An Approach to Improve the Architecture of ART-2 Artificial Neural Network Based on Multi-Level Memory

D G Bukhanov[1] **and V M Polyakov**[1]

[1] Department of software for computers and operating systems, Institute of energy, information technologies and operating systems, Belgorod State Technological University named after V.G. Shouhov, Russia

Abstract. This paper describes research of artificial neural networks based on adaptive resonance theory. Discovered the shortcomings and problems of applying the existing structures of these networks in real-time systems. To solve the problem, a new model of the recognition field F2 of the ART-2 network is proposed, which is a tree structure, with a recurrently changing similarity parameter for each subsequent level. At each level, the similarity parameter increases, that leads to a sequential search for an active resonating neuron. Computational experiments have been carried out to prove the effectiveness of the proposed approach in comparison with the classical realization of networks of adaptive resonance theory.

1. Introduction

Nowadays, more and more tasks of automated and automatic recognition of an object technical conditions are solved using intelligent methods of data analysis [1]. One of such methods are artificial neural networks (ANN). The development of the ANN application is going on in several directions [2]: technology and telecommunications [3], information technologies for working with texts, methods for text recognizing, also, neural networks dominating in determining of the keynote of the text. The next direction is the application of ANN in the field of economics and finance, advertising and marketing.

ANN can be used to recognize images represented by both graphic forms and numeric parameters. To recognize patterns represented by numerical parameters, dominating network described by Grossberg and Carpenter [4, 5]. Their proposed network of adaptive resonance theory (ART) with discrete inputs takes place in diagnosis of digital devices [6]. Due to the flexibility of the network structure, it is used to control automatic systems [7]. There are some modifications of ART, which can work with fuzzy inputs [8]. The network architecture, which works with continuous values of inputs, has short name ART-2. Structures of such networks consist of three types of fields: the input field of comparison, the output field of recognition and the decision field (the reset module), which forms the control action to the recognition field. The basic principle of the of these networks includes finding the correspondence between the rising sensor signal to the expected descending signal. One of the possible applications of such network type is the recognition of current state of the computer network for diagnosing possible failures [9].

2. Main idea of the ART-2 network

Fig. 1 shows the structure of ART-2. It consists of three main fields F1, F2 and G [6]. The field F1 can be represented by a tuple F1 = {S, W, X, V, U, Q, P, WN, VN, PN}, where the first seven elements are the layers of neurons processing ART-2, and the last three – normalizing elements. Field F2 consists of recognizing neurons Y_j (j = 1..m, where m is the number of different classes of images), F2 = {Y}. Field G is a tuple consisting of control neurons R_i (where i = 1..n, and n – number of input parameters), G = {R, RN}, where RN normalizing element.

Theorems that describe main principles of ART networks are proved in [4, 5]. The main corollaries of the theorems are:

- the process of searching for a previously trained image is stable, i.e. after determining the winning neuron in the network there will be no stimulations of other neurons, Only the reset signal can activate them;
- the learning process is stable, training weights of the winner neuron will not lead to switching to another neuron.

