Quantitative determination of fault tolerance of memristor-based artificial neural networks

S N Danilin, S A Shchanikov, I A Bordanov and A D Zuev

Vladimir State University named after Alexander and Nickolay Stoletovs, 87, Gorkogo Street, Vladimir, 600000, Russia

E-mail: seach@inbox.ru

Abstract. The authors have reviewed interpretations of the terms “dependability” and “fault-tolerance” in Russian and interstate standards. A new quantitative criterion of fault-tolerance of the memristor-based artificial neural networks is proposed and substantiated. The authors have also proposed and provided rationalization for a revised definition of fault-tolerance as a property of the memristors-based artificial neural networks, which most fully conforms to the new version of its quantitative criterion. An example of the application practice for the fault-tolerance criterion during the design stage of an artificial neural network of a test degree of complexity is given.

1. Introduction

Today, all the world’s largest manufacturers of computing machinery are actively engaged in developing exaflop supercomputers, essential to nuclear, chemical and biological processes modeling and required for creating industrial and military systems of national security. An extensive work in this direction has been launched in Russia following the decisions of the President and the Government.

The publications of the leading foreign and Russian experts and researchers indicate that the recently produced nano-sized memristors [1-6] have the highest technical potential for aiding the implementation of supercomputers. The main reason for such development is that the new electronic components allow for creating giant-powered computers, the parameters of which are incomparably better than any available today [1-6].

One of the main scientific research trends is finding design, engineering, and process solutions for implementation of the memristors-based ANNs (ANNMs), which would have a benefit of bringing the networks’ key parameters and characteristics more in line with the potentially attainable values.

It may seem that with the level of development the world has reached today in electronics, circuit engineering, micro- and nanotechnologies, design and production automation facilities, the said task would not be extremely difficult to solve. In reality, it proved to be quite different. The research and manufacturing giants: HP, IBM, Intel, Sony, and Samsung, having estimated their resources and following the Moore’s law, had planned to create specialized and universal memristors-based exaflop neural computers by 2017. Yet, today, their plans remain unfulfilled [2, 7].

According to a number of reputable researchers, the main reasons for this are the following:

- imperfection of the ANNM theory [8];
- imperfection of technology of nano-sized electronic elements production [7];
- the unique electrophysical properties of memristors, which are understudied [9].
Apart from the above-stated reasons, the authors of this article have established and proved in a number of their works [6] that the ANNM should be studied, designed, produced and operated as united physical and informational objects implemented through hardware and software learning tools. Information and physical internal and external factors destabilizing normal ANNM performance generally act as dependent cofactors.

This article discusses the development of the ANNM theory component, which is referred to as a priority in the work [8]: it is a synthesis of approaches, methods and algorithms of ensuring the ANNM dependability at all stages of its life cycle in compliance with the interstate and Russian standards.

2. Approach
Performance accuracy is the main parameter of the ANNM as information transducers. In the theory of tolerances, the limits of performance accuracy variation, along with parameters of any computing tool elements, are usually defined through a margin tolerance, which meets the technical requirements for design, production and operation. Applied to the ANNM, values of the metrics characterizing accuracy may be divided into 3 types: the $X_{tr}$ achieved while training, $X_{a}$ allowable when solving specific practical tasks, and $X_{i}$ obtained during destabilizing influence. Figure 1 shows the margin tolerances for each type.

![Figure 1. Margin tolerance for the ANNM.
Left to right, top to bottom, by blocks: achieved while ANN training ($X_{tr}$), error – $X_{tr}$, obtained during destabilizing influence ($X_{i}$), error – $X_{tr}$, $X_{i}$, allowable when solving specific practical tasks ($X_{a}$), error – $X_{a}$, $X_{tr}$, $X_{i}$, ANN failure $X_{i}>X_{a}$.]

The ANNM operation tolerance is the maximum permissible deviation of the actual value of a quality parameter, which characterizes the network’s functional properties, from the reference value. The ANNM failure is an event occurring during operation, when the ANNM performance value expanded beyond the limits of the specified operation tolerances. Based on the phenomena and processes discussed above, a more complete definition of the ANNM fault-tolerance (FT) may be provided.

FT is the property of an ANNM to maintain the required quality (accuracy) of its performance within
the specified tolerances for the required service hours (duration and (or) work progress) with any variations in parameters of the elements, structures, information, mathematical, algorithmic and software support under the influence of internal and (or) external information and (or) physical factors.

A review of scientific and engineering sources done by the authors of the article indicates that there are no publications on the practical definition of the ANNM dependability and quantitative criteria for the FT definition.

In [10], the authors of the article proposed a general approach and a quantitative criterion for the ANNM fault-tolerance, which they referred to as $U$.

