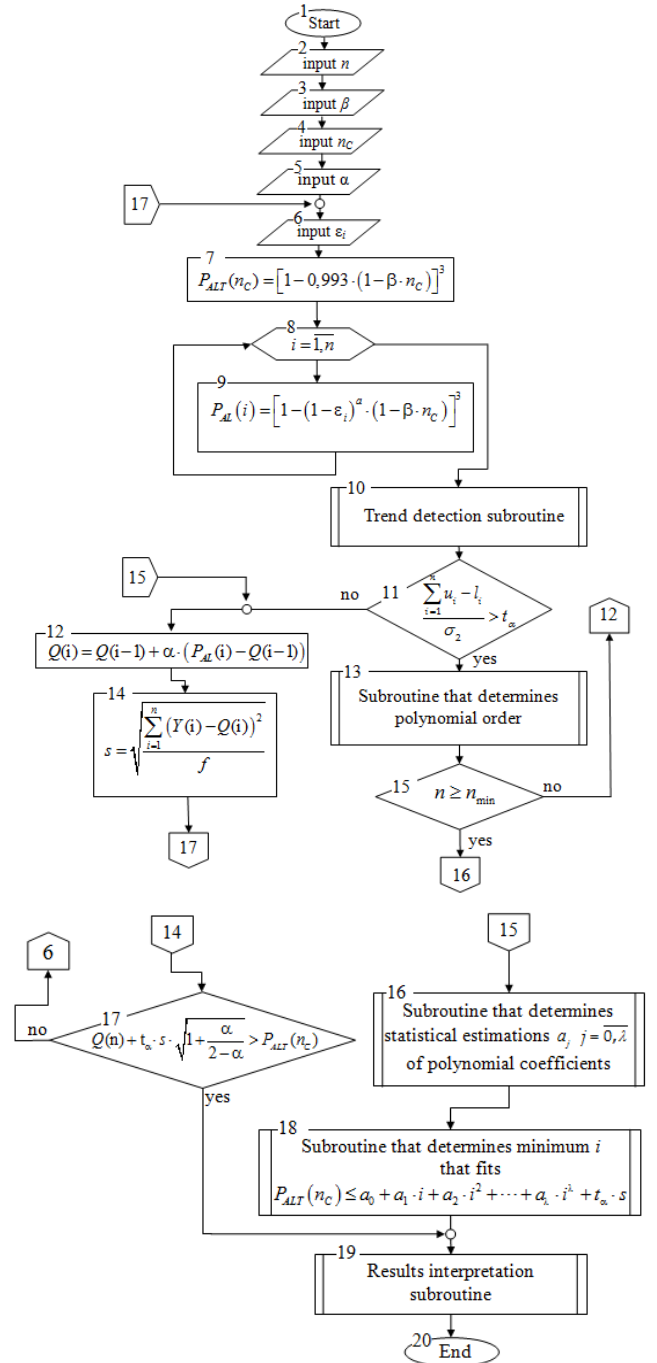
Fig. 2. Algorithm of long-term forecasting of frame alignment losses using bit error rate monitoring in a mobile backhaul network with CESoETH.

The algorithm uses exponential moving average extrapolation till the 23-rd second of simulation. During this period frame alignment loss has not been detected. After the 23 seconds of simulation the algorithm has the quantity of P_{AL} values sufficient for the polynomial extrapolation and trend has been already proven. At this time the algorithm obtains the order $\lambda = 3$ of the polynomial. Then the following statistical estimations of the polynomial coefficients are obtained: $a_0 = 3{,}184 \times 10^{-10}$; $a_1 = 3{,}616 \times 10^{-11}$; $a_2 = 2{,}037 \times 10^{-12}$; $a_3 = 1{,}430 \times 10^{-13}$. This allows to predict frame alignment loss after 68-th second. The actual frame alignment loss happens between 68-th and 69-th second of simulation as it is shown in the Fig. 3.

Fig. 3. Long-term forecasting of frame alignment losses.
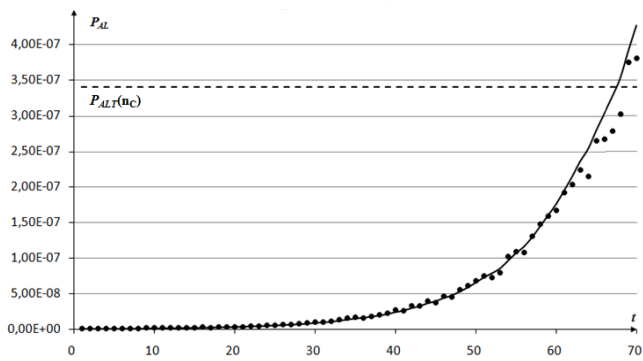
V. CONCLUSION

This article has proposed and formalized with the algorithm the method for long-term forecasting of frame alignment losses using bit error rate monitoring. Being the part of frame alignment losses predictive monitoring implementation for mobile backhauls using CESoETH, the developed method for long-term forecasting of frame alignment losses allows to detect in a long-term perspective the time instants when frame alignment losses are supposed to happen. This allows engineering personnel to make a forehanded decision about the necessary actions preventing the emergency situation or easing possible consequences. Besides, the results of long-term forecasting of frame alignment losses can correct a decision about the volume of TDM data encapsulated into an Ethernet frame based solely on the results of the short-term forecasting of frame alignment losses in mobile backhaul using CESoETH.

The proposed algorithm of long-term forecasting of frame alignment losses along with the method of short-term forecasting of frame alignment losses based on monitoring the bit error rate is acting as an implementation of predictive monitoring of frame alignment loss probability in the case of CESoETH applied in a mobile backhaul network.

REFERENCES

[1] ITU-T Recommendation G.8001 Terms and definitions for Ethernet frames over Transport. (2008). Approved 2008-03-29. Geneva: ITU, 12.

[2] ITU-R Recommendation V.662-3 Terms and definitions. (2005). Approved 2005. Geneva: ITU, p. 19.

[3] ITU-R Recommendation BT.1790 Requirements for monitoring of broadcasting chains during operation. (2007). Approved 2007. Geneva: ITU, p. 6.

[4] ITU-T Recommendation I.113 Vocabulary of terms for broadband access of ISDN. (1997). Approved 1997-06-20. Geneva: ITU, p 35.

[5] E. Voronova, I. Kislukhina, "The Study of the Trends in Social-Economic Systems Based on Predictive Monitoring," Feb. 2012 Available: http://www.uecs.ru/uecs-38-382012/item/1052-2012-02-20-05-48-08.

[6] J. Levitt, "Complete Guide to Predictive and Preventive Maintenance," New York: Industrial Press, Inc., 2005, p. 256.

[7] L. Clifton, D.A. Clifton, M.A.F. Pimentel, P.J. Watkinson, L. Tarassenko "Predictive Monitoring of Mobile Patients by Combining Clinical Observations With Data From Wearable Sensors," IEEE Journal of Biomedical and Health Informatics, Vol.:18, Issue: 3, 225, pp. 722 – 730. DOI: 10.1109/JBHI.2013.2293059.

[8] J. R. Moorman, C.E Rusin, L. Hoshik, L.E. Guin, M.T. Clark, J.B. Delos, J. Kattwinkel, D.E. Lake, "Predictive monitoring for early detection of subacute potentially catastrophic illnesses in critical care," Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2011. Boston, USA Aug.–Sep. 2011. DOI: 10.1109/IEMBS.2011.6091407.

[9] A. Metzger, R. Franklin, Y. Engel, "Predictive Monitoring of Heterogeneous Service-Oriented Business Networks: The Transport and Logistics Case," Annual SRII Global Conference, San Jose, USA July 24-27, 2012. DOI: 10.1109/SRII.2012.42.

[10] M. Franceschinis, F. Mauro, C. Pastrone, M.A Spirito, M. Rossi, "Predictive monitoring of train wagons conditions using wireless network technologies," The 14th International Symposium on Information, Communication and Automation Technologies (ICAT), Sarajevo, Bosna i Hercegovina, Oct-Nov 2013. DOI: 10.1109/ICAT. 2013.6684032.

[11] S.M Zain, Kien Kek Chua, "Development of a neural network Predictive Emission Monitoring System for flue gas measurement," IEEE 7th International Colloquium on Signal Processing and its Applications (CSPA), Penang, Malaysia March 2011. DOI: 10.1109/ CSPA.2011.5759894.

[12] A. Metzger, P. Leitner, D. Ivanovic, E. Schmieders, R. Franklin, M. Carro, S.Dustdar, K. Pohl, "Comparing and Combining Predictive Business Process Monitoring Techniques," IEEE Transactions on Systems, Man, and Cybernetics," Vol.:PP, Issue: 99, 1. DOI: 10.1109/TSMC.2014.2347265.

[13] Y. Babich, L. Nikityuk, "Forecasting changes of frame alignment loss probability in convergent networks," Eastern-European Journal of Enterprise Technologies, 2014, 5(9(71)), pp. 19-24. Available: http://journals.uran.ua/eejet/article/view/28005/25489.

[14] MEF 3 Circuit Emulation Service Definitions, Framework and Requirements in Metro Ethernet Networks, 2004, The Metro Ethernet Forum. Available: http://www.metroethernetforum.org/ Assets/Technical_Specifications/PDF/MEF3.pdf.

[15] MEF 8 Implementation Agreement for the Emulation of PDH Circuits over Metro Ethernet Networks, Oct. 2004, The Metro Ethernet Forum. Available: http://www.metroethernetforum.org/Assets/ Technical_Specifications/PDF/MEF8.pdf.

[16] Y. Babich, L. Nikityuk, "Analysis and optimization of parameters of circuit emulation service in mobile networks," Eastern-European Journal Of Enterprise Technologies, 2013, (9(64)), pp. 59-62. Available: http://journals.uran.ua/eejet/article/view/16396.

[17] ITU-T Recommendation G.826 Error performance parameters and objectives for international, constant bit rate digital paths at or above the primary rate. (2002). Approved 2002-12-14. Geneva: ITU, p. 34.

[18] E. Sukachev, "Sotovie seti radiosviazi s podvizhnymy obektami: handbook," Odessa: ONAS n.a. A. S. Popov, 2013, p. 256.

[19] E. Chetirkin, "Statisticheskie metodi prognozirovania," M.: Statistika, 1977, p. 200.

[20] ITU-T Recommendation G.706 Frame Alignment and Cyclic Redundancy Check (CRC) Procedures Relating to Basic Frame Structures Defined in Recommendation G.704, Approved 1991-04-05. Geneva: ITU, 18.

[21] Nokia FlexiHopper Technical Specifications, June 2014, TelecomConsulting. Available: http:// http://telekom.org.ru/katalog-naimenovanii-res/nokia-flexihopper-7/

Comparative Analysis of Signals Restoration by Different Kinds of Approximation

Irina Strelkovskaya, Olena Lysiuk, Roman Zolotukhin

Odessa National A.S. Popov Academy of Telecommunications –

Institute of Infocommunications and Software Engineering

Kuznechnaya Str, 1, 65029, Odessa, Ukraine

E-mail: dekanat.ik@onat.edu.ua, lysyuk_elena@mail.ru, zolotuhin-roman@mail.ru

Abstract—**The comparative analysis of continuous signals restoration by different kinds of approximation is performed. The software product, allowing to define optimal method of different original signals restoration by Lagrange polynomial, Kotelnikov interpolation series, linear and cubic splines, Haar wavelet and Kotelnikov-Shannon wavelet based on criterion of minimum value of mean-square deviation is proposed. Practical recommendations on the selection of approximation function for different class of signals are obtained.**

Keywords: **restoration, signal, approximation, Kotelnikov series, spline functions, wavelet functions, cubic spline, deviation, software product, mean-square deviation.**

I. INTRODUCTION

Significant attention is paid to the signals restoration in telecommunications. It is connected with the fact that the substantial part of analog signals requires the restoration at different stages of analog-to-digital conversion. Furthermore, to use a single approach to signals restoration is not always possible. Indeed, original signals can differ in forms, make hops or "splashes" on the amplitude. The traditional method of signal restoration is the application of Kotelnikov series. However, the application of Kotelnikov series as the approximation function is effective in the infinite interval of time, while in the real systems it is necessary to examine a signal in a certain limited interval. In this case, there appears an error in signals restoration by means of Kotelnikov series. Therefore, to examine other kinds of approximation is reasonable. The following: Lagrange exponential polynomial, spline function, wavelet function and many others are examined as the approximation functions [1]-[5]. In certain cases [1]-[3] the application of spline functions and wavelet functions allows to reach the best results of continuous signal restoration in the finite time interval, rather than with the application of Kotelnikov series.

The signals restoration by means of Kotelnikov interpolation series, cubic spline and wavelet functions were examined by the authors in works [1]-[5]. In works [1]-[3] the signals of linear frequency modulation, the random signals, as well as the signals, which are characterized by the amplitude steepness were investigated. Restoring such signals by Kotelnikov series and cubic splines, the estimations of restoration error are obtained and it is shown that under specific conditions the application of cubic splines is more reasonable. In works [4][5] it is found that for some signals the application of wavelet functions as approximation ones is more preferencial in comparison with the use of Kotelnikov series and cubic splines. The task to select the optimum method of signals restoration, depending on their original kind, emerges.

In this work let us consider the restoration of different signal kinds by means of the above mentioned methods, that will allow us to give recommendation on the application of one or another restoration method depending on the examined signal.

The objective of this work is the comparison of the signals restoration methods with the use of different approximation functions for their more precise restoration.

To achieve the set objective let us find errors of original signal restoration by means of different approximation functions. As the criterion we will use a minimum value of mean-square deviation between the original signal and the approximation function. We will consider Kotelnikov interpolation series, Lagrange polynomial, spline function and wavelet function as the approximation functions. For this purpose let us develop the program realization, which will allow to investigate different methods of signals restoration. Based on the developed program, let us carry out investigation of different signals restoration and develop practical recommendations on the selection of the approximating device.

II. CONTINUOUS SIGNALS RESTORATION BY KOTELNIKOV SERIES AND SPLINE FUNCTIONS

Let us consider the continuous signal $f(x)$ in the interval $0 \le x \le 1$, that is given by the function of the form [2][3]:

$$f(x) = 2\frac{\sin(16\pi(x-\frac{1}{2}))}{16\pi(x-\frac{1}{2})} - 3\frac{\sin(8\pi(x-\frac{1}{2}))}{8\pi(x-\frac{1}{2})}, \qquad (1)$$

It is considered that $f(x)$ is limited on the spectrum $F_{\max} = 8$ kHz, then the sample spacing $\tau = \dfrac{1}{2F_{\max}}$.

Let us consider samples of the original signal $f(x)$ as interpolation points. Let be given function values as $f_k = f(x_k)$, $k = \overline{0, N-1}$, in the interval [0;1] in mesh points Δ: $0 = x_0 < x_1 < ... < x_N = 1$. The given original signal $f(x)$ belongs to the class $|f(x)| < \dfrac{A}{|x^2|}$, where $f(x)$ is original signal, A is a certain constant. To restore the signals of such

class, Kotelnikov interpolation series and cubic spline were examined in works [2][3].

