

# Analog-to-digital conversion that allows to increase the dynamic range a standard ADC using the properties of electrical signals

Stanislav Gritsutenko<sup>1,\*</sup>

<sup>1</sup>Peter the Great Saint Petersburg Polytechnic University, Polytechnicheskaya 29, St. Petersburg, 195251, Russian Federation

**Abstract.** A method is proposed to significantly increase the dynamic range of current and voltage meters in the traction networks of railway transport. The method is based on the low entropy of signals of this type and is based on the calculation of samples whose amplitude was distorted due to the output beyond the ADC dynamic range. The approach with the solution of system of linear equations is considered. An approach without solving a system of linear equations is presented. The accuracy of the method is discussed at the end of the article.

## 1 Introduction

Consider the following practical problem. It's well known, that an electric locomotive, depending on the type and mode of operation, can consume up to several kiloamperes at a relatively stable level of supply voltage. Basically, such power consumption takes place during a movement of the electric locomotive. In this case, it is reasonable to put the electric energy meter, which begins to work (self-propelled cutoff mode), for example, with 50 amps. This is due to both economic and purely technical reasons. However, the practice of operation of electric locomotives shows that a sufficiently high percentage of consumed electricity is spent at a time when the electric locomotive stands and, it would seem, should not consume it in principle. The reason for this lies in the service personnel conducting routine maintenance and, accordingly, consuming electricity from the electric locomotive, mainly for heating. But in this case, consumption current is several amperes, which is obviously less than the cut-off of the counter. Therefore, the meter does not record this consumption of electricity, but since the consumption occurs for a long time, its unaccounted cost can be very significant [1-4]. Actually, the task is to organize a more accurate accounting, and currently it is not solved.

The second frequently encountered in practice problem is fixing a random surge. As it's known in modern power networks voltage is quite stable [5-9]. Naturally, it may be slightly reduced due to the load and other factors. But this reduction is usually only a percentage of the face value. Therefore, it would be logically to use a measurement scheme, where the

---

\* Corresponding author: [st256@mail.ru](mailto:st256@mail.ru)

upper limit of the dynamic range of the ADC would correspond to the nominal voltage, which is designed for this power network.

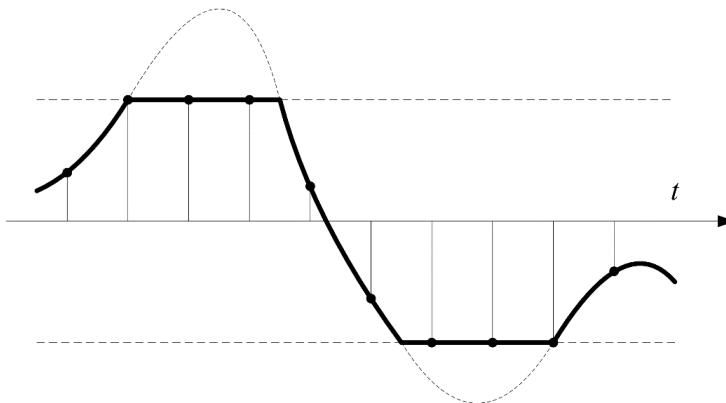
However, quite rarely, but nevertheless regularly, such network can be exposed to external influences (for example, a lightning strike and like this). It leads to voltage surges that are not recorded by the ADC because the output quantized signal is beyond its dynamic range. The signal is simply "clipped" in amplitude. Obviously, the task of solving the problem of fixing random voltage surges in power networks that supply very expensive equipment should also be considered relevant.

Unfortunately, both of these methods are quite difficult to implement. The first solution is too cumbersome. It leads either to the need to switch the buses from which the meter receives signals for measurement, or to the need to protect counter inputs from multiple overloads when the measurement data is received from one bus. The second solution is too expensive, as it involves the use of a counter with an excessively wide dynamic range. Let's show it.

Indeed, at present, most electricity meters are based on the principles of digital signal processing, that is, the measured current (or voltage) is supplied after the standard buffering on the ADC. But the characteristics of the ADC are strongly dependent on its bit length or (as is commonly stated in the technical documentation) the dynamic range. Let's calculate the dynamic range of the ADC, designed to measure the current from 1 A to 1 kA with an error of 1% throughout the measurement range.