Based on numerical values of this criterion, the ANNM dependability graph may be constructed and the main dependability factors may be calculated and subsequently optimized using the current standards. The main factors and characteristics of a technical object dependability depend on the service hours (duration $t$ and (or) volume $v$ of work progress). The criterion $U$ proposed by the authors of [10] is statistic and univariate, which is its drawback and an obstacle for its full integration into the existing Russian and interstate standardized methodology for defining and ensuring dependability of the IS, information and technological systems and (or) software, NC, ANN and ANNM. The authors removed this drawback of the $U$ criterion. The modified multivariate dynamic criterion $U_{a(i,j,k,l,p,r,s,f)}(t,v)\geq 0$ complies with the revised definition of FT. The practice of its application is discussed below.

A systemic, multivariate, dynamic version of the criterion of the ANN fault-tolerance, $U_{a(i,j,k,l,p,r,s,f)}(t,v)$ is applicable at each level of the structural-functional hierarchy with the set operation tolerances, and being invariable with respect to the structure and type of problems solved by the ANN:

$$U^{a,a}_{i,j,k,l,p,r,s,f}(t,v) = 1 - \frac{X_{i,j,k,l,p,r,s,f}^{a,a}(t,v) - X_{i,j,k,l,p,r,s,f}^{a,a}(t,v)_{\text{tr}}}{X_{i,j,k,l,p,r,s,f}^{a,a}(t,v)_{\text{tr}} - X_{i,j,k,l,p,r,s,f}^{a,a}(t,v)},$$

where $X_{i,j,k,l,p,r,s,f}^{a} > X_{i,j,k,l,p,r,s,f}^{a}(t,v)_{\text{tr}}$

$$U^{a,a}_{i,j,k,l,p,r,s,f}(t,v) = 1 - \frac{X_{i,j,k,l,p,r,s,f}^{a,a}(t,v)_{\text{tr}} - X_{i,j,k,l,p,r,s,f}^{a,a}(t,v)}{X_{i,j,k,l,p,r,s,f}^{a,a}(t,v)_{\text{tr}} - X_{i,j,k,l,p,r,s,f}^{a,a}(t,v)_{\text{tr}}},$$

where $X(t,v)_{\text{tr}} < X(t,v)$

where $X_{i,j,k,l,p,r,s,f}$ is an $n$-th ($n = 1,2,...N$) factor of the ANN performance quality of the $a$-th level of the structural-functional hierarchy ($a = 1,2,...A$);

$X_{i,j,k,l,p,r,s,f}^{a}$ is an allowable value (tolerance) of the ANN performance metric;

$X_{i,j,k,l,p,r,s,f}^{a}(t,v)_{\text{tr}}$ is a value of the ANN performance metric achieved while training;

$X_{i,j,k,l,p,r,s,f}$ is the value of the ANN performance quality factor with variation of parameters of the $i$-th structure ($i = 1,2,...I$), $j$-th neuron ($j = 1,2,...J$), the $k$-th neural component ($k = 1,2,...A = 1,2,...K$), the $l$-th characteristic of the IO ($l = 1,2,...L$), $p$-th quality criterion of the MO ($p=q,1,2,...P$), $r$-th parameter of the AO ($r = 1,2,...R$), the $s$-th parameter of the PO ($s = 1,2,...S$), $f$-th noise parameter of the reference value ($f = 1,2,...F$).

3. Experiment

To recognize a digital signal with QAM256 modulation, the ANN model is synthesized and trained [10]; it should be prepared for its technical implementation based on memristors with the FT definition when operating under noise conditions (Figure 3).

The studied device instruments are the recognition of digital radio signals with QAM256 modulation. A sampled quadrature radio signal is sent to the ANNM input in a complex form, delayed by 32 cycles. The number of the ANNM inputs is 64, the number of neurons in the first layer is 128 with the antitangent activation function, and 1 neuron is in the second layer with the threshold function of activation (Figure 1). The test set consists of a pseudo-random signal sequence with FM4, QAM16 and QAM256 modulations. The estimated probability equal to $33.333\%$ corresponds to the lack of the QAM256 modulated signal recognition. Using the test set, the trained ANNM model ensures a probability of correct signal recognition with QAM256 modulation in the noise environment equal to $99.3154\%$. 


For clarity, we use the technology for determining the quantitative level of the ANNM FT in accordance with the expression (2) when \( t = v = const \) for the 2nd level of the hierarchy \( (a=2) \) and the 1st ANNM performance quality factor – the probability of the set signal recognition.