Fig.1 illustrates the obtained results of given above signals restoration, examined in the works [2][3], particularly: diagram of the given signal $f(x)$ (diagram 1), values of function $f(x)$ in the mesh points (diagram 2), the diagram of the restored signal by Kotelnikov series (diagram 3), the diagram of the restored signal by the cubic spline (diagram 4).
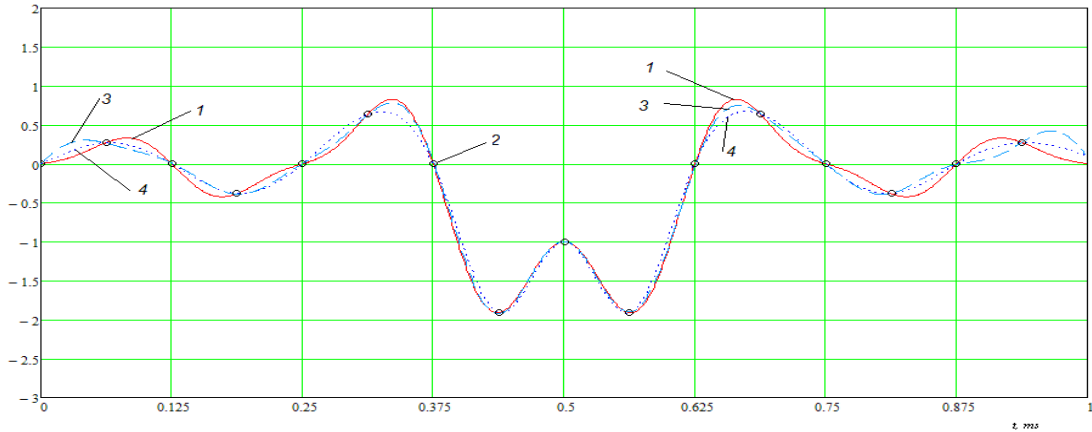

Fig. 1 – Original continuous signal $f(x)$ and its restoration by Kotelnikov series and c-spline.

The obtained results [2][3] show that in the intervals of interpolation "with amplitude smooth change", where the form of the signal continuously changes without "amplitude steepness" the best results are given by spline approximation, and in the intervals with "amplitude steepness" – by Kotelnikov interpolation series. As it is shown in [2], the approximation error by application of Kotelnikov series and cubic spline can reach 6.2% and 9.2% in the first and the second cases respectively due to the fact, that with the uniform sampling the sample points can not coincide with the amplitude maximums or the minimums. It is found that the use of cubic spline as the approximation function during continuous signals restoration is reasonable for the signals, which do not have bumping parameters changes. At the same time, it is much more difficult to select the approximation function for the signals, which possess the amplitude hops and "rapid" oscillations.

Therefore it would be reasonable to consider another restoration method, for which this approximation error would be less. In our opinion, such method is wavelet transform device.

III. CONTINUOUS SIGNALS RESTORATION BY WAVELET FUNCTIONS

Recently special development was given to the wavelet analysis device in information theory, theory of coding, theory of signals and images. To the number of the major advantages of the wavelet functions application as the approximation function during the signals restoration can be referred the following [6][7]:

– the signals restoration with the low information losses, in other words, distortion in the restored signal, appearing as

a result of quantization, while using the wavelet functions can be made comparatively small and thus to elliminate the noise interferences;

– the ability of wavelet functions to reveal the local special features of the investigated signal, in particular for the pulse and digital signals and the images (point of discontinuity, peaks, sharp drops and self-similarity), which miss other methods of signals restoration, allows to reproduce the most important characteristics of the original signal;

– the use of wavelet functions during the signals restoration allows to obtain information about the nature of the parameters change of the investigated function, that is particular important in examining the non-stationary signals, for example, consisting of different components, which act in different time intervals, the modulated signals, etc., which nowadays find much wider application than stationary or quasi-stationary signals, as well as processes and systems generating them;

– the possibility of the independent analysis of the function on the different scales of its change, because, each frequency signal component is additionally described by the scaling function, that allows to determine the time interval, where a certain signal frequency characteristic changed.

Wavelet transform allows to perform the signals restoration both in the temporary parameters with the basic wavelet function $\psi(x)$, and in the frequency parameters with the scaling function $\varphi(x)$ [6]:

$$\psi(x) = \sqrt{2} \sum_k g_k \varphi(2x - k), \ \varphi(x) = \sqrt{2} \sum_k h_k \varphi(2x - k), \quad (2)$$

where

$$h_k = \sqrt{2} \int \varphi(x) \overline{\varphi}(2x-k)dx , \; g_k = (-1)^k h_{(N-1)-k-1} -$$

expansion coefficient, k, N – natural numbers.

All the wavelets are obtained from the basic wavelet with the scaled transformation $1/2^k$ and shifts $j/2^k$ [6]:

$$\psi_{j,k} = 2^{j/2} \psi(2^j x - k) , \; \varphi_{j,k} = 2^{j/2} \varphi(2^j x - k) . \quad (3)$$

Then, according to [6], any function $f(x)$ from Hilbertian space $L^2(R)$ can be expressed in the series of the form:

$$f(x) = \sum_k s_{j_n,k} \varphi_{j_n,k}(x) + \sum_{j \geq j_n,k} d_{j,k} \psi_{j,k}(x) , \quad (4)$$

where

$f(x)$ – temporary function of the original signal;

$\varphi_{j,k}(x)$ – scaling function (father wavelet);

$\psi_{j,k}(x)$ – basic wavelet function (mother wavelet);

$s_{j,k}$ and $d_{j,k}$ – wavelet coefficients;

j_n – scaling level;

k, j, j_n – natural numbers.

Wavelet coefficients $s_{j,k}$ and $d_{j,k}$ for the expression (4) are calculated as follows [6]:

$$s_{j-1,k} = \sum_m h_m s_{j,2k+m} , \; d_{j-1,k} = \sum_m g_m s_{j,2k+m} , \quad (5)$$

where

h_m, g_m – approximation and expansion coefficients.

On the most detailed level of scaling $j = j_{max}$ there left only s coefficients and the signal is given by the scaling function [6]:

$$f(x) = \sum_k s_{j_{max},k} \varphi_{j_{max},k}(x) . \quad (6)$$

In the case of signal restoration on the most detailed level, i.e. by using the formula (6), coefficients $s_{j_{max}}$ are calculated as follows [6]:

$$s_{j max,n} = f(n\tau) / 2^{\frac{j max}{2}} . \quad (7)$$

where

τ - sampling interval.

Let us consider the restoration of the continuous signal $f(x)$ of the form (1) by wavelet function. To restore the signal let us use Kotelnikov-Shannon wavelet for which basic wavelet function $\psi(x)$ and scaling function $\varphi(x)$ take the following form (Fig. 2, a) and (Fig. 2, b) relatively:
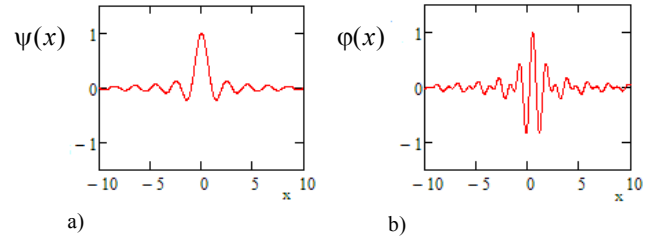

a) b)

Fig. 2 The basic wavelet function $\psi(x)$ a) and scaling function $\varphi(x)$ b) for Kotelnikov-Shannon wavelet

For Kotelnikov-Shannon wavelet, according to [7], analytic note for its basic function $\psi(x)$ and scaling function $\varphi(x)$ t akes the form:

$$\psi(x) = 2 \frac{\sin 2\pi(x-1/2) - \sin \pi(x-1/2)}{\pi(x-1/2)} ,$$

$$\varphi(x) = \begin{cases} \sin \pi x / \pi x, & x \neq 0, \\ 1, & x = 0, \end{cases} \quad (8)$$

Mother wavelet $\psi(x)$ and scaling function $\varphi(x)$ are connected by the relations (2), besides, for Kotelnikov-Shannon wavelet [7]:

$$h_k = \begin{cases} 1/\sqrt{2}, & k = 0, \\ (-1)^{(k-1)/2} \sqrt{2} / \pi k, & k - \text{odd}, \\ 0, & \text{for the rest } k \end{cases} \quad (9)$$

$$g_k = (-1)^k h_{N-1-k}, \; k = \overline{0, N-1} ,$$

where

N – natural number, which defines the coefficients h_k and g_k number.

The necessary condition of the signal restoration by the wavelet function is its definition by samples number that is equal to $N = 2^j$, where $j \geq 1$ defines the maximum possible number of scaling levels [6]. In our case $j = j_{max} = 4$, $N = 2^4 = 16$.

Lets us calculate the h_k and g_k expansion by the formula (9).

The obtained values are given in the table 1.

TABLE I

VALUES OF h_k AND g_k COEFFICIENTS

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
h_k	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{2}}{\pi}$	0	$-\frac{\sqrt{2}}{3\pi}$	0	$\frac{\sqrt{2}}{5\pi}$	0	$-\frac{\sqrt{2}}{7\pi}$	0	$\frac{\sqrt{2}}{9\pi}$	0	$-\frac{\sqrt{2}}{11\pi}$	0	$\frac{\sqrt{2}}{13\pi}$	0	$-\frac{\sqrt{2}}{15\pi}$
g_k	h_{15}	0	h_{13}	0	h_{11}	0	h_9	0	h_7	0	h_5	0	h_3	0	h_1	$-h_0$

Table 2 shows the number values of $s_{j,k}$ and $d_{j,k}$ coefficients, obtained according to (5), besides, $s_{0,k}$ coefficients are values of function $f_k = f(x_k)$ in the interpolation points.

Lets us perform $f(x)$ signal restoration by the formula (4) [6], using Kotelnikov-Shannon wavelet, for which $\psi(x)$ and $\varphi(x)$ are defined by the formula (8) [7].

TABLE II
VALUES OF $s_{j,k}$ AND $d_{j,k}$ COEFFICIENTS

	k	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
j=1	$s_{j,k}$	0,217	0	0	-0,478	-1,714	0	0	0,123
	$d_{j,k}$	-0,048	0,068	-0,113	0,338	0,161	-0,113	0,068	-0,048
j=2				k	0	1	2	3
				$s_{j,k}$	0,134	-0,026	-1,065	-0,095
				$d_{j,k}$	0,082	0,318	-0,443	-0,082
j=3						k	0	1
						$s_{j,k}$	0,097	-0,796
						$d_{j,k}$	0,103	-0,191
j=4							k	0
							$s_{j,k}$	-0,29
							$d_{j,k}$	0,062

The results of the original signal $f(x)$ restoration by Kotelnikov-Shannon wavelet [4][5] are obtained in the Mathcad 15 environment and are shown on the Fig. 3. As it is shown on the Fig. 3 the original signal $f(x)$ restoration by Kotelnikov-Shannon wavelet allowed to obtain the result that corresponds to the kind of the original signal [4][5]. It is allows to make the conclusions that it is the most reasonable to use wavelet functions as the approximation function during the signals restoration, which are characterized by the presence of rapid oscillations.

It is possible intuitively to see that the signals with the abrupt changes must be analyzed rather by the wavelet functions than by the classical approach on the basis of Kotelnikov series, that is used as the basic $\sin x$ function.
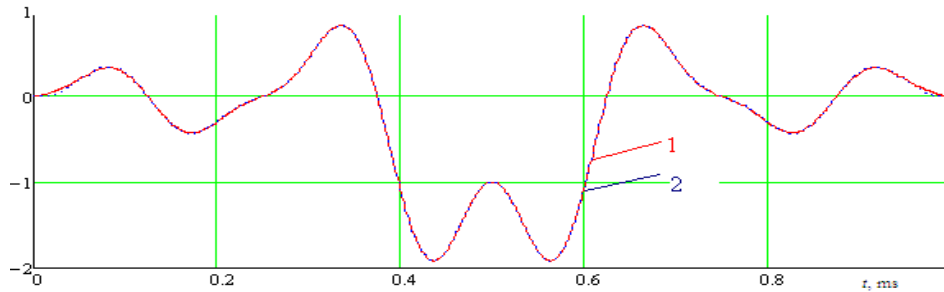

Fig. 3. Restoration of the original signal $f(x)$ by Kotelnikov-Shannon wavelet.

Let us compare the results obtained during the signal $f(x)$ restoration, that is characterized by the amplitude higher steepness, using Kotelnikov series, cubic spline and Kotelnikov-Shannon wavelet.

IV. ANALYSIS OF THE RESULTS OF CONTINUOUS SIGNALS RESTORATION BY DIFFERENT APPROXIMATION FUNCTIONS

To compare the results of the original $f(x)$ signal restoration by Kotelnikov-Shannon wavelet [4][5] let us consider the results obtained earlier in works [2][3], the approximation of the same signal by Kotelnikov series and cubic spline.

Fig. 4 illustrates [4][5] the signal restoration of the form (1) in the interval $[0;1]$ by means of different approximation functions, and Fig. 5 – in the interval $[0,625; 0,75]$: the diagram of the given signal $f(x)$ (diagram 1), the approximation of the original signal $f(x)$ by Kotelnikov series (diagram 2), the approximation of the original signal $f(x)$ by cubic spline (diagram 3), the approximation of the original signal $f(x)$ by Kotelnikov-Shannon wavelet (diagram 4).