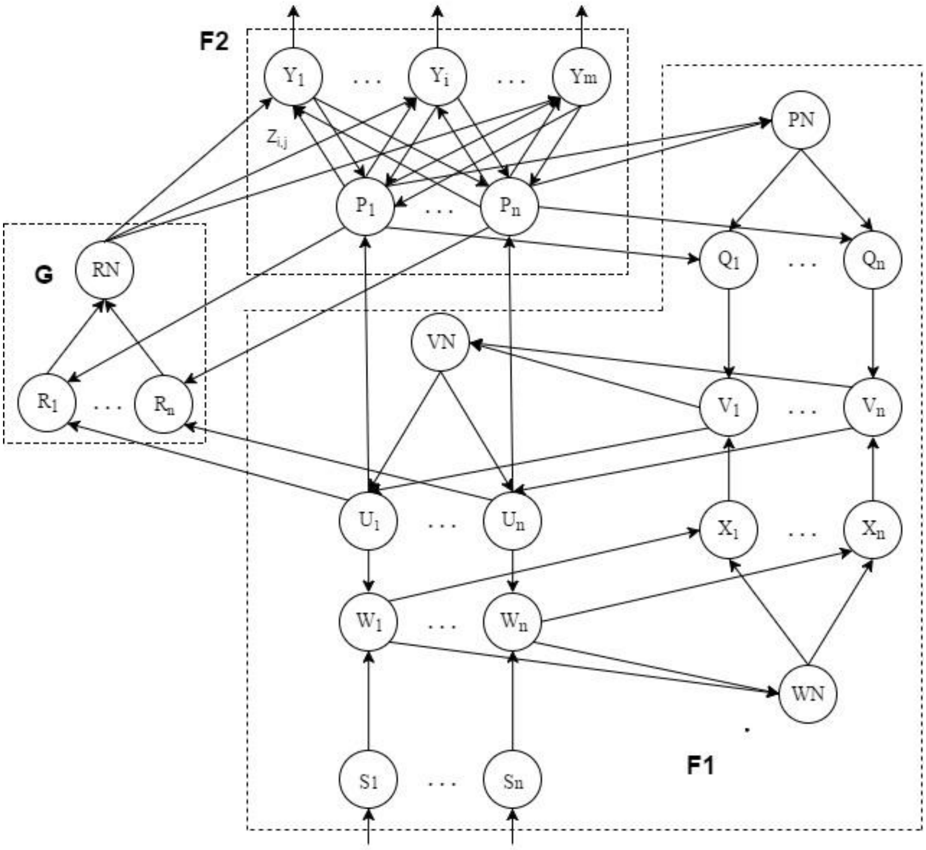


Figure 1. Structure of the ART-2 network.

There are no fundamental differences in learning and recognition processes with such structure of a network [10]. The only difference is that there is a change in the training weights of connections z_{ij}, made by recognizing neurons. And in recognizing process the change in weights does not happen.

The algorithm of the network working consists of the following stages.

Stage 1. The input of S-layer receives a vector of signals consisting of n elements. Neurons of W-layer perceive the signals of S-layer and add them to the output signals of the U-layer.

$$w_i = s_i + au_i;$$
$$i = \overline{1..n}.$$

Stage 2. Output signals w_i of W-layer neurons proceed to the inputs of X-layer and into the normalizing element WN, which calculates the norm:

$$\|w\| = \sqrt{\sum_{i=1}^{n} w_i^2};$$

$$x_i = \frac{w_i}{e + \|w\|};$$
$$i = \overline{1..n};$$
where e – small positive constant that prevents division by zero.

Stage 3. The output signals of V-layer are determined by expression:
$$v_i = f(x_i) + bf(q_i);$$
$$i = \overline{1..n};$$
where *f()* – threshold function, to suppress noise signals:
$$f(x) = \begin{cases} x, & if \ \ x > \theta; \\ 0, & else; \end{cases}$$
where θ – noise threshold.

Stage 4. Output signals v_i of V-layer neurons proceed to the inputs of U-layer and into the normalizing element VN, which calculates the norm:
$$\|v\| = \sqrt{\sum_{i=1}^{n} v_i^2};$$
$$u_i = \frac{v_i}{e + \|v\|};$$
$$i = \overline{1..n}.$$

Stage 5. The signals p_i are uniquely determined by u_i. The output signals of the P-layer neurons proceed into the normalizing element PN and to the group of Q-neurons.
$$\|p\| = \sqrt{\sum_{i=1}^{n} p_i^2};$$
$$q_i = \frac{p_i}{e + \|p\|};$$
$$i = \overline{1..n}.$$

Then repeat stages 3-5. Stable state in F1-field is set after two iterations.

Stage 6. The output signals from the P-layer proceed to the recognition layer Y, and the maximum element y_i.is calculated
$$y_j = \sum_{i=1}^{n} z_{ij} p_i; \quad (1)$$
$$j = \overline{1..m};$$
$$i\max = \max(Y);$$
then an output signal from the Y-layer is calculated:
$$p_i = u_i + z_{i\max i} d;$$
$$i = \overline{1..n};$$
where z – weight relationships of neurons in P-layer to the neurons of Y-layer.

Stage 7. The signals of the elements of the control layer are calculated.
$$r_i = \frac{u_i + cp_i}{e + \|u\| + c\|p\|}. \quad (2)$$

The normalizing element RN calculates an output signal.

$$\|r\| = \sqrt{\sum_{i=1}^{n} r_i^2}. \tag{3}$$

Stage 8. If $\|r\| \geq \rho$, (where ρ – the parameter of the correspondence of the expected result to the current one, which changes in the interval [0;1]), then it is assumed that the currently active neuron of Y-layer with an index *imax* of F2-field, is the winner. Then there is additional training of its weights.

$$z_{i\max i} += d(1-d)\left(\frac{u_i}{1-d} - z_{i\max i}\right);$$

$$i = \overline{1..n}.$$

If $\|r\| < \rho$, then the current active neuron of Y-layer is frozen and not involved in further competition.