Since 1 kA exceeds 1% of 1 A 100000 times, the ADC should be about 18 bits, which is difficult to implement at the present time. Therefore, the question arises: is it possible to significantly increase the dynamic range of the meter for standard 16-bit ADC using properties specific to the current and voltage in power networks? Below is a method that solves this problem.

Consider an analog-to-digital conversion adapted to signals with high dynamic range. When working with standard analog-to-digital converters often a following problem expects engineer. The quantized signal can go beyond the dynamic range of the ADC [10, 11]. In this case, the signal count is "cut off" by amplitude, as shown in figure 1. That is, there is a loss of information, which in turn can lead to deterioration of the characteristics of the designed device. To avoid this, it is necessary to expand the dynamic range of the ADC.

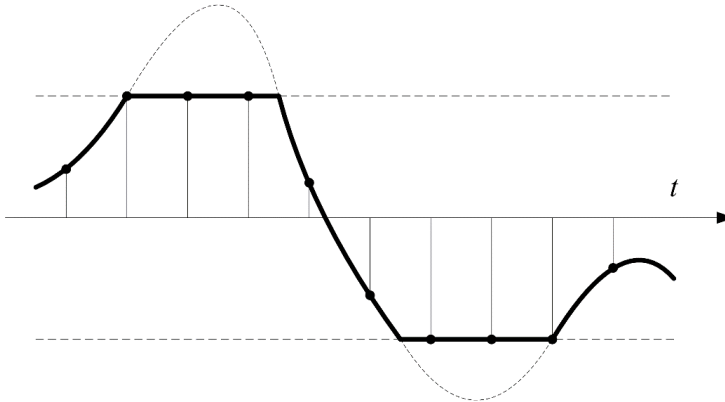


**Fig. 1.** Limiting the signal beyond the dynamic range of the ADC.

In this example, we consider a new method of analog-to-digital conversion, which allows us to expand the dynamic range of the standard converter, taking into account rounding errors

that occur during calculations. Let's first describe how the algorithm works if rounding is not applied.

The basic idea [12-14] is to use non-uniform sampling. Signal samples are taken only at those points where the signal does not go beyond the discharge grid of the ADC. But these samples are taken quite often (Fig. 2) so that the signal can be recovered at points where it is outside the specified dynamic range by means of an interpolation procedure [15, 16].



**Fig. 2.** Limiting the signal beyond the dynamic range of the ADC.

The interpolation is done this way. As you know, any analog signal  $f(t)$  can be represented by a series expansion in some basic functions  $\phi_k(t)$ :

$$f(t) = \sum_{k=1}^{\infty} C_k \phi_k(t) . \tag{1}$$

But since it is impossible to work with infinite sums, this series should be truncated to  $K$  terms:

$$f(t) \approx \sum_{k=1}^K C_k \phi_k(t) . \tag{2}$$

Performing the described procedure, the interpolation is executed on the existing samples of the function  $f(t)$  at the points  $t_n$  (where they don't exceed the ADC range). Let's make a column of values  $f(t_n)$ . Then at the same points the values of the basis functions  $\phi_k(t)$  are calculated and the matrix  $\phi_k(t_n)$  is obtained. Now, the expansion coefficients  $C_k$  can be found when solving the following system of equations:

$$\begin{bmatrix} \phi_1(t_1) & \phi_2(t_1) & \dots & \phi_k(t_1) & \dots & \phi_K(t_1) \\ \phi_1(t_2) & \phi_2(t_2) & \dots & \phi_k(t_2) & \dots & \phi_K(t_2) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \phi_1(t_n) & \phi_2(t_n) & \dots & \phi_k(t_n) & \dots & \phi_K(t_n) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \phi_1(t_N) & \phi_2(t_N) & \dots & \phi_k(t_N) & \dots & \phi_K(t_N) \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \dots \\ C_k \\ \dots \\ C_K \end{bmatrix} = \begin{bmatrix} f(t_1) \\ f(t_2) \\ \dots \\ f(t_n) \\ \dots \\ f(t_N) \end{bmatrix} , \tag{3}$$

Finally, the value at the point of interest is calculated using the formula (2). Since interpolation gives the best approximation usually in the middle of the function segment

where the interpolation nodes  $t_n$  are located. If we take the middle of the section as 0, it is possible to write that

$$f(t_i) \approx \sum_{n=1}^K C_n \phi_n(t_i) ,$$

where  $t_i$  are nearest to the 0 points.

