# About Some Characteristics Of Computers Of New Generation

Oleg Granichin[*], Sergey Sysoev[†]
Department of Mathematics and Mechanics

St. Petersburg State University

*Abstract* — **The paper deals with the problem of creating of computers of the new generation. Some abstract approach of constructing the hardware is considered and the basic scheme of functioning is proposed. The new generation is supposed to be the generation of computers with artificial intelligence features. In this paper term "intelligence" is understood as ability of system to adapt to the real world conditions by the effective choosing of the task to solve in current moment.**

## 1. Introduction

For more than 50 years history of practical using of computers by mankind the main disappointment for many people became impossibility to resolve the problem of creation of artificial intelligence. In spite of an impetuous progress in the sphere of the creation of VLSIC (Very Large Scale Integration Circuit) it is still impossible to speak about essential progress in understanding of the more exact problem setting and of appropriate directions of the solution of this task.

Most of the research on artificial intelligence, known to authors, is devoted to solving of some particular task or claiming on generality of the approach by realisation of some heuristic behavior algorithm which successfully consults with test examples but collides during operations in real conditions.

Before moving to discussion about approaches of the creation of artificial intelligence let's look at some achievements in the contemporary mathematics and computer engineering of the past century which allow us to look with a great optimism at the solution of the problem in the future. The first computers were specialized for solution of the concrete tasks. It was based both on the economic reasons and on the weak development of electronic components. For a short time the engineering progress has allowed to develop the universal computer and theoretical development has made it possible

[*]e-mail: oleg_granitchin@mail.ru
[†]e-mail: sysoev_s@mail.ru

to effectively solve many of the tasks set. The final conception of the universal programming language appeared in the middle of 70th and during the 80-90th years the programming technology has been on a top of its development. However we are still left with considerable problems tied with the machine's ability to make a quick and effective decision in tasks with a large data dimension (matrix transformation, large number factorization, Fourier or Laplace transformation, calculation of function convolution and so on). A lot of them form a minimal base of an artificial intelligence system.

Today the most practical way is a creation of fast computing devices for solving any concrete tasks based on possibility of development of "parallel" fastest algorithms for processing the large sizes of data. Two directions are the most perspective. The first one (almost developed) is to create the classical electronic dievices as VLSIC using the high-speed technologies. Now the "0.13mc" technology is already widely available and some advanced corporations already use process engineering "0.11mc" and less. The chips created with this technology can simultaneously execute hundreds thousands of calculations in one step. "IBM" has declared that within a couple years it will be possible to produce a device with millions parallel calculations per cycle. For example, such devices will allow in the real time to convert matrixes of the large dimensionalities which is necessary in many tasks of recognition.

Second and even more perspective direction in development of computing devices in the future is the use of "quantum computers" [1,2]. Channels for data transfer in such computer will be more informative. For implementation of a defined structure of parallel calculations the defined set of atoms (molecule) will be used. Probably in future it will be possible to use equivalence of the descriptions of many appearances in micro and macro world.

Despite of wide use of specialized parallel processors many reseachers agree that we still don't have the common methodology for development of appropriate algorithms and devices. Probably we move to the necessity to comprehensively analyze a lot of artificial intelligence tasks. We need to somehow classify their settings and possible algorithms for their solution and to create a convenient formal logical means for the description of the tasks, algorithms and data.

## 2. One approach of creating the intelligent system

Let's take a look at one of the possible solutions of creating some kind of intelligent system. The term "intelligent" means here the ability of system to adapt to the real world conditions by making effective choice of the task to solve in the current moment. Such system could consist of two parts : internal one and ensuring interaction with the external world. Set of sensor controls and power installations controlling various organs of the device will link it with the external world. The internal part will match the following conditions:

- for each considered particular task the system has the particular device which is able to optimize the solving of this task by an appropriate choosing of the system's parameter from some final set,

- information from external world must be delivered to all such devices simultaneously.

By this conditions the system is able to interpret and to process the data received from external world through different devices simultaneously (in parallel).

