Захист інформації та кібернетична безпека

UDC 004.93 DOI: 10.30748/soi.2019.156.11

V. Krasilenko, A. Lazarev, A. Sheremeta

Vinnytsia National Technical University, Vinnytsia, Ukraine

DESIGNING OF ARRAY NEURON-EQUIVALENTORS WITH A QUASI-UNIVERSAL ACTIVATION FUNCTION FOR CREATING A SELF-LEARNING EQUIVALENT-CONVOLUTIONAL NEURAL STRUCTURES

In the paper, we consider the urgent need to create highly efficient hardware accelerators for machine learning algorithms, including convolutional and deep neural networks, for associative memory models, clustering, and pattern recognition. We show a brief overview of our related works the advantages of the equivalent models (EM) for designing bio-inspired systems. Such EM-paradigms are very perspective for processing, clustering, recognition, storing large size, strongly correlated, highly noised images and creating of uncontrolled learning machine. And since the basic nodes of EM are such vector-matrix (matrix-tensor procedures with continuous-logical operations as: normalized vector operations "equivalence", "nonequivalence", and etc., we consider in this paper new conceptual approaches to the design of full-scale arrays of such neuron-equivalentors (NEs) with extended functionality, including different activation functions. Our approach is based on the use of analog and mixed (with special coding) methods for implementing the required operations, building NEs (with number of synapsis from 8 up to 128 and more) and their base cells, nodes based on photosensitive elements and current mirrors. Simulation results show that the efficiency of NEs relative to the energy intensity is estimated at a value of not less than 10^{12} an. op. / sec on W and can be increased. The results confirm the correctness of the possibility of creating NE and MIMO structures on their basis.

Keywords: *self-learning equivalent-convolutional neural structure, neuron-equivalentor, current mirror, hardware accelerators, equivalent model, continuous-logic, activation function, nonlinear processing, recognition.*

Introduction

For creation of biometric systems, machine vision systems are necessary to solve the problem of object recognition in images. Discriminant measure of the mutual alignment reference fragment with the current image, the coordinate offset is often a mutual 2D correlation function. In paper [1] it was shown that to improve accuracy and probability indicators with strong correlation obstacle-damaged image, it is desirable to use methods of combining images based on mutual equivalently 2D spatial functions and equivalence models (EMs), nonlinear transformations of adaptive-correlation weighting. For the recognition, clustering of images, various models of neural networks (NN), auto-associative memory (AAM) and hetero-associative memory (HAM) are also used [2–3]. The EM has such advantages as a significant increase in the memory capacity and the possibility of maintaining strongly correlated patterns of considerable dimensionality. Mathematical models and implement of HAM based on EMs and their modification described in papers [3–4]. For of analysis and recognition should be solved the problem of clustering of different objects [4]. Hardware implementations of these models are based on structures, including matrix-tensor multipliers, equivalentors [5]. And the latter are basic operations in the most promising paradigms of convolutional neural networks (CNN) with deep learning [6–9]. Scientists create algorithms to identify previously unknown structures in data, including those whose complexity exceeds human understanding. In paper [10] we showed that the self-learning concept works with directly multi-level images without processing the bitmaps. But, as will be explained below, for all progressive models and concepts, nonlinear transformation of signals, image pixel intensities are necessary. For implementation SLECNS [9], we need certain new or modified known devices capable of calculating normalized spatial equivalence functions (NSEqFs) with the necessary speed and performance. In papers [4; 9–10] we showed models for the recognition and clustering of images that combine the process of recognition with the learning process. For all known convolutional neural networks, as for our SLECNS, it is necessary to calculate the convolution of the current fragment of the image in each layer with a large number of templates which are used, selected or formed during the learning process. But, as studies show, large images require a large number of filters, and the size of the filters can also be large. Therefore, the problem of increasing the computing performance of hardware implementations of such CNNs and their neurons-cells is acute. We proposed a new structure [11–12]. It consists of a micro-

display dynamically displaying current fragments, an optical node in the form of a micro-lens array (MLA) with optical lenses (not shown!) and a 2D array of equivalentors (Eqs) or non-equivalentors (NEq) with optical inputs. Simulation on 1.5μm CMOS in different modes has shown that the Eq and their base units can operate correctly in low-power modes and high-speed modes, their energy efficiency is estimated to be not less than 10^{12} an.op / sec per W the produced and can be increased by an order, especially considering FPAA. But much depends on the accuracy of the current mirrors (CM) and their characteristics. Thus, at the inputs of each Eq we have two arrays of currents representing the compared fragment and the corresponding filter, and the output of the Eq is an analog signal, nonlinearly transformed in accordance with the activation function. As will be shown work [10], non-linear component-wise transformations allow even without WTA network to allocate the most Eq with the greatest activity. And the use of an array of cells that perform hardware, nonlinear transformations adequate to auto-equivalence operations, allows in maps to automatically select these pixel-winners using only several transformations steps.

Setting goal and objectives of the work. From described it follows that for implementations of SLECNS, an important issue is the design of parallel nonlinear transformations of intensity levels. **The goal of the work is to design** and simulate specialized neural accelerators in the form of an array of neuron-equivalentors that effectively calculate the comparison function of two two-dimensional arrays without using multiplication operations, and using a nonlinearly transformed function of normalized equivalence of arrays.

The main material

1. Description and simulation of analog neuron-equivalentors and its nodes

The Fig. 1–2 shows the neu structure and block diagram of the main unit of SLECNS, allowing parallel, with a high rate, equal to the speed of selection from the processed image of its shifted current fragment, to compute a set of stream components (elements) immediately one-cycle all the equivalents convolutions of the current fragment with the corresponding filters. The matrix **X** forms a certain number of convolutions in the form of matrices **e** using a set of defined filters-templates **W** which, in our case, are multilevel values, in contrast to the binary ones we used earlier. Thus, we compare each filter with a current fragment in the matrix **X**. As a measure of the similarity of the fragment of the matrix **X** and the filter the equivalent measures of proximity or other measures such as a histogram can be used. Thus, we compare for each filter similar fragments in the matrix. Each **NEq** is implemented in a modular hierarchical manner and can consist of similar smaller sub-pixel, also 2D type, base nodes. The **NEq** has a matrix (ruler) of photo-detectors, on which a halftone image of the fragment is projected, and the number of electrical analog inputs equal to the number of photo-detectors, to which by means of any known method. Each **NEq** has its own filter mask from a set of filters formed by training. Thus, at the inputs of each **NEq** we have two arrays of analog currents representing the compared fragment and the corresponding filter, and the output of the **NEq** is an analog current signal, nonlinearly transformed in accordance with the activation function and representing measure of their similarity. In our case, this measure is a normalized equivalence (eq) and nonequivalence (neq).

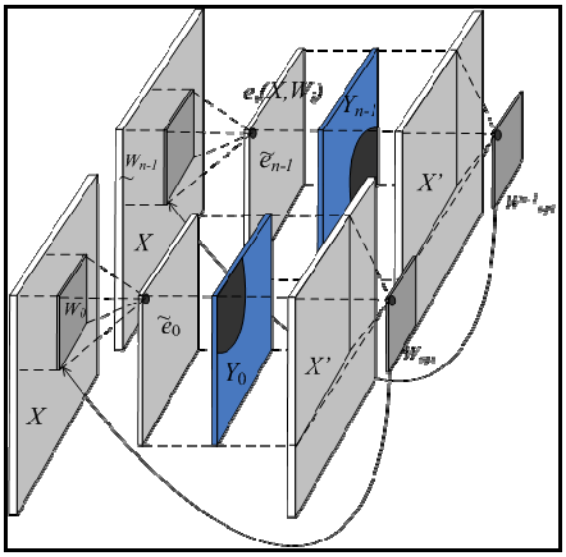


Fig. 1. The structure of the basic unit of the SLECNS, which explains the principle of its functioning; Figure explains the principles of learning neural network model based on the multi-port memory to find centroid cluster elements

The base node, see Fig. 3, contains *N* two input counters of maximum and minimum currents and one normalizer on CM, which forms two output signals corresponding to eq and neq from two vectors.

