



✉ Correspondence author

Yu.V. Opotyak

yurii.v.opotyak@lpnu.ua

Article received 15.10.2024 p.

Article accepted 19.11.2024 p.

UDC 004.8

I. G. Tsmots, Yu. V. Opotyak, B. V. Shtohrinets, T. B. Mamchur, O. O. Oliinyk*Lviv Polytechnic National University, Lviv, Ukraine*

OPERATIONAL BASIS OF ARTIFICIAL NEURAL NETWORKS AND EVALUATION OF HARDWARE CHARACTERISTICS FOR ITS IMPLEMENTATION

The tasks performed by the intelligent components of mobile robotic systems (MRS) are analyzed and their features are determined. The operational basis for the implementation of hardware accelerators of artificial neural networks (ANN) is defined and divided into three groups of neurooperations: preprocessing, processing and calculation of transfer functions. It is shown that the operations of the first group provide the transformation of the input data to the form that will give the best results, the operations of the second group (multiplication, addition, group summation, calculation of the dot product, calculation of a two-dimensional convolution, multiplication of the matrix by a vector) are performed directly in the neural network itself in the process of training and functioning, operations of the third group provide calculation of transfer functions. It is determined that the specialized hardware of the intelligent components of the MRS should provide real-time operation and take into account the limitations in terms of dimensions and power consumption. It is proposed to carry out the development of specialized hardware of intelligent components of the MRS on the basis of an integrated approach, which covers the capabilities of the modern element base, parallel methods of data processing, algorithms and structures of hardware and takes into account the requirements of specific applications. For the development of hardware accelerators ANN, the following principles were chosen: modularity; homogeneity and regularity of the structure; localization and reduction of the number of connections between elements; pipeline and spatial parallelism; coordination of intensities in the receipt of input data, calculation and issuance of results; specialization and adaptation of hardware structures to algorithms for the implementation of neurooperations. It is proposed to use the following characteristics to evaluate specialized hardware: hardware resources, operation time and equipment utilization efficiency. Analytical expressions and a simulation model for evaluating the characteristics of specialized hardware have been developed, the results of which are used to select the most effective accelerator and elemental structure for the implementation of intelligent components of the MRS. The method of selection of the element base for the implementation of intelligent components of the MRS has been improved, which, by taking into account the results of the assessment of the characteristics of hardware accelerators, the requirements of a specific application and the existing element base for their implementation, ensures the selection of the most effective of the existing ones.

Keywords: artificial neural network, operational basis, specialized hardware, method of selection of element base, parallel algorithms, simulation model, real time, element base.

Introduction / Вступ

The current stage of development of artificial neural networks (ANNs) is characterized by the expansion of their applications, a significant part of which requires the processing of intensive data streams in real-time by means that must simultaneously take into account the limitations in terms of size, weight, power consumption, and therefore have high efficiency in the use of equipment. Such applications include mobile robotic systems (MRS), in which intelligent components are implemented on the basis of ANN, used to solve the following tasks: recovery of lost data, parameters measurement accuracy improvement for ground-based MRS in conditions of interference and incomplete information, forecasting of spatial data, prediction of MRS movement, neurofuzzy control of MRS movement, neuro-like cryptographic data protection, obstacle recognition, neurofuzzy control of a group of ground-based MRS.

The real-time mode in the intelligent components of the MRS is provided through the use of specialized hardware

accelerators that implement the most complex basic operations of algorithms. For the implementation of hardware accelerators, it is necessary to allocate the operational basis of the ANN. Such an operational basis can consist of the following groups of neurooperations: preprocessing, processing, and computation of transfer functions.

To ensure a wide range of applications, intelligent components of MRS must have a variable composition of equipment and be implemented on the basis of a problem-oriented approach, which involves the use of universal processor cores, supplemented by specialized hardware accelerators. Such hardware accelerators should provide high performance by parallelizing the process of calculating basic ANN operations, easily adapt to the requirements of specific applications, and be used to synthesize a wide range of intelligent components of MRS that operate in real-time. The creation of hardware accelerators for the implementation of basic ANN operations with high equipment utilization efficiency is carried out on the basis of an integrated approach, which includes a modern element base,

technologies of ultra-large integrated circuits (VLSI), methods, algorithms, and VLSI structures for parallel calculation of basic ANN operations.

When developing hardware ANN accelerators, an urgent task is to assess their main characteristics: hardware resources, operation time, and equipment utilization efficiency. Based on the results of the evaluation, the VLSI structure of the accelerator is compared and selected for its hardware implementation based on programmable logic integrated circuits of the FPGA (Field-Programmable Gate Array) type. When implementing an FPGA accelerator, it is advisable to take the unit of measurement of hardware resources as a logic gate that implements the operations *NOT*, *AND-NOT*, *OR-NOT*. The execution time of the basic operation in the accelerator is estimated by the sum of the delays on the gates during the passing of data from the input to the output. Based on the results of the assessment of hardware resources, the time of execution of the basic operation, and its complexity, the efficiency of the equipment utilization is determined, which gives an assessment of the resulting structure in terms of productivity. The results of the evaluation of the main characteristics of hardware accelerators and the requirements of a specific application are taken into account when choosing the element base for the implementation of intelligent components of the MRS. The selection of the most effective element base should be carried out on the basis of an integrated assessment of the effectiveness of each of its types.

Therefore, the urgent task is to determine the operational basis of the ANN, to develop and evaluate the main characteristics of the hardware accelerators of the ANN and to select the element base for the implementation of intelligent components of the MRP.

The object of the research is the processes of determining the operational basis of the ANN, the development and evaluation of the characteristics of the hardware accelerators of the ANN, and the selection of the element base for the implementation of intelligent components of the ANN.

The subject of the research is the methods of development and evaluation of the characteristics of hardware accelerators of ANN and the selection of the element base for the implementation of intelligent components of the MRS.

The aim of the work is to improve the method of selecting the element base, to develop simulation models for evaluating the characteristics of hardware accelerators of the ANN, and the selection of the element base, which will allow the creation of intelligent components for the MRS with high equipment utilization efficiency.

To achieve this goal, the following main *tasks of the study are defined*:

- determine the operational basis;
- formulate requirements and select the principles for the development of hardware accelerators for ANN;
- to develop analytical expressions for evaluating the characteristics of hardware ANN accelerators;
- to improve the method of selecting the element base for the implementation of intelligent components of the MRS;
- to develop a simulation model for evaluating the characteristics of hardware accelerators of ANN;

- to develop a simulation model for the selection of the element base for the implementation of intelligent components of the MRS.

Analysis of the latest research and publications.

Intelligent components of mobile robotic systems (MRS) are an integral part of ensuring their autonomous operation and adaptability. Recent research in this area has largely focused on the use of artificial neural networks (ANNs) to improve tasks such as navigation, pattern recognition, and real-time decision-making. Advances have been made in the application of deep learning and reinforcement learning to solve problems of autonomous navigation and decision-making by robots under uncertain conditions [1]. An important aspect is the integration of ANN with other artificial intelligence algorithms, which increases the overall efficiency of systems [2].

The operational basis of ANN is based on mathematical operations, such as multiplication and addition, which form the basis of computational processes in neural networks [3]. ANN algorithms are optimized to improve performance at the hardware level, allowing for efficient processing of large amounts of data [4]. At the same time, studies show the need to improve hardware accelerators to optimize energy efficiency and computing speed [5].

The implementation of ANN requires significant computing resources to ensure speed and low latency, which is important when working with large neural networks. For this purpose, specialized hardware accelerators are being developed to optimize the performance of basic operations of neural networks [6].

Studies of hardware ANN accelerators demonstrate that their design principles are based on the spatio-temporal mapping of neural networks and parallel data processing [7]. This makes it possible to significantly increase machining efficiency and reduce energy costs, making such solutions important for a wide range of applications [8].

The performance evaluation of hardware accelerators is based on analytical models that allow you to calculate the performance, power efficiency, and latency associated with data processing. Such models are used to compare hardware accelerators with software implementations on CPU and GPU processors [9]. Studies show that the use of hardware accelerators provides significant advantages in performance and energy efficiency compared to traditional solutions [10].

The choice of the element base for the implementation of hardware accelerators is an important step that affects their performance and cost. Research shows that FPGAs are the most flexible solution for implementing hardware accelerators, as they allow a balance between energy efficiency and performance, while ASICs provide maximum performance at higher costs [11].