As we can see from the Fig. 4 and 5, the proposed method of continuous signals restoration, which are characterized by the amplitude abrupt changes, based on wavelet functions allows to obtain the best results of restoration (curve 4), rather than restoration methods by Kotelnikov series (curve 2) and by cubic spline (curve 3) [4][5].
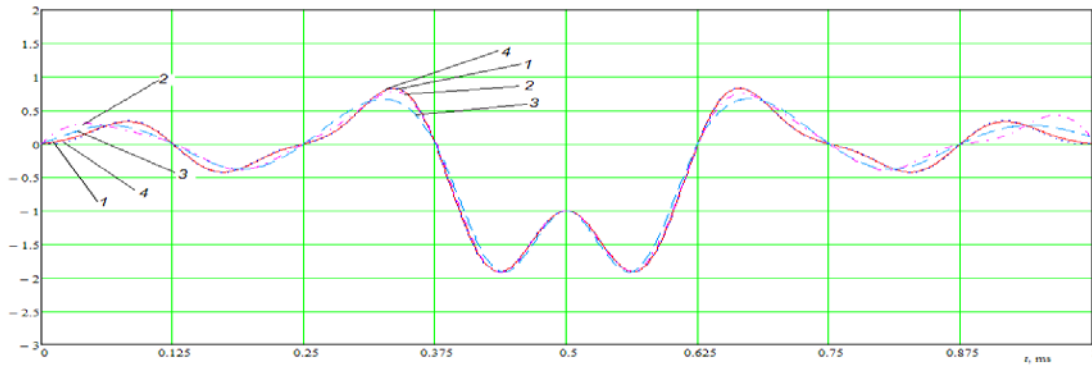
Fig. 4. Restoration of the original signal $f(x)$ by Kotelnikov series, cubic spline and Kotelnikov-Shannon wavelet in the interval $[0;1]$.
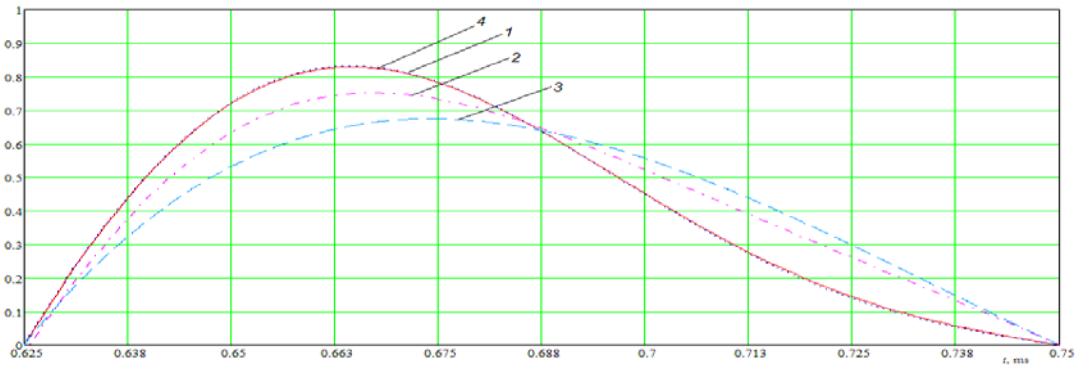

Fig. 5. Restoration of the original signal $f(x)$ by Kotelnikov series, cubic spline and Kotelnikov-Shannon wavelet in the interval $[0,625;0,75]$.

To estimate the fact how precisely one or another method (Kotelnikov series, cubic spline or Kotelnikov-Shannon wavelet) restores the original signal, let us define the value of deviation ε_n of the restored signal values form the original signal.

The results are given on the Fug.6: diagram ε_n with the approximation of the original signal $f(x)$ by Kotelnikov series (curve 1), diagram ε_n with the approximation of the original signal $f(x)$ by cubic spline (curve 2), diagram ε_n with the approximation of the original signal $f(x)$ by Kotelnikov-Shannon wavelet (curve 3).
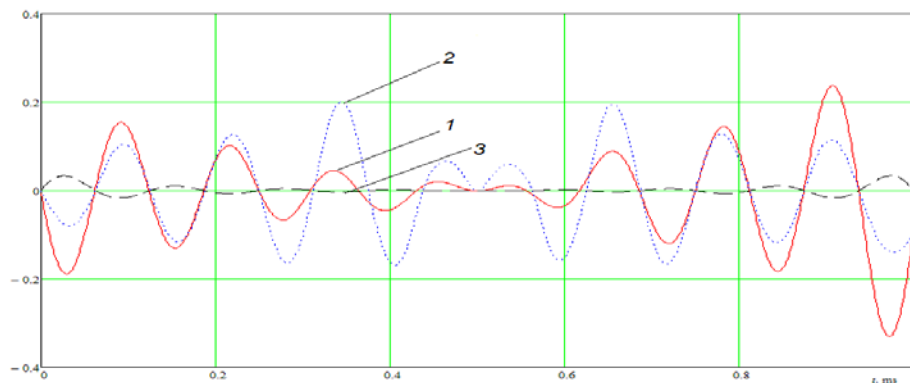

Fig. 6. Deviation of the restored signal from the original one.

Similarly to [2], let us define the mean-square deviation (MSD) of the restored and original signals difference. The results of the MSD values of the restored and original signals are given in table 3. As we can see from Fig. 4-6, the application of Kotelnikov series for continuous signals restoration, which are characterized of the higher steepness of the amplitude and a sharp function increase, leads to the global signal smoothing and therefore local peaks at the moment of abrupt changes in the signal amplitude are restored badly.

The analogous approximation results are given by the application of cubic spline. An approximation error in this case reaches 6% [2]. However, the wavelet transform application does not prevent the restoration of these peaks due to the property of scalability.

Fig. 7 illustrates the dependence of mean-square deviation on the number of scaling level in the signal $f(x)$ restoration by Kotelnikov-Shannon wavelet.

TABLE III

DEVIATION VALUES ε_n AND MSD σ_n OF THE RESTORED SIGNAL BY APPROXIMATION FUNCTIONS (KOTELNIKOV SERIES, CUBIC SPLINE AND KOTELNIKOV-SHANNON WAVELET)

№	Interval	Number limits of the interval	ε_n Kotelnikov series	ε_n cubic spline	ε_n Kotel'nikov-Shannon wavelet	σ_n Kotelnikov series	σ_n cubic spline	σ_n Kotel'nikov-Shannon wavelet
1	$[t_0, t_1]$	[0; 0,063]	0,119	0,05	0,021	0,059	0,025	0,01
2	$[t_1, t_2]$	[0,063; 0,125]	0,098	0,066	-0,011	0,048	0,032	0,0051
3	$[t_2, t_3]$	[0,125; 0,188]	0,083	0,074	0,0065	0,041	0,036	0,0032
4	$[t_3, t_4]$	[0,188; 0,25]	0,066	0,082	0,0044	0,031	0,038	0,0021
5	$[t_4, t_5]$	[0,25; 0,313]	0,039	0,103	0,0029	0,023	0,051	0,0014
6	$[t_5, t_6]$	[0,313; 0,375]	0,023	0,128	0,0019	0,02	0,061	0,0092
7	$[t_6, t_7]$	[0,375; 0,438]	0,027	0,106	0,0011	0,017	0,053	0,00055
8	$[t_7, t_8]$	[0,438; 0,5]	0,012	0,041	0,00035	0,0072	0,022	0,00022
9	$[t_8, t_9]$	[0,5; 0,563]	0,0048	0,036	0,00035	0,0049	0,02	0,00022
10	$[t_9, t_{10}]$	[0,563; 0,625]	0,022	0,099	0,0011	0,015	0,049	0,00053
11	$[t_{10}, t_{11}]$	[0,625; 0,688]	0,059	0,121	0,0018	0,026	0,062	0,00095
12	$[t_{11}, t_{12}]$	[0,68; 0,75]	0,079	0,107	0,0029	0,035	0,05	0,0014
13	$[t_{12}, t_{13}]$	[0,75; 0,813]	0,091	0,081	0,0043	0,046	0,04	0,0021
14	$[t_{13}, t_{14}]$	[0,813; 0,875]	0,117	0,076	0,0012	0,056	0,036	0,0031
15	$[t_{14}, t_{15}]$	[0,875; 0,938]	0,149	0,072	-0,01	0,075	0,037	0,0054
16	$[t_{15}, t_{16}]$	[0,938; 1]	0,225	0,102	0,02	0,092	0,038	0,0065

Thus, we can state, that the smallest restoration error by the wavelet functions can be obtained with the maximum scaling level, and an increase of the maximum scaling level is possible with the samples number increase.
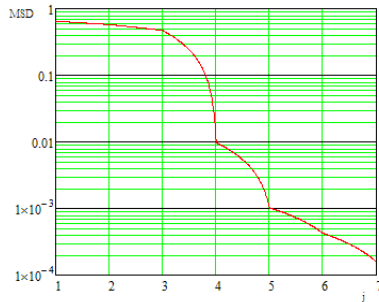

Fig. 7. Dependence of the MSD on the number of scaling levels.

One of the tasks be solved with the use of wavelet functions to restore the continuous signals, is what samples number is necessary to take so that the original and restored signals deviation would be minimum with the set computational volume.

V. SOFTWARE REALIZATION OF THE SIGNALS RESTORATION BY DIFFERENT APPROXIMATION FUNCTIONS

To compare the methods of signals restoration by means of different approximation functions (Kotelnikov series, Lagrange polynomial interpolation, linear spline functions, cubic spline functions, as well as wavelet functions) the software realization "*Signal Restoration*", which realizes the mentioned methods of the signals restoration is developed [5]. Software product is executed by means of framework Q_t with using the C++ programming language. The control-flow chart is given in Fig. 8.
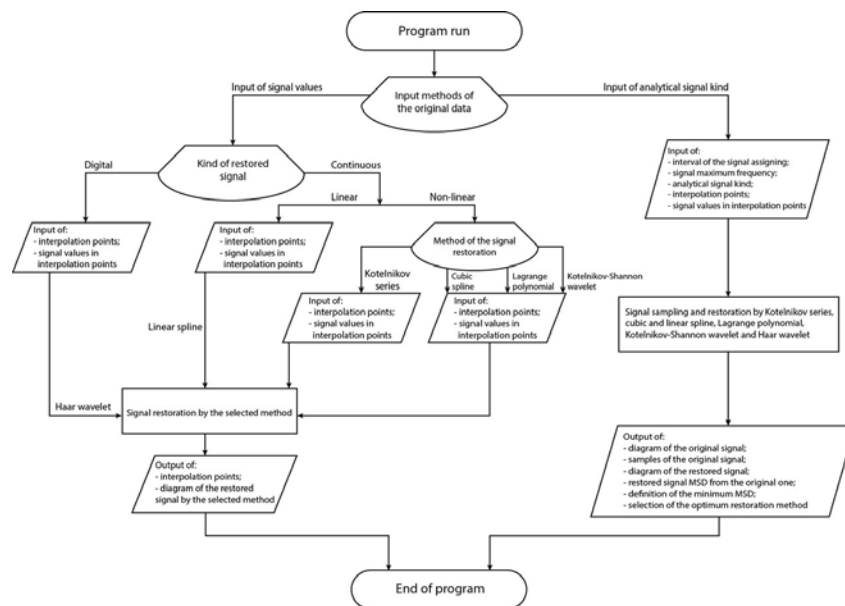
Fig. 8. Algorithm diagram of the *«Signal Restoration»* program.

Functionally, the given product allows to perform the signals restoration of different classes (both continuous and discrete) by all indicated methods, as well as to obtain the estimation of restoration error according to the MSD criterion of the restored signal from the original one. The best method of signal restoration based on the obtained MSD is chosen.

The algorithm of the given software product consists in the following:

1. Input of the original data about the signal (to set a signal by means of the samples or in the analytical form).

2. Selection of the signal restoration method (by means of Kotelnikov series, Lagrange polynomial interpolation, linear spline functions, cubic spline functions, Haar wavelet or Kotelnikov-Shannon wavelet).

3. The signal restoration by the selected method.

4. MSD definition for each restoration method.

5. Conclusion of the obtained results.

Linear frequency-modulated signal (LFM signal), LFM signal with the linearly changing amplitude (modified LFM signal), frequency-modulated signal (FM signal) and the signal, which is characterized by the amplitude higher steepness and sharp function increase were considered for the analysis.

In particular, Fig. 9 illustrates the results of frequency-modulated signal restoration by means of Kotelnikov series, Lagrange polynomial interpolation, linear spline functions, cubic spline functions, Haar wavelet and Kotelnikov-Shannon wavelet [5].
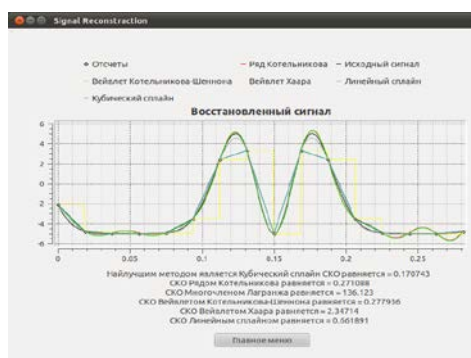

Fig. 9. Restoration of FM signal.

By means of the software product the following results and recommendations regarding the use of the specific approximation functions depending on the original signal kind, i.e., its class, on the MSD criterion are obtained [5]:

− continuous FM signal is reasonable to restore by means of cubic spline, for which MSD is approximately equal to 0,171;

− during the LFM signal restoration, it turned out that it is reasonable to use Kotelnikov-Shannon wavelet, in this case MSD is approximately equal to 0,169;

− in case of the restoration of the LFM signal with the linearly changing amplitude, it is recommended to use Kotelnikov series as the approximation function (MSD is approximately equal to 0,125);

− for the continuous signals, which are characterized by the presence of a large number of oscillations and sharp function increase, it is recommended to use wavelet function based on Kotelnikov-Shannon wavelet;

− as for Haar wavelet, it is necessary to use only for the signals restoration of the "squared pulse" type.

VI. Conclusion

1. As a result of carried out analysis of the continuous signals restoration on their samples by Kotelnikov series, cubic spline and Kotelnikov-Shannon wavelet it is shown, that the first two restoration methods do not consider such special features of signal as the amplitude abrupt change, which leads to the restored signal values deviation from its original value. The use of Kotelnikov-Shannon wavelet allows to decrease this deviation. This implies, that the mathematical device of wavelet analysis can be used for the signals restoration, which are characterized by the presence of rapid oscillations.

2. The estimations of the restoration error of different signal classes by the functions, pointed out above, are obtained, which allowed to speak about the advantage of a certain method of the signals restoration depending on its kind, i.e., the class of signals.

3. The program, which allows to find the restoration errors of different signals based on the MSD minimum value criterion by means of Lagrange polynomial, Kotelnikov interpolation series, linear and cubic splines, Haar wavelet and Kotelnikov-Shannon wavelet is developed.

4. The practical recommendations regarding the selection of the approximation function based on the mean-square functions deviation criterion for different signal kinds are obtained. The results of work with the software realization "*Signal Restoration*" allow to carry out the substantiated selection of the optimum restoration method of a certain original signal.

References

[1] I. Strelkovskaya, D. Bukhan "Continuous signals restoration based on Kotelnikov series and cubic splines," Radioelectronics: All-Ukr. Interdep. scien.-resear. coll. – 2007. – Iteration. 151. – P. 181-185.

[2] I. Strelkovskaya, O. Lysiuk, R. Zolotukhin, "Comparative analysis of continuous signals restoration by Kotelnikov series and spline-functions," Eastern-European Journal of Enterprise Technology. – 2013. – №2/9(62). – P. 12-15.

[3] I. Strelkovskaya, O. Lysiuk, R. Zolotukhin, "Investigation of continuous signals restoration methods," The 7th International Scientific Research Conference "Telecommunications Problems": Conference Proceedings., April 16-19, 2013,–K.NTUU"KPI" Kyiv, 2013. – P. 82–84.

[4] I. Strelkovskaya, O. Lysiuk, R. Zolotukhin, "Continuous signals restoration based on wavelet-functions," The 1st International Scientific Research Conference "Infocommunications Problems. Science and Technologies" (PIC S&T-2013). October 9-11, 2013,- KNURE, TCN dep. – p. 223-225.