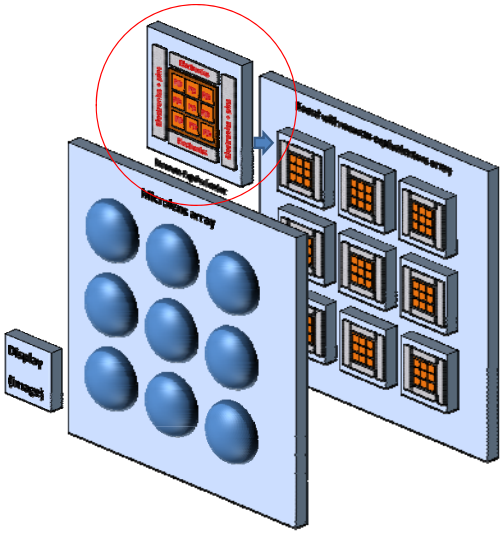


Fig. 2. The array of neuron-equivalentors, which is the main basic node of the SLECNS

The basic unit for calculating the normalized Eq (NEq) by averaging the component peak and minima of currents on the basis of CMs and the schemes of the limited difference is shown in Fig. 3. Sources of analog currents are shown as current generators for modeling in OrCAD. The dimension of the vector inputs is 9, which corresponds to the filter size 3x3. The results of modeling this basic unit with a additional nonlinear transformation are shown in Fig. 3 (right). At the instants of 11–12μs and 13–14μs, the output signals of equivalence and nonequivalence testify to the coincidence of the input vectors. The results of modeling the base unit for the filter size 3×3 (with 9 inputs) showed, that processing time is from 1μs to 0.1μs for currents Imax=5μA, consumption power is from 200μW to 50μW.

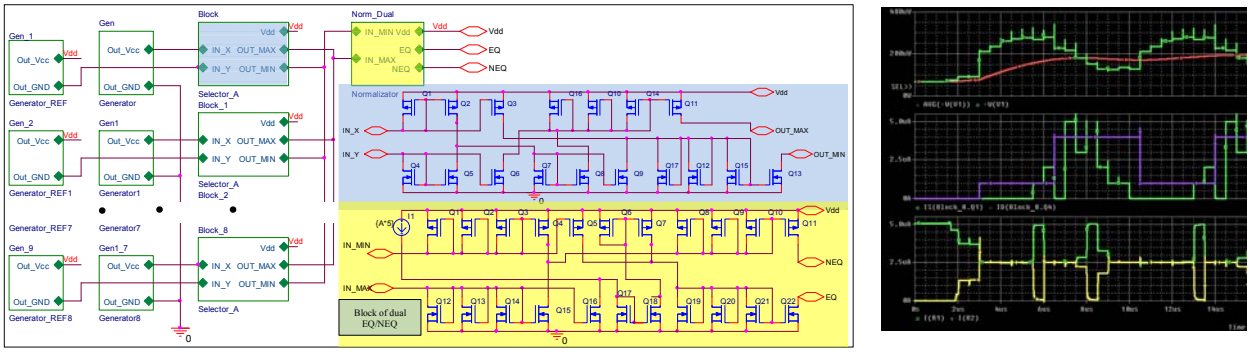


Fig. 3. On the left: The basic unit for calculating the normalized Eq (NEq) by averaging the component peak and minima of currents on the basis of current mirrors and the schemes of the limited difference. On the right: The results of modeling the base unit for the filter size 3×3 (with 9 inputs): on the left for current *Imax*=5μA, *T*=0.5μs, *V*=1.8V, *P*=200μW. On the right for current *Imax*=2.5μA, *T*=1μs, *P*=100μW. Red line shows power consumption, input (green) and reference (lilac) signals are showed on the midle graphs, on the bottom graph normalized eq (green) and neq (yellow) are showed

In addition to simulating the base node on 9 inputs, we additionally synthesized a neuron circuit having 8 such nodes, each of which compares 8 input vectors, resulting in a neural circuit having 2 vector inputs of 64 dimensions. For a non-linear transformation, we used a node of a circuit, which realizes a piecewise linear approximation of the power-law activation function (auto-equivalence). The simulation results of 64-input NE with nonlinear output conversion showed that such a NE provides good time characteristics and has a power consumption 2mW, a low supply voltage, contains less than 1000 CMOS transistors. To test the functioning of the developed NEs within the network, we created a mini-network of eight 9-input NEs, shown in Fig. 4. And the results of modeling it are shown in Fig. 5. Simulation in different modes has shown that the base unit (9-input NE) can operate correctly in low-power modes (1st and 2nd) and high-speed modes (3rd, 4th):1) *Imax* = 0,5μA, *T*=1μs, *Vdd* =1.8V, (*P* = 20μW); 2) *Imax* = 0,25μA, *T*=2,5μs, *Vdd*=1.8V, (*P*=10μW); 3) *Imax* = 5μA, *T*=0,1μs, *Vdd*=1.8V, (*P*=200μW); 4) *Imax* = 10μA, *T*=0,05μs, (*P*=500μW). If we take into account that at least 20 analog operations (9×2 comparisons by limited subtractions, current additions, their divisions and nonlinear transformations (2)) are performed by the NE for the one tact *T*, the relative to the energy efficiency of NEs is estimated to be not less than 20×10^6 an. op./sec: 20μW = 10^{12} an. op / sec per *W* and can be increased by an order. To check the quality of such neural-equivalentor circuits in networks, in the SLECNS, we also performed simulations in Mathcad. The simulation results are shown in Fig. 5 (right).

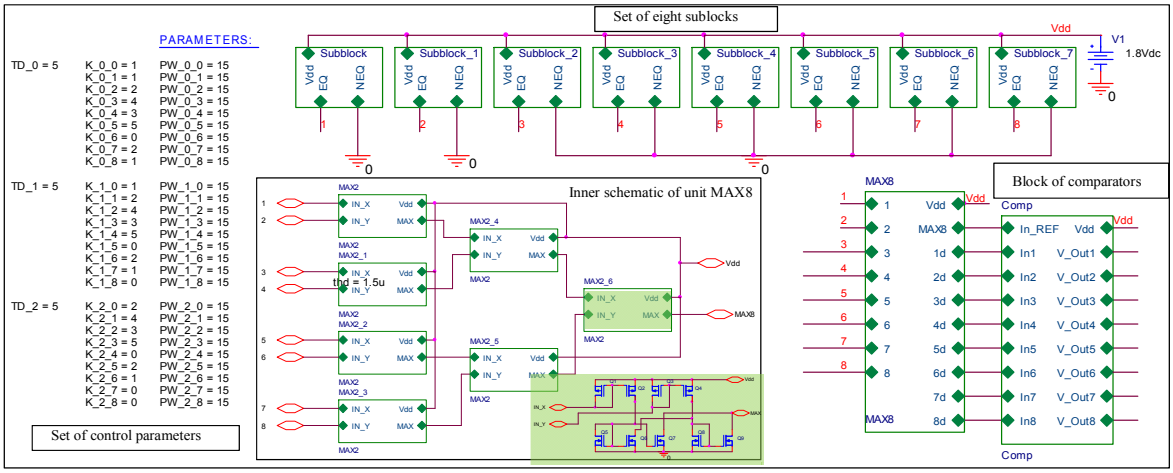


Fig. 4. A network of 8 9 input NEs based on 9 basic modules for calculating the normalized eq (neq)