Modern simulation models make it possible to evaluate the effectiveness of various configurations of hardware components for ANN [12]. Such models help to select the optimal element base for the implementation of systems, which increases the overall efficiency of mobile robotic systems adapted to work in complex environments [13].

Research also highlights the importance of developing models that take into account power consumption, processing speed, and compatibility with mobile system components [14]. This makes it possible to provide flexibility

and scalability of solutions for various tasks of autonomous robotic systems.

Research results and their discussion / Результати дослідження та їх обговорення

1. Intelligent components of the MRS and determination of the operational basis of the ANN. The current stage of MRS development is characterized by the widespread use of ANN for the implementation of intelligent components for data processing, obstacle detection, platform traffic control, and data protection. The main requirements for such components are real-time operation and high technical and operational characteristics, the main of which are restrictions on power consumption, dimensions, and weight.

The analysis of the tasks that are implemented by the intellectual components of the MRS has shown that they have the following features:

- high intensity and consistency of incoming data flows;
- constant complication of processing algorithms and increased requirements for the accuracy of results;
- the ability to parallelize data processing both in time and space;
- ability to generalize and abstract;
- learning, self-learning, and self-organization under the influence of the external environment.

The development of highly efficient intelligent components of MRS requires the widespread use of modern element base (microprocessors, microcontrollers, a SoC (System on Chip), programmable logic integrated circuits such as FPGA (Field-Programmable Gate Array), the development of new methods and algorithms for real-time data processing and structures focused on the modern element base. In practice, intelligent components of MRS can be implemented in software, hardware, or software and hardware.

The software implementation of intelligent components of the MRS involves the use of universal means (microprocessors, microcontrollers). With the software implementation of intelligent components, computing processes are mainly deployed in time with a large amount of information transfer between RAM and operating devices. When using software, the problem arises of minimizing the volume of programs and the time of their implementation with a given accuracy of calculations. These tools are characterized by high flexibility in terms of the possibility of modifying and replacing operating algorithms along with low speed.

Advances in the development of FPGAs make it possible to increasingly shift the implementation of neuroalgorithms to hardware that deploys the computing process both in time and space. The structural organization of such hardware is based on the principle of adequate hardware mapping of graphs of neuroalgorithms. Hardware is characterized by high speed along with the complexity of modifying and changing data processing algorithms.

In most cases, intelligent components of MRS are implemented on the basis of SoC, which combines universal and special approaches, software, and hardware. At the same time, the development of intelligent components of

the MRS with specified technical parameters on the SoC comes down to supplementing the universal computing core with specialized hardware.

Specialized hardware of intelligent components of the MRS is based on the operational basis of the ANN, which is shown in Fig. 1.

The operational basis of ANN consists of three groups of basic operations: the first group is neurooperations of preprocessing, the second group is neurooperations of processing, and the third group is the calculation of transfer functions.

The first group is preprocessing neurooperations. The operations of this group ensure that the input data is converted to the form that produces the best results. The learning vector contains one value for each input of the neural network and one value for each output of the network, depending on the type of training (unsupervised or supervised). As a rule, training a network on a “raw” set does not give quality results. To improve the quality of the neural network usage, input data is pre-processed, which boils down to performing the following operations: normalization, quantization, and filtration.

Normalization is a procedure for pre-processing input data (training, testing, and working samples), in which the values of the features that form the input vector are reduced to a certain specified range. After normalization, all values of the input features will be reduced to some narrow range $[0, 1]$ or $[-1, 1]$.

Normalization of input data to the range $[0, 1]$ is performed as follows:

$$x_i^{\times} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (1)$$

where x_i is the input data, x_{\max} is the maximum value of the input data, x_{\min} and is the minimum value of the input data.

Normalization of input data to the range $[-1, 1]$ is as follows:

$$x_i^{\times} = \frac{x_i}{|x|_{\max}}. \quad (2)$$

These kinds of normalization do not require complex calculations and are widely used for x_i inputs that tightly fill a certain gap.

After normalizing the input data in the RBF and GRNN networks, the Euclidean distance from each input vector to all the others must be calculated. This calculation of the Euclidean distance is performed using the operation:

$$y = \|x_i^e - x_i^b\|^2 = (x_1^e - x_1^b)^2 + (x_2^e - x_2^b)^2 + \dots + (x_N^e - x_N^b)^2. \quad (3)$$

For other types of neural networks, filtering can be used, which is performed on noisy input data and is reduced to discarding values that are invalid. In addition, quantization is performed on continuous quantities, which involves the determination of a finite set of discrete values.

The second group is processing neurooperations. This group includes operations that are performed directly in the neural network itself in the process of training and functioning. The group of processing neurooperations includes the following: multiplication, addition, group summation, calculation of a dot product, calculation of a two-dimensional convolution, and multiplication of a matrix by a vector.

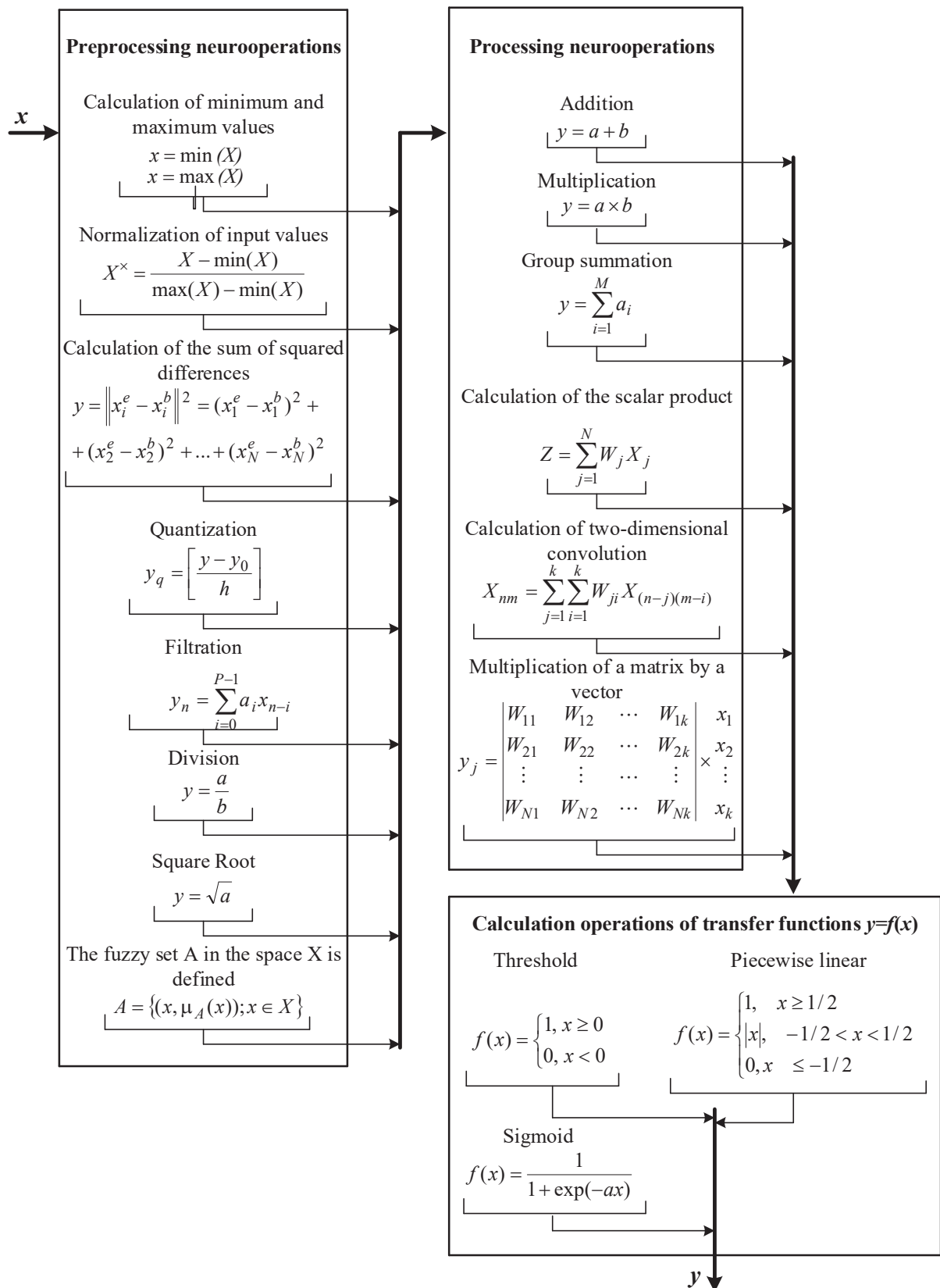


Fig. 1. The operational basis of ANN / Операційний базис ШНМ

The third group consists of the transfer function computing operations. Neuroelement transfer functions are mathematical functions that determine the response of a neuroelement to input signals. In a neuroelement, the result of calculating a dot product is

converted into an output signal through an algorithmic process known as the transfer function. The group of transfer function computing operations provides the following transfer functions: threshold, sigmoidal, and piecewise linear.

