



A Mathematical Method for Determining the Parameters of Functional Dependencies Using Multiscale Probability Distribution Functions

Ilya E. Tarasov

Article



Citation: Tarasov, I.E. A Mathematical Method for Determining the Parameters of Functional Dependencies Using Multiscale Probability Distribution Functions. *Mathematics* **2021**, *9*, 1085. https://doi.org/10.3390/math9101085

Academic Editor: Liliya Demidova

Received: 27 March 2021 Accepted: 7 May 2021 Published: 12 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Institute of Informational Technologies, RTU MIREA, Vernadsky pr. 78, 119454 Moscow, Russia; tarasov_i@mirea.ru

Abstract: This article discusses the application of the method of approximation of experimental data by functional dependencies, which uses a probabilistic assessment of the deviation of the assumed dependence from experimental data. The application of this method involves the introduction of an independent parameter "scale of the error probability distribution function" and allows one to synthesize the deviation functions, forming spaces with a nonlinear metric, based on the existing assumptions about the sources of errors and noise. The existing method of regression analysis can be obtained from the considered method as a special case. The article examines examples of analysis of experimental data and shows the high resistance of the method to the appearance of single outliers in the sample under study. Since the introduction of an independent parameter increases the number of computations, for the practical application of the method in measuring and information systems, the architecture of a specialized computing device of the "system on a chip" class and practical approaches to its implementation based on programmable logic integrated circuits are considered.

Keywords: statistics; multiscale analysis; data analysis; system on chip

1. Introduction

Currently, data to be analyzed are subject to the action of many factors, which is due to both the increasing complexity of objects and systems that are the sources of such data, and the increasing requirements for the quality of analysis of complex multi-parameter systems. An additional factor is the effect of interference, including outliers, which complicates the analysis in an automated mode, forcing researchers to perform additional operations to identify data that incorrectly describe the process under study and to eliminate them from the analyzed samples.

On the other hand, advances in computing technology and increased computing performance make it attractive to use digital computing systems to implement advanced data analysis techniques that could benefit from such increased performance. Therefore, the search for new methods of data analysis can improve the quality of information, measuring and analytical systems, if modern high-performance computing systems can be used for their implementation.

The presented method is an original study inspired by the positive results and experience of using multiscale analysis for processing experimental results in applied physics. Practical results include the application of the method in a number of precision measuring devices, for example, a series of dielectric loss tangent meters (Tangent-M3, RU.C.34.010A reg. No. 52972, MEP-6IS, reg. No. 44621-10). The noted disadvantage of the method, as will be shown below, is the large number of computations for its implementation; therefore, during the research process, it had limited application. At present, the development of computer technology, especially FPGA, makes it possible to revise the practical aspects of the presented method and re-iterate it taking into account the new possibilities of computer technology. An approach based on the synthesis of hypotheses with subsequent verification of their quality can be significantly more effective compared to analytical approaches based on the use of predetermined statistical parameters. The Anscombe quartet [1] is a wellknown example illustrating the ambiguous nature of regression analysis. It represents four sets of data (pairs of points) with the same statistical characteristics, but visually significantly different.

The literature notes that all sets of points have the same function, determined based on a regression analysis:

$$y(x) = 0.5x + 3,$$
 (1)

Thus, the application of the least squares method does not allow one to distinguish datasets either by the characteristics of the regression dependence, by the correlation coefficient or by the main statistical characteristics of the datasets, which also coincide. In general, the role of the Anscombe quartet is to demonstrate the imperfection of regression analysis when applied to data that are the result of various processes that cannot be taken into account in the formulation of the regression analysis problem. The modern reference to the Anscombe quartet is usually made in the context of the importance of data visualization [2,3]. In this article, the Anscombe quartet will be used as a widespread example to illustrate the analysis results.

2. Theoretical Description of the Method of Approximating the Parameters of Functional Dependencies Using Multiscale Probability Distribution Functions

The analysis method considered below, first described in [4], consists of using the Bayes' theorem for data analysis, in which a free parameter "scale of the probability density function" is introduced, and the posterior probabilities are calculated for different values of this parameter. This approach is intended to partially compensate for the well-known drawback of Bayes' theorem in the form of insufficient substantiation of the choice of the a priori error distribution law.