[5] I. Strelkovskaya, O. Lysiuk, R. Zolotukhin, "Software development of the different signal restoration methods," Proc. of the 8th International Scientific Research Conference "Telecommunications Problems". April 22-25, 2014,–K.NTUU"KPI" Kyiv, 2014. – P. 128–130.

[6] I. Dremin, O. Ivanov, V. Nechitailo, "Wavelets and their application," Progress of physic sciences. – 2001. T.171. – №5. – p. 465-501.

[7] N. Smolentsev, "Fundamentals of wavelets theory," Wavelets in MATLAB – M.: DMK Press, 2005. – 304.

Software for Explicitly Parallel Memory-Centric Processor Architecture

Goce Dokoski, Danijela Efnusheva, Aristotel Tentov, Marija Kalendar

SS. Cyril and Methodius University - Faculty of Electrical Engineering and Information Technologies

Karpos II bb, PO Box 574, 1000 Skopje, Macedonia

E-mail: {gocedoko, danijela, toto, marijaka}@feit.ukim.edu.mk

Abstract— **Advances in computer memory technology justify research towards new and different views on computer organization. This paper proposes a novel memory-centric computing architecture with the goal to merge memory and processing elements in order to provide better conditions for parallelization and performance. The paper introduces the architectural concepts and afterwards shows the design and implementation of a corresponding assembler and simulator.**

Keywords: **Explicit parallelism, memory-centric, high-performance computing, assembler, simulator.**

I. INTRODUCTION

The most notable characteristic of a classical Von-Neumann- and Harvard – based computer architecture is the clear separation between memory and processing: a computer system consists of memory that keeps program code and data; and processing units that read program code, load data from memory and process it. In this context, it is important to note that the difference between memory and processing speed has always imposed a bottleneck in computing throughput [1], which has led to development of various caching solutions. However, memory speed is significantly improved now. For example [2] shows a DDR3 memory that achieves a working frequency of 1.8 GHz.

This justifies research towards new concepts in computer architectures that would make closer ties among computing and memory resources. It should be mentioned that current research is mainly focused toward parallel processing as processor technology slowly reaches its upper bounds on chip complexity and speed.

Our research focuses on achieving stronger merge between memory and processing units, i.e. incorporating processing hardware directly in memory. As part of the memory is always used to store program code and the rest is used for data, it makes sense to add execution hardware into the code parts. Also if the need for virtual memory is taken into account, and the memory hardware is observed as a set of blocks, then it would be convenient to have an execution unit in each memory block that contains program code. This allows completely concurrent execution of all programs that fit in memory at a given time. The only issues that need to be considered in this organization are data inter-dependence and synchronization among running threads, and they can be resolved by using an appropriate interconnection system.

A single self-executing unit is solely responsible to fetch and decode instructions from its code block, and then issue data movement and arithmetic commands over the bus so that they are performed on the data blocks. Since commands are executed directly over the memory data, there is no need for processor registers and they are not used. The removal of processor registers makes a significant simplification in the way programs are written, compiled and executed. This paper presents the influence these changes will make on processor performance.

The explicitly parallel instruction computing (EPIC) architecture as a concept implies that compiler is responsible for detection of instruction-level parallelism (ILP) in the high-level program code, embedding this information into machine-code executables and possibly even preparing a complete execution plan [3]. This relieves the burden on hardware for various processor features such as: dependence checking, branch prediction, pipeline hazards, out-of-order execution etc. Although, this implies great simplification for the EPIC hardware, at the same time achieving great performance improvements, still its most notable implementation – Intel IA-64, hasn't achieved a significant success. The presented architecture will incorporate some of the EPIC concepts, probably resulting in more intensive data exchange between the memories and computing elements, thus forcing better usage of computational hardware elements. Finally, that will hopefully result in better acceptance of the EPIC philosophy.

This paper presents and examines an instruction set architecture appropriate for the suggested processor architecture. As a proof of concept we show an instruction-level simulator design for the architecture and evaluate its performance. The following chapter presents the state-of-the-art in the field of parallel computing architectures. The next two chapters describe the memory-centric computing architecture and its instruction set architecture. The fourth chapter presents the corresponding simulator, followed by simulation results and their analyses. The paper ends with conclusion and future work.

II. STATE OF THE ART

The research in the field of computer architectures is usually focused on increasing their throughput. The most common ways to deal with this issue include efficient data

organizations and optimal interconnection between memory and processing units.

Memory-centric computer architectures with special attention on connectivity issues are proposed in [4]-[5]. Here the computer system is viewed as a reconfigurable array of processing units on one side, and memory blocks on the other side, all interconnected by direct communication links. The processing units and the memory blocks have 1-1 mapping, and their interconnection is realized by interconnection circuit (crossbar switch). This organization shows an improvement of around 75% in regard to classical computer organizations.

In [6] centralized memory pools are proposed that are designed for transaction-based applications. In order to allow optimal data exchange, they introduce specialized three-dimensional interconnection network.

Large numbers of papers also investigate adjustable models for memory-centric architectures [7]-[12][13] discusses the advantages and the reasons to use memory-centric architectures, and proposes various implementation options. [14]-[16] on the other hand give proposals for hardware realizations of different memory-centric architectures.

III. ARCHITECTURE DETAILS

As previously mentioned, the main idea of the proposed architecture is to merge the processing and memory units. This is done by implementing the processor's fetch and decode phases directly in the program part of the memory i.e. exactly where the instructions are located. Furthermore, the complete fetch/decode circuitry may be added to every program memory block, and in case this is too expensive, it can also be added to a group of blocks. So, one block or a group of blocks with execution capabilities will be called a *self-executing memory block* in the rest of the text.

Therefore, a memory-centric EPIC architecture as proposed in this paper consists of self-executing memory blocks, data memory blocks and their interconnection network.
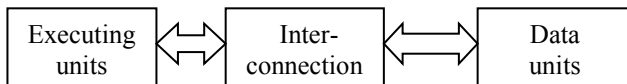

Fig. 1. Memory-centric computer architecture.

Every active self-executing block fetches and decodes its instructions and issues them to the memory blocks and to the ALUs. Whenever a new process is created, the system scheduler will assign a free block if available, or will free one in case all are busy.

Given that instructions are executed in-memory, the need for classical register becomes obsolete, and the instructions will work directly with memory locations. In order to decrease instruction length, it is a good idea to use a concept similar to segmentation. In this case, every executing block can work with one associated data block at a time, and the memory operands will represent address offsets in the data page instead of complete addresses. When data from another page is needed, the re-association will be performed by a special instruction. It should be noted that in this context, the terms block and page are used interchangeably.

Having only one associated data page at a time will impose problems when a thread accesses data parts that are interleaved over several data pages, because this will result in frequent data page re-associations, so it is a compiler's responsibility to provide an efficient data-to-page mapping and data manipulation algorithm. This is not a trivial task, and there is a lot of research that needs to be done in order to resolve this issue.

As mentioned previously, the goal of this research is to observe memory as a set of memory pages with execution capabilities that are interconnected by an appropriate bus system. It should be noted that this organization is especially convenient for FPGA implementation, because FPGA technology is already designed as a reconfigurable network of small memory and processing blocks. The only issue is the limit on achievable working frequency of the FPGA technology as well as its relatively high cost per implemented logic. Therefore, FPGA should be used for prototyping and afterwards the final product should be produced as integrated circuit.

IV. INSTRUCTION SET ARCHITECTURE

The initial instruction set has simple MIPS-like instructions for arithmetical-logical operations, program flow control as well as auxiliary operations. Instruction operands always represent memory addresses that can address data either directly or indirectly. The instruction formats and their functions are shown in Table 1.

TABLE I
INSTRUCTION SET ARCHITECTURE

	Operation	Instruction format
arithmetical-logical	Addition	add dest, op1, op2
	Subtraction	sub dest, op1, op2
	Multiplication	mul dest, op1, op2
	Division	div dest, op1, op2
	Logical Bit-wise AND	and dest, op1, op2
	Logical Bit-wise OR	or dest, op1, op2
	Logical Bit-wise XOR	xor dest, op1, op2
	Logical Bit-wise NOT	not dest, op1
program flow control	Branch to dest if op1 is equal to op2	beq dest, op1, op2
	Branch to dest if op1 is not equal to op2	bne dest, op1, op2
	Branch to dest if op1 is greater than op2	bgt dest, op1, op2
	Branch to dest if op1 is less than op2	blt dest, op1, op2
	Branch to dest if op1 is greater than or equal to op2	bge dest, op1, op2
	Branch to dest if op1 is less than or equal to op2	ble dest, op1, op2
	Branch to dest	b dest
auxiliary	Terminate running thread and deactivate page	hlt
	Load byte op to dest	lb dest, op
	Load 2 bytes op to dest	lh dest, op
	Load 4 byte word op to dest	lw dest, op

All operations by default use direct addressing. Additionally, any operand may be set to use indirect addressing, by surrounding it in square brackets. For example, the instruction

add [000F0000h], 00FF0004h, [000F0008h]

will perform addition of the word on address 00FF0004h, and the word on address that is specified by the word on address 000F0008h. The result than will be written to the address specified by the word that resides on address 000F0000h.

Additionally, every instruction may accept an immediate value as a second operand. This is specified by adding *i* to the opcode (*addi*, *subi*, *muli* etc.) For example,

muli 000F0000h, 00FF0004h, 000F0008h

will calculate the product of the word on address 00FF0004h and the immediate value 000F0008h, and will store the result on address 000F0000h.

The assembler supports the data memory directives .space, .byte, .half and .word that allocate and initialize the data segments.

V. INSTRUCTION-LEVEL SIMULATOR

In order to make initial analysis, an instruction-level simulator is developed in the Python 2.7 programming language. It consists of several modules: *parser*, which contains an assembler grammar definition, parses the code, and returns python objects representing the instructions; *program memory units* that contain the code and execute it instruction by instruction; and *data memory units* that contain the data segments. It should be noted that the parser is developed by using the *pyparsing* library.

The main work of the simulator is done in the program memory units. Every unit has a program counter, page number, program number, time tag and a flag that indicates whether the unit is in use. The time tag is used to identify the least recently used unit, in the case when all units are active and there is a new thread that needs to be executed.

The simulator maintains an array of all code pages and performs cycle by cycle execution of all active units. Aside from showing the transition of memory states, it keeps information about the number of executed cycles, page faults and data page re-allocations. As such it is a valuable tool for performance estimation of the proposed architecture.

Currently the simulator doesn't simulate the interconnection among memory units and assumes constant time for data exchange. This will not be the case in real scenarios, so for more accurate results a simulation on FPGA prototype is needed.

VI. SIMULATION RESULTS

In order to estimate and compare the performance benefits of a memory-centric architecture, we perform simulations and analysis for a typical computing problem: Fibonacci array generation. The program is written in assembler for both a classical MIPS assembler language as described in [17]; and the proposed MIPS-like assembler presented in the previous chapters of this paper. The former is simulated in the MARS simulator [17].

The tested program calculates a Fibonacci array of 4-byte elements and stores it in memory. The program code is first written for the memory-centric MIPS-like architecture, and is simulated for variable array lengths. Then it is rewritten for the original MIPS assembler and simulated for the same array lengths. Figure 1 shows a comparison of the total number of executed instructions. It can be easily noticed that the number of instructions is almost 50% lower with the memory-centric architecture. This means that even with half of the working frequency, the memory-centric architecture can potentially achieve the same performance.

This would also mean a 50% decreased power consumption.
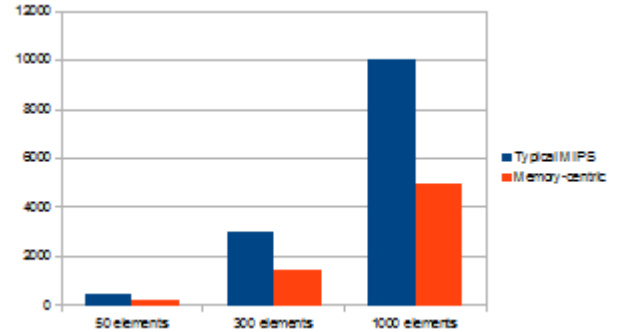

Fig. 1. Comparison of total number of instructions.

In order to make a fair comparison of execution time, we equalize the number of cycles for the two instruction set architectures. For a typical multi-cycle MIPS architecture the loads execute in 5 cycles; stores and R-type instructions in 4 cycles; whereas branching occurs in 3 cycles [16]. Since the memory-centric architecture accesses memory in every instruction we can safely assume that it doesn't take more than 6 cycles for each instruction. Figure 2 shows a comparison of the equivalent number of execution cycles. When using this model the number of executed instructions is still smaller although not with the same degree.
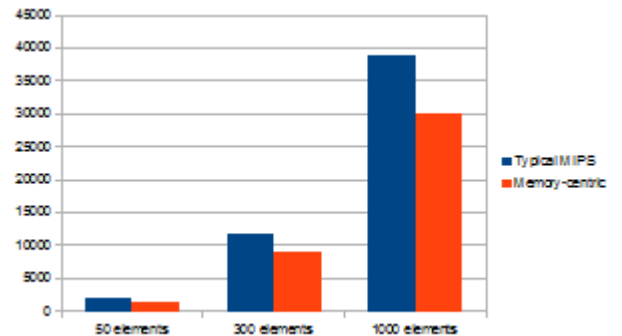

Fig. 2. Comparison of equivalent number of cycles.

VII. CONCLUSION AND FUTURE WORK

This paper shows that a memory-centric architecture has the potential for achieving much better performance and lower power consumption than that of a traditional computer architecture. One reason for this improvement is the fact that programs contain significantly smaller number of instructions when written for the proposed architecture and require less clock cycles to perform the same operations.

If we also consider the better parallelization opportunities, it should become clear that this architecture is very

promising, and may even revolutionize the parallel computing paradigm.

It should be noted however, that presented analyses do not take into account cache block misses. This should not influence the results significantly, as the size of the self-executing blocks is similar to the cache memory blocks, and the time penalty for a miss should be of the same order of magnitude. Comparison that takes this parameter into account is left for future work.

Another important consideration for this architecture is the interconnection network, because its characteristics will have a crucial impact on the system performance. The influence will depend on many factors regarding the interconnection system, such as its structure and implementation, and this is a crucial part of our future research.

The proposed architecture is suitable for FPGA implementation due to its similarity to the FPGA technology – both are structured as a set of data and execution units. This solution however is limited when it comes to execution speeds due to the nature of the FPGA. Higher speeds will be achieved by using ASIC implementation and this opens up a whole new field of research.