Analysis of the operational basis of the ANN shows that neural network operations can be divided into one-operand (square root, transfer functions), two-operands (addition, division, multiplication), and multi-operands (determination of minimum and maximum numbers, group summation, calculation of dot product, calculation of the sum of squares of differences, calculation of two-dimensional convolution, multiplication of the matrix by a vector). Existing hardware neuroelements and neural networks are implemented mainly on one- and two-operand operations, this is due to the capabilities of the element base. The evolution of the development of hardware for the implementation of basic operations is closely related to the structural unit of processing, that is, to the number of bits and the number of operands that the operating device simultaneously processes. With the development of integral technology, there is a tendency to change the structural unit of processing from one- and two-operand to multi-operand, which is performed in parallel.

The peculiarity of multi-operand neurooperations is that they are performed on a set of operands and the result of the operation is one number. Multi-operand neurooperations are proposed to be performed on the basis of a multi-operand approach, in which the process of calculating a neurooperation is considered as the performance of a single operation based on elementary arithmetic operations.

2. Requirements and principles for the development of hardware accelerators ANN. *The main requirements* for specialized hardware of intelligent components of MRS are minimization of dimensions, power consumption, reliability, flexibility, and real-time operation. The creation of such specialized hardware requires the widespread use of modern element base, the development of new pipeline methods and neural algorithms for processing data streams of different intensities in real-time. Real-Time Mode imposes a limit on the time t_{SP} for solving the problem, which should not exceed data input period t_{DI} , i.e.:

$$t_{SP} \leq t_{DI}. \quad (4)$$

The exchange time depends on the volume N , the number of bits n and the frequency F_{DI} of the incoming data, as well as on the number of k channels and their bit width n_k . This time is determined by the formula:

$$T_E = \frac{Nn}{F_{DI}kn_k}. \quad (5)$$

To ensure the processing of data streams in real time, the performance of specialized tools must be:

$$P = \frac{\beta R F_{DI} k n_k}{N n}, \quad (6)$$

where R is the complexity of algorithms for solving problems; β is the coefficient of taking into account the features of the means of implementing the algorithm.

It is also possible to ensure the operation of conveyor specialized hardware in real-time by matching the intensity of data intake with the intensity of their processing. The intensity of data receipt P_D depends on the number and bit width of data channels and the frequency of data intake:

$$P_D = knF_{DI}. \quad (7)$$

The intensity of data processing in conveyor specialized hardware is defined as follows:

$$D_k = \frac{m_m n_m}{T_k}, \quad (8)$$

where m_m is the number of data channels in the conveyor steps; n_m is the number of bits of data channels in the pipeline steps; T_k is a conveyor cycle of data processing.

To ensure the processing of intensive data streams in real time, the intelligent components of the MRS use conveyor-specialized hardware that implements the basic operations of the ANN. Each such basic operation can have several variants of its hardware implementation. To select a specific variant of the hardware implementation of the basic operation of the ANN, it is proposed to use the criterion of equipment utilization efficiency E , which links performance to hardware costs and evaluates hardware by performance. The quantitative value of the equipment utilization efficiency is determined as follows:

$$E = \frac{R}{t_{SP} W_{su}}, \quad (9)$$

where W_{su} is the hardware resources for the implementation of specialized unit, R is the complexity of algorithms for the operation of specialized hardware, t_{SP} is the time for solving the problem.

Principles of construction of specialized ANN hardware. The development of specialized hardware for intelligent components of the MRS is proposed to be carried out on the basis of an integrated approach, which is based on the capabilities of the modern element base, covers parallel methods of data processing, algorithms, and structures of hardware for the implementation of basic ANN operations and takes into account the requirements of specific applications. For the fullest use of the advantages of the modern element base, the development of structures for the hardware implementation of basic operations is proposed ANN algorithms should be carried out according to the following principles:

- use of the basis of elementary arithmetic operations for the implementation of basic operations of ANN algorithms;
- modularity, which involves the development of specialized hardware for the implementation of basic operations of ANN algorithms in the form of functionally complete devices;
- localization and reduction of the number of connections between elements of structures for the implementation of basic operations of ANN algorithms;
- pipeline and spatial parallelism in the development of structures for the implementation of basic operations of ANN algorithms;
- homogeneity and regularity of the hardware structure;
- consistency of the intensity of data intake with the intensity of calculations in hardware;
- specialization and adaptation of hardware to the structure of algorithms for the implementation of the basic operation and the intensity of data inflow.

3. Evaluation of the characteristics of hardware ANN accelerators. The main characteristics that are used to evaluate the hardware for the implementation of the basic

operations of the ANN algorithms are hardware resources, operation execution time, and equipment utilization efficiency.

Hardware resources are the amount of equipment that is required to build specialized hardware, expressed in certain units. The unit of measurement of hardware resources can be: the number of blocks of standard sizes; the number of printed circuit boards; the number of ultra-large integrated circuits (VLSIs) of the same type or conditional packages reduced to one type of package; the number of gates or transistors when implementing specialized hardware in the form of VLSI. For specialized hardware that implements the basic operations of ANN algorithms using FPGAs, it is advisable to take a logic gate that implements *NOT*, *AND-NOT*, *OR-NOT* operations as a unit of measurement of hardware resources. Such hardware is implemented on the basis of functional nodes (triggers, registers, adders,

switches, decoders, multiplication devices, memory elements, etc.), which are characterized by the speed and hardware resources for their implementation. To estimate the hardware resources (in the number of gates) and speed (which is determined by the sum of delays on the stages of logic gates) of individual functional nodes of MRS intelligent components, analytical expressions have been developed, which are given in Table 1, where n is the number of bits of functional nodes, m is the number of inputs, τ is the data delay time when passing through the gate.

These values were obtained by modeling functional nodes in the implementation of specialized VLSI processors for fast cosine and sine Fourier transforms [15] and devices for parallel-flow calculation of scalar products [16]. With the use of the developed analytical expressions (Table 1), the estimation of hardware resources for the implementation of specialized unit is carried out.

Table 1. Analytical expressions for estimating hardware resources and performance of functional units /
Аналітичні вирази для оцінювання витрат обладнання та швидкодії функціональних вузлів

Salary No.	Names of functional units	Hardware resources (gates)	Number of delay stages (τ gates)
1	Trigger	6	3
2	Register	$7n$	3
3	Single-digit adder	18	7
4	Single-digit subtractor	18	7
5	Single-digit adder-subtractor	20	8
6	n -bit Adder	$20n$	$7 \log_2 n$
7	n -digit Subtractor	$21n$	$8 \log_2 n$
8	n -digit Adder-Subtractor	$23n$	$8 \log_2 n$
9	m -input n -bit Adder	$(m-1)20n$	$7 \log_2 n \log_2 m$
10	m -input n -digit Conveyor Adder	$27(m-1)n$	10
11	Multiplication Device	$18n_2$	$14n$
12	Square Elevation Device	$9n_2$	$12n$
13	Division Device	$20n_2$	$16n$
14	Comparison Scheme	$7n$	$3 \log_2 n$
15	Binary Counter	$12n$	$5 \log_2 n$
16	Decoder $m \times l$	$(2m+2\log_2 l)$	m
17	m -input n -bit Switch	$3mn$	m
18	m -input n -bit ROM	$2_m n$	$(m+3)$
19	m -input n -bit RAM	$2_m 3n$	$(m+3)$

Estimation of hardware resources for the implementation of the r -th component of the ANN is carried out according to the formula:

$$W_{CANNr} = \sum_{j=1}^{k_r} W_{FNrj} q_{rj}, \quad (10)$$

where W_{CANNr} is the hardware resources for the implementation of the r -th component of the ANN; k_r is the number of types of functional nodes in the r -th component of the ANN, W_{FNrj} is the hardware resources for the j -th type of the functional node of the r -th component of the ANN, q_{rj} is the number of functional nodes of the j -th type of the r -th component of the ANN. Estimation of hardware resources for implementation of ANN is carried out according to the formula:

$$W_{ANN} = \sum_{r=1}^M \sum_{j=1}^{k_r} W_{FNrj} q_{rj}, \quad (11)$$

where M is the number of ANN components.