Stage 9. If not all neurons Y-layer are frozen then proceed to Stage 6, otherwise a new neuron is created in the Y-layer, m increases by 1. It is considered that this neuron will resonate with new image, so weights are calculated as follows:

$$z_{mi} = d(1-d)\frac{u_i}{1-d}.$$

Figure 2 shows a graph of the change in the weight value z_1 of the resonant neuron of the Y-layer, with a fixed approximate output signal in the F1-field (U1≈0,62).

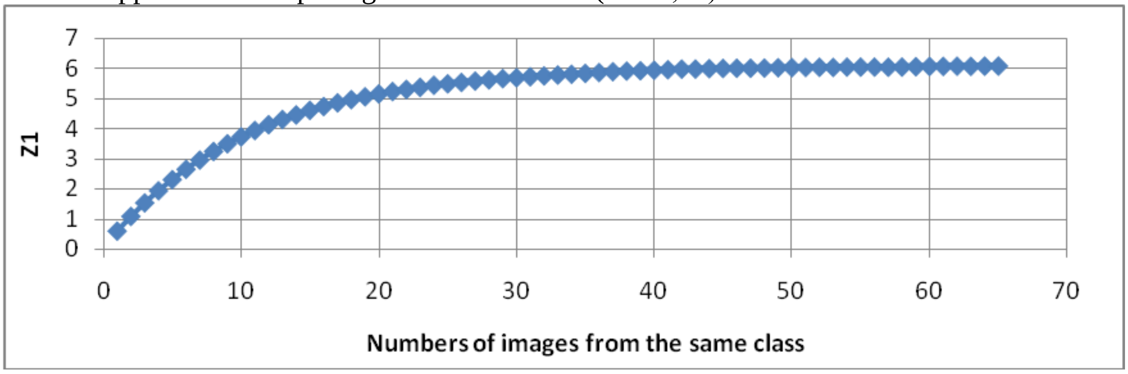


Figure 2. Changes of z_1 weigths.

This graph shows that the changes of weights in the memory of F2-field aims to a limit, and it confirms theorem formulated in [4], about the finiteness and stability of the learning process.

When using the classical implementation of the network the following problems have been identified.

1. Significant increase in recognition time when learning a network with a large number of different data vectors with a great similarity parameter, since for each class of images a separate neuron is created, in the Y-layer.
2. The impossibility of parallelization of the recognition process, in view of the fact that the memory, which is represented by the matrix of weight coefficients z, is sensitive to the input sequence of recognizable images.
3. Long search for the active Y-layer neuron, since the search is performed only based on the maximum y_i (1), which characterizes the frequency of input images.

3. Structure of ART-2 with self-organizing memory

Consider the proposed structure of ANN ART-2 with a modified memory (ART-2m). It consists of three main fields F1, F2, G. Field F1 consists of the same elements as in the classical implementation ART-2, as described in Section 1 above.

As follows from the theory of adaptive resonance, after checking all neurons of the Y-layer new element y_m is added. Verification of the correspondence of the rising signal to the active descending signal from the Y-layer occurs as follows:

$$\frac{\sqrt{\sum_{i=1}^{n}(u_i + cp_i)^2}}{\sqrt{(u_i)^2} + c\sqrt{(p_i)^2}} > \rho;$$

where $p_i = u_i + z_{i\max i}d;\ z_{i\max i} += d(1-d)\left(\frac{u_i}{1-d} - z_{i\max i}\right);\ i = \overline{1..n}.$

Hence it follows that the resonance between the memory and the rising signal will be if and only if u_i will be proportional to z_i:

$$u_i \equiv z_i;\ i = \overline{1..n}.$$

This statement and the theorem, experimentally confirmed above, allow us to organize a multilevel memory with different parameters on each of the levels. Thus, the F2-field can be represented as layers of recognition Y-neurons with different levels detailing of the images and *M*-layers – semantic neurons connecting Y different levels, F2 = {Y, M}. Then, the G-field is a tuple consisting of control neurons R_i (where i = 1..n, n – number of input parameters), *RN* is a normalizing element, $Riter_p$ – threshold element ($p = \overline{1..kc.}, kc$ – number of memory levels), *G = {R, Riter, RN}*. Fig. 3 shows the structure of the fields F2 and G with modified memory.