REFERENCES

[1] C. Carvalho, „The Gap between Processor and Memory Speeds",3rd Internal Conference on Computer Architecture (ICCA'02), Braga, 2002

[2] Memory Performance of Xeon E5-2600 v2 (Ivy Bridge-EP) based Systems, White Paper, Fujitsu, 2013

[3] Schlansker, M.S.; Rau, B.R., "EPIC: Explicitly Parallel Instruction Computing," Computer , vol.33, no.2, pp.37,45, Feb 2000

[4] P. Sirisuk et. al. (Eds.): ARC 2010, LNCS 5992, pp. 400-405, Springer-Verlag, Berlin Heidelberg, 2010

[5] Kiyoung Choi, „Reconfigurable Computing: Architectures, Tools and Applications", 6th International Symposium, ARC 2010, Bangkok, Thailand, March 17-19, 2010

[6] Wassal, A.G.; Sarhan, H.H.; ElSherief, A., „Novel 3D memory-centric NoC architecture for transaction-based SoC applications", Electronics, Communications and Photonics Conference (SIECPC), 2011 Saudi International, pp.1,5, 24-26, April 2011

[7] Yamin Li, Sanli Li, and Wanming Chu, „Memory Centric Interconnection Mechanism for Message Passing in Parallel Systems", Third International Conference on Massively Parallel Computing Systems (MPCS98), Colorado Springs, Colorado, Apr. 6–9, 1998

[8] ChunYi Su, Dong Li, Dimitrios S. Nikolopoulos, Kirk W. Cameron, Bronis R. de Supinski, Edgar A. Le´on, Model-Based, „Memory-Centric Performance and Power Optimization on NUMA Multiprocessors, Workload Characterization (IISWC)", 2012 IEEE International Symposium on , vol., no., pp.164,173, 4-6 Nov. 2012

[9] D. Li, B. de Supinski, M. Schulz, K. Cameron, and D. Nikolopoulos, „Hybrid MPI/OpenMP Power-Aware Computing", in IEEE International Symposium on Parallel Distributed Processing, 2010

[10] K. Singh, M. Curtis-Maury, S. A. McKee, F. Blagojevi´c, D. S. Nikolopoulos, B. R. de Supinski, and M. Schulz, „Comparing Scalability Prediction Strategies on an SMP of CMPs", Proc. of the 16th international Euro-Par conference on Parallel processing, 2010.

[11] Berić, A.; Van Meerbergen, J.; de Haan, G.; Sethuraman, R., „Memory-Centric Video Processing", Circuits and Systems for Video Technology, IEEE Transactions on , vol.18, no.4, pp.439,452, April 2008

[12] Ivan Lunteren, J., „Towards memory centric computing: a flexible address mapping scheme", Electrical and Computer Engineering, 1999 IEEE Canadian Conference on , vol.1, no., pp.385,390 vol.1, 9-12 May 1999

[13] Burger, D.; Goodman, J.R., „Memory-centric architectures: why and perhaps what", Innovative Architecture for Future Generation High-Performance Processors and Systems, 1997 , vol., no., pp.92,, 22-24 Oct 1997

[14] Bonatto, A.C.; Susin, A.A., „Memory subsystem architecture design for multimedia applications", VLSI (ISVLSI), 2013 IEEE Computer Society Annual Symposium on , vol., no., pp.213,214, 5-7 Aug. 2013

[15] Gwangsun Kim; Kim, J.; Jung Ho Ahn; Jaeha Kim, „Memory-centric system interconnect design with Hybrid Memory Cubes", Parallel Architectures and Compilation Techniques (PACT), 2013 22nd International Conference on , vol., no., pp.145,155, 7-11 Sept. 2013

[16] David A. Patterson, John L. Hennessy, "Computer Organization and Design: The Hardware/Software Interface (The Morgan Kaufmann Series in Computer Architecture and Design)", Paperback, Morgan Kaufmann Publishers, Oct 2013

[17] Kenneth Vollmar, Pete Sanderson, "MARS: An Education-Oriented MIPS Assembly Language Simulator", SIGCSE'06 Houston, USA, March 2006

Double-Exponential Model of a Photovoltaic Cell with Temperature Dependence

Alena Okhorzina[1, 2], Norbert Bernhard[2], Alexey Yurchenko[1]

[1] National Research Tomsk Polytechnic University – Department of Information-Measuring Engineering
Lenin Ave. 30, 634050, Tomsk, Russia

[2] Anhalt University of Applied Sciences – Faculty of Electrical, Mechanical and Industrial Engineering
Bernburger Str. 57, 06366, Köthen, Germany
E-mail: ohra.avit@gmail.com, n.bernhard@emw.hs-anhalt.de, niipp@inbox.ru

Abstract—**The purpose of the research is the creation of mathematical models in MATLAB based on the double exponential model of the photovoltaic cell. The developed model allows for different physical and environmental parameters. An equivalent circuit of the model includes a photocurrent source, two diodes, and a series and parallel resistance. The paper presents the simulation results for each parameter. The simulation data are displayed graphically and numerical results are saved in a file.**

Keywords: **double-exponential model, photovoltaic cell, Simulink.**

I. INTRODUCTION

MATLAB is a mathematical package of great potential. MATLAB works with matrix data and enables to create a custom calculated program. MATLAB is an operating environment and a programming language. Simulink is a standalone toolbox in MATLAB. Simulink can be used together with the MATLAB and other toolboxes, as well as individually. Simulink implements the principle of visual programming: the user uses the library to construct the model, configures the solver and calculation step. MATLAB/Simulink allows creation of blocks, programs and libraries. MATLAB and Simulink are graphical displays of simulation results.

The paper focuses on the double-exponential model of the photovoltaic cell (PV). The mathematical model is described in section two. Section three considers the model based on the Simulink toolbox. Section four presents the results of the model operating. The results for the effect of solar radiation, the PV cell temperature, diode saturation current, diode quality factor, and serial and parallel resistances are presented.

This study is part of the energy installation model and examines only the work of the photovoltaic. Attention is paid to the influence of temperature, because the installation will also run with the thermal energy.
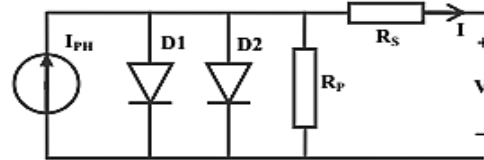

Fig.1. Equivalent circuit for a double-exponential model.

II. DOUBLE-EXPONENTIAL MODEL OF PV

The circuit (Fig.1) consists of a current source, diodes, and a serial and a parallel resistance. The current source simulates the occurrence of the photocurrent in a cell under illumination. The first diode is connected in forward direction in parallel to the current source. The saturation current can flow through the diode when affected by the bias due to excessive concentrations of electrons in the n- region of the cell and the concentration of holes in its p-region. The parallel photovoltaic cell resistance arises due to the reverse resistance of the n-p-transition and various conductive films or contamination on the cell surface or local small shunts. It is also connected in parallel to the current source. Series resistance represents contacts (mainly semiconductor-metal contact resistance) and the resistance of the semiconductor material from which the solar cell is manufactured (the resistance of each of the p- and n- regions of the cell) [1].

The single diode equation assumes a constant value for the quality factor N. In actual practice, the quality factor is a function of the voltage across the device. At high voltage, when the recombination in the device is dominating in the surface layer and bulk regions, the quality factor is close to one. At lower voltages, recombination in the junction dominates and the ideality factor approaches two [1]. The mathematical equation of the model is [1]:
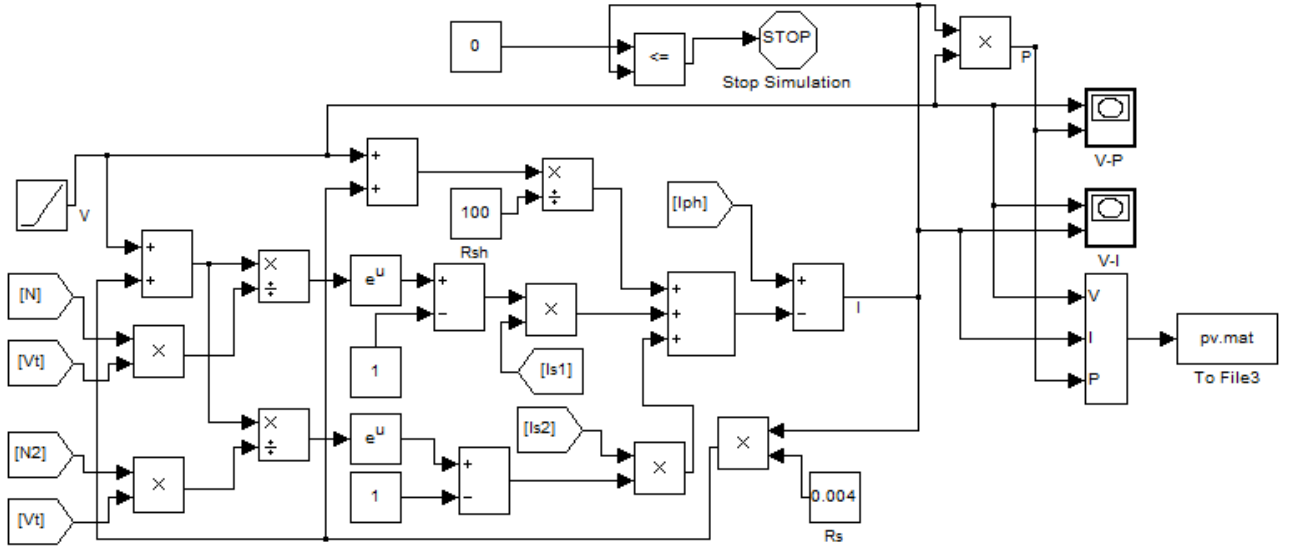
Fig.2. Simulink-model.

$$I = I_{ph} - I_S \left(\exp\left(\frac{V + I \cdot R_S}{N \cdot V_T} \right) - 1 \right) -$$
$$- I_{S2} \left(\exp\left(\frac{V + I \cdot R_S}{N_2 \cdot V_T} \right) - 1 \right) - \frac{V + I \cdot R_S}{R_P} \quad , \quad (1)$$

where

Iph is the solar-induced current;

Is is the saturation current of the first diode;

I_{S2} is the saturation current of the second diode;

V_T is the thermal voltage;

N is the quality factor (diode emission coefficient) of the first diode;

N2 is the quality factor (diode emission coefficient) of the second diode;

Rs and Rp are the serial and parallel resistances of the cell;

V is the voltage across the solar cell electrical ports.

The parameters in the double-exponential model depend on the solar radiance and temperature of the cell.

III. SIMULINK MODEL

The Simulink-model (Fig. 2) implements equation (1) and builds the output V-I and V-P characteristics (Fig.3).

Model features:

The voltage is set by uniformly time-varying signals. The current values are calculated for a definite voltage value at each step. The calculation stops when the current value reaches 0. This function is performed by the block «Stop Simulation». «Stop Simulation» compares the current value with 0 at each calculation step. When reaching the equality, the simulation ends.

The data is stored in the pv.mat file. The data is stored in an array of four lines: the first line is the simulation time, the second line is voltage, the third line is current, and the fourth line is power. The stored data can be loaded in the Workspace for later processing.
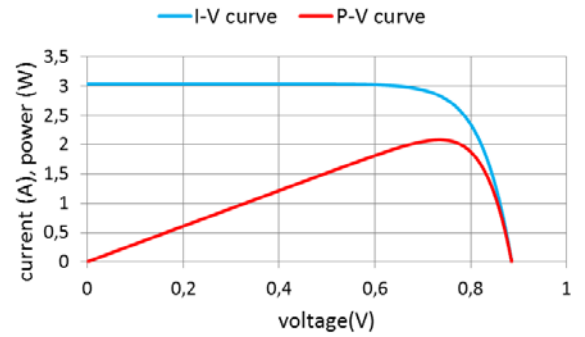

Fig.3. I-V and P-V curves.

IV. SIMULATION RESULTS

A. Effect of Solar Radiation Variation

The effect of solar radiation is determined by equation (2) [2][3]:

$$I_{ph} = \frac{S}{1000} \cdot \left(I_{SC} + K_i (Tc - Tr) \right), (2)$$

where

Iph is a light-generated current or photocurrent;

S is the value of solar radiation (in W/m^2 , to be inserted without dimension);

Isc is the cell short-circuit current at 25°C and 1kW/m^2;

Tc is the cell working temperature;

Tr is the cell reference temperature (25°C);

K_i is a constant which reflects the temperature dependence of the photocurrent (due to the slightly varying bandgap with temperature).

The realization of equation (2) as a subsystem is shown in Fig. 4. The effect of solar radiation is shown in Fig.5.

The effect of the variation of solar radiation on the current of the PV module is greater than that on the voltage.
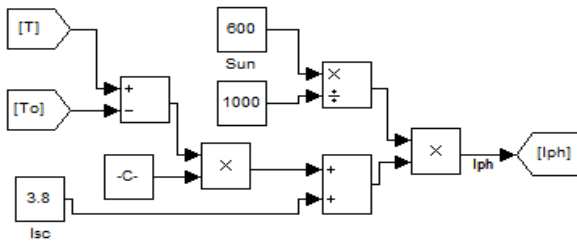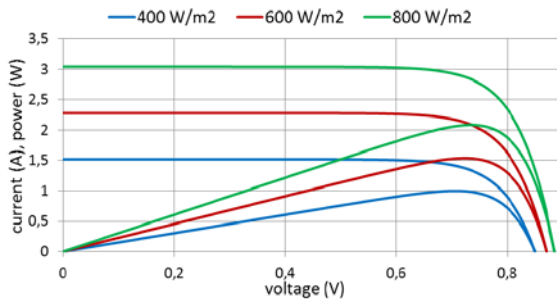
Fig.4. Subsystem realizes Iph.

Fig.5. I-V and P-V curves for different solar radiations (S).

B. Effect of Diode Quality Factor Variation

As can be seen from equation (1), other conditions being equal, the value of the diode quality factor has the same effect on the result. Therefore, it is important to consider this value as they may be unequal. The results of the simulation are shown in Fig.6.

Reducing the value of the diode quality factor has a significant effect on the PV module voltage. For different values of the diode quality factor for each diode, the voltage is found to decrease as well.

It is generally accepted that N is 1 and N2 is 2 [2][3]. But in real conditions quality factor is dependent on solar cell.
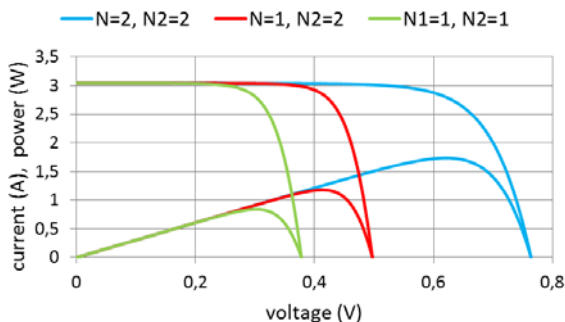

Fig.6. I-V and P-V curves for different quality factors.