The speed of an asynchronous ANN is determined by the delay time when data passes through components that are on the longest path of execution of the neuroalgorithm and is estimated using the following formula:

$$t_{ANN} = \sum_{r=1}^M t_{CANNr}, \quad (13)$$

where t_{CANNr} is the delay time for data to pass through the r -th component in the ANN.

Estimating the execution time of the basic ANN operation. In the asynchronous (single-clock) mode of operation,

the speed of the r -th hardware component of the ANN is determined by the delay time when data passes through the functional nodes that lie on the longest path of execution of the r -th component algorithm and is estimated by the following formula:

$$t_{CANNr} = \sum_{s=1}^{P_r} t_{rjs}, \quad (12)$$

where t_{rjs} is the delay time when data passes through the s -th functional node j -th of the r -th component of the ANN, P_r is the number of functional nodes that lie on the longest path of execution of the algorithm of the r -th component of the ANN.

In the synchronous (conveyor) mode of operation, the speed of the r -th hardware component and the ANN is determined by the conveyor cycle of operation T_k , which is equal to the greatest delay in the data passage in the conveyor steps and is estimated by the formula:

$$T_k = \sum_{l=1}^Z \max t_{jl}, \quad (14)$$

where t_{jl} is the delay time when the data in the conveyor step passes through the l -th functional node of the j -th type, Z is the number of functional nodes through which the data passes through the pipeline step.

Evaluation of the equipment utilization efficiency. To process continuous, intensive data streams, it is advisable to use pipelined specialized hardware. The efficiency of using the equipment by conveyor specialized hardware that implements the r -th basic operation (r -th component) of the ANN is determined as follows:

$$E_{CANNr} = \frac{R}{\sum_{l=1}^Z \max_{j=1} t_{jl} \sum_{r=1}^{k_r} W_{FNrj} q_{rj}} = \frac{R}{T_k W_{CANNr}}. \quad (15)$$

The equipment utilization efficiency by conveyor specialized hardware that implements ANN as a whole is determined as follows:

$$E_{ANN} = \frac{R}{\sum_{l=1}^Z \max_{j=1} t_{jl} \sum_{r=1}^M \sum_{j=1}^{k_r} W_{FNrj} q_{rj}} = \frac{R}{T_k W_{ANN}}. \quad (16)$$

When evaluating the equipment utilization efficiency for components that are implemented in the form of VLSI, it is necessary to take into account the number of interface pins, geometric, dynamic and other parameters of active elements and the relationships between them. The equipment utilization efficiency by conveyor components, which are implemented in the form of VLSI, is determined by:

$$E_{VLSI} = \frac{R}{T_k k_1 k_2 \sum_{j=1}^{k_r} W_{FNrj} q_{rj} + k_3 Y}, \quad (17)$$

where k_1 is the coefficient of taking into account the homogeneity of the structure, k_2 is the coefficient of taking into account the regularity and locality of connections, k_3 is the coefficient of taking into account the number of pins of the communication interface $k_3 = f(Y)$.

Taking into account the coefficients k_1 , k_2 and k_3 is due to the fact that the cost of the VLSI component is largely determined by the area of the crystal. Reducing the size of active elements leads to a proportional increase in their speed and a decrease in the length of communication lines.

Improvement of the method of selection of the element base for the MRS intelligent components implementation.

A promising element base for the implementation of intelligent components for the MRS is SoC (System on Chip), which includes built-in processor cores and an FPGA field with high integration of gates on a chip (more than 10 million gates). The use of SoC for the implementation of intelligent components of the MRS will provide:

- integration of software and hardware;
- increase in productivity due to the hardware implementation of basic ANN operations;
- reducing the cost of components;
- reduction of power consumption due to the ability to turn off the power supply to the FPGA field;
- reduction of development time due to the availability of a large set of development and debugging tools.

For the selection of the element base (SoC) in the implementation of intelligent components of the MRS, the method has been improved, which is based on the application of the theory of multi-criteria analysis and takes into account the requirements of a specific application (performance of the processor core, memory capacity, power consumption of the processor core, number of FPGA gates, clock speed of the FPGA, power consumption and cost of the FPGA, weight, dimensions, temperature range, reliability, resistance to special factors, etc.).

The basis of the method for selecting the element base for the MRS intelligent components implementation is the calculation of an integrated assessment of its efficiency on the basis of partial efficiency criteria, which are formed for each specific application.

The calculation of the integrated performance assessment will be carried out according to the scheme of trade-offs. According to this scheme, the integrated assessment of the efficiency of the j -th element base is calculated in accordance with the expression:

$$E_{IPAj} = \sum_{i=1}^n \lambda_i E_{nC_i} \Rightarrow \max, \quad (18)$$

where $i = 1, \dots, n$ is the number of partial performance criteria of the hardware and software component included in the convolution; λ_i – i -th weighting coefficient; E_{nC_i} – normalized assessment of the effectiveness of the i -th partial criterion.

The method of selecting the element base for the MRS intelligent components implementation requires the following stages:

- to form a list of partial criteria on which the effectiveness of the element base depends;
- determine the scale of changes in numerical values of partial criteria for the effectiveness of the element base;
- determine the set of elements that meet the requirements of the terms of reference;

- determine the values of the weighting coefficients that determine the relative importance of the i -th partial criterion;
- calculate the values of i -th partial normalized efficiency criteria for the element base;
- calculate an integrated assessment of the effectiveness of each j -th element base;
- compare and select the element base for the implementation of intelligent components of the MRS.

At the first stage of selection of the element base for the MRS intellectual components implementation, a list of par-

tial criteria for the effectiveness of the element base is formed. The list of partial criteria for the efficiency of the element base for the MRS intellectual components implementation is given in Table 2.

At the second stage of selection of the element base for the MRS intellectual components implementation, the scale of changes in the numerical values of the partial efficiency criteria is determined. The formation of the scale of changes in these numerical values is carried out on the basis of the terms of reference for the development of intelligent components of the MRS.

Table 2. A list of partial criteria for the equipment utilization efficiency of the elementary base for the MRS intellectual components implementation / Перелік часткових критеріїв ефективності елементної бази реалізації інтелектуальних компонентів MPC

Criterion Name	Denomination
Performance of the j -th processor core	$\Pi_{\Pi j}$
Memory capacity of the j -th SoC	Q_j
Clock speed of the j -th processor core	$F_{\Pi j}$
Power consumption of the j -th SoC	$P_{\Pi j}$
Clock speed of the j -th FPGA	$F_{FPGA j}$
Number of logic elements of the j -th FPGA	$K_{\Pi j}$
Memory capacity of the j -th FPGA configuration	$Q_{FPGA j}$
Power consumption of the j -th FPGA	$P_{FPGA j}$
External interface of the j -th SoC	$R_{SoC j}$
J -th SoC Enclosure	$Y_{SoC j}$
Availability of tools for the development of the j -th SoC	$B_{SoC j}$
Maximum operating temperature of the j -th SoC	$t_{max j}$
Minimum operating temperature of the j -th SoC	$t_{min j}$
The cost of the j -th SoC	$C_{SoC j}$
The cost of tools for the development of the j -th SoC	$C_{ISoC j}$
Reliability of the j -th SoC	$H_{SoC j}$

At the third stage of the selection of the element base for the MRS intellectual components implementation, a set of SoCs that meet the requirements of the terms of reference is determined. Threshold coefficients are used to select such a set. The selection of a set of possible SoCs for the MRS intellectual components implementation is carried out according to the following formulas:

$$W_{E_{SoC}} = \sum_{j=1}^N E_{SoC j} n_j q_j p_j c_j m_j s_j f_j d_j r_j y_j b_j t_j h_j \gamma_j, \quad (19)$$

where $E_{SoC j}$ is the j -th SoC; N is the number of elements of the set; $n_j, q_j, p_j, v_j, d_j, f_j, c_j, m_j, s_j, r_j, y_j, b_j, t_j, h_j, \gamma_j$ – threshold coefficients j -th SoC respectively in terms of performance, memory capacity, power consumption, data transfer rate, bit width of the processor core, clock speed, cost, weight, dimensions, interface, die area, development tools, temperature range, reliability, resistance to special factors. Threshold coefficients can take one of two values: 1 – when a particular parameter meets the requirements of the terms of reference or 0 – when a particular parameter does not meet the requirements of the terms of reference.