We can imagine many ways of realization of such system. Further we'll discuss one of them. Suppose that we want to have a system able to solve any problem of some final set:

$$\{A_1, A_2, \dots, A_n\}.$$

Here the $n$ is some integer (may be equal $1, 2, 3$ or sufficiently large $100000, 1000000$). For example our devices can be presented as some big complicated molecule consisting of smaller parts

$$\{D_1, D_2, \dots, D_n\}$$

curled up in spiral or ball one by another. Such form of spatial representation allows to suppose that some biological or physical influence acts to the all of devices simultaneously. Considered small parts of a big molecule are the particular devices in above sense. It is possible to consider each of them as some quantum computer. We'll assume that the behaviour of each of them is determined by their internal structure, input flow of data and some parameters $\theta i$ from some set $\Theta i$ with final dimension

$$D_i = D_i(\theta_i),\ \theta_i \in \Theta_i,\ i = 1, 2, \dots, n.$$

The general system's parameter is denoted by

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix},\ \theta \in \Theta = \Theta_1 \otimes \dots \otimes \Theta_n.$$

This scheme satisfies both of above conditions. The parameters of data processing can be passed to devices among with the input data. For simplicity one also can present each device as particular neural network. The advantage of such presentation is an existing of the wide range of developed algorithms for attuning the system parameters. To fulfill the second condition the data should be duplicated and passed to the inputs of all particular networks.

For better clarity of presentation let's consider the simple example of the task that such system is intended to solve. Let's imagine that we have a number of devices each configured to recognize its own object from the image data coming from the digital camera. One of these devices can have a priority on controlling the camera (PTZ control) and each of them have some parameters to set up for better recognition. Assuming that the image is passed to the input of all these devices we have a system with all above conditions fulfilled. We also suppose that in the real world its impossible to recognize an object (such as a cat) with complete certainty. So each device will calculate some probability of recognition of the object. One can set up the barrier value of such probability (0.7 for example) that will separate positive and negative answers.

Here we came to the most important part of proposed solution — the concept of the informational resonance. After the data is delivered from all the sensors to the all devices, these devices begin to work and solve each its own task. We say that device came into resonance with the data if it succeeded in solving the task. In our example the device should give us a positive answer. There can be three different situations:

- only one device has resonance,
- several devices have resonance,
- none of the devices has resonance.

In the first case the resonant device accepts control on the system. In the second case the system should determine which device to use. It can be done by choosing the device with the greatest value of some parameter (probability of recognition in a last example). The third case means that none of the devices has resonance with the data but we need to choose the strong rule of a system behaviour. Assume that this rule is determined by setting the system parameter $\theta \in \Theta \subset \mathrm{R}^d$. We have a problem of the stochastic multidimensional optimization. There is an effective SPSA algorithm with one measurement per iteration proposed in [3–6] for solving this problem.

## 3. SPSA algorithm

Let's take a look at the method:

$$x_t = \theta^{t-1} + \beta_t \Delta_t,\ y_t = f(x_t) + v_t,$$

$$\theta^t = \theta^{t-1} - \frac{\alpha_t}{\beta_t} \Delta_t y_t.$$

Here $\{x_t\} t = 1, 2, \ldots$ is a sequence of parameters have been choosing at the time moment $t$; $\{\theta^t\}$ is the sequence of estimates of the adjusted vector $\theta$; $\{\alpha_t\}$, $\{\beta_t\}$ are some numerical sequences, matching some conditions; $\{\Delta_t\}$ is an observable sequence of independent random vectors with the same dimensionality as $\theta$ and with the distribution of Bernoulli (each component equals $\pm 1$ with probability 1/2) which are called *trial simultaneous perturbation*; $\{v_t\}$ is a sequence of disturbances in the measurements channel and $f(\cdot)$ is the functional on $\theta$. In the previous example it can be of the following type:

$$f(\theta) = 1 - \max(\mathrm{prob}_i(\theta_i)).$$

It has its minimum when the maximum of probabilities has its maximum.

This SPSA algorithm was designed to solve the problem of seeking the minimum point of the mean risk functional under almost arbitrary disturbance in measurement. Contrary to the many other optimizations it takes only one measurement (calculation) of $f(\cdot)$ per iteration. That is very useful property in our case because the type of the input data changes representation of $f(x_t)$ in time. Note that this method is pseudo gradient algorithm because the mean value (in probability sense) of expression

$$\frac{\alpha_t}{\beta_t} \Delta_t (f(\theta^{t-1} + \beta_t \Delta_t) + v_t)$$

is a "good" approximation of gradient vector of function $f(\cdot)$ at the point $\theta^t$.