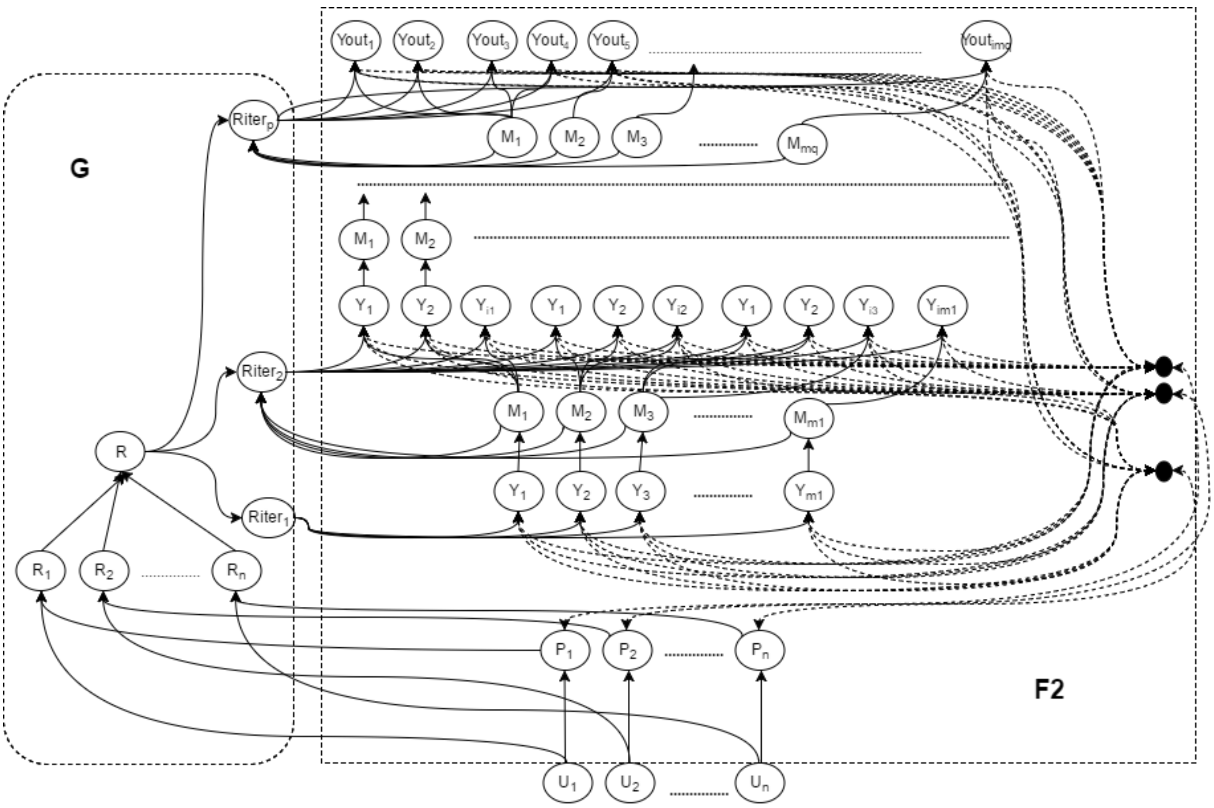


Figure 3. F2 and G field structure with self-organized memory.

To reduce the search time of an active neuron y_i, F2-field structure can be represented as a tree of memory with a concrete measure of similarity at each of the levels of the tree Riter, moreover, the learning of memory at the same level does not affect the weight of the above elements.

The algorithm of the network consists of the following steps.

Stages 1-5. Working in the F1-field, there in no differences from the classical implementation described above.