At the fourth stage of selecting the element base for the MRS intellectual components implementation, we determine the values of the weighting coefficients λ_j for the partial criteria of the elements' efficiency. The value of the weighting coefficients is determined by the importance of the criterion for the functioning of the intellectual components of the MRS. When determining the weighting coefficients, it must be borne in mind that the sum of all

weighting coefficients must be equal to one $\sum_{i=1}^n \lambda_j = 1$. The

determination of weighting coefficients is carried out by means of an expert survey. In the process of developing the intellectual components of MRS, the method of attribution of points or the method of ranking is often used.

At the fifth stage of the selection of the element base for the MRS intellectual components implementation, the normalization of partial efficiency criteria is performed. For each j -th variant of the SoC, normalization of performance efficiency criteria E_{PEj} , memory capacity E_{Mj} , power consumption E_{Pj} , data transfer rate E_{DRj} , cost E_{Cj} , weight E_{Wj} , dimensions E_{Dj} and reliability E_{Rj} is performed by dividing the corresponding parameter by its value specified in the terms of reference.

At the sixth stage of selecting the element base for the MRS intellectual components implementation, an integrated assessment of the efficiency of the j -th SoC is calculated. The calculation of the integrated efficiency score of the j -th SoC is performed using the formula:

$$E_{IES_{SoC j}} = \lambda_{PE} E_{PEj} + \lambda_M E_{Mj} + \lambda_P E_{Pj} + \lambda_{DR} E_{DRj} + \lambda_C E_{Cj} + \lambda_W E_{Wj} + \lambda_D E_{Dj} + \lambda_R E_{Rj} \quad (20)$$

At the seventh stage of the selection of the element base, the SoC is determined, which will be used to implement the intelligent components of the MRS. From the set of SoCs that meet the requirements of the terms of reference, the SoC whose integrated performance assessment is the largest $E_{IES_{SoC}} \max$ is selected.



Fig. 2. The user interface of the software tools for evaluating the characteristics of the ANN components / Користувачський інтерфейс програмних засобів для оцінювання характеристик компонентів ШНМ

The improved method makes it possible to automate the selection of the optimal element base for the MRS intelligent components implementation in accordance with the requirements of the terms of reference.

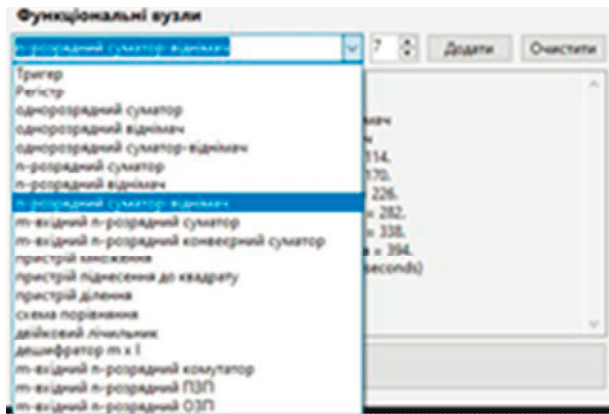


Fig. 3. Addition of functional nodes to the designed structure of the ANN component / Додавання функціональних вузлів у проєктовану структуру компоненти ШНМ

5. Development of a simulation model for evaluating the characteristics of hardware accelerators of ANN. To implement the simulation model for evaluating the characteristics of hardware accelerators of ANN, software tools have been developed to evaluate the characteristics of hardware for the implementation of basic operations of ANN algorithms. The developed software tools make it possible to calculate the hardware resources and the execution time of an individual component of the ANN, taking into account the basic characteristics of the constituent elements of the hardware implementation, which are presented in Table 1. These data are further used to evaluate the equipment utilization efficiency. The user interface of the developed software tools is presented in Fig. 2.

In the process of use, the program offers: select the bit size of the created ANN component, set the value of the number of inputs, and other basic parameters. With the help of the interface (Fig. 3), individual functional units are added to the designed structure of the ANN component.

After determining all the functional nodes of the designed structure of the ANN component, the calculation of hardware resources and data processing delays is carried out, and the corresponding graphs are built, which make it possible to visualize the obtained characteristics of the ANN component for different parameters of the input data bit depth and the number of inputs.

Microsoft Excel tools can be used to visualize the results of the assessment of hardware resources, speed, and equipment utilization efficiency used in the specialized hardware development stage for the implementation of basic operations of ANN algorithms. The performed calculations data is exported to a CSV file format, which provides ample opportunities for creating various graphs and diagrams and helps in evaluating the parameters of the created specialized hardware of the ANN components.

Let us consider the evaluation of the characteristics of hardware accelerators of ANN on the example of a device for calculating a dot product with the formation of a group of partial products [15, 16]. The parallel-flow structure of the device for calculating the dot product with the formation of group partial products is shown in Fig. 4, where TI_1 are the first clock pulses; TI_k – conveyor clock pulses; N is the number of inputs X_j and weights W_j , $j = 1, \dots, N$; k is the number of digits of the factors X_j , which are analyzed to calculate the group partial products of P_{jh} ; $h = 1, \dots, m$;

$m = \left\lceil \frac{n}{k} \right\rceil$, m is the number of conveyor steps, $\lceil \cdot \rceil$ is the rounded sign to a larger integer; CS_h – h -th stage of the conveyor; Rg – register; Ad – adder; GPPFB – block for the formation of group partial products; PPF is a partial product former, which is implemented on the basis of n logical elements I ; kAd is the k – input adder; Z is the output of the dot product.

The calculation of the dot product in this device involves dividing the factors of X_j into groups of k digits ($k \geq 3$) [16]. As a result of this partition, we get m groups. For each h -th group of digits of the multiplier X_j , the group partial product P_{jh} is calculated using the following formula:

$$P_{jh} = \sum_{s=1}^k 2^{-(s-1)} W_j X_{jhs} . \quad (21)$$

After the group partial products P_{jh} are formed, the h -th macropartial product is calculated using the following formula:

$$P_{Mh} = \sum_{j=1}^N P_{jh} . \quad (22)$$

The calculation of the dot product Z with the formation of group partial products is performed as follows:

$$Z_h = 2^{-k} Z_{h-1} + P_{Mh} . \quad (23)$$

The device works in such a way that with each clock pulse TI_1 , which enters the clock inputs of the registers of the data format converter, the input data X_j and the weighting factors W_j are written to the registers RgW₁ and RgX₁. After n cycles in the registers of the format converter, we get N input data X_j and N weighting factors W_j , which are rewritten by the front edge of the clock pulse TI_k into the registers RgW₁, ..., RgW_N, RgX₁, ..., RgX_N of the first step of CS_1 .

In each h -th cycle of operation, data from the outputs ($h-1$) of the conveyor stage CS_{h-1} are recorded into registers RgW₁, ..., RgW_N, RgX₁, ..., RgX_N and RgZ₁ of the h -th conveyor stage CS_h . In the block h -th conveyor stage CS_h for the h -th group of digits $X_{jh1}, X_{jh2}, \dots, X_{jkh}$ of the multiplier at the outputs of PPF₁, ..., PPF_k is formed k partial products according to the formula $P_{jhs} = W_j X_{jhs}$.