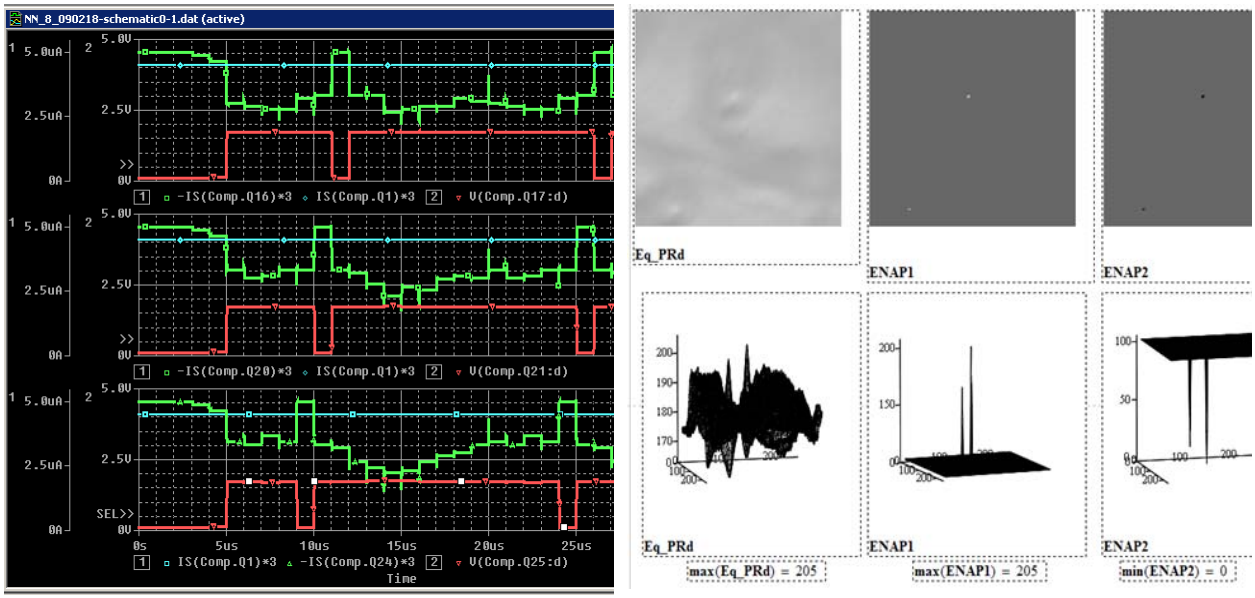


Fig. 5. On the left: The result of a network simulation of 8 9-input NEs, a fragment of the successive activity of three neighboring NEs, green – current outputs, blue – additional threshold, red – NE outputs (voltage, potential). On the right: The Mathcad windows on which the module of the program with results of recognition of fragments (left eye – filter-template) on the image are shown, where in 2D and 3D from left to right: the computed NE equivalent, non-linear (after activation) equivalent, non-linear non-equivalent (part) functions

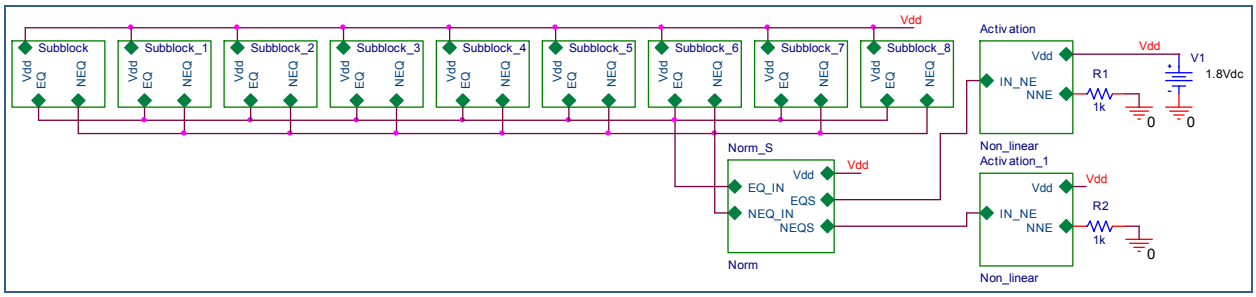


Fig. 6. Scheme of 81-input NE with dual outputs based on 9 basic units for computing normalized eq (in the diagram EQS) and neq (in the diagram NEQS) and their transformed responses (NNE outputs of activation circuits)

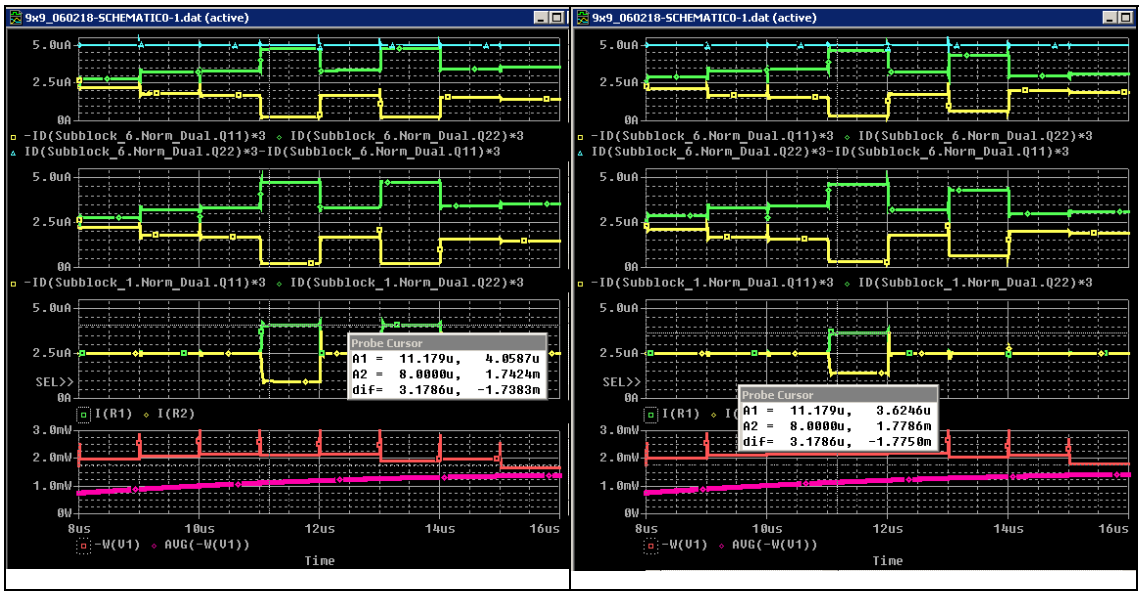


Fig. 7. The results of modeling the 81-inputs NE for the filter size 9×9 (with 81 inputs) for current *Imax* = 5μA, *T*=1μs, *V*=1.8V, *P*=1–2mW (bottom): on the left and right when comparing different input arrays (the measured levels are shown), green and yellow are the dual outputs of the sub-blocks and the entire NE