The essence of the approach considered here is to accept a probabilistic hypothesis of the form "the quantity has a real value equal to X_i ." Then, the probability that the observed (measured) value will have a value x is determined by the probability density function of the error distribution, where the argument is the deviation value $X_i - x$. In accordance with Bayes' theorem, the posterior probability that a series of measurements $x_1, x_2, ..., x_n$ appeared in the process of measuring a quantity with a real value Xi is determined by the formula:

$$f_A(x,\sigma) = \frac{1}{N} \sum_{i=1}^n P(x_i | x_{real}, \sigma),$$
(2)

There is a known maximum likelihood method, which is based on a similar formula, but estimates the product of probability density functions, not the sum:

$$L(x,\sigma) = \prod_{i=1}^{n} P(x - x_i, \sigma),$$
(3)

There is a significant difference between the method of maximum likelihood and the considered method of statistical analysis. There is an obvious difference in the mathematical formulation, which consists of the summation of the probabilities in the considered method and in their multiplication in the maximum likelihood method. From the point of view of probability theory, the multiplication of the probability of independent events corresponds to the probability of their simultaneous occurrence. In relation to the measurement process, this means that all measured values objectively reflect the state of the measured object. The presence of disturbances, partial inconsistency between the object and the measuring system, interference in the measuring channel, etc. calls into question the correctness of such an assumption. At the same time, the addition of probabilities corresponds to a situation where some (but not necessarily all) measured values can be used to assess the state of an object. In this case, there is an implicit filtering of misses (although in reality they all participate in determining the probability), since the analysis technique provides for the determination of such a value of the measured quantity x, at which the appearance

of all quantities in the analyzed sample can be explained taking into account any possible interference in the measuring system.

At the same time, the appearance of a single slip for the maximum likelihood method will lead to the appearance of a sample element with a large deviation from the considered variant of the approximating dependence, for which the probability density function will be close to zero.

The provisions of the considered method can be applied to modify the regression analysis. It is known that regression analysis uses the method of least squares, which minimizes the sum of squares of deviations of experimental points from the corresponding values obtained from the analytical dependence:

$$P(\beta) = \sum_{i=1}^{N} (f(x_i, \beta) - y_i)^2,$$
(4)

where β is the vector of dependence parameters.

Let us introduce the function $\rho(x_1, x_2) = (x_1 - x_2)^2$ for the least squares method. In this case, Equation (4) can be represented as:

$$P(\beta) = \sum_{i=1}^{N} \rho(f(x_i, \beta), y_i),$$
(5)

Since the function $\rho(x_1, x_2)$ is a function of the distance between the points x_1, x_2 in some space $R(x, \rho)$, the least squares method can be considered as a method that minimizes the sum of distances in the space R from the approximating dependence to the experimental points. Moreover, for the LSM the function R is quadratic.

An idea of using a distance function other than quadratic is known. For example, the least absolute deviation method is based on the linear distance. Alternatives to the least squares method are known, such as quantile regression, least trimmed squares and others, and this shows that the disadvantages of the least squares method are significant and practical applications require alternative methods of analysis.

In Equation (5), the distance function can be replaced. It is easy to imagine that with an arbitrary choice of distance functions, the results of the analysis will be radically different, but no particular case will be preferable. The choice and justification of the distance function in this work is proposed to be made on the basis of the known or assumed distribution law of the measurement error. Taking into account the mentioned analysis method based on Bayes' theorem, it is proposed to carry out statistical analysis for a series of functions with different values of σ . For this, a relationship should be established between the distance function and the probability function. One can introduce the operator $g : L(x) \rightarrow \rho(x)$, which maps the probability of a dependence with these parameters to the distance from the approximating dependence to the next experimental point in the space R.