Stage 6. Required to calculate Riter for each level. If the difference between the similarity parameter and $\|r\|$ will be smaller, then тем more number of different neurons of Y-layer will be created, Accordingly, it is required to construct a recurrence relation of finding the similarity parameter for each level of the memory tree. In this research used the following recurrence relation with the initial measure of similarity equal to 0.5:

$$Riter_1 = 0.5;$$
$$Riter_{k+1} = Riter_k + 0.75(1 - Riter_k);$$
$$i = \overline{1..kc}.$$

Stage 7. The output signals from the P-layer are proceeded to the recognition Y^k-layer and the maximum element is calculated y^k_j.

$$y_j^k = \sum_{i=1}^{n} z_{ij}^k p_i;$$
$$i\max^k = \max(Y^k);$$
$$p_i = u_i + z^k_{i\max i} d;$$
$$i = \overline{1..n};$$

where z^k – weight relationships of neurons in layer P-layer to the neurons Y^k -layer.

Stage 8. Then, the signals of the elements of the control layer are calculated by (2.3).

Stage 9. If $\|r\| \geq Riter_k$, then it is assumed that the current active neuron of Y^k -layer F2-field, is the winner. Then there is a further training of his weights and proceed to stage 10.

$$z^k_{i\max i} += d(1-d)\left(\frac{u_i}{1-d} - z^k_{i\max i}\right).$$

If $\|r\| < Riter_k$, then the current active neuron of Y^k -layer frozen and further in the competition is not involved. If all neurons of the current Y^k -layer are frozen or completely absent, then proceed to stage 11.

Stage 10. Jump to the next memory level, associated with the current active neuron of Y^k -layer by M^k relationships.

M^k -layer stores semantic relationships of Y^k -layer neurons with Y^{k+1}. If $k \leq kc$, then proceed to stage 7, otherwise, it is considered that the active neuron of Y^{k+1}-layer is the one in whose weights the final result of recognition is stored.

Stage 11. Next, a new neuron of Y^k -layer is created, m is increased and his weights are trained according to the following rule:

$$z^k_{mi} = d(1-d)\frac{u_i}{1-d}.$$

The proposed structure of the ART-2 network has the following features.

1. The search of an active neuron, in which the result image is stored, now not only based on the frequency of the images of the corresponding classes, but also based on the semantic relationships that are formed at the network learning time. This fact significantly reduces the number of checks, since all memory is now represented in the form of a tree structure.
2. The possibility of parallelizing the calculations in the field F2, which is due to the impossibility of the Y^k -layer neuron stimulation, which is not semantically related to the current.

4. Experimental research of ART-2m network working

The experiment was carried out with 10 inputs. As inputs to the analysis, were generated random numbers belonging to the interval (0;1000]. The speed of network training was assessed.

Figure 4 shows the graphs of the dependence of the learning rate on the number of different input images for the classical implementation of the neural network ART-2 and proposed ART-2m, with multi-level memory.

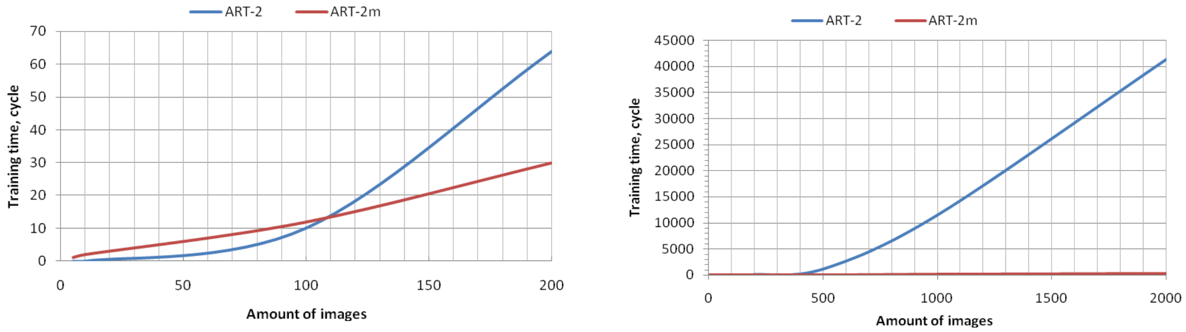
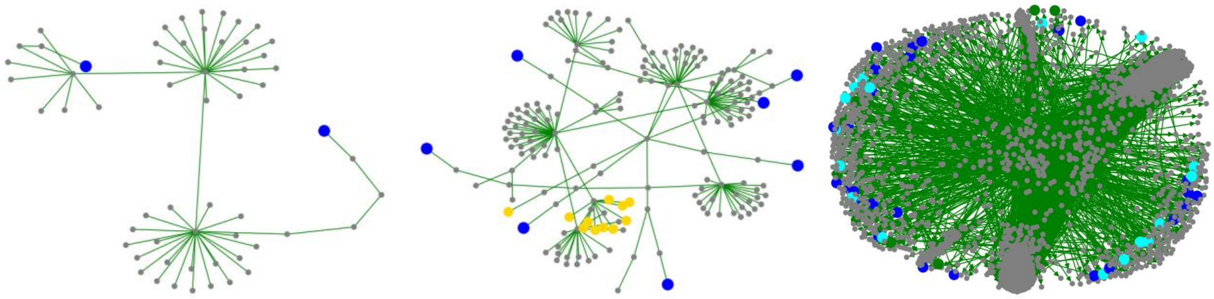