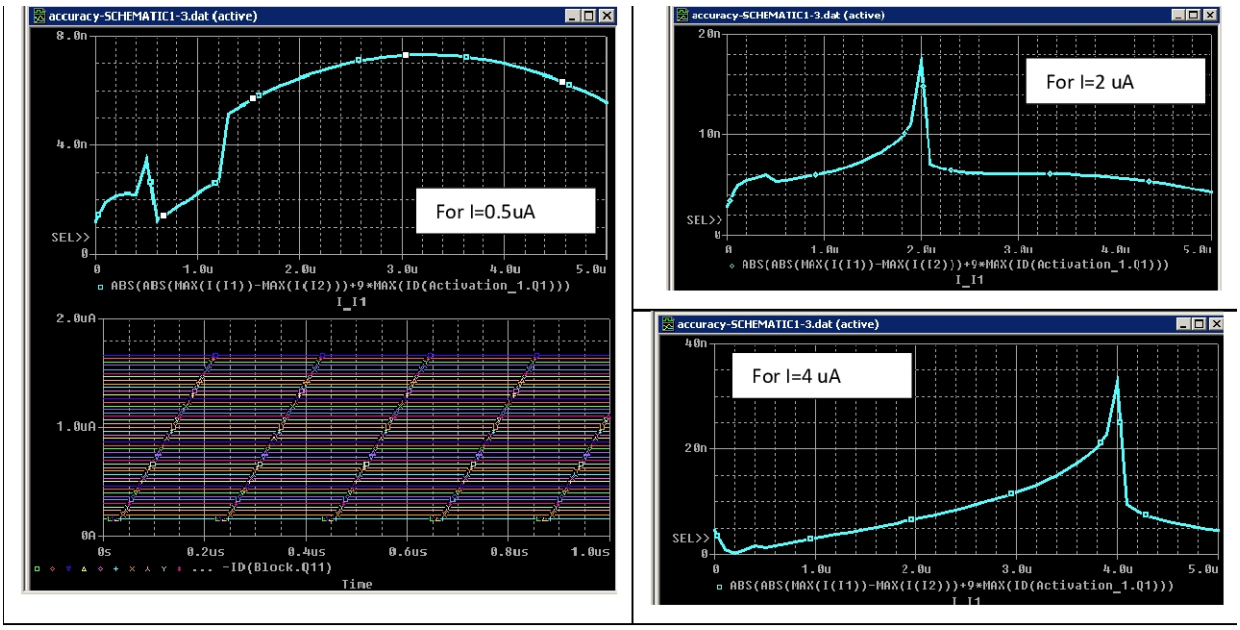


Fig. 8. The result of the calculation accuracy check for normalized eq for various reference DC currents with linearly increasing currents at the second vector inputs (for *Imax* = 5μA, *T*=1μs, *V*=1.8V)

In addition, on the basis of combining nine 9-input NEs, NEs were designed and modeled for two 81-component inputs, i.e. for convolution by a 9×9 filter. It has 2 bus analog inputs. Scheme of 81-input NE with dual outputs is shown in Fig. 6, and the results of its modeling in Fig. 7 and 8. As can be seen from Fig. 8, the maximum deviations of the generated values of the currents at the NE outputs from the mathematically calculated required values do not exceed 50nA, which corresponds to a relative error at the level of 1% and allows to hope for the possibility of obtaining, in the produced NEs, taking into account technological dispersion, no more than 5 %. For preliminary rough definitions of neuron-winners in neural and CNN networks this is enough. In addition, as will be shown below, non-linear componentwise transformations allow even without WTA network to allocate the most NEs with the greatest activity.

We modeled on 1.5μm CMOS, and therefore, especially considering the produced FPAA [13–16], there are all possibilities for this, but much depends on the accuracy of the current mirrors and their characteristics. The analysis of the obtained results confirms the correctness of the chosen concept and the possibility of creating NE and MIMO structures [17–18] on their basis, as hardware accelerators for compact high-performance systems of machine vision, CNN and self-learning biologically inspired devices.

2. Synthesis and modeling array universal cells for image intensity functional transformation

A lot of work has been devoted to the design of hardware devices that realize the functions of activation of neurons, but they do not consider the design of exactly the auto-equivalent transformation functions for EMs and the most common arbitrary types and types of nonlinear transformations. Therefore, the goal of this section is the design of cells for hardware parallel transformation of image intensity levels. In work [10], the question of the simplest approximations of auto-equivalence functions (three-piece approximation with a floating threshold) was partially solved. The basic cell of this approximation consisted of only 18–20 transistors and allowed to work with a conversion time of 1 to 2.5μs. At the same time, the general theoretical approaches to the design of any nonlinear type of intensity transformation were not considered, and this is the object of the paper. We will note that on a current mirror more easily to execute these operations of addition or subtraction of currents.

2.1. Simulation of image intensity transformation with Mathcad

Using both the basic components for the composition of the lambda function (LF) **fspΔs2** , shown in Fig. 9 and described by expression:

$$\mathbf{fsp\Delta s2}(\mathbf{xs},\mathbf{p\Delta x},\mathbf{p\Delta},\mathbf{k}) = \\ = \mathbf{k}\times\mathbf{obs}\left(\mathbf{obs}(\mathbf{xs},\mathbf{p\Delta x}),\mathbf{obs}(\mathbf{xs},\mathbf{p\Delta})\times\mathbf{2}\right),$$

where **xs** – argument, **pΔx** – parameter indicating the lower bound-level **xs** (beginning), **pΔ** – the second parameter indicating the level for the maximum, k – is the third parameter indicating the scalar gain multiplier; and

$$obs(a,b) = a \div b = \begin{cases} a-b, if\ a>b \\ 0, if\ a\le b \end{cases}$$

we proposed a LF-composition **fspΔsS** , which is calculated by:

$$\mathbf{fsp\Delta sS}(\mathbf{xs},\mathbf{\Delta k},\mathbf{VK}) = \\ = \sum_{\mathbf{i=1}}^{\mathbf{\Delta k}} \mathbf{fsp\Delta s2}\left[\mathbf{xs},\frac{\mathbf{255}}{\mathbf{\Delta k}}\times(\mathbf{i-1}),\frac{\mathbf{255}}{\mathbf{\Delta k}}(\mathbf{i}),\mathbf{VK_i}\right],$$

where **Δk** – number of components (LFs), **xs** – argument of the LF, **VK** – vector of gain factors. The result of constructing some types of transfer characteristics (TC) using these functions in the Mathcad environment is shown in Fig. 9. To approximate auto-equivalence, we also offer simpler (2-step) basic N-functions:

$$\mathbf{af}(\mathbf{xs},\mathbf{xp}) = \left[\mathbf{obs}\left(\mathbf{xs},\mathbf{obs}(\mathbf{xs},\mathbf{xp})\right)+\mathbf{obs}\left(\mathbf{xs},(\mathbf{DP}-\mathbf{xp})\right)\right]\cdot\left(\frac{\mathbf{DP}}{\mathbf{xp}\cdot\mathbf{2}}\right),$$

and triple their composition:

$$\mathbf{afS}(\mathbf{xs},\mathbf{VaF},\mathbf{KaF}) = \sum_{\mathbf{iv}=\mathbf{0}}^{\mathbf{2}} \mathbf{af}\left(\mathbf{xs},\mathbf{VaF}_{\mathbf{iv}}\right)\cdot\left(\mathbf{KaF}_{\mathbf{iv}}\right).$$

In general, the number of components in a composition can be arbitrary, but for modeling we used 8 and 16 component compositions and adjustment vectors. Examples of such functions and compositions for the synthesis of TC are shown in Fig. 10 (left).