If the distance function ρ (Δx) \in [0; ∞), and the probability density function P (Δx) [0; P_{max}] and, obviously, the zero distance in the space R correspond to the maximum probability, then g(x) can be taken as:

$$g: P_{max} - P(x) \tag{6}$$

When using such a mapping, the minimum (zero) distance corresponds to the maximum probability density. This is the easiest way to convert the maximum probability to the minimum distance and should be reviewed and improved upon in the future. Consequently, the condition for the maximum of the probability function defined by Equation (2) can also be written as an adequate condition for minimizing the total distance in some space R between the assumed dependence and the points of the original sample.

From the point of view of the assessment method by the probability criterion, Equation (5) can be written as:

$$P(\beta) = \sum_{i=1}^{N} p(f(x_i, \beta) - y_i, \sigma)$$
(7)

When using this criterion, it is necessary to determine the set of parameters β , upon substitution of which in Equation (5) the function $P(\beta)$ reaches its maximum value.

From the above, it follows that for a set of experimental readings, a space is constructed with a function of the distance between points determined by the characteristics of the measuring channel (including the digital part). For this space, the fulfillment of the triangle axiom is not obvious (moreover, it can be shown that this axiom does not hold in the case of a Gaussian distribution of the error); thus, in the general case, one cannot speak of its metric. For such spaces, where the triangle axiom does not hold (however, the rest of the axioms characteristic of metric spaces hold), the terms "symmetric" or "pseudo metrics" are used. Due to this replacement, it is incorrect to designate the resulting function as "regression dependence," since this implies a search for a minimum of deviations using the least squares method. In this case, it is more correct to use the more general term "approximating dependence."

If we consider the widely used Gaussian function and construct a distance function based on it, it will look compared to the quadratic distance function, as shown in Figure 1.



Figure 1. Comparative representation of the quadratic distance function and the distance function constructed under the assumption that the error has a Gaussian probability distribution.

Thus, in the considered approach, the procedure for analyzing the parameters of the regression dependence is replaced by the procedure for synthesizing a possible vector of parameters and estimating its probability in accordance with a modified Bayesian estimate, in which the independent parameter "scale of the probability density function" is

introduced. Based on this function, a space is formed in which the sum of the distances from the hypothetical dependence to the available points of the analyzed sample is minimized.

An obvious objection to the proposed approach is the sharply increasing volume of computational operations. Indeed, in addition to the synthesis of hypotheses about the possible values of the parameters of the approximating dependence, it is also required to change the parameter "scale of the probability distribution function," which forms the corresponding distance function. Thus, for n parameters of the dependence to be analyzed, there is n + 1 parameter, which in this work is proposed to be obtained not analytically, but by methods of computational mathematics.

The development of computer technology and the emergence of new architectures, including a great deal of attention to parallel computing, makes it attractive to use numerical methods of analysis if the use of high-performance computing systems allows new possibilities to be obtained. For example, in [5], the author considered the practical issues of building a high-performance computing system with parallel computing nodes based on an FPGA. The current state of the architecture of high-performance computing devices demonstrates the widespread use of parallel operating nodes and systems [6–9].

As an example, consider the process of analyzing the parameters of a linear relationship. Among the possible options for representing a linear function, it seems advisable to choose the option with the direction vector angle and the distance to the origin of coordinates (θ , p). This form of representation, in particular, allows one to find vertical and horizontal lines, and also provides a more uniform distribution of hypotheses when the angle of the direction vector changes linearly.

The straight line equation is represented by the formula:

$$R = X\cos\theta + Y\sin\theta - p \tag{8}$$

In this representation of straight line, R is a distance between point with (X; Y) coordinates and a line described by (θ , p). This is very useful for the described method, because Equation (8) may be easily converted to a distance in any other space, including quadratic, Gaussian etc.

If we introduce the criterion of the quality of an approximation in the form of a function that takes the maximum value when the distance between the pixel and the generated line representation is zero, then the search for a set of parameters for a straight line can be performed by maximizing Equation (7):

$$S(\theta, p) = \sum_{X, Y} f(X \cos\theta + Y \sin\theta - p)$$
(9)

In a more general form, Equation (9) can be represented as:

$$S(\vec{a}) = \sum_{X,Y} f(X,Y,\vec{a})$$
(10)

where \vec{a} is the vector of the line parameters.