One of the important features of the intellectual system is it's ability to the real time adaptation. In our case this means the changing the optimal $\theta$ vector at real time. There are two ways of using the SPSA method for solving this problem. The SPSA method is suitable for this kind of tasks. If the optimal value of $\theta$ does not change too fast it is possible to choose a sufficiently long time intervals and to make a several iterations with decreasing to zero sequences $\{\alpha_t\}$ and $\{\beta_t\}$. This enables us to get a "good" approximation of an optimal value of $\theta$. In many other cases we can choose (as in [7]) a sufficiently small constants $\alpha$ and $\beta$ instead of sequences.

## 4. Example of the intellectual system at real time

We'll consider the example of the real time intellectual system working on the above principles. Let's imagine some robot receiving information about the real world by using a number of devices like cameras, microphones, termometers, chemical analyzers, radiolocators etc, and having some devices allowing it to move, shoot, and so on. The central computer of this robot will consist of thousands of simple devices processing the same data and producing the internal representation of the external world in the robots brains. Each external device (like hand or leg) will be supplied by its own simple computer that will be running some driver of that device and will supply the central computer by the data about some corrections in the work of the device. This corrections will take place because of the inconsistency of the external world and its internal representation. The discussed inconsistency can be considered as some kind of disturbance in the real world measurements. So all the data coming to the inputs of the central computer is produces by external devices.

Central computer uses the internal representation of the world in its predictive calculations trying to produce some optimal trajectory to avoid the dangers and to carry out its mission. Some of the processed events are not involved in the building of internal representation. This concerns all types of the tasks that can be solved in the background not affecting the choosen trajectory (like small stones and branches on the road etc). These background calculations are performed by resonant devices of the central computer. In the discussed example there can work several resonant devices as they are solving different tasks and use different resources.

Most of devices of the central computer are involved in the calculating of the internal representation of the real world. This internal representation must be simple, structured, free of worthless detailes that can be processed in the background. It is made by performing image and sound recognitions, interpreting the radiolocation data, etc. Some of the produced facts can be treated as a signal to some immidiate reaction. For example it can be a loud noise making the robot to turn its cameras to the direction of the sounds. This tells us that for processing the internal representation we need to have a number of parallel calculations which can be performed by another chain of simple devices.

All the parameters of the used algorithms can be adjusted by the discussed above methods (like SPSA). For this reason we have to measure somehow the quality of our internal representation during the work. This can be made by calculating the number and the structure of corrections, made by external devices, quality parameters in recognition algorithms, etc.

In the more complicated case robot can have the connection to some satellites. In fact it means that the satellites are the robot external devices. One can name a lot of other possible devices, like road cameras, controlled through the Internet, other simplier robots, distributed everywhere in the world and so on.

## 5. Conclusion

The proposed type of system is based on assumption that it will be possible in the visible future to set up thousands of devices to work in parallel. This assumption is based on the great advance made by developers of quantum and DNA computers. It seems obvious to authors that the future of computer science lies in the field of parallel calculations where a great numbers of basic devices will work on solving of various particular problems.

## References


1. *P.W. Shor* Quantum computing. Proc. of the 9th Int. Math. Congress. Berlin, 1998, www.math.nine.edu/documenta/xvol-icm/Fields/Fields.html

2. *D. Deutsch, R. Jozsa* Rapid solutions of problems by quantum computer // Proc. Roy. Soc. London, Ser. A, 449, 1992, p. 553–558.

3. *O. N. Granichin* A stochastic recursive procedure with dependent noises in the observation that uses sample perturbations in the input // Vestnik Leningrad Univ. Math., 1989, 22: 1(4), p. 27–31.

4. *B.T. Polyak, A.B. Tsybakov* On stochastic approximation with arbitrary noise (the KW case) / In: Topics in Nonparametric Estimation. Khasminskii R.Z. eds.. // Advances in Soviet Mathematics, Amer. Math. Soc.. Providence, 1992, 12, p. 107–113.

5. *J.C. Spall* A one–measurement form of simultaneous perturbation stochastic approximation // Automatica, 1997, 33, p. 109–112.

6. *O. N. Granichin* Randomized algorithms for stochastic approximation under arbitrary disturbances // Automation and Remote Control, 2002, 63: 2, p. 209–219.

7. *O. N. Granichin* Nonminimax filtering in unknown irregular constrained observation noise // Automation and Remote Control, 2002, 63: 9, p. 1482–1488