For a quantitative estimation of the general level of the ANNM fault-tolerance, we calculate the average level of the FT, which is determined using the expression (3), and the relative level of the FT, which is determined using the expression (4):

\[
U_{\text{mean}} = \frac{1}{N} \sum_{i=1}^{N} U_i, \quad (3)
\]

where \( N \) is a number of neurons (structural components) in the ANNM.

\[
U_{\text{rel}} = \frac{N_{\text{FT}}}{N_{\text{gen}}}, \quad (4)
\]

where \( N_{\text{FT}} \) is a number of fault-tolerant neurons (structural components) in the ANNM; \( N_{\text{gen}} \) is a total number of neurons (structural components) in the ANNM.

\[\text{FT is determined as follows:}\]

- We train a neural network to achieve the best result with respect to the criterion of the sum of squared errors (SSE).
- We record the obtained criteria values, i.e. mean squared error (MSE), mean absolute error (MAE), maximum absolute error \( \varepsilon_{\text{abs, max}} \) or relative error \( \varepsilon_{\text{rel, max}} \) of network operation or problem solving.
- We simulate critical failures of neurons (“0” or “1”).
- We record the values obtained for the SSE, MSE, MAE, \( \varepsilon_{\text{abs, max}} \) criteria during simulation.

\[X = 1.052\]

- Using the expressions (2)-(4) with the ratio \( X \approx 0.95 \) we calculate the values of the ANNM fault-tolerance level \( U_1, U_{\text{mean}}, U_{\text{rel}} \), which are entered in tables and then graphically expressed as intolerable to the failure of 16% of memristors-based neurons with respect to SSE – 2%, MSE – 2%, MAE – 4%.

Therefore, when the same neurons of each ANNM fail, the FT level \( U_i \) has different values depending on the selected criterion (factor) of quality or accuracy of the ANNM performance.

4. Conclusion
- The authors have reviewed interpretations of the terms “dependability” and “fault-tolerance” in
Russian and interstate standards. They have revealed their similarities and differences.

- A new revised definition of the ANNM fault-tolerance, which best complies with the new version of its quantitative criterion, is proposed and substantiated.
- A new version of the n-variate quantitative fault-tolerance criterion of the ANNM $U(t,v)$ as a function of the service hours (duration and (or) volume of work progress) is proposed and rationalized.
- Unlike the previously developed criteria, the $U(t,v)$ criterion completely conforms with and may be integrated into the system of intergovernmental standards on dependability of the IS and PO technology, and hardware and software systems.
- The authors give an example of the application practice for the $U(t,v)$ criterion during the design stage of the ANNM for radio technical signal recognition of a test degree of complexity.
- Consideration of a fault-tolerance as a dynamic property complies with the nature of change of the X parameters at all stages of the ANNM life, expanding the range of scientific and practical application of the $U(t,v)$ criterion.

![Figure 3](image)

**Figure 3.** Level of fault tolerance of the neural network realizing recognition of a signal with modulation of QAM256 for refusals of neurons of the “0” type on MSE value.

**Acknowledgments**

This study was conducted with the support of the Russian Foundation for Basic Research, grants No. № 19-07-01215.

**References**

[1] Antonov I N, Belov A I, Mikhaylov A N, Morozov O A and Ovchinnikov P E 2018 Formation of Weighting Coefficients in an Artificial Neural Network Based on the Memristive Effect in Metal–Oxide–Metal Nanostructures Journal of Communications Technology and Electronics 63(8) 950-57

[2] Bayat M F, Prezioso M, Chakrabarti B, Nili H, Kataeva I and Strukov D 2018 Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits Nature Communications 9(1)

[3] Adhikari et al 2015 A Circuit-Based Learning Architecture for Multilayer Neural Networks With Memristor Bridge Synapses IEEE Transactions on Circuits and Systems I: Regular Papers 62(1) 215-23

[4] Yang C, Kim H, Adhikari S P and Chua L O 2017 A Circuit-Based Neural Network with Hybrid Learning of Backpropagation and Random Weight Change Algorithms Sensors 17(1)
[5] Koryazhkina M N et al 2018 Bipolar resistive switching in metal-insulator-semiconductor nanostructures based on silicon nitride and silicon oxide Journal of Physics: Conf. Series 993

[6] Galushkin A I, Danilin S N and Shchanikov S A 2015 The research of memristor-based neural network components operation accuracy in control and communication systems 2015 International Siberian Conference on Control and Communications (Tomsk: Institute of Electrical and Electronics Engineers Inc.) pp 1-6

[7] Burr G W et al 2017 Neuromorphic computing using non-volatile memory Advances in Physics: X 2(1) 89-124

[8] Galushkin A I 2007 Neural Networks Theory (Heidelberg: Springer-Verlag Berlin) p 396

[9] Chua L O 2018 Five non-volatile memristor enigmas solved Applied Physics A 124 563-606

[10] Galushkin A I, Danilin S N and Panteleev S V 2014 Quantitative determination of fault tolerance for neuronetwork devices of infocommunication 24th Int. Crimean Conf. Microwave and Telecommunication Technology (Moscow: Institute of Electrical and Electronics Engineers Inc.) pp 328-29

[11] Zhiganov S N, Zhiganova E A and Mikheev K V 2019 Development of the Method for Constructing Trajectories of Air Objects 2018 Int. Multi-Conf. on Industrial Engineering and Modern Technologies (Vladivostok: Institute of Electrical and Electronics Engineers Inc.)