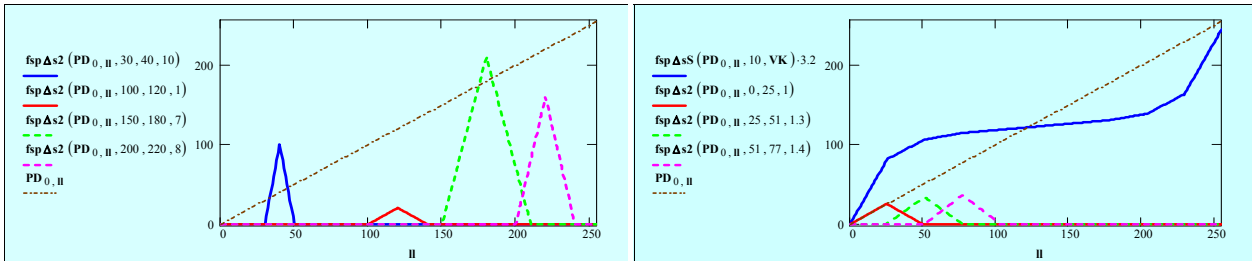


Fig. 9. Graphs of synthesized transformation functions

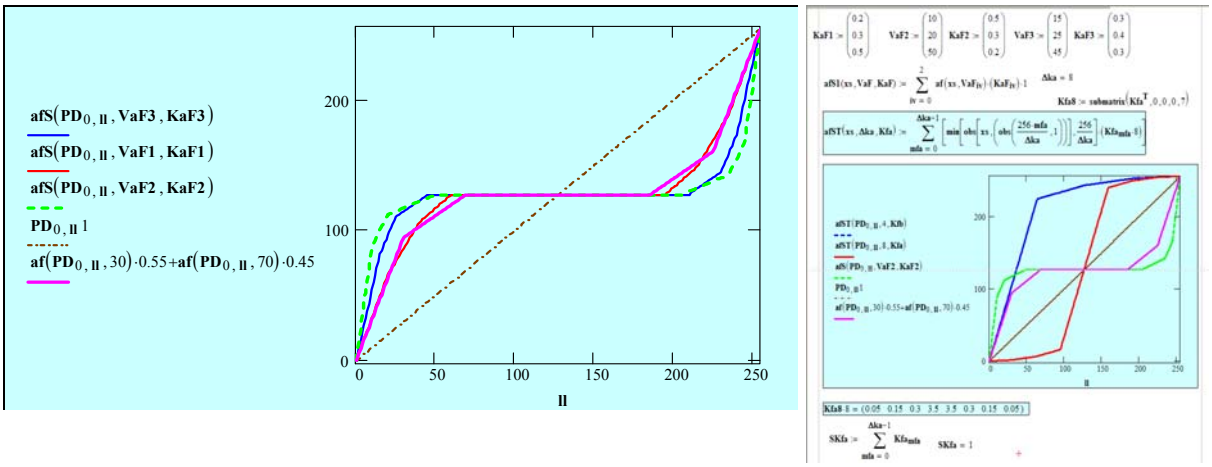


Fig. 10. Examples of synthesized transfer characteristics for auto-equivalence functions (left), Mathcad windows with the formulas and graphs of synthesized functions transformation (right)

Another variety of functions is shown in Fig. 10 (right), and the results of using such TCs to prepare the original PIC image are shown in Fig. 11.

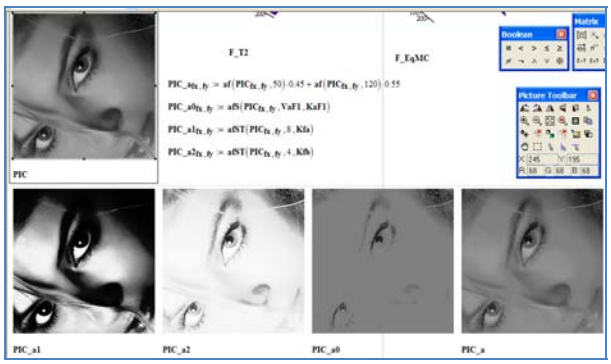


Fig. 11. Mathcad windows on which the formulas and results image intensity transformation are shown, where in 2D from left to right: input image PIC, the computed auto-equivalence functions, non-linear (after activation) output images (bottom row)

2.2. Design and simulation of array cells for image intensity transformation with Orcad Pspice

Let us first consider the design and simulation of a single base cell for the image intensity of an arbitrary transformation, using the example of a 4-piece approximation by triangular signals. In Fig. 12 shows the scheme used for modeling. To form 4 triangular signals from the input signal, we use 4 identical sub-nodes, each of which consists of 14 transistors (Ts) and an additional CM (2 transistors), and for propagation of the input photocurrent (IPhC) and threshold levels, the auxiliary circuit consists of 14 Ts. The IPhC was simulated by a current generator I2. In general, the cell layout consisted of 68 Ts. In this scheme for simulation, we used four fixed different gain values for each triangular signal. The simulation results for various signals are shown in Fig. 13. The power consumption of the cell is 150μW at a supply voltage of 2.5V, $Imax = D = 8\mu A$, $N = 4$, $p = 2\mu A$, and the periods of the input signals are 200μs and

100μs. The general scheme of the cell realizing the dynamic intensity conversion with eight piecewise linear approximation is not shown. This circuit contains from 170 to 200 Ts, consists of eight basic nodes. The Node A consists of 8 (7) Ts and generates a triangular signal from the input signal at a given threshold for each subband. The processes of formation output nonlinearly transformed signal and simulation results of this circuit will be shown and discussed in the report. For a supply voltage of 2.5V, *Imax* = *D* = 8μA, *N* = 8, *p* = 1μA and the period of the input linearly increasing-decreasing triangular signal equal to 1000μs. Removing only one transistor in the nodes A of the circuit allows it to modify and implement on its basis tunable nonlinear transformations in accordance with the help of $s_i = (x \div (i-1)p) \div (x \div i \cdot p)$, but not t_i for required function by means of triangular signals:

$$y_a = \sum_{i=1}^{N} k_i \cdot t_i = \sum_{i=1}^{N} k_i \left[(x \div (i-1) \cdot p) \div 2(x \div i \cdot p) \right].$$

The results confirm the possibility of synthesizing cells with required accuracy characteristics of the transformation laws and, in particular, auto-equivalence functions, the microvolt level of power consumed by them, high speed. For the simplest and approximate approximation functions, but often quite sufficient for the selection of the winning function by the activation function, the cell circuits consist of only 17–20 transistors, have a very high speed (*T* = 0.25us), a small power consumption (less than 100 microwatts).