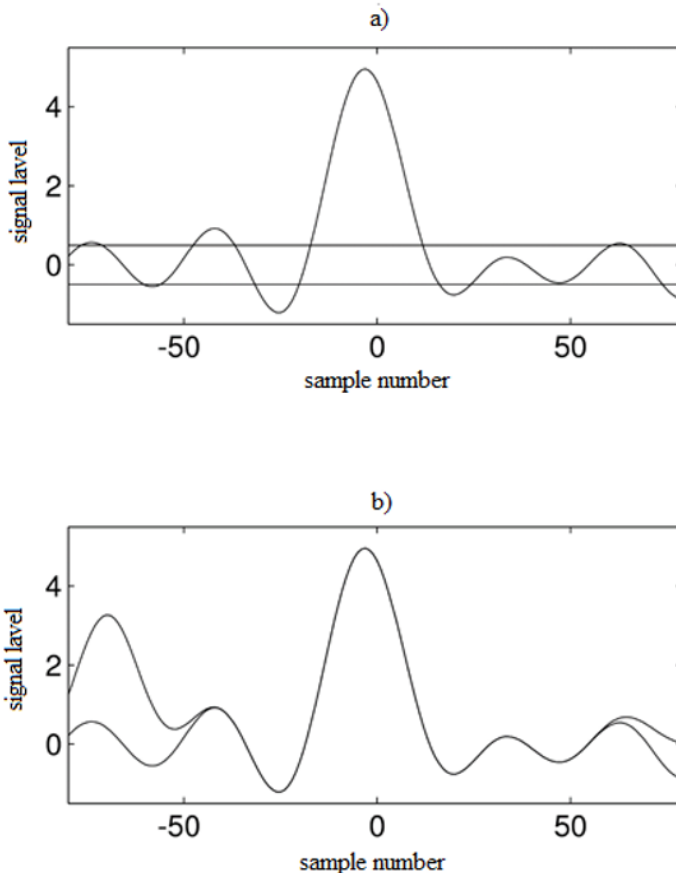
Thus, to extend the dynamic range of the ADC it means to solve the system of equations (3) with respect to the coefficients  $C_k$ .

## 2 The case of the pulse signal

Let we have the signal.

$$f(t) = \sin(9.3\pi t + 0.112) + \sum_{k=0}^{k=3} (-1)^k \sin((2k+1)\pi t + 0.112) .$$

This signal has a form is shown in Fig. 3. This signal is rather a short pulse. Exceeding the dynamic range of the ADC in the case of such pulse is typically for example, when the electrical bus is exposed to lightning.



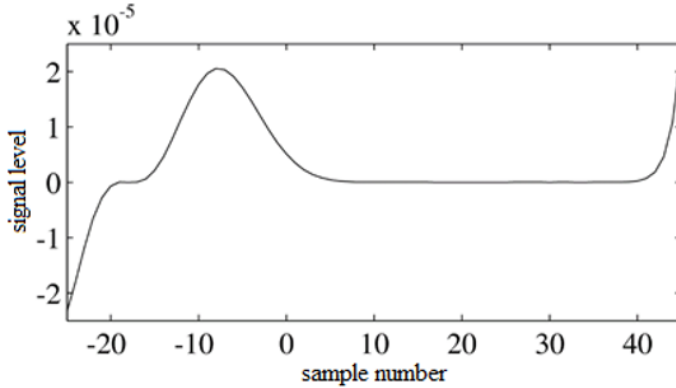
**Fig. 3.** Interpolation of the signal beyond the limits of the dynamic range of the ADC: a) the original signal; b) interpolated and the original signal.