As follows from Equations (8)–(10), an important role is played by the choice of the function f, which is used to determine the quality criterion of the approximation. It can be pointed out that similar approaches were previously used in problems of approximating functions from experimental data. For example, the Hough transformation [10] assumes, instead of analyzing image pixels, the synthesis of hypotheses about the presence of a line with certain parameters and counting the number of image pixels belonging to each of these lines. Similarly, "voting for hypotheses" is implemented, which is mathematically similar to performing a probabilistic estimate.

For the Hough transform, the function S can be represented as:

$$f = \begin{cases} 0 \text{ with } |X \cos\theta + Y \sin\theta - p| > 1\\ 1 \text{ with } |X \cos\theta + Y \sin\theta - p| \le 1 \end{cases}$$
(11)

From Equation (11) it can be seen that Equation (10) can be reduced to the Hough transformation by passing to the limit if the approximation quality function is presented as a delta function (according to the principle of the presence or absence of a pixel). The possibility of using nearby pixels for analysis appears when using functions that make sense of the probability of deviation of pixels in an image from an idealized line representation with specified parameters. This property seems to be significant in the analysis of high-resolution images, where graphic objects, visually perceived as straight lines, have deviations from the idealized straight line when they are presented in a discrete form using a video matrix subject to interference.

It can be noted that a number of publications devoted to the Hough transformation provide for the transformation of the original image with blurring on the basis of the corresponding convolution kernels [11,12]. The proposed approach is mathematically close to this effect, though it is not the original image that is blurred, but rather the analyzing function. In this case, the function used for such blurring is based on the known form of the error distribution law, which can be qualitatively obtained from the analysis of the subject area and its characteristic processes. Additionally, the introduction of an independent parameter "distribution function scale" allows a number of analysis procedures to be performed with a comparison of the results.

3. Illustration of the Characteristics of the Method Using the Example of the Anscombe Quartet

The data presented in the Anscombe quartet were analyzed using the method described in this paper. The purpose of this analysis was to identify the ability of the method to distinguish between datasets based on the quantitative characteristics obtained. We used the representation of a straight line in Equation (8), with several options for the values of the scale σ . A probabilistic representation of the result was used, since it involves finding the maximum probability (and not the minimum distance in a nonlinear distance space), which provides a better visual representation of the results. The analysis results are shown in Figure 2, where the left side of the figure visualizes the value of the probability function depending on the parameters (θ , p), and the right side shows the corresponding datasets of the Anscombe quartet with superimposed straight lines corresponding to the maximum of the probability function.

The plotting algorithm is quite simple. For each pair of values, the probability density function is calculated according to Equation (7) and is represented as a color in the hot palette. The horizontal axis represents the θ value, and the vertical axis represents the p value.



Figure 2. Cont.



Figure 2. Results of the analysis of the data presented in the Anscombe quartet ($\sigma = 1$).

The data in Figure 2 clearly show the cardinal difference of the proposed method, which provided the identification of the assumed straight lines, which can be determined by taking into account the presence of specially introduced outliers in cases 3, 4. For case 2, the initial data are a parabola; therefore, the search for the parameters of the straight line is obviously inappropriate.

It can be noted that as the parameter "distribution scale" increases, the results of the analysis tend to the results obtained for the regression dependence. This is illustrated in Figure 3, where the analysis was carried out with $\sigma = 10$.



Figure 3. Results of the analysis of the data presented in the Anscombe quartet, with an increase in the scale of the probability distribution function ($\sigma = 10$).

Thus, the method considered in the article can be reduced to the well-known method of regression analysis by the passage to the limit ($\sigma \rightarrow \infty$). At the same time, the analysis with other values of σ , including in the range determined based on the analysis of the experimental distribution of the error in the initial data, is able to give results that adequately reflect the dependences in the analyzed data, being stable to the single outliers in the sample.

4. Examples of Using the Method for Analyzing Samples of Various Types

The positive effect of sample analysis using the developed method is manifested primarily for situations where the data generally correspond to the assumed analytical law describing the relationship between them. In this section of the article, the results of the analysis of a number of synthetic and experimentally obtained samples are considered.