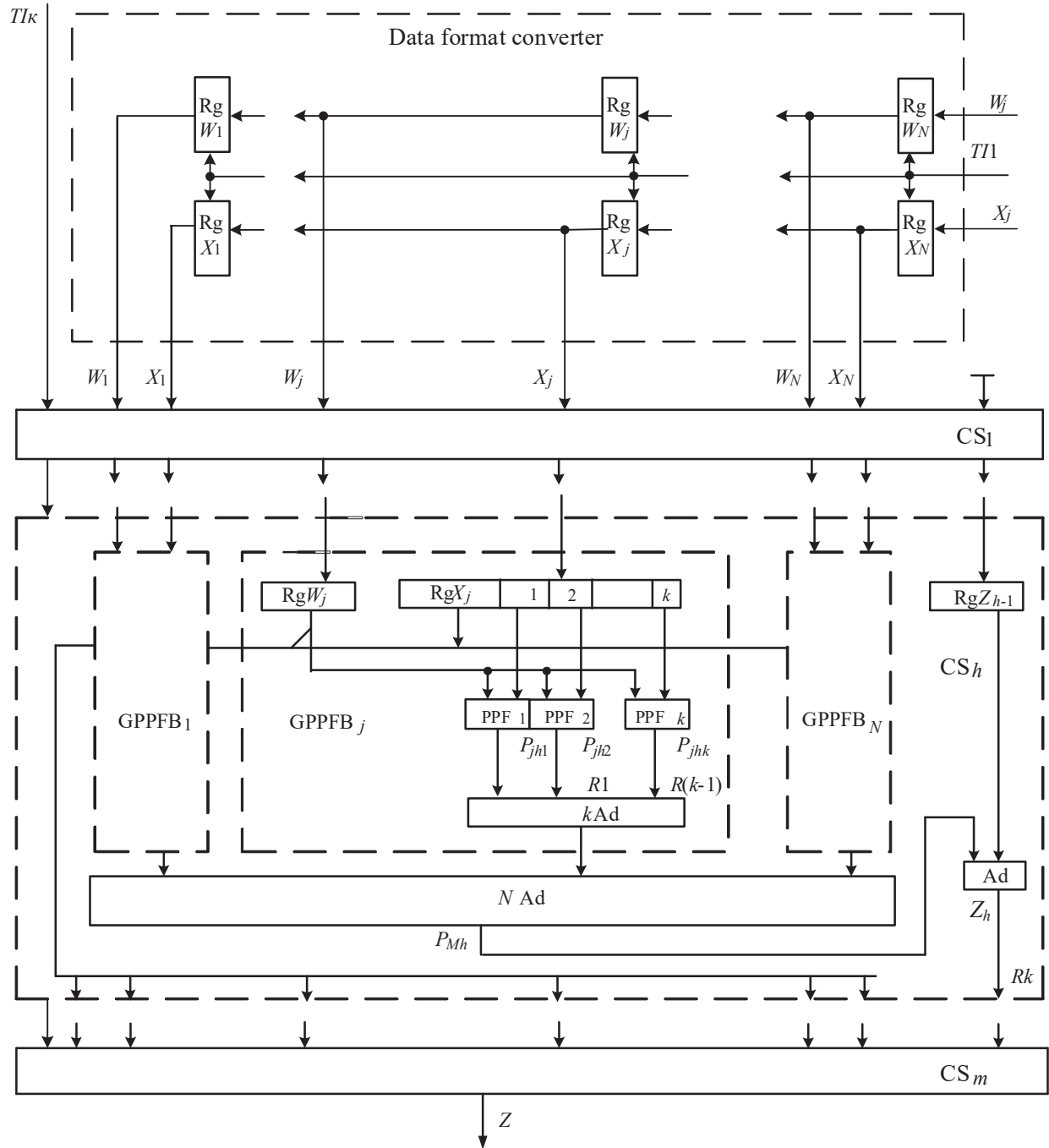


Fig. 4. The parallel-flow structure of the device for calculating the scalar product with the formation of group partial products / Паралельно-потоківна структура пристрою обчислення скалярного добутку з формуванням групових часткових добутків

The formed partial products go to the input of the k -input adder, and the s -th ($s = 1, \dots, k$) partial product $W_j X_{jhs}$ is shifted relative to $(s-1)$ -th of the partial product by one digit to the right. By adding the partial products at the output $W_j X_{jhs(s-1)}$ of the k -input adder, we obtain the group partial product P_{jh} according to the formula (21).

The formed group partial product P_{jh} goes to the j -th input of the N -input adder NAd , at the output of which, in accordance with the formula (22), we obtain the h -th macro-partial product P_{Mh} . The computed h -th macropartial product P_{Mh} enters the input of the adder Ad , where it is added to the $(h-1)$ -th partial result Z_{h-1} according to formula (23).

The result of the calculation of the first dot product is obtained at the output of the device after the m -th clock pulse TI_k . In each subsequent cycle of the TI_k , the output of the device will receive the results of the calculation of the following dot products.

This device works with a conveyor cycle, which is calculated using the formula:

$$T_k = t_{Rg} + t_l + t_{kAd} + t_{NAd} + t_{Ad} = [6 + 1 + 7 \log_2 n \times \log_2 k + 7 \log_2 n \times \log_2 N + 7 \log_2 n]^\tau, \quad (24)$$

where t_{Rg} is the time of writing/reading from the register; t_l is the data delay time on logical element I; t_{kAd} and t_{NAd} are the time of addition on the k -input and N -

input adders, respectively; t_{Ad} is the time when two numbers were added.

The hardware resources for the implementation of the dot product calculation device are determined by the expression:

$$\begin{aligned} W_{CD} &= W_{PPF} + m[NW_{GPPFB} + W_{NAd} + W_{Ad} + W_{Rg}] \\ &= 2NW_{Rg} + m[N(2W_{Rg} + kW_{LogI} + W_{kAd}) + W_{NAd} + W_{Ad} + W_{Rg}] =, \quad (25) \\ &= 14Nn + m[N(14n + kn + (k-1)20n) + (N-1)20n + 20n + 7n] \end{aligned}$$

where W_{PPF} is the hardware resources for the dot product calculation device; W_{GPPFB} is the hardware resources for the block of formation of group partial products; partial product former; W_{Rg} – hardware resources for the implementation of the register; W_{LogI} – hardware resources for the implementation of logical element I ; W_{kAd} and W_{NAd} are the hardware resources for implementation, respectively, for k – input and N – input combiners; W_{Ad} is the hardware resources for the implementation of the adder.

The equipment utilization efficiency by a conveyor device for calculating the dot product is determined by:

$$E_{CD} = \frac{R}{T_k W_{CD}}, \quad (26)$$

where $R = N(n+1)$, R is the complexity of the algorithm for calculating the dot product in the number of addition operations.

An example of the evaluation of the ANN component developed above, and visualization of characteristics is shown in Fig. 5, 6.

A family of graphs for estimating the hardware resources of the developed parallel-flow structure of the device for calculating the dot product with the formation of group partial products is presented on Fig. 5. The dependencies of the calculated hardware resources on the number of inputs for the values of 8, 16, 24, and 32 bits and the bit width of the data in the range from 4 to 32 bits are given.

Graphs for estimating data processing delays for the developed parallel-stream structure device (Fig. 6) depending

on the number of inputs and data bit depth are given similarly to Fig. 5.

6. Development of a simulation model for the selection of the element base for the MRS intelligent components implementation.

On the basis of the improved method of selecting the element base for the MRS intelligent components implementation, a simulation model is developed. The algorithm of the simulation model for selecting the element base consists of the following steps:

Step 1: Initialization of the simulation model.

Step 2: Initialization of the connection with the database on the structure of the hardware accelerator and the MRS intelligent component, the composition of the available element base, and the requirements of the terms of reference.

Step 3: Reading the technical characteristics of the element base.

Step 4: Reading the search criteria and limitation data.

Step 5: Reading information about the structure of the hardware accelerator and the MRS intelligent component.

Step 6: Filtering the element base according to the min/max values of the element criteria.

Step 7: Normalization of weights for each of the criteria.

Step 8: Normalization of the partial criteria of each of the filtered elements.

Step 9: Calculation of the integrated performance score for each of the elements.

Step 10: Sorting items in descending order of integrated efficiency value.

Step 11: Synthesize the alternative connections of a subset of elements into a module. Validate component interfaces and select alternatives that satisfy interface compatibility.

Step 12: Compute an integrated performance score for each of the synthesized modules.

Step 13: Sort items in descending order of integrated performance value.