The dynamic range of the ADC is indicated in the figure by horizontal lines (ADC can process the signal values only from -0.5 to +0.5). In this case, the dynamic range of the signal exceeds the dynamic range of the ADC by 10 times. As a basis, we use functions of the form

$$\phi_k(t) = \frac{\sin \frac{2\pi}{T}(t-kT)}{\frac{2\pi}{T}(t-kT)} .$$

We choose from the samples that do not exceed the dynamic range of the ADC, twenty, that are closest to the interpolation points and build a system of equations of type (3) of the 20th order. As a result of the solution of this system, we obtain the function shown in Fig. 3.b.

As we can see, the interpolation results are quite accurate in the center of the selected interval. In Fig. 4 the interpolation error is shown, which is calculated as the difference of signals in Fig. 3.b.



**Fig. 4.** Interpolation error.

We consider another example.

### 3 The case of the sinusoidal signal

Let we have the signal

$$f(t) = \sin(3\pi t + 0.112) .$$

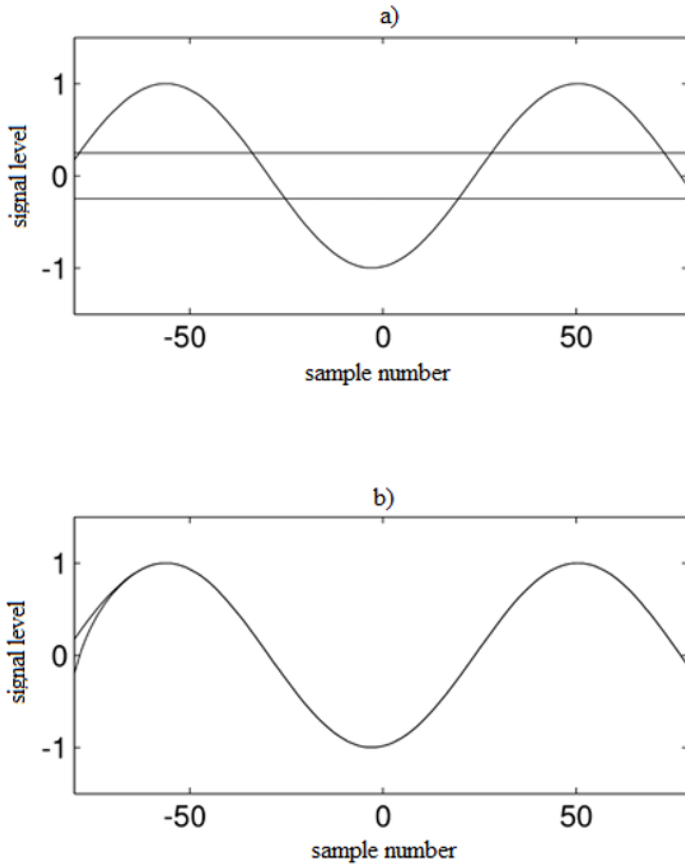
That is, this is the case when it is necessary to measure sinusoidal currents with a large range of amplitudes. It is proposed to adjust the ADC so that it does not limit only the signals with low amplitude.

Let the dynamic range of the signal exceeds the dynamic range of the ADC by 4 times. As a basis, select the functions  $\phi_k(t) = t^{k-1}$ . Just as in the first example, let's build a system of the 20th order. After its solution, we obtain the interpolating signal shown in Fig. 5 and the interpolation error shown in Fig. 6.

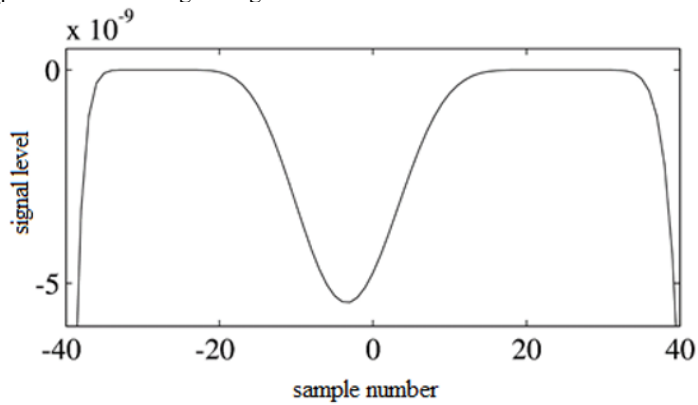
### 4 Discussion

The system of equations (3) cannot be solved exactly due to rounding errors. Therefore, an algorithm is needed to minimize the error and guarantee the accuracy of the solution. Otherwise, accumulating rounding errors can lead to catastrophic loss of accuracy or, for example, to the interruption of the algorithm caused by the practical degeneration of the

system matrix. Therefore, the expansion of the dynamic range of the ADC requires the use of mathematical apparatus of the theory of guaranteed accuracy.