C. Effect of Temperature Variation

The diode reverse saturation current varies as a cubic function of the temperature [2][3][4] multiplied by an exponential temperature activation expression and it can be expressed as:

$$I_{S1}(T) = I_S \cdot \left(\frac{Tc}{Tr}\right)^{3/N} \cdot \exp\left(\left(\frac{Tc}{Tr} - 1\right) \cdot \frac{E_g}{N \cdot V_T}\right)$$

$$I_{S2}(T) = I_{S2} \cdot \left(\frac{Tc}{Tr}\right)^{3/N2} \cdot \exp\left(\left(\frac{Tc}{Tr} - 1\right) \cdot \frac{E_g}{N2 \cdot V_T}\right) , (3)$$

where

Is is the diode reverse saturation current of fist diode;

I_{S2} is the diode reverse saturation current of second diode;
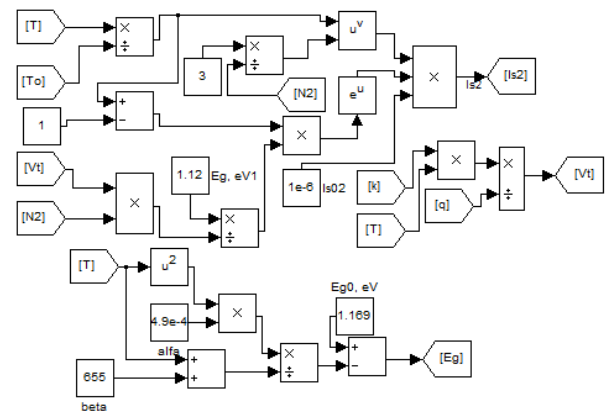

Fig.7. Subsystems realizes temperature depends.

Eg is the band gap energy;

Vt is the thermal voltage ($Vt=kT_C/q$, where k is the Boltzmann constant, T_C is the cell´s working temperature, q is the elementary charge on an electron);

N is the quality factor (diode emission coefficient) of the first diode;

N_2 is the quality factor (diode emission coefficient) of the second diode;

Tc is the cell's working temperature;

Tr is the cell's reference temperature.

The energy gap of silicon as a function of temperature [3] is presented below:

$$Eg(T) = Eg(0) - \frac{\alpha T^2}{T + \beta} \quad (4)$$

where

Eg is the band gap energy;

Eg(0) is the band gap energy of semiconductor at 0 K;

ß and α are is specific value to each semiconductor.

Since this is a double model, the subsystem it used for each diode. In this case, all the diode parameters are equal for both diodes, except the reverse saturation current Is0, which is generally different from Is02.

The realization of equations (3) and (4) is shown in Fig.7. The results of the simulation are shown in Fig.8.

The temperature variation has a significant impact on the VI curve and leads to a decrease in energy production. The effect of the temperature on Voc is one order of magnitude stronger than that on Isc.

D. Effect of Diode Saturation Current Variation

The model enables to predict the behavior of the PV cell for different reverse saturation currents of the diodes.

The results of the simulation are shown in Fig.9.

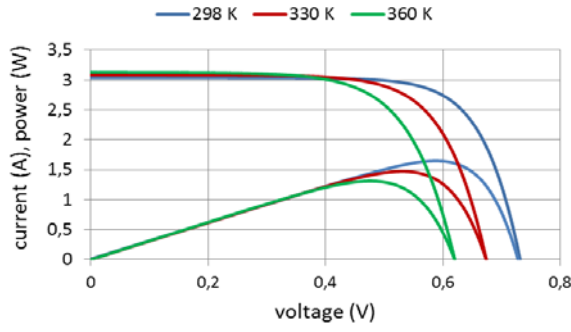The variation of the diode saturation current affects significantly Voc.

43

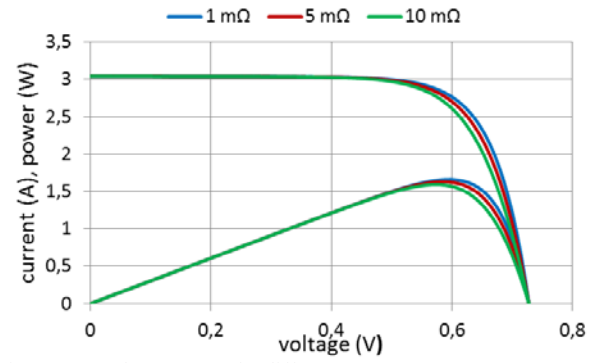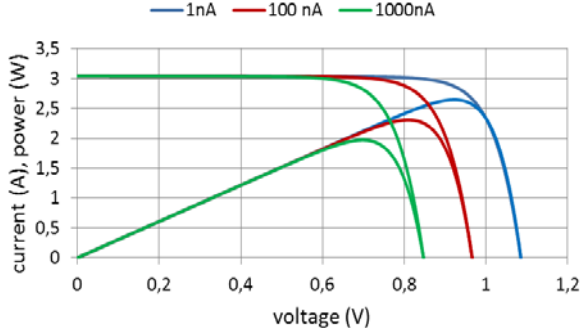Fig.8. I-V and P-V curves for different temperature.

Fig.9. I-V and P-V curves for different value of diode saturation current.

Usually the diode saturation currents of the tow diodes are not equal. This case was simulated and the results are shown in Fig.10. I_S is varied, Is2 is constant.

As one diode saturation current increases, Voc decreases, which, in turn, reduces the power output.
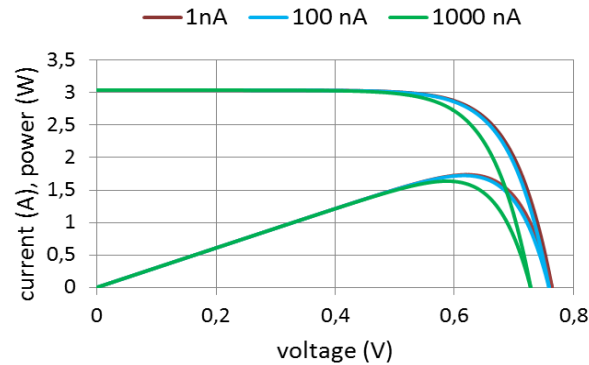

Fig.10. I-V and P-V curves for different values of diode saturation currents.

E. Effect of Serial Resistance Variation

The power P = IV produced by the cell is shown in Fig. 1. The cell generates the maximum power P_{max} at V_m voltage and I_m current, and the fill factor FF [1][4] can be defined as

$$FF = \frac{I_m \cdot V_m}{I_{sc} \cdot V_{oc}} = \frac{P_{max}}{I_{sc} \cdot V_{oc}},$$

The simulation results for Rs variation are shown in Fig.11.

As can be seen in Fig.11, the variation of Rs affects the slope angle of the I-V curves resulting in a deviation of the maximum power point. The fill factor appreciably decreases for higher value of Rs.

Fig.11. I-V and P-V curves for different Rs.

F. Effect of Parallel Resistance Variation

The parallel resistance of any cell should be large enough for higher power output and fill factor. In fact, for low Rp, the PV cell current collapses more steeply which means higher power loss and lower fill factor [5].

The simulation results for Rp variation are shown in Fig.12.
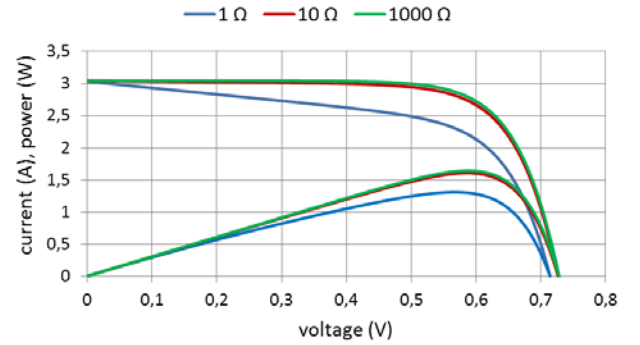
Fig.12. I-V and P-V curves for different Rp.

V. CONCLUSION

The Simulink PV model has been created in Simulink toolbox based on the double exponential model of the photovoltaic cell. This model takes into account the non-linearity of PV. The developed model is based on mathematical equations and an equivalent circuit, which includes a photocurrent source, two diodes, series and parallel resistors.

The model for the proposed range of the equivalent circuit elements, irradiance and temperature as model inputs, with the corresponding values of voltages, currents, and power as outputs is presented.

The temperature effect is taken into account when modeling the photocurrent and diode saturation current. Possible differences in the parameters of the diodes in the equivalent circuit are taken into account, as they have a significant impact on Voc and power output.

Simulation data are displayed graphically and numerical results are saved in a file.

The double-exponential model of a photovoltaic cell with temperature dependence will be used to creation of the energy installation. Accounting for the effects of temperature on the output characteristics is of interest for further work.

REFERENCES

[1] T. Markvart, L. Castafier, "Practical Handbook of Photovoltaics: Fundamentals and Applications," Elsevier Science Ltd. p. 985, 2003.

[2] Tsai, H., L., Tu, C., S., and Su, Y., J.", "Development of Generalized Photovoltaic Model Using MATLAB/SIMULINK" Proceedings on the world congress on Engineering and Computer Science. 2008, 6p.

[3] Zhao Ran et al., "A Simplified Double-Exponential Model of Photovoltaic Module in Matlab™," Energy and Environment Technology, 2009. ICEET '09. International Conference on (Volume:3) 2009.

[4] Tarak Salmi et al., "MATLAB/Simulink Based Modelling of Solar Photovoltaic Cell". International journal of renewable energy research 2012, Vol.2, No.2, 6p.

[5] Savita Nema, R.K. Nema, Gayatri Agnihotri, "MATLAB/Simulink based study of photovoltaic cells / modules / array and their experimental verification," International journal of Energy and Environment, vol.1, No.3, pp.487-500, 2010.

A Novel Memory-centric Architecture and Organization of Processors and Computers

Danijela Efnusheva, Goce Dokoski, Aristotel Tentov, Marija Kalendar

SS. Cyril and Methodius University - Faculty of Electrical Engineering and Information Technologies

Karpos II bb, PO Box 574, 1000 Skopje, Macedonia

E-mail: {danijela, gocedoko, toto, marijaka}@feit.ukim.edu.mk

Abstract—**The modern computer systems that are in use nowadays are mostly processor-dominant, which means that their memory is treated as a slave element that has one major task – to serve execution units data requirements. This organization is based on the classical Von Neumann's computer model, proposed seven decades ago in the 1950ties. This model suffers from a substantial processor-memory bottleneck, because of the huge disparity between the processor and memory working speeds. In order to solve this problem, in this paper we propose a novel architecture and organization of processors and computers that attempts to provide stronger match between the processing and memory elements in the system. The proposed model utilizes a memory-centric architecture, wherein the execution hardware is added to the memory code blocks, allowing them to perform instructions scheduling and execution, management of data requests and responses, and direct communication with the data memory blocks without using registers. This organization allows concurrent execution of all threads, processes or program segments that fit in the memory at a given time. Therefore, in this paper we describe several possibilities for organizing the proposed memory-centric system with multiple data and logic-memory merged blocks, by utilizing a high-speed interconnection switching network.**

Keywords: **Explicit parallelism, Field Programmable Gate Array (FPGA), high-performance computing, processor architecture and organization, processing in memory.**

I. INTRODUCTION

Computers are an important part of the modern human life, which cannot be imagined without the use of these electronic devices. The purpose of these complex systems is to perform data processing, data storage, data movement to and from the computer, and control of the whole system operation, [1]-[3]. These functionalities are provided by several basic computer components, including: central processing unit (CPU), memory (which is generally hierarchically organized), input/output devices and interconnection busses responsible for movement of data, address and control signals.

The central processing unit is one of the most complicated parts of the computer system that has ever been created by the human beings. The processor has the main role in the computer system, since it handles the instruction and data flow, controls the communication with the memory and input/output devices and thus coordinates the whole system operation, [4]-[6]. As a result, computer architects constantly face with the challenge to develop novel architectural solutions that can maximize the computer performance, while retaining the cost, power and functional requirements. Regarding this, they should consider three aspects of computer architecture design, such as: instruction set architecture, organization (memory system, memory-processor interconnect, internal processor), and hardware logic design.

The constant race of the various computer technologies resulted in a wide range of processor architectures, including CISC, RISC, Superscalar, VLIW, EPIC, Vector, [6]-[13], and Data Flow, [14]-[20]. These architectures provide various benefits and drawbacks, and are characterized with different ways of parallel programs execution, organization and instruction set architecture. Each of them is developed with the intention to overcome some of the problems of its predecessors and thus to provide better computing performances. However, besides the great advances in computer systems technology, their architecture and organization, the evolvement of multi-cores and various parallelization techniques for program execution, current computer architectures are still dominantly based on the classic Von Neumann's model, [21]-[23]. Main focus and prime role in this type of computer architectures and organizations is dedicated to the execution units of different type, and the memory is treated as slave element which main function is to serve the execution units data requirements.

The existing model of processor-centric computer architecture allows performance scale only if these two conditions hold: the processing core has sufficient work to do, so it can mitigate the cache miss latencies, and the processor has enough bandwidth to load the changes into the cache memory without excessive delay. However, the contemporary technologies for memory production can't cope with processor's requirements for data speed and bandwidth. As a consequence, there exists definitely a substantial gap of more than a couple of times between the processor working frequency, and the available memory data transfer speed, [24]-[27]. As a result, superscalar processors, [28], which are capable to execute several instructions per clock cycle, always lack of data, due to lower memory working frequency and reduced number of internal processor registers. Moreover, the Itanium EPIC processor, [29], did not manage to achieve the expected success, because of the problems with the memory speed.

A few decades ago, in the 1990[ties], some researchers predicted that the memory behavior would be preponderant over the global performance of the computer system. Their proposals suggested integration of the memory and processing elements into a single chip, creating memories with processing capacity. This merged chip is known as: smart memory, computational memory (C-RAM), [30]-[34], processor in memory (PIM), [35]-[41], intelligent RAM (IRAM), [42]-[49], etc. Recent work in this area lead to several architectural approaches, which can be classified based on the role of the merged chip: main processor(s) in the system, special-purpose processor, co-processor or intelligent memory system, [24]. For example, IRAM, [48] is implemented as a vector co-processor to a general purpose MIPS processor into the VIRAM single-chip computer.

The aim of this paper is to propose a novel memory-centric processor architecture that provides a stronger merge between memory and processing elements. This is achieved by adding a processing hardware directly to the memory blocks used for storing programs, thus allowing simpler instruction decode and execution, easier management of data requests, and direct communication between the program (code) and data memory blocks, without the use of registers. These logic-memory merged chips are named as self-executing units. The memory-centric architecture should be organized to work with multiple self-executing units, in order to provide concurrent execution of all threads, processes or program segments that fit in the memory, at a given time. Therefore, in this paper we propose and evaluate several models of computer system design with multiple data memory blocks and self-executing units, connected via high-speed interconnection switching network.

The paper is organized in five sections. Section two presents the current state, discussing a variety of modern processor architectures and organizations that are in use today. Section three describes the novel memory-centric architecture and its basic building blocks, providing details about the hardware design and its verification. The next section proposes several ways of organizing the proposed memory-centric architecture with multiple self-executing units. The paper ends up with conclusion, stated in the last section.