Figure 4. Comparative analysis of the ART-2 and ART-2m network training time.

To visualize the results of the network working, used the matplotlib [11] library for Python. Figure 5 shows a part of the memory structure when outputting a different number of Y-neurons. The track vertices are the number of vertices from the beginning to the last neuron, which is highlighted in the figure. The final vertices determine the current recognition class.



10 track vertices
130 track vertices s
10000 track vertices

Figure 5. Memory structure of ART-2m network.

From the results of the proceeded experiments it follows that when using the proposed memory structure, the network operation time at the selected intervals varies linearly. This is due to the existence of semantic relationships between different levels of the Y-layers in the search of an active neuron. The learning time for a small number of images for ART-2m networks is longer, since more time is required to build a memory tree. But as soon as the number of different filed images exceeds the product of the height of the tree by the average number of elements in the layer, ART-2m shows the best time.

It can be seen from the graphs presented in Figure 4 that as the number of input images increases, the learning time also increases linearly. This allows the proposed modification of the ART-2m network to be used to recognize more different data vectors than when using ART-2.

5. Conclusion

To eliminate the shortcomings of the classical implementation of ART-2, a modified model of the network of adaptive resonance theory with multilevel memory – ART-2m was proposed. Due to its tree structure of memory, it allows to reduce the search time of a previously stored image. Thus, it is possible to improve recognition accuracy by adding new memory levels with a more tough similarity parameter. The architecture of the ART-2m network allows you to perform image classification much faster, which makes it possible to apply it in real-time systems.

In the future, it is planned to develop an algorithm for parallel ART-2m network and carry out an experiments. It is also planned to apply it in diagnosing of the state of a computer network.

6. References

[1] Horsova A V 2016 Data mining *Theory and practice of modern science* vol 7 pp 356–359

[2] Melnikova A A and Mykhaylychenko Zh V 2017. Development of artificial neural networks. *BBK* 60 **27** p 128.

[3] Platonov V V and Semenov P O 2016 Detection of network attacks in computer networks using the method of data mining *Intellectual technologies in transport* vol 4(8) pp 16–21

[4] Carpenter G A and Grossberg S 1987 ART 2: Self-organization of stable category recognition codes for analog input patterns *Applied optics* vol 26(23) pp 4919–4930

[5] Carpenter G A, Grossberg S and Rosen D B 1991 ART 2: An adaptive resonance algorithm for rapid category learning and recognition *Neural networks* vol 4(4) pp 493–504

[6] Dmitrienko V D, Terjohina V M and Zakovorotnyj A Ju 2004 Computing device for recognition of operation modes of dynamic objects // *Bulletin of the National Technical University Kharkov Polytechnic Institute. A series of informatics and modeling* vol 34 pp 70–81

[7] Grosspietsch K E and Silayeva T A 2012 Modified ART network architectures for the control of autonomous systems *International Conference on Product Focused Software Process Improvement* (Springer Berlin, Heidelberg) vol 1 pp 300–319

[8] Majeed S, Gupta A, Raj D and Rhee F C H 2018 Uncertain fuzzy self-organization based clustering: interval type-2 fuzzy approach to adaptive resonance theory *Information Sciences*, **424** pp 69–90

[9] Bukhanov D G, Poljakov V M and Smakaev AV 2017 Diagnosis of the state of a computer network based on the ART neural networks *Bulletin of BSTU named after V.G. Shouhov* vol 7 pp 157–162

[10] Grossberg S 2013 Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world *Neural Networks* vol 37 pp 1–47

[11] McClarren R G 2018 NumPy and Matplotlib *Computational Nuclear Engineering and Radiological Science Using Python* chapter 4 pp 53–74