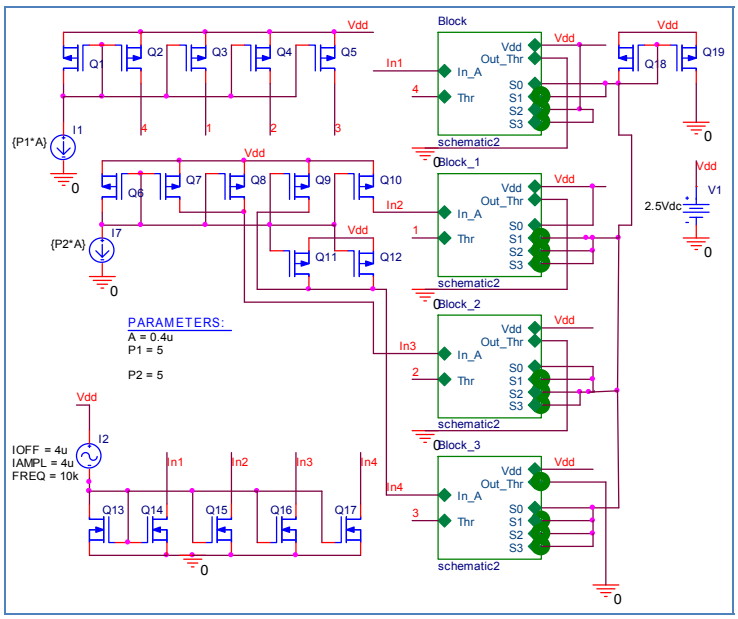
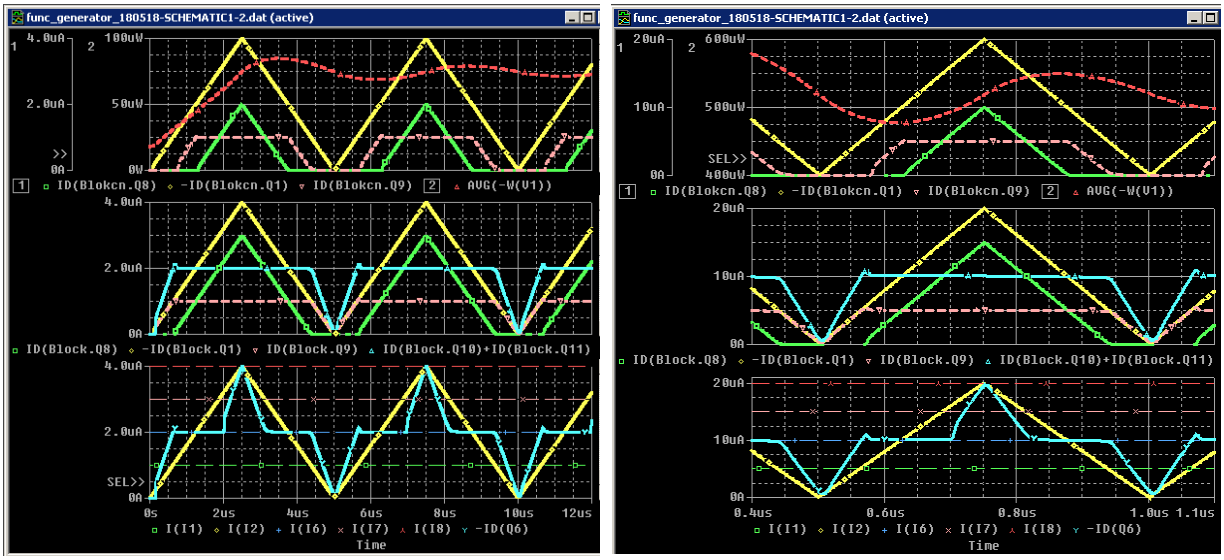


Fig. 12. Circuit for simulation of nonlinear converter cell on the base of four piece-linear approximation and four base sub-nodes



4uA, 70uW, 2.5us, Vdd=2.5V 20uA, 550uW, 0.25us, Vdd=3.3V

Fig. 13. Simulation result for four levels approximation, the realized non-linear transformation is the normalized auto-equivalence function for self-learning convolutional networks (for different input currents and transformation periods): input signal – yellow line, output signal – blue line, power consumption – red line

Conclusions

The paper proposes the mathematical foundations of design of array neuron-equivalentors using continuously logical cells (CLC) based on current mirrors (CM) with functions of preliminary analogue processing for image intensity transformation for construction of mixed image processors (IP) and neural networks (NN). Several effective schemes have been developed and modeled of CLC and optoelectronic complement dual analog neuron-equivalentors as hardware accelerators SLECNS. The proposed CLC have a modular hierarchical construction principle and are easily scaled. Their main characteristics were measured. NEs have a processing-conversion time of 0.1–1μs, low supply voltages of 1.8–3.3V, minor relative computational errors (1–5%), small consumptions of no more than 1–2mW, can operate in low-power modes less than 100μW) and high-speed (10–20MHz) modes. The efficiency of NEs relative to the energy intensity is estimated at a value of not less than 10^{12} an. op. / sec on W and can be increased by an order of magnitude. The obtained results confirm the correctness of the chosen concept and the possibility of creating NE and MIMO structures on their basis. They can become the basis for the implementation of CNN and self-learning biologically inspired devices with the number of such NEs equal to 1000, to realize the calculation of equivalent convolutions with filter sizes up to 32×32.

Список літератури

1. Неперервно-логічні еквівалентністні моделі архітектур нейронних мереж Хеммінга з адаптивно-кореляційним зважуванням / В.Г. Красиленко, Ф.М. Салецький, В.І. Яцковський, К. Конате // Праці SPIE. – 1998. – Т. 3402. – С. 398-408.

2. Красиленко В.Г. Багатопортова оптична асоціативна пам'ять на основі матрично-матричних еквівалентів / В.Г. Красиленко, А.Т. Магас // Праці SPIE. – 1997. – Том 3055. – С. 137-146.

3. Красиленко В.Г. Проектування та моделювання гетеро-асоціативної пам'яті багатопортової нейронної мережі для оптичного розпізнавання образів / В.Г. Красиленко, О.О. Лазарев, С.К. Грабовляк // Праці SPIE. – 2012. – Том 8398. – 83980N-1.

4. Красиленко В.Г. Експериментальні дослідження методів кластеризації та вибору фрагментів зображень з використанням просторових інваріантних еквівалентністних моделей / В.Г. Красиленко, О.О. Лазарев, Д.В. Нікітович // Праці SPIE. – 2014. – Том 9286. – 928650.

5. Красиленко В.Г. Структури оптичних нейронних мереж на основі нових матричних еквівалентністних моделей (МТЕМ) і результати їх моделювання / В.Г. Красиленко, О. І. Нікольський, Ю.А. Флавицька // Оптична пам'ять і нейронні мережі (інформаційна оптика). – 2010. – Т. 19 (1). – С. 31-38.

6. Лекунь Ю. Згорткові мережі для зображень, мови та часових рядів / Ю. Лекунь, Ю. Бенжіо // Довідник теорії мозку і нейронних мереж. – MIT Press, 1995.

7. Shafiee A. ISAAC: прискорювач згорткової нейронної мережі з внутрішньо-мережевою аналоговою арифметикою в перехрестях / A. Shafiee та інш. // 2016 ACM / IEEE 43-й щорічний міжнародний симпозіум з комп'ютерної архітектури (ISCA). – Сеул, 2016. – С. 14-26. https://doi.org/10.1109/ISCA.2016.12

8. Vehicle license plate recognition using visual attention model and deep learning / Di Zang, Zhenliang Chai, Junqi Zhang, Dongdong Zhang, Jiujun Cheng // Journal of Electronic Imaging. –2015. – Вип. 24(3). – С. 033001. http://dx.doi.org/10.1117/1.JEI.24.3.033001.

9. Красиленко В.Г. Моделювання та можлива реалізація самонавчальних еквівалентністно-згорткових нейронних структур для автокодування-декодування та кластеризації зображень / В.Г. Красиленко, О. О. Лазарев, Д.В. Нікітович // Праці SPIE. – 2017. – Том 10453. – 104532N.

10. Красиленко В.Г. Моделювання біологічно мотивованих самонавчальних еквівалентно-згорткових рекурентно-багатошарових нейронних структур (BLM_SL_EC_RMNS) для кластеризації та розпізнавання фрагментів зображення / В.Г. Красиленко, О. О. Лазарев, Д.В. Нікітович // Праці SPIE, MIPPR 2017: Pattern Recognition and Computer Vision. – 2018. – Том 10609. – 106091D. https://doi.org/10.1117/12.2285797.