II. State of the Art

The complexity of modern processor architectures and the constant race of various computer technologies resulted in a wide range of computer architectures, each with its own advantages and disadvantages, but with the ultimate goal to increase the overall computer system performances. Therefore, the research of computer architects was aimed at developing various mechanisms for parallel computing that will provide efficient utilization of the system hardware resources. Generally, there are three different forms of parallel computing that have been created, including: instruction- (execution of more than one instructions in a single processor cycle), data- (execution of single instruction stream on multiple data streams) and thread-level parallelism, (concurrent execution of unrelated and distinct tasks), [1]. Most modern computer systems support several types of parallel processing in order to achieve better computing performances.

One of the first computer architectures, such as the Intel IA-32, belongs to the Complex Instruction Set Computer (CISC) design which takes advantage of microcode and supports a wider range of variable-length instructions, [6]. In order to reduce the complexity of these instructions and to provide hardware-based control of their execution, Reduced Instructions Set Computing (RISC) was introduced, [7]. Further research led to the idea that dividing the work of a single processor to multiple execution units would speed up the instructions execution. This resulted with superscalar and Very Long Instruction Word (VLIW) architectures that were designed to take advantage of Instruction Level Parallelism (ILP). A superscalar architecture consists of a number of pipelines that are working in parallel, and relies on hardware to detect and overcome data dependencies. On the other hand, Very Long Instruction Word (VLIW) architecture, [11], uses software solution (compiler) to mark independent operations that can be executed simultaneously. The limits of the parallelism, defined by the length of the VLIW instruction is an issue that has caused development of explicitly parallel instruction computing (EPIC) architecture, [9]. Despite the advantages of EPIC over VLIW, the IA-64 Itanium architecture could not manage to solve all of VLIW's problems. Other alternative to the conventional control flow architecture in providing concurrency in execution of programs is the dataflow architecture, [20]. This architecture is only a concept that has never been implemented in a real hardware.

Each of the discussed processor architectures is described in table 1. The given table shows that pipelining is one of the most used parallelization techniques. This ILP method allows parallel execution of N different instructions in N diverse pipeline stages, so the pipeline length is proportional to the theoretical increase in speed, [2]. Further performance improvements are achieved when multiple pipelines are simultaneously executed on multiple execution units, like in superscalar processors, [20]. Pipelining as an ILP method can be also combined with vector processing, thus allowing data-parallel vector operations to be executed on multiple pipeline execution units, [5]. The achievable parallelism in such processor is dictated by the number of execution units, which also applies to other processor architectures that can be organized to work with multiple execution units, like: Superscalar, VLIW, EPIC and Data Flow.

The performance of computer systems primary depends on the CPU execution time, which is secondary related to the average memory access time, [25]. As a result, computer architects are faced up to the problem of decreasing the CPU execution time, while improving the memory bandwidth and keeping the processor busy all the time. There are several mechanisms that have been developed to target this problem, including: multi-level cache memories, separation of memories for storing programs and data (Harvard architecture), speculative and re-order execution, branch prediction algorithms, etc, [6]. Further improvements are achieved by hardware parallelization i.e. use of multi-core or multi-processor systems that support multithreading, [3]. This approach introduces more intensive work with the memory resources, causing a bottleneck in the system. A possible solution to this problem is integration of processing modules into memory, such as the IRAM-based approach.

TABLE I
COMPARISON OF DIFFERENT PROCESSOR ARCHITECTURES

	Features				
	Program execution	Parallel Processing	Parallelization Techniques	Number of operations in cycle	Instruction Format
CISC Architecture	Control flow	Yes, instruction level	Each instruction executes more operations	Depends on the instruction complexity	Variable length complex operations
RISC Architecture	Control flow	Yes, instruction level	Pipeline that implements several stages	Depends on the pipeline depth (4 or 5 stages)	Fixed length, (usually: 16,32,64)
Superscalar Architecture	Control flow	Yes, instruction level	Several parallel pipelines	Depends on the pipeline depth (8 - 10 stages) and the number of execution units	Fixed length, (usually: 16,32,64)
VLIW Architecture	Control flow	Yes, instruction level	Fixed number of mini-operations in instruction word (implicit parallelism)	Depends on the number of execution units and mini-operations	Fixed length: more mini-operations in one word
EPIC Architecture	Control flow	Yes, instruction level	Variable, but limited number of mini-operations in instruction word (explicit parallelism)	Depends on the number of execution units and mini-operations	Variable length: more mini-operations in one word
Vector Architecture	Control flow	Yes, data level	Same operation is executed on different fixed length vectors	Depends on the number of execution units and vector length	Fixed length, (usually: 16,32,64)
Data-flow Architecture	Data flow	Yes, instruction level	Variable, but limited number of instructions (depends on the availability of input operands)	Depends on the number of execution units and active instructions	Packet instruction format

III. DESIGN AND VERIFICATION OF THE NOVEL MEMORY-CENTRIC ARCHITECTURE

The standard Von Neumann based architecture specifies a model of computer system, which consists of memory and processing parts that are strictly separated. In such a system, the memory is used to store both data and instructions, and the central processing unit is purposed to read and decode program code, load/store data to/from cache or/and registers and execute arithmetical-logical operations over the loaded operands. This organization suffers from limited throughput (bottleneck) on the processor-memory interface, caused by the memory and processor speeds discrepancy. Actually, it is well known that the memory system operates at an average access time that is much greater than the processor execution time, [49]. This dissimilarity leads to many wasted processing cycles, since the memory system is not capable to constantly feed the CPU and keep it busy.

Assuming that the processor technology slowly reaches its upper bounds on chip complexity and speed, we suggest that the bottleneck problem should be targeted by introducing novel concepts in computer architectures that will provide closer tie between computing and memory resources, and as well will allow higher utilization of parallel computing. Therefore, our research is directed towards the development of a novel memory-centric computer architecture that organizes the memory system in a completely different way from the long enduring Von Neumann-based systems. In our approach, the memory system is observed as a set of blocks, quite similar to the virtual memory concept. Actually, in the first steps, the memory is separated into many data blocks and a code block that is enhanced with some processing and control capabilities (see Figure 1b). We have named the logic-memory merged block a self-executing unit, since it adds some execution hardware into the code memory, used for storing programs. Having this organization, the memory system has complete control over the program execution (instruction decode, data transfers and ALU operations).

a) Von Neumann computer architecture
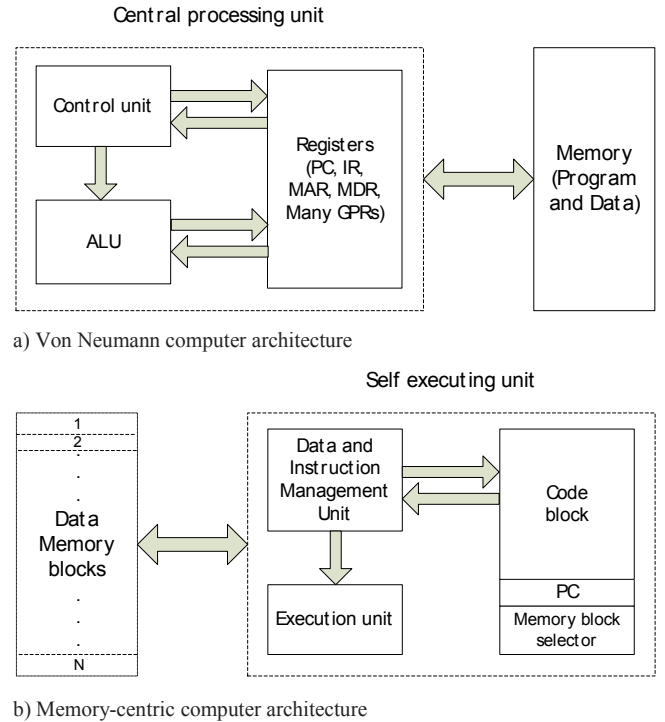

b) Memory-centric computer architecture

Fig. 1. Comparison between the classical processor-centric and the novel memory-centric computer architecture

Figure 1 presents the novel memory-centric architecture; consisting of data memory blocks and a self executing unit, connected through a bus interconnect. The data memory blocks are responsible for storing the data used by the programs. The self-executing unit is purposed to immediately fetch and decode the instructions located into the code memory block, and then issue data movement and arithmetic commands over the bus. The operations are performed over the selected data operands from a certain data memory block, which is specified by a specialized register, named memory block selector. Given that the commands are directly executed, the system doesn't need to include intermediate memory resources, such as processor registers or cache memory, so they are not part of the initial memory organization proposal.

The data are stored in the data memory blocks as 32-bit signed integers, represented in second complement notation. This abstraction can be further extended to allow use of 32-bit floating-point numbers with single-precision, according to IEEE-754 format, [50]. The data operands are directly accessed via the interconnection bus, and then served to the execution unit. The operations and the functionalities that the system provides are defined in its instruction set. Basically, the instruction set architecture of the proposed system is RISC-like, and includes arithmetical-logical, shift, branching and auxiliary operations. Each instruction is fixed-length, typically specifying the operation code and the operands, given as direct addresses or immediate values. This is presented in Figure 2.

Operation code	Op1 Addr. offset/ Immediate value	Op2 Addr. offset/ Immediate value	Addr. offset of the Result
6 bit	10 bit	10 bit	10 bit

Fig. 2. Instruction format of the memory-centric ISA

In order to decrease the instruction length, the system utilizes the memory segmentation concept, allowing each data memory block to be represented as a separate memory segment. Accordingly, the self-executing unit is associated with only one memory segment at a given moment. This results with some simplifications in instructions formatting, since the memory operands are specified as address offsets, instead of complete addresses to the data memory block (segment). The identification number of the active memory segment is set by a special instruction that affects the content of the memory block selector hardware. Therefore, it is compiler's responsibility to provide efficient mapping of memory segments, and thus to allow handling of data-intensive operations without difficulties.

The memory-centric system allows several functionalities, which are specified into its instruction set architecture. Generally, the instruction set is divided into four basic groups, each having several simple instructions, with similar structure and function. The arithmetical-logical group includes arithmetical instructions for: addition with overflow detection, subtraction, multiplication, integer division (div) and modulo division (mod), and as well instructions for performing logical operations: logical bit-wise AND, logical bit-wise OR, logical bit-wise XOR and logical bit-wise NOT. The shifting group comprises several instructions for left and right logical shift and rotations. The branching group includes instructions for conditional and unconditional change of the program flow. There are several options for conditional branching, including: equal (=), not equal (!=), less (<), greater (>), less or equal (<=), and greater or equal (>=). The last group is the auxiliary group, consisting of control instructions for program termination and system halt, as well as memory instructions for: constants loading, data transfers between memory blocks, and updating of the memory block selector value.

Data transfer between memory and registers

LOAD R1, Mem-addr1
LOAD R2, Mem-addr2
ADD R1,R2,R3
STORE R3, Mem-addr3

RISC-based processor

Direct transfer of memory data

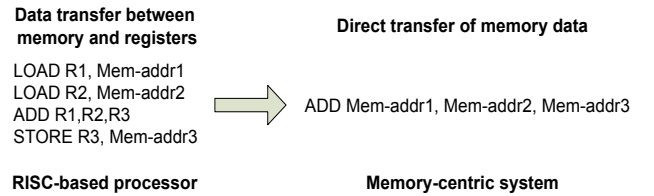ADD Mem-addr1, Mem-addr2, Mem-addr3

Memory-centric system

Fig. 3. Program segment written in ISA of the RISC-based load-store processor and the novel memory-centric system

Figure 3 presents a simple program segment that performs addition of two integer numbers. According to that, it is obvious that the memory-centric system reduces the number of program instructions in comparison with the RISC-based processor. This essentially comes from directly addressing the memory data, without using processor registers. Besides that, its instructions length is not longer than 40 bits, meaning that the program size is significantly reduced. This has a very positive impact on lowering the power dissipation of the memory-centric system. Other benefit, which has to be mentioned, is that the self-executing unit immediately fetches and decodes the instruction, thus speeding up the instruction execution. The only drawback of the novel proposed system is that it has to operate on memory working frequency. However, the approach of removing the processor registers brings significant improvement and simplification in the way programs are written, compiled and executed. This is very suitable for applications that perform data-intensive computations, such as: digital signal processing, multimedia, network processing etc.

The proposed memory-centric architecture is implemented in VHDL, by means of Xilinx ISE Design Suite tool. This software environment includes ISim simulator, used for functional analysis of VHDL models. In addition to that, there are several other tools for hardware synthesis and FPGA implementation. The FPGA technology is utterly suitable for research purposes, due to its advantage in terms of speed, cost, flexibility and ease of re-programmability, [51]. Therefore, in this research we make use of the XUPV505-LX110T Virtex5 FPGA board.

Figure 4 presents simulation results for the execution unit, while performing five arithmetic operations over two integer numbers: addition, subtraction, division, modulo division and multiplication. The execution unit, which is part of the self-executing unit, is described in VHDL as a module that receives two 32-bit operands, operation code and chip enable signal as an input and produces 32-bit result value and additional bit for overflow detection as an output. The simulation results shown in Figure 5 verify that the VHDL model is completely functional. In addition to that, the module has been synthesized and implemented in Virtex 5 logic, utilizing 1% of the slice registers, 4% of the slice LUT resources, which is 5% of the occupied FPGA slice resources. The synthesis results demonstrate that the module can achieve maximal working frequency of 952.381MHz.
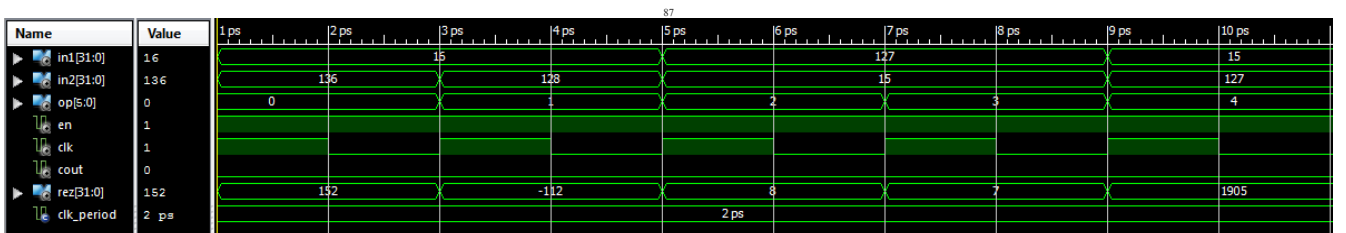

Fig. 4. Simulation results of the arithmetical operations performed by the execution unit: addition, subtraction, integer division, modulo division and multiplication

The self-executing unit is responsible to directly fetch, and decode instructions from the code memory block, and then issue commands that are executed over data operands, selected from a certain data memory block. Each data memory block is described in VHDL as an array of 1Kx4B data words, represented in second complement notation. The 32-bit data words can be also defined in IEEE-754 single-precision format as floating-point numbers, but this would require some additional modifications into the execution unit hardware. The code memory block is also designed in VHDL, as an array of 1K instruction words, wherein the first location is used to preserve the address pointer value of the current instruction, which is actually the program counter value. The code memory block is associated with only one data memory segment at a given moment, which is configured through special block selector hardware. The instruction execution and the data transfers are controlled by the instruction and data management unit, which has complete control over the execution unit, providing direct communication with the associated data memory block.
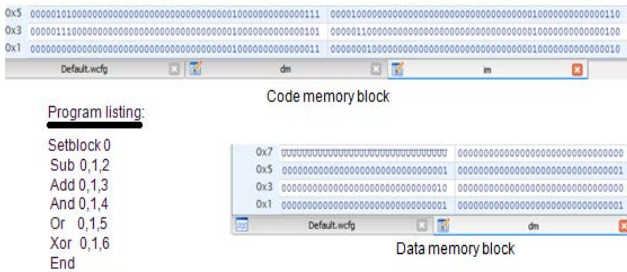

Fig. 5. Simulation results of a simple program execution on the novel memory-centric system

The functionality of the complete system is analyzed through the Xilinx ISE Design Suite environment, which allows monitoring of the memory blocks state in each clock cycle. Therefore, for simulation purposes, we have created a test scenario with a simple program, presented in Figure 5 and an arbitrary data set. The program instructions are first filled into the code memory block, and then executed over the associated data memory block. The simulation results presented in Figure 5 show that each instruction affects the data memory block with some changes, every clock cycle. Therefore, it is verified that the memory-centric system operates properly.