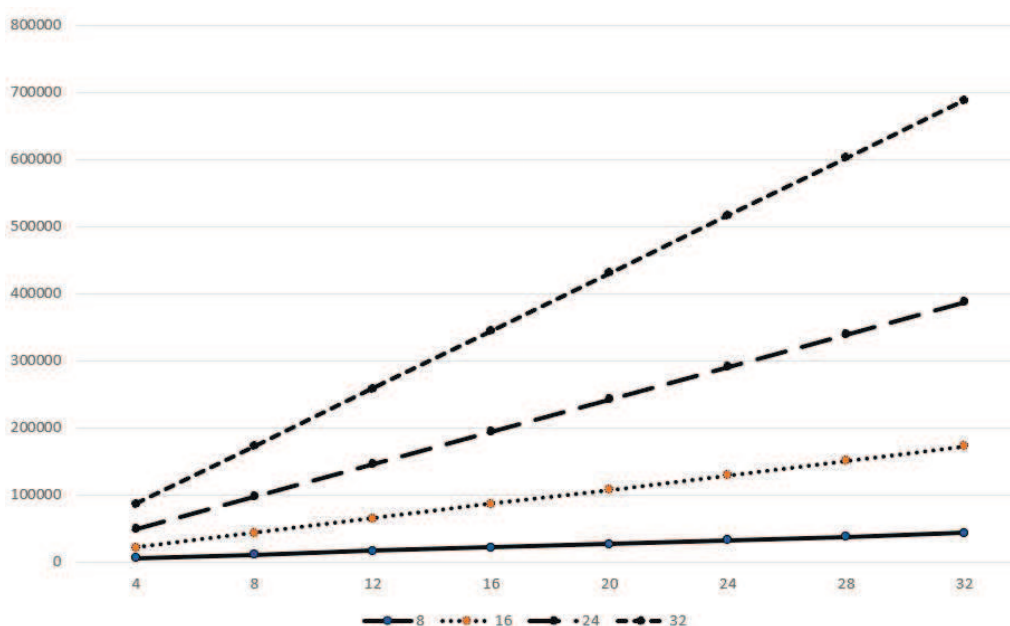


Fig. 5. Evaluation of hardware resources depending on the number of inputs and bit rate of the developed ANN component / Оцінка витрат обладнання залежно від кількості входів та розрядності розробленої компоненти ШНМ

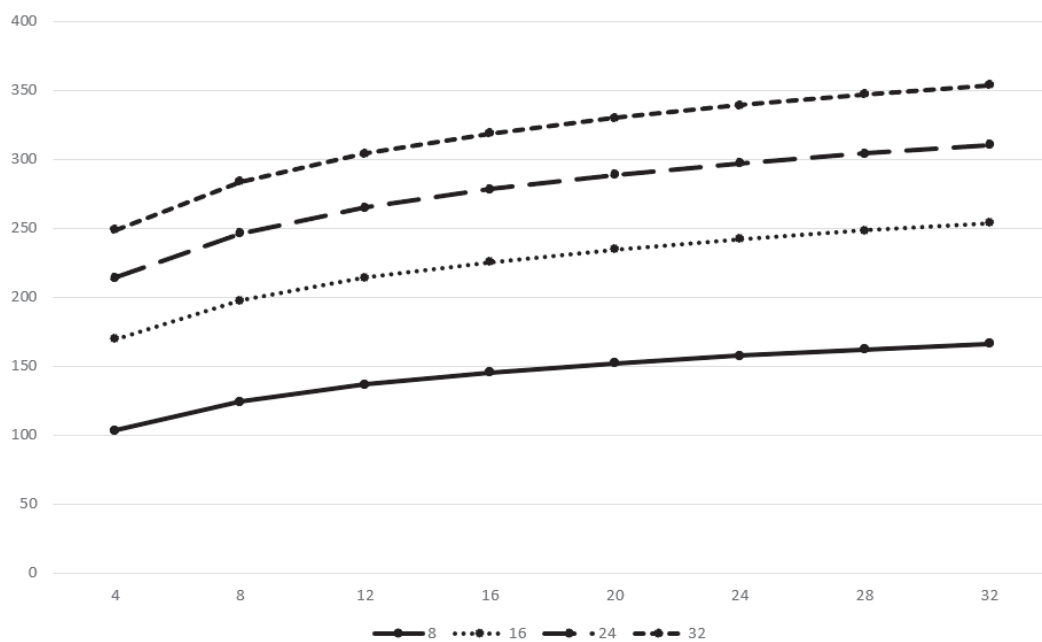


Fig. 6. Evaluation of data processing delays depending on the number of inputs and bit rate for the developed ANN component / Оцінка затримок оброблення даних залежно від кількості входів та розрядності для розробленої компоненти ШІМ

Step 14: Output the results to the user.

The developed simulation model provides the selection of the element base for the MRS intelligent components implementation. It takes into account the structure of specialized components, the results of the performance assessment, and the requirements of specific applications in accordance with the terms of reference.

Discussion of the results obtained. Analysis of the tasks of intelligent components of mobile robotic systems makes it possible to formulate an operational basis for the implementation of hardware accelerators of artificial neural networks. In general, it is worth distinguishing three groups of neurooperations: preprocessing, processing, and computing of transfer functions. The operations of the first group provide the transformation of input data, the operations of the second group, for example, multiplication, addition, group summation, matrix multiplication by vector, etc., are performed in neural networks, and the operations of the third group provide the calculation of transfer functions.

The specialized hardware of the intelligent components of the MRS must provide real-time operation while taking into account dimensional and power consumption constraints. Therefore, it is expedient to implement such means on a certain element base. Since there can be several variants of circuit implementation, and they are determined by the capabilities of a specific element base, the question arises of evaluating the characteristics of specialized hardware. The results of such evaluation are used to select the most effective accelerator structure and element base for the implementation of intelligent components of the MRS.

It is proposed to use the characteristics of hardware resources, operation time and equipment utilization efficiency to evaluate specialized hardware in comparison with the one given in [17], for which appropriate analytical expressions and a simulation model have been developed. The advantage of the improved method of selecting the element base is taking into account the results of the assessment of the characteristics of hardware accelerators, the requirements of a specific application, and the element base in the

implementation of intelligent components of the MRS. A feature of the improved method is that it is adapted to the requirements of a particular application by using a hardware SoC with a hardware implementation of basic ANN operations on FPGAs.

The scientific novelty of the obtained results of the study is that the method of selecting the element base for the MRS intelligent components implementation has been improved, which, by taking into account the results of the assessment of the characteristics of hardware accelerators, the requirements of a specific application and the existing element base for their implementation, ensures the selection of the most effective of the existing ones.

Practical significance of the research results – the use of the results of the evaluation of the characteristics of hardware accelerators provides the choice of the most effective structure for its implementation on FPGAs. The use of the developed simulation model for the selection of the element base for the MRS intelligent components implementation ensures the selection of the most effective element base for their implementation.

Conclusions / Висновки

The operational basis of the ANN has been determined, the requirements have been formulated, the principles of development have been selected, analytical expressions have been proposed, and a simulation model for evaluating the characteristics of hardware accelerators has been developed, the method for the selection of the element base for the MRS intelligent components implementation has been improved, and a corresponding simulation model has been developed.

Based on the results of the research, the following main conclusions can be drawn.

The operational basis for the implementation of hardware accelerators of the ANN has been determined, which consists of the following groups of neurooperations: preprocessing, processing and calculation of transfer functions.

It is proposed to carry out the development of intelligent components for the MRS on the basis of an integrated approach, which is based on the capabilities of the modern element base, covers parallel methods of data processing, algorithms, and structures of hardware for the implementation of basic operations of the ANN and takes into account the requirements of a specific application.

The following principles have been defined for the development of hardware accelerators ANN: modularity; homogeneity and regularity of the structure; localization and reduction of the number of connections between elements; pipeline and spatial parallelism; coordination of intensities of the receipt of input data, calculation and issuance of results; specialization and adaptation of hardware structures to algorithms for the neurooperations implementation.

Analytical expressions and a simulation model for evaluating the characteristics of hardware accelerators have been developed, the results of which are used to select the most effective accelerator structure and element base for the implementation of intelligent components for the MRS.

The method of selection of the element base for the MRS intelligent components implementation has been improved. By taking into account the results of the assessment of the characteristics of hardware accelerators, the requirements of a specific application, and the existing types of the element base for their implementation, it ensures the selection of the most effective of the existing ones.