11. Красиленко В.Г. Проектування та моделювання оптоелектронних нейрон-еквівалентів як апаратних прискорювачів самонавчальних еквівалентністних згорткових нейро-структур (СНЕЗНС) / В.Г. Красиленко, О. О. Лазарев, Д.В. Нікітович // Праці SPIE. – 2018. – Том 10689. – 106890C.

12. Красиленко В.Г. Проектування та моделювання елементів масиву для трансформації інтенсивності зображення та кодування, що використовуються в процесах обробки змішаних зображень та нейронних мереж / В.Г. Красиленко, О.О. Лазарев, Д.В. Нікітович // Праці SPIE. – 2018. – Том 10751. – 1075119.

13. Schlottmann C.R. A Highly Dense, Low Power, Programmable Analog Vector-Matrix Multiplier: The FPAA Implementation / C.R. Schlottmann, P.E. Hasler // IEEE Journal on Emerging and Selected Topics in Circuits and Systems. – 2011. – Том 1(3). – С. 403-411. https://doi.org/10.1109/JETCAS.2011.2165755.

14. 1 TERA-OPS/WATT analog deep machine learning engine with floating gate storage in 0.13-um CMOS [Electronic resource] / J. Lu, S. Young, I. Arel, J. Holleman // IEEE JSSC. – 2015. – Vol. 50. – P. 270-281. –Available at: ieeeexplore.ieee.org.

15. Lu J. Аналогова пам'ять з плаваючою заслінкою з двонаправленими сигмоподібними оновленнями в стандартному цифровому процесі / J. Lu, J. Holleman // Праці ISCAS'13. – 2013. – С. 1600-1603.

16. Lu J. Аналогова онлайн кластеризація в 130 нм CMOS / J. Lu, S. Young, I. Arel, and J. Holleman // Праці A-SSCC. – 2013. – С. 177-180.

17. Сімейство оптоелектронних фото-струмових реконфігурованих універсальних (або багатофункціональних) логічних елементів (ОПР ULE) на основі неперервно-логічних операцій (НЛО) і віддзверкалювачів струму (ВдС) / В.Г. Красиленко, К. В. Огородник, О.І. Нікольський, В.М. Дубчак // Праці SPIE. – 2011. – Том 8001. – 80012Q.

18. Красиленко В.Г. Проектування та моделювання інтелектуального багатофункціонального неперервно-логічного пристрою як базової комірки сучасних високопродуктивних сенсорних систем з MIMO-структурою / В.Г. Красиленко, О.І. Нікольський, О.О. Лазарєв // Праці SPIE. – 2015. – Том 9450. – 94500N.

References

1. Krasilenko, V.G., Saletsky, F.M., Yatskovsky, V.I. and Konate, K. (1998), "Неперервно-логічні еквівалентністні моделі архітектур нейронних мереж Хеммінга з адаптивно-кореляційним зважуванням" [Continuous logic equivalence models of Hamming neural network architectures with adaptive-correlated weighting], *Proceedings of SPIE*, Vol. 3402, pp. 398-408.

2. Krasilenko, V.G. and Magas, A.T. (1997), "Bahatoportova optychna asotsiatyvna pamiat na osnovi matrychno-matrychnykh ekvivalentoriv" [Multiport optical associative memory based on matrix-matrix equivalentors], *Proceedings of SPIE*, Vol. 3055, pp. 137-146.

3. Krasilenko, V.G., Lazarev, A. and Grabovlyak, S. (2012), "Proektuvannia ta modeliuvannia hetero-asotsiatyvnoi pamiati bahatoportovoi neironnoi merezhi dlia optychnoho rozpiznavannia obraziv" [Design and simulation of a multiport neural network heteroassociative memory for optical pattern recognitions], *Proc. of SPIE*, Vol. 8398, 83980N-1.

4. Krasilenko, V.G., Lazarev, A.A. and Nikitovich, D.V. (2014), "Експериментальні дослідження методів кластеризації та вибору фрагментів зображень з використанням просторових інваріантних еквівалентністних моделей" [Experimental research of methods for clustering and selecting image fragments using spatial invariant equivalent models], *Proceedings of SPIE*, Vol. 9286, 928650.

5. Krasilenko, V.G., Nikolskyy, A.I. and Flavitskaya, J.A. (2010), "Struktury optychnykh neironnykh merezh na osnovi novykh matrychnykh ekvivalentnistnykh modelei (MTEM) i rezultaty yikh modeliuvannia" [The Structures of Optical Neural Nets Based on New Matrix_Tensor Equivalently Models (MTEMs) and Results of Modeling], *Optical Memory and Neural Networks (Information Optics)*, Vol. 19 (1), pp. 31-38.

6. LeCun, Y. and Bengio, Y. (1995), Convolutional networks for images, speech, and time-series, *The Handbook of Brain Theory and Neural Networks*, MIT Press.

7. Shafiee, A. (2016), ISAAC: A Convolutional Neural Network Accelerator with In-Situ Analog Arithmetic in Crossbars, *ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA)*, Seoul, pp. 14-26. https://doi.org/10.1109/ISCA.2016.12.

8. Di Zang, Zhenliang Chai, Junqi Zhang, Dongdong Zhang and Jiujun Cheng (2015), Vehicle license plate recognition using visual attention model and deep learning, *Journal of Electronic Imaging* 24(3), 033001. http://dx.doi.org/10.1117/1.JEI.24.3.033001.

9. Krasilenko, V.G., Lazarev, A.A. and Nikitovich, D.V. (2017), "Modeliuvannia ta mozhlyva realizatsiia samonavchalnykh ekvivalentnistno-zghortkovykh neironnykh struktur dlia avtokoduvannia-dekoduvannia ta klasteryzatsii zobrazhen" [Modeling and possible implementation of self-learning equivalence-convolutional neural structures for auto-encoding-decoding and clusterization of images], *Proceedings of SPIE*, Vol. 10453, 104532N.

10. Krasilenko, V.G., Lazarev, A.A. and Nikitovich, D.V. (2018), "Modeliuvannia biolohichno motyvovanykh samonavchalnykh ekvivalentno-zghortkovykh rekurentno-bahatosharovykh neironnykh struktur (BLM_SL_EC_RMNS) dlia klasteryzatsii ta rozpiznavannia frahmentiv zobrazhennia" [Modeling of biologically motivated self-learning equivalent-convolutional recurrent-multilayer neural structures (BLM_SL_EC_RMNS) for image fragments clustering and recognition], *Proc. SPIE 10609, MIPPR 2017: Pattern Recognition and Computer Vision*, 106091D. https://doi.org/10.1117/12.2285797.

11. Krasilenko, V.G., Lazarev, A.A. and Nikitovich, D.V. (2018), "Proektuvannia ta modeliuvannia optoelektronnykh neiron-ekvivalentoriv yak aparatnykh pryskoriuvachiv samonavchalnykh ekvivalentnistnykh zghortkovykh neiro-struktur (SNEZNS)" [Design and simulation of optoelectronic neuron equivalentors as hardware accelerators of self-learning equivalent convolutional neural structures (SLECNS)], *Proceedings of SPIE*, Vol. 10689, 106890C.

12. Krasilenko, V.G., Lazarev, A.A. and Nikitovich, D.V. (2018), "Proektuvannia ta modeliuvannia elementiv masyvu dlia transformatsii intensyvnosti zobrazhennia ta koduvannia, shcho vykorystovuiutsia v protsesakh obrobky zmishanykh zobrazhen ta neironnykh merezh" [Design and simulation of array cells for image intensity transformation and coding used in mixed image processors and neural networks], *Proceedings of SPIE*, Vol. 10751, 1075119.