Figures 4 and 5 show the results of the analysis of a sample containing two sets of points belonging to different straight lines with the same slope, but with a different constant coefficient.



Figure 4. Analysis results of the data sample, $\sigma = 0.1$.



Figure 5. Analysis results of the data sample, $\sigma = 5$.

In Figure 4 it can be seen that the assumed straight lines corresponding to the analytically specified functions y = x and y = x + 2 are reflected in the graph of the function P as two bright points corresponding to the local maxima of the probability function for these hypotheses. At the same time, an increase in the scale of σ to a value of 5, exceeding the distance between the straight lines, led to the obtaining of the "averaged line." This allows us to talk about the possibility of managing the results of the analysis by choosing such a scale of the probability density function that most adequately reflects the processes occurring in the system under study and measuring devices.

Figure 6 shows the results of an experimental study of the transient process in a test RC circuit connected to a source of rectangular voltage pulses. The sample is the voltage readings across the capacitor measured with a digital oscilloscope. In Figure 6, artifacts can be seen in the form of digital noise caused by interference in the measurement path. The presence of horizontal sections and fragments where the investigated function decreases will not allow investigating the characteristics of the RC chain using only local analysis, since the time constant calculated for such fragments will tend to infinity or go beyond the domain of the logarithmic function.



Figure 6. Results of the analysis of a sample of data describing the transient process in the RC chain.

At the same time, an expression can be chosen in the form of an approximating function:

$$V(t) = V_0(1 - exp(-t/\tau)),$$
(12)

where V_0 is the amplitude voltage value and τ is the time constant of the RC chain.

In Figure 6, it can be seen that the discovered function passes through the main experimental points, making it possible to quantitatively determine the parameters of the dependences approximating the experimental data.

Interestingly, in Figure 6 there are artifacts originating from the measuring equipment. The measurement mode with low voltage resolution was deliberately chosen so that the experimental data contained horizontal segments that cannot be adequately described by an exponential function. At the same time, the considered method makes it possible to find the analytical parameters of the exponential function that adequately describes the experimental data and is robust to artifacts.

5. Implementation of the Method in Measuring and Information Systems

As mentioned above, the considered method is demanding on the performance of the computing device for its implementation. The synthesis of n parameters of the approximating dependence, performed in the range of their possible values, supplemented by a change in the "scale" parameter, in the direct calculation of the probabilities of hypotheses has a computational complexity of O (n + 1). This significantly distinguishes the considered analysis method from averaging and even digital filtering. Therefore, for the practical application of the method, it is necessary to consider the issues of its implementation in measuring and information systems.

At present, for high-performance computing, along with general-purpose processors (CPUs), graphic processors (GPUs) [13–15] and programmable logic integrated circuits with FPGA architecture [16–20] are also widely used.

Taking into account the fact that the calculations of individual probabilities of individual hypotheses about the parameters of the approximating function are independent processes, parallel computing architectures can be effectively used for this method. Therefore, GPUs and FPGAs appear to be promising hardware platforms, since they provide a large number of computational nodes, the capabilities of which are sufficient to calculate the probability of hypotheses about the parameters of the approximating dependence.

At the same time, the use of a GPU for this task is fraught with certain difficulties. For example, when using a GPU with a program based on the CUDA SDK, the measured data preparation time in the CPU was 0.2–0.3 s, and the calculation time was 27 ms (for 100 CUDA cores and 1000 iterations in each core) with a tabular presentation error probability density function. Computing the Gaussian function in the GPU increases the computation time to 61 ms. Thus, up to 90% of the analysis time is spent on data transfer operations between the CPU and GPU. Therefore, one should also consider computational architectures that process data as it is received.

A specialized computing device of the "system on a chip" (SNC) class, combining data reception and processing, will eliminate the operations of sending data and probability functions, replacing them with storing (or generating) probability functions in the SNC and processing the incoming values in real time (or in combination with their storage in the buffer memory). Prototyping of such a device is now widely performed using FPGAs. The feasibility of implementing a specialized VLSI is determined by technical and economic factors; however, the high cost of preparing VLSI production determines the widespread use of FPGAs in systems manufactured in small quantities.