References

1. Lee, D., Park, M., Kim, H., & Jeon, M. (2021). AI-based mobile robot navigation using deep neural networks and reinforcement learning. *IEEE Access*, 9, 329–345. <https://doi.org/10.1109/ACCESS.2021.3102345>
2. Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics: A review. *Cognitive Robotics*, 3, 54–70. <https://doi.org/10.1016/j.cogr.2023.04.001>
3. Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12), 2295–2329. <https://doi.org/10.1109/JPROC.2017.2761740>
4. Chen, J., Lin, K., Yang, L., & Ye, W. (2024). An energy-efficient edge processor for radar-based continuous fall detection utilizing mixed-radix FFT and updated blockwise computation. *IEEE Internet of Things Journal*, 11(19), 32117–32128. <https://doi.org/10.1109/IJOT.2024.3422251>
5. Gu, J., & Joseph, R. (2024). Perspective Chapter: Dynamic timing enhanced computing for microprocessor and deep learning accelerators. In *Deep Learning – Recent Findings and Research*. <https://doi.org/10.5772/intechopen.113296>
6. Sabareeshwari, V., & S. K. C. (2025). Artificial Intelligence in Communications. In Z. Hammouch & O. Jamil (Eds.), *Convergence of Antenna Technologies, Electronics, and AI* (pp. 209–238). IGI Global. <https://doi.org/10.4018/979-8-3693-3775-2.ch008>
7. Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2020). A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 31(7), 2227–2249. <https://doi.org/10.1109/TNNLS.2020.2996649>
8. Hussain, M. (2024). Sustainable machine vision for Industry 4.0: A comprehensive review of convolutional neural networks and hardware accelerators in computer vision. *AI*, 5(3), 1324–1356. <https://doi.org/10.3390/ai5030064>
9. Wang, C., & Luo, Z. (2022). A review of the optimal design of neural networks based on FPGA. *Applied Sciences*, 12(21), 10771. <https://doi.org/10.3390/app122110771>
10. Gilbert, M., Wu, Y. N., Emer, J. S., & Sze, V. (2024). LoopTree: Exploring the fused-layer dataflow accelerator design space. *IEEE Transactions on Circuits and Systems for Artificial Intelligence*, 1(1), 97–111. <https://doi.org/10.1109/TCASAI.2024.3461716>
11. Taherdoost, H. (2023). Deep learning and neural networks: Decision-making implications. *Symmetry*, 15(9), 1723. <https://doi.org/10.3390/sym15091723>
12. Xu, Y., Luo, J., & Sun, W. (2024). Flare: An FPGA-based full precision low power CNN accelerator with reconfigurable structure. *Sensors*, 24, 2239. <https://doi.org/10.3390/s24072239>
13. Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2022). A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12), 6999–7019. <https://doi.org/10.1109/TNNLS.2021.3084827>
14. Goel, S., Kedia, R., Sen, R., & Balakrishnan, M. (2024). EXPRESS: A framework for execution time prediction of concurrent CNNs on Xilinx DPU accelerator. *ACM Transactions on Embedded Computing Systems*, 24(1), 11. <https://doi.org/10.1145/3697835>
15. Tsmots, I., Rabyk, V., Kryvinska, N., Yatsymirskyy, M., & Teslyuk, V. (2022). Design of processors for fast cosine and sine Fourier transforms. *Circuits, Systems, and Signal Processing*, 41(9), 4928–4951. <https://doi.org/10.1007/s00034-022-02012-8>
16. Tsmots, I., Teslyuk, V., Kryvinska, N., Skorokhoda, O., & Kazymyra, I. (2023). Development of a generalized model for parallel-streaming neural element and structures for scalar product calculation devices. *Journal of Supercomputing*, 79(5), 4820–4846. <https://doi.org/10.1007/s11227-022-04838-0>
17. Цмоць, І. Г., Скорохода, О. В., & Теслюк, В. М. (2013). Пристрій для обчислення скалярного добутку. Патент України на винахід №101922, 13.05.2013, Бюл. №9.

І. Г. Цмоць, Ю. В. Опотяк, Б. В. Штогрінець, Т. Б. Мамчур, О. О. Олійник

Національний університет “Львівська політехніка”, м. Львів, Україна

ОПЕРАЦІЙНИЙ БАЗИС ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ ТА ОЦІНЮВАННЯ ХАРАКТЕРИСТИК АПАРАТНИХ ЗАСОБІВ ДЛЯ ЙОГО РЕАЛІЗАЦІЇ

Проаналізовано завдання, які виконують інтелектуальні компоненти мобільних робототехнічних систем (МРС), і визначено їхні особливості. Визначено операційний базис для реалізації апаратних прискорювачів штучних нейронних мереж (ШНМ) і розділено його на три групи нейрооперацій: попереднього оброблення, процесорних та обчислення передатних функцій. Показано, що операції першої групи забезпечують перетворення вхідних даних до вигляду, який дасть найкращі результати, операції другої групи (множення, додавання, групового підсумовування, обчислення скалярного добутку, обчислення двовимірної згортки, множення матриці на вектор) виконуються безпосередньо у самій нейромережі у процесі навчання та функціонування, операції третьої групи забезпечують обчислення передатних функцій. Визначено, що спеціалізовані апаратні засоби інтелектуальних компо-

ментів МРС повинні забезпечувати роботу в реальному часі та враховувати обмеження стосовно габаритів і енергоспоживання. Запропоновано розроблення спеціалізованих апаратних засобів інтелектуальних компонентів МРС здійснювати на основі інтегрованого підходу, який охоплює можливості сучасної елементної бази, паралельні методи оброблення даних, алгоритми та структури апаратних засобів і враховує вимоги конкретних застосувань. Для розроблення апаратних прискорювачів ШНМ вибрано принципи: модульності; однорідності та регулярності структури; локалізації та зменшення кількості зв'язків між елементами; конвеєризації та просторового паралелізму; узгодження інтенсивностей під час надходження вхідних даних, обчислення та видавання результатів; спеціалізації та адаптації апаратних структур до алгоритмів реалізації нейрооперацій. Запропоновано для оцінювання спеціалізованих апаратних засобів використовувати такі характеристики: витрати обладнання, час виконання операції та ефективність використання обладнання. Розроблено аналітичні вирази та імітаційну модель оцінювання характеристик спеціалізованих апаратних засобів, результати оцінювання яких використовують для вибору найефективнішої структури прискорювача й елементної для реалізації інтелектуальних компонентів МРС. Вдосконалено метод вибору елементної бази для реалізації інтелектуальних компонентів МРС, який завдяки врахуванню результатів оцінювання характеристик апаратних прискорювачів, вимог конкретного застосування та наявної елементної бази для їх реалізації забезпечує вибір найефективнішої з наявних.

Ключові слова: штучна нейронна мережа, операційний базис, спеціалізовані апаратні засоби, метод вибору елементної бази, паралельні алгоритми, імітаційна модель, реальний час, елементна база.

Інформація про авторів:

Цмоць Іван Григорович, д-р техн. наук, професор, кафедра автоматизованих систем управління.

Email: ivan.tsmots@gmail.com; <https://orcid.org/0000-0002-4033-8618>

Опотяк Юрій Володимирович, канд. техн. наук, доцент, кафедра автоматизованих систем управління.

Email: yurii.v.opotyak@lpnu.ua; <https://orcid.org/0000-0001-9889-4177>

Штогрінець Богдан Володимирович, аспірант, кафедра автоматизованих систем управління.

Email: bohdan.v.shtohrinets@lpnu.ua; <https://orcid.org/0009-0001-4956-3862>

Мамчур Тарас Борисович, аспірант, кафедра автоматизованих систем управління.

Email: taras.b.mamchur@lpnu.ua; <https://orcid.org/0009-0006-0593-7937>

Олійник Олександр Олександрович, аспірант, кафедра автоматизованих систем управління.

Email: oleksandr.o.oliinyk@lpnu.ua; <https://orcid.org/0009-0000-5093-7334>

Цитування за ДСТУ: Цмоць І. Г., Опотяк Ю. В., Штогрінець Б. В., Мамчур Т. Б., Олійник О. О. Операційний базис штучних нейронних мереж та оцінювання характеристик апаратних засобів для його реалізації. *Український журнал інформаційних технологій*. 2024, т. 6, № 2, С. 125–138.

Citation APA: Tsmots, I. G., Opotyak, Yu. V., Shtohrinets, B. V., Mamchur, T. B., & Oliinyk, O. O. (2024). Operational basis of artificial neural networks and evaluation of hardware characteristics for its implementation. *Ukrainian Journal of Information Technology*, 6(2), 125–138. <https://doi.org/10.23939/ujit20.02.125>