13. Schlottmann, C.R. and Hasler, P.E. (2011), A Highly Dense, Low Power, Programmable Analog Vector-Matrix Multiplier: The FPAA Implementation, *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, Vol. 1, No. 3, pp. 403-411. https://doi.org/10.1109/JETCAS.2011.2165755.

14. Lu, J., Young, S., Arel, I., and Holleman, J. (2015), A 1 TOPS/W analog deep machine-learning engine with floating-gate storage in 0.13 μm CMOS, *IEEE JSSC*, Vol. 50, available at ieeeexplore.ieee.org.

15. Lu, J. and Holleman, J. (2013), A floating-gate analog memory with bidirectional sigmoid updates in a standard digital process, *Proc. ISCAS'13*, pp. 1600-1603.

16. Lu, J., Young, S., Arel, I. and Holleman, J. (2013), An analog online clustering in 130 nm CMOS, *Proc. A-SSCC*, pp. 177-180.

17. Krasilenko, V.G., Ogorodnik, K.V., Nikolskyy, A.I. and Dubchak, V.N. (2011), "Simeistvo optoelektronnykh fotostrumovykh rekonfihurovanykh universalnykh (abo bahatofunktsionalnykh) lohichnykh elementiv (OPR ULE) na osnovi neperervno-lohichnykh operatsii (NLO) i viddzverkaliuvachiv strumu (VdS)" [Family of optoelectronic photocurrent reconfigurable universal (or multifunctional) logical elements (OPR ULE) on the basis of continuous logic operations (CLO) and current mirrors (CM)], *Proceedings of SPIE*, Vol. 8001, 80012Q.

18. Krasilenko, V.G., Nikolskyy, A.I. and Lazarev, A.A (2015), "Проектування та моделювання інтелектуального багатофункціонального неперервно-логічного пристрою як базової комірки сучасних високопродуктивних сенсорних

систем з MIMO-структурою" [Designing and simulation smart multifunctional continuous logic device as a basic cell of advanced high-performance sensor systems with MIMO-structure], *Proceedings of SPIE*, Vol. 9450, 94500N.

Received by Editorial Board 25.01.2019
Signed for Printing 19.02.2019

Відомості про авторів:

Красиленко Володимир Григорович
кандидат технічних наук доцент
старший науковий співробітник
Вінницького національного технічного університету,
Вінниця, Україна
https://orcid.org/0000-0001-6528-3150

Лазарєв Олександр Олександрович
кандидат технічних наук доцент
Вінницького національного технічного університету,
Вінниця, Україна
https://orcid.org/0000-0003-1176-5650

Шеремета Олександр Петрович
кандидат технічних наук доцент
Вінницького національного технічного університету,
Вінниця, Україна
https://orcid.org/0000-0002-5782-6554

Information about the authors:

Vladimir Krasilenko
Candidate of Technical Sciences Associate Professor
Senior Research of Vinnytsia National
Technical University,
Vinnytsia, Ukraine
https://orcid.org/0000-0001-6528-3150

Alexander Lazarev
Candidate of Technical Sciences Associate Professor
of Vinnytsia National Technical University,
Vinnytsia, Ukraine
https://orcid.org/0000-0003-1176-5650

Alexander Sheremeta
Candidate of Technical Sciences Associate Professor
of Vinnytsia National Technical University,
Vinnytsia, Ukraine
https://orcid.org/0000-0002-5782-6554

ПРОЕКТУВАННЯ МАСИВУ НЕЙРО-ЕКВІВАЛЕНТОРІВ З КВАЗІУНІВЕРСАЛЬНОЮ ФУНКЦІЄЮ АКТИВАЦІЇ ДЛЯ СТВОРЕННЯ САМОНАВЧАЛЬНИХ ЕКВІВАЛЕНТНІСТНО-ЗГОРТКОВИХ НЕЙРОННИХ СТРУКТУР

В.Г. Красиленко, О.О. Лазарєв, О.П. Шеремета

У статті розглядається необхідність створення апаратних прискорювачів для машинного навчання, згорткових нейронних мереж, асоціативної пам'яті, кластеризації, розпізнавання образів. Зроблено огляд робіт про переваги біологічно-натхненних еквівалентністних моделей (ЕМ) для проектування. Показано, що ЕМ-парадигми дуже перспективні для кластеризації, розпізнавання, великорозмірних, корельованих зображень з високим рівнем завад і для неконтрольованого навчання. Для реалізації ЕМ необхідні векторно-матричні процедури з неперервно-логічними нормалізованими операціями: "еквівалентність", "нееквівалентність" і відповідні схеми. Розглядаються нові підходи до проектування апаратних масивів таких нейрон-еквівалентів (НЕ) з розширеною функціональністю за рахунок нелінійних постперетворень та довільно вибраних видів функцій активації. Підходи засновані на використанні аналогових і змішаних (із спеціальним кодуванням) методів для реалізації необхідних операцій, побудові на основі ієрархічних вузлів з базових комірок на віддзеркалювачах струму і фотоприймачах нейрон-еквівалентів з числом синапсів від 8 до 128 і більше і мереж на основі масиву таких НЕ. Результати моделювання показують, що ефективність НЕ з універсальною перебудовованою функцією активації по відношенню до енергоємності оцінюється значенням не менше 10^{12} і вище ан. оп./сек. на W і може бути збільшена. Отримані результати підтверджують можливість та перспективність створення запропонованих НЕ та на їх основі масивів, структур з багатьма входами і багатьма виходами (так званих на англійській мові MIMO-систем).

Ключові слова: *самонавчальна еквівалентністно-згорткова нейронна структура, нейрон-еквівалентор, віддзеркалювач струму, апаратні прискорювачі, еквівалентністна модель, безперервна логіка, функція активації, нелінійна обробка, розпізнавання.*

ПРОЕКТИРОВАНИЕ МАССИВА НЕЙРО-ЭКВИВАЛЕНТОРОВ С КВАЗИУНИВЕРСАЛЬНОЙ ФУНКЦИЕЙ АКТИВАЦИИ ДЛЯ СОЗДАНИЯ САМООБУЧАЕМЫХ ЭКВИВАЛЕНТНОСТНО-СВЕРТОЧНЫХ НЕЙРОННЫХ СТРУКТУР

В.Г. Красиленко, А.А. Лазарев, А.П. Шеремета

Рассматривается необходимость создания аппаратных ускорителей для сверточных нейронных сетей. Сделан обзор работ о преимуществах эквивалентностных моделей (ЕМ), которые очень перспективны для распознавания коррелированных зашумленных изображений и обучающей машины. Основными для ЕМ являются векторно-матричные процедуры с непрерывно-логическими нормализованными операциями: "эквивалентность", "неэквивалентность" с их нелинейной постобработкой. Рассматриваются подходы к проектированию массивов многофункциональных нейрон-эквивалентов (НЭ) с любыми функции активации. Подходы основаны на использовании аналоговых методов для реализации необходимых операций, построения на основе иерархических узлов из базовых ячеек на токовых зеркалах НЭ с числом синапсов от 8 до 128 и более и сетей на основе массива таких НЭ. Моделирования показывают, что эффективность НЭ с перестраиваемой функцией активации оценивается на уровне 10^{12} и выше ан. оп./сек. на W. Результаты подтверждают возможность и перспективность создания структур НЭ и MIMO на их основе.

Ключевые слова: *самообучающаяся эквивалентностно-сверточная нейронная структура, нейрон-эквивалентор, токовое зеркало, аппаратные ускорители, эквивалентностная модель, непрерывная логика, функция активации, нелинейная обработка, распознавание.*