When considering the architecture and implementation of a specialized computing device, it is necessary to be guided by the following considerations. Since GPUs are effective in the tasks of building three-dimensional graphics and are structurally similar to them, most likely, it will not be possible to exceed their technical and economic characteristics when implementing similar architectures or repeating existing ones. Therefore, for a dedicated FPGA-based device, the following must be done:

- 1. Reduce the functionality of the computational units compared to the CUDA cores used in a GPU.
- 2. Provide storage or generation of probability density functions in a computing device, thereby eliminating the need to transfer them from a control device (for example, a PC).

A variant of the structural diagram of a specialized computing device for determining the parameters of the approximating dependence is shown in Figure 7.

Additional advantages of the proposed approach from the point of view of specialized computing systems can be indicated. For example, in Figure 7, the analyzed data and the error probability distribution function can be generalized for a number of computing units. However, it is more important that there are no requirements for constant transmission of input data and results, since for the operation of such a computing system, it is sufficient to load the analyzed sample and the table representing the error probability distribution function. This significantly reduces the load on the peripheral devices of such a computer.

Prototyping of the computing device was carried out on the basis of an FPGA with APSOC Xilinx Zynq-7020 architecture. The obtained characteristics of the module for calculating the probability are shown in Table 1. The module has an internal table for storing the values of the probability density function and an external interface for entering individual values of the sample under study in integer format. The clock frequency of the module was 150 MHz for the XC7Z20 FPGA.

From the data in the table, it can be seen that when describing a highly specialized module in the hardware description language (HDL), a compact implementation of a computational node can be obtained. Thus, for the considered FPGA ZC7020, related to the initial-middle level, it is possible to estimate the number of parallel working nodes in 80–90 (which is limited by the number of memory blocks). For other FPGAs, the number of channels will be correspondingly higher.

Table 2 shows estimates of the comparative performance characteristics of some FPGAs, assuming that the number of placed modules is determined by the number of

memory blocks, and the practically achievable clock frequency is 150 MHz for the Zynq-7000 and 200 MHz for the more productive UltraScale+ and Versal FPGAs.



Figure 7. Block diagram of a specialized computing device for determining the parameters of the approximating dependence.

Table 1. Using the resources of FPGA XC7Z7020 when implementing one module for calculating the probability of a hypothesis.

Resource	Utilization	Available	Utilization %
LUT	214	53,200	0.40
FF	158	106,400	0.15
BRAM	1.50	140	1.07
DSP	2	220	0.91

Table 2. Comparative characteristics of FPGA performance when calculating the probability of hypotheses.

FPGA	Number of Channels	Performance, 10 ⁹ Samples/Sec
Zynq-7000 XC7Z20	90	13.5
Kintex UltraScale+ XCKU3P	240	48
Kintex UltraScale+ XCKU19P	1000	200
Zynq MPSOC ZU2CG	100	15
Zynq MPSOC ZU19EG	600	120
Versal VM1102	100	20
Versal VM2902	1300	260

As can be seen from Table 2, FPGA performance ranges from approximately 10 to 250 billion samples per second. Sample processing means calculating the probability that a given sample belongs to one of the sets of approximating dependencies and adding this probability to the accumulator. In this case, the digital interfaces of the FPGA provide direct transfer of the analyzed data to the computing device, and the probability density functions are specified in the considered implementation by tables.

6. Conclusions

The method presented in the article is proposed for processing data exposed to unpredictable impulse noise. The noted disadvantage of the method is the large number of computations required for its implementation, which can be compensated for by using a high-performance element base, such as FPGA. Several issues outlined in the article require additional research and development. For example, the construction of the distance function is of great theoretical interest, and this publication presents it only briefly. Similar to other methods of multiscale analysis, the study and classification of distance functions and their properties opens up great opportunities for research.