IV. SEVERAL PROPOSAL FOR ORGANIZING THE NOVEL MEMORY-CENTRIC ARCHITECTURE WITH MULTIPLE SELF-EXECUTING UNITS

The proposed memory-centric architecture that utilizes one self-executing unit and a set of data memory segments can be organized to work with multiple self-executing units. For this purpose, the system is expanded with multiple code memory segments that have execution capabilities (see Fig. 6). The integration of processing hardware into memory is very suitable for FPGA implementation, because the FPGA technology is already designed as a reconfigurable network, [52], of small memory and processing blocks.

The operating system is responsible to load the parallelizable programs and their data into various code and data memory blocks of the system. This means that each code memory block holds and executes a separate thread, process or program segment at any given moment. As a result, the system achieves concurrent execution of all the programs that fit into the code blocks memory.
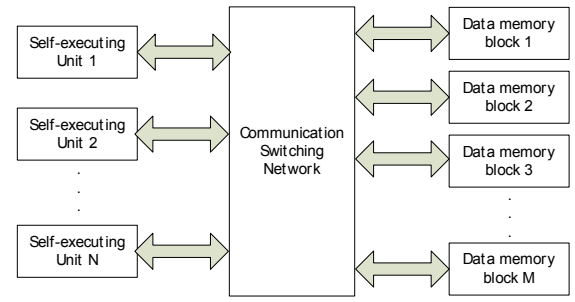

Fig. 6. Organization of the memory-centric system with multiple self-executing units

Figure 6 presents the proposed organization of the memory-centric system with multiple self-executing units and many more data memory blocks (M>>N), connected through a communication switching network. The switching network provides direct communication of N self-executing units with N data memory blocks, thus allowing parallel execution of N independent programs.

There are several interconnection mechanisms, such as AMBA, CoreConnect and Wishbone system-on-chip buses that can be used when implementing communication between multiple master and slave elements. According to the analyzes given in [53] the Wishbone bus is most suitable for the purposes of this research, basically because it is an open-source interconnect that can operate in several different modes, including: shared bus, pipeline and crossbar switch. However, each of the proposed solutions provides N-to-N mappings, which means that each code memory block is associated with only one data memory block. This can be very tricky if a program needs to operate on a larger data set, spread out over several memory segments. In older to resolve this problem, we propose a system that allows reconfigurable use of data memory blocks (see Figure 7).
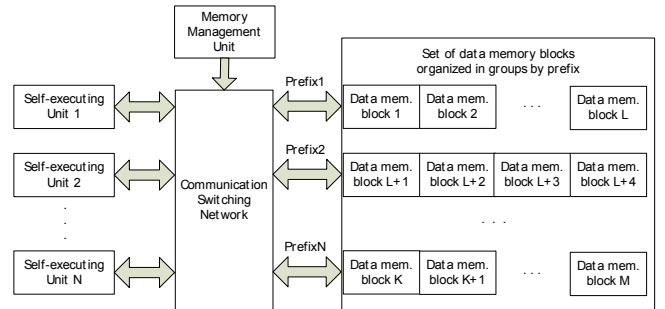

Fig. 7. Proposed model of a memory-centric system, that allows reconfigurable use of data memory blocks

The proposed model, presented in Figure 7 introduces several changes to the way data memory blocks are organized and used by the programs. This system is capable to assign a configurable number of data memory blocks to each program segment. The basic idea behind this approach is the use of prefixes that select a group of multiple data memory blocks. These prefixes are very similar to the IP prefixes used in computer networks for routing or IP assignment purposes. The management of the prefixes is performed by a special hardware (MMU), whose operation is controlled by the operating system. Once the prefixes are set, the communication switching network performs matching of each self-executing unit with a group of data memory blocks, selected by a given prefix. This approach and its applicability are still subject of research, and there are still some issues that need to be resolved.

V. Conclusion

The architecture and the organization of the computer systems haven't evolved much in comparison with its beginnings. Therefore, recent computer architectures are still dominantly based on the classical Von Neumann's computer model, which suffers from limited throughput on the processor-memory interface. In order to solve this problem, in this paper we propose a novel memory-centric architecture that adds processing hardware directly to the code memory blocks used for storing programs, thus allowing immediate instruction fetch, decode and execution, easier management of data requests, and direct communication between the data and code memory blocks, without the use of registers. The paper shows that the memory-centric system reduces the number of program's instructions and speeds up the instructions execution, thus providing much better performance characteristics and lower power consumption than traditional computer architectures. The proposed memory-centric system is designed in VHDL, and its correct operation is verified through simulations.

In order to provide explicitly parallel execution, the proposed memory-centric system can be organized to work with multiple logic-memory merged modules (i.e self-executing units) that will operate on separate threads, processes or program segments in parallel. A critical part of such multi-block (unit) organization is the communication switching network, which should provide connectivity of N self-executing units with N data memory blocks. In this paper we observe several solutions, and furthermore propose an approach that allows reconfigurable use of data memory blocks, based on prefixes. The applicability and the performance characteristics of these proposals depend on the system structure and implementation, which is a crucial part of our future research. However, we believe that the proposed memory-centric architecture has the potential to create a new generation of computers with increased portability, reduced size and power consumption, without compromising the performance and efficiency.

References

[1] David A. Patterson, John L. Hennessy, "Computer Organization and Design: The hardware/software interface," 5th ed. Elsevier, 2014.

[2] W. Stallings, "Computer organization and architecture: Designing for performance," 8th edition, Prentice Hall, 2009.

[3] John L. Hennessy, David A. Patterson, "Computer Architecture: A Quantitative Approach," 4th ed., Morgan Kaufmann Publishers, 2007.

[4] G. McFarland, "Microprocessor design: a practical guide from design planning to manufacturing," The McGraw-Hill Companies, 2006.

[5] Sivarama P. Dandamudi, "Fundamentals of Computer Organization and Design," New York: Springer, 2002.

[6] D. Jakimovska, et al., "Modern Processor Architectures Overview," Proc. ICEST, Bulgaria, June 2012, pp. 239-242.

[7] Sivarama P. Dandamudi, "Guide to RISC processors: for programmers and engineers," Springer, 2005.

[8] Vojin G. Oklobdzija, "Reduced instruction set computers," Technical Paper, University of California, 1999.

[9] J. Huck, D. Morris, et al., "Introducing the IA-64 Architecture," Proc. IEEE Micro, vol. 20, no. 5. pp. 12-23., Sept/Oct 2000.

[10] C. Kozyrakis, "Scalable vector media-processors for embedded systems," PhD Thesis, University of California, Berkeley, 2002.

[11] T. M. Conte, "Superscalar and VLIW Processors," Handbook, 1996.

[12] N. FitzRoy-Dale, "The VLIW and EPIC processor architectures," Master Thesis, New South Wales University, 2005.

[13] Michael J. Mahon, et al. "Hewlett - Packard Precision Architecture: The Processor," Hewlett-Packard journal, 1986.

[14] A. L. Davis, R. M. Keller, "Data flow program graphs," Proc. IEEE Trans. On Computers, February 1982.

[15] J. Silc, B. Robic and T. Ungerer, "Asynchrony in parallel computing: From dataflow to multithreading," Journal of Parallel and Distributed Computing Practices, 1998.

[16] Ben Lee and A. R. Hurson, "Issues in dataflow computing," Journal of Advances in Computers, 1993.

[17] G. M. Papadopoulos, "Implementation of a general-purpose dataflow multiprocessor," Tech. Report TR-432, MIT Laboratory of Computer Science, Cambridge, August 1988.

[18] R. Buehrer, K. Ekanadham, "Incorporating dataflow ideas into von Neumann processors for parallel execution," Proc. IEEE Trans. On Computers, December 1987.

[19] R. A. Iannucci, "Toward a dataflow/von Neumann hybrid architecture," Proc. 15th ISCA, May 1988.

[20] J. Silc, B. Robic, T. Ungerer, "Processor architecture: From Dataflow to Superscalar and Beyond," Springer, 1999.

[21] Zomaya, A.Y.H, "Parallel and Distributed Computing Handbook," McGraw-Hill, 1996.

[22] M. Smotherman, "Understanding EPIC Architectures and Implementations," Proc. ACM Southeast Conference, 2002.

[23] Ravikanth Ganesan, Kannan Govindarajan, Min-You Wu, "Comparing SIMD and MIMD programming modes," Journal of Parallel Distributed Computing, 1996.

[24] Carlos Carvalho, "The gap between processor and memory speeds," Proc. ICCA 2002, Braga Portugal, 2002.

[25] N. R. Mahapatra, B. Venkatrao, "The processor-memory bottleneck: problems and solutions," ACM Crossroads, 1999.

[26] Christianto C. Liu, Ilya Ganusov, et al., "Bridging the processor-memory performance gap with 3D IC technology," IEEE Design & Test of Computers, vol. 22, no. 6., 2005, pp. 556-564.

[27] Damian Miller, "Reconfigurable systems: a potential solution to the Von Neumann bottleneck," Senior Thesis, Liberty University, 2011.

[28] Christoforos Kozyrakis, David Patterson, "Vector vs. superscalar and VLIW architectures for embedded multimedia benchmarks," Proc. 35th International Symposium on Microarchitecture, November 2002.

[29] Harsh Sharangpani, Ken Arora, "Itanium processor microarchitecture," Proc. IEEE Micro, 2000.

[30] C. Cojocaru, "Computational RAM: implementation and bit-parallel architecture," Master Thesis, Carletorn University, Ottawa, 1995.

[31] Peter M. Nyasulu, "System design for a computational-RAM logic-in-memory parallel-processing machine," PhD Thesis, Carletorn University, Ottawa, 1999.

[32] D. Elliott, et al., "Computational RAM: the case for SIMD computing in memory," Proc. ISCA '97, June 1997.

[33] Duncan G. Elliott, Michael Stumm, et al., "Computational RAM: implementing processors in memory," Journal IEEE Design & Test. vol. 16, issue 1, January 1999.

[34] Duncan G. Elliott, W. Martin Snelgrove, Michael Stumm, "Computational RAM: A memory-SIMD hybrid and its application to DSP," Proc. Integrated Circuits conference, 1992.

[35] Peter M. Kogge, Jay B. Brockman, et al., "Processing in memory: chips to petaflops," Technical report, Proc. International Symposium on Computer Architecture, June 1997.

[36] Daescu, Ovidiu, Peter M. Kogge, Danny Chen, "Parallel content-based image analysis on PIM processors," Proc. IEEE Workshop on Content-Based Access to Image and Video Databases, June 1998.

[37] Jeffrey Draper et al., "Implementation of a 256-bit WideWord processor for the data-intensive architecture (DIVA) processing-in-memory (PIM) chip," Proc. 28th European Solid-State Circuit Conference. September 2002.

[38] Maya Gokhale et al., "Processing in memory: the Terasys massively parallel PIM array," IEEE Computer, 1995.

[39] Jeff Draper, Jacqueline Chame, et al., "The architecture of the DIVA processing in memory chip," Proc. 16th international conference on Supercomputing ICS'02, USA, 2002.

[40] Thomas L. Sterling, Huns P. Zimu, "Gilgamesh: a multithreaded processor-in-memory architecture for petaflops computing," Proc. ACM Supercomputing, 2002.

[41] T. Sterling, M. Brodowicz, "The "MIND" scalable PIM architecture," Proc. High Performance Computing Workshop, 2004.

[42] D. Patterson et al., "Intelligent RAM: chips that remember and compute," Proc. Solid-State Circuits Conference, 1997.

[43] David Patterson, Thomas Anderson, et al., "A case for intelligent RAM: IRAM," Proc. IEEE Micro, April 1997.

[44] D. Patterson, et al., "Intelligent RAM (IRAM): the industrial setting, applications, and architectures," Proc. International Conference on Computer Design: VLSI in Computers & Processors, University of California. Berkeley, USA, 1997.

[45] João Paulo Portela Araújo, "Intelligent RAM: a radical solution?," Proc. 3rd Internal Conference on Computer Architecture, 2002.

[46] Brian R. Gaeke, Parry Husbands, et al., "Memory-intensive benchmarks: IRAM vs. cache-based machines," Proc. International Parallel and Distributed Processing Symposium (IPDPS), April. 2002.

[47] Joseph Gebis, Sam Williams, et al., "VIRAM1: a media-oriented vector processor with embedded DRAM," 41st Design Automation Student Design Contenst, San Diego CA, June 2004.

[48] David Martin, "Vector extensions to the MIPS-IV instruction set architecture, the V-IRAM architecture manual," Technical paper, March 2000.

[49] Danijela Efnusheva and Aristotel Tentov, "Integrating processing in RAM memory and its application to high speed FFT computation," Proc. International Conference on Information Society and Technology, Serbia, March 2014.

[50] IEEE, "754-2008 - IEEE Standard for Floating-Point Arithmetic," Technical paper, 2008.

[51] D. Efnusheva, et al., "Efficiency comparison of DFT/IDFT algorithms by evaluating diverse hardware implementations, parallelization prospects and possible improvements," Proc. Second International Conference on Applied Innovations in IT, Germany, March 2014.

[52] Andre De Hon, "Reconfigurable Architectures for General-Purpose Computing," Technical Report, 1996.

[53] Milica Mitić and Mile Stojčev, "A Survey of Three System-on-Chip Buses:AMBA, CoreConnect and Wishbone," Proc. of ICEST, 2006.