An important practical issue is the improvement of algorithms for finding the extrema of the distance or probability function. The examples shown use brute force to construct an entire surface that illustrates the behavior of the probability function. The amount of computation can be significantly reduced using known or new algorithms for finding the extremum of a function. In this area, it is possible to create software implementations of the method, frameworks and libraries, as well as hardware accelerators for high-performance data analysis.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

- 1. Anscombe, F.J. Graphs in Statistical Analysis. Am. Stat. 1973, 27, 17–21.
- 2. Li, Q. Overview of data visualization. In *Embodying Data*; Springer: Singapore, 2020. [CrossRef]
- 3. Embarak, O. The importance of data visualization in business intelligence. In *Data Analysis and Visualization Using Python;* Apress: Berkeley, CA, USA, 2018. [CrossRef]
- 4. Tarasov, I.E. Estimation of measurements data with use of probability distribution functions with variable scale Zavodskaya Laboratoriya. *Diagn. Mater.* **2004**, *70*, 55–61.
- Tarasov, I.E.; Potekhin, D.S. Real-time kernel function synthesis for software defined radio and phase-frequency measuring digital systems. *Russ. Technol. J.* 2018, *6*, 41–54. (In Russian) [CrossRef]
- Jouppi, N.P.; Young, C.; Patil, N.; Patterson, D.; Agrawal, G.; Bajwa, R.; Bates, S.; Bhatia, S.; Boden, N.; Borchers, A.; et al. In-Datacenter Performance Analysis of a Tensor Processing Unit. In Proceedings of the 44th International Symposium on Computer Architecture (ISCA), Toronto, ON, Canada, 26 June 2017.
- 7. Hennessy, J.L.; Patterson, D.A. *Computer Architecture*, 6th ed.; A Quantitative Approach (The Morgan Kaufmann Series in Computer Architecture and Design); Morgan Kaufmann: Burlington, MA, USA, 2017; 936p.
- Olofsson A 2016 Epiphany-V: A 1024 Processor 64-Bit RISC System-on-Chip. Available online: https://www.parallella.org/wp-content/uploads/2016/10/e5_1024core_soc.pdf (accessed on 2 February 2021).
- 9. Japan's Own Processor "PEZY-SC2" and Equipped with Supercomputers "Gyoko" Details. Available online: https://pc.watch. impress.co.jp/docs/news/1091458.html (accessed on 2 February 2021).
- 10. Hough, P.V.C. Method and Means for Recognizing Complex Patterns. U.S. Patent 3,069654, 18 December 1962.
- 11. Hassanein, A.S.; Mohammad, S.; Sameer, M.; Ragab, M.E. A Survey on Hough Transform, Theory, Techniques and Applications. *arXiv* **2015**, arXiv:1502.02160.
- 12. Barbosa, W.O.; Vieira, A. On the Improvement of Multiple Circles Detection from Images using Hough Transform. *TEMA* **2019**, 20, 331–342. [CrossRef]
- 13. Perepelkin, E.E.; Sadovnikov, B.I.; Inozemtseva, N.T. *Computing on Graphics Processing Units (GPU) in Problems of Mathematical and Theoretical Physics*, 3rd ed.; Lenand: Moscow, Russia, 2019; 240p, ISBN 978-5-9710-6490-9.

- 14. Boreskov, A.V.; Kharlamov, A.A. *Basics of Working with CUDA Technology*; DMK Press: Moscow, Russia, 2019; 232p, ISBN 978-5-9760-715-2.
- 15. Sanders, J.; Kendroth, E. CUDA Technology in Examples: An Introduction to GPU Programming: Translate from English; Slinkina, A.A., Boreskov, A.V., Eds.; DMK Press: Moscow, Russia, 2018; 232p, ISBN 978-5-97060-581-3.
- 16. Tarasov, I.E. FPGA Xilinx. In VHDL and Verilog Hardware Description Languages, CAD, Design Techniques; Hotline-Telecom: Moscow, Russia, 2019; p. 538. ISBN 978-5-9912-0802-4.
- 17. Available online: www.xilinx.com (accessed on 2 February 2021).
- 18. Available online: https://www.xilinx.com/products/silicon-devices/fpga/kintex-ultrascale.html (accessed on 2 February 2021).
- 19. Available online: https://www.xilinx.com/products/silicon-devices/soc.html (accessed on 2 February 2021).
- 20. Available online: https://www.xilinx.com/products/silicon-devices/acap/versal.html (accessed on 2 February 2021).