**Fig. 5.** Interpolation of the signal beyond the limits of the dynamic range of the ADC: a) the original signal; b) interpolated and the original signal.



**Fig. 6.** Interpolation error.

The foundations of the theory of guaranteed accuracy in solving linear algebra problems were laid in such works as [17, 18]. It was further developed in the research of academician S.K. Godunov [19].

A key role in assessing the accuracy of the SLEEP solution is played by the condition number of the system of equations  $\mu(A)$ , which in the case of a square matrix  $A$  is  $\mu(A) = \|A\| \|A^{-1}\|$ . It follows, in particular, that the condition number is the greater the  $A$  is closer to the degenerate matrix. In the case of a rectangular matrix, the condition number is expressed as singular numbers  $\mu(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$ .

A priori estimation essentially depends on a specific computational algorithm [20]. However, it is possible to estimate the order of error of the result. The following statement is used [19, 20]

Theorem. Let  $A$  and  $\Delta A$  are square matrices of size  $N \times N$ ,  $f$  and  $\Delta f$  are vectors of size  $N$ , and

$$\|\Delta A\| \leq \alpha \|A\|, \quad \|\Delta f\| \leq \phi \|f\|, \quad \alpha \mu(A) < 1,$$

then from

$$Ax = f, \quad (A + \Delta A)(x + \Delta x) = (f + \Delta f)$$

we get

$$\frac{\|\Delta x\|}{\|x\|} \leq (\alpha + \phi) \frac{\mu(A)}{1 - \alpha \mu(A)}.$$

For a number of algorithms for solving linear systems of error of execution of arithmetic operations can be interpreted as perturbations of the original matrix and right part (inverse error analysis [20], Chapter 2). In this case, the constants  $\alpha$  and  $\phi$  have an order  $K\varepsilon$ , where  $K$  is a constant that depends on the algorithm and on the size of the system,  $\varepsilon$  is the relative error of arithmetic operations on the computer. Thus, the expected error of the solution

$$\frac{\|\Delta x\|}{\|x\|} \approx \varepsilon \mu(A) K.$$

For example, if the relative error of arithmetic operations is  $\varepsilon \approx 10^{-16}$  (so-called “double precision”) and the condition number  $\mu(A) \approx 10^{10}$ , then the a priori estimate of accuracy is

of order  $\frac{\|\Delta x\|}{\|x\|} < \varepsilon \mu(A) \approx 10^{-4}$ . If  $\mu(A) \approx 10^{14}$  and more, then from a computational point of view, such matrix is practically degenerate, and as a result there is no way to guarantee any significant figures.

A posteriori estimate can be found based on the same theorem. Let  $C_{calc}$  is the calculated solution of the system of equations;  $A$  is a square matrix with elements  $\phi_k(t_n)$ . After calculating the residual of the solution  $r = f - AC_{calc}$ , we obtain the following estimate:

$$\frac{\|C_{calc} - C\|}{\|C\|} \leq \mu(A) \frac{\|r\|}{\|f\|}.$$

This estimate is usually overestimated, but in some cases it turns into equality. In order to take advantage of these estimates in the case of a non-square matrix, it is convenient to apply a QR decomposition to the system and reduce it to the case of a square matrix system.

The interpolation method based on the solution of systems of equations is quite promising, but it has one feature that in some cases prevents its application. As shown by the simulation, and as shown above, the system of equations 3 can sometimes be degenerated. As an estimate of the probability of such an event, we can say the following, when solving a system of equations of the 20th order with basic functions such as sin and sinc, approximately every 30th time system was degenerated. To be precise, the fact that the matrix is degenerated was reported by Matlab, in which the simulation was performed. To be precise, the message did take place, but in fact the accuracy of the interpolation remained high enough, which indicates still quite good reliability of the proposed method.

However, recognizing the problem, an attempt was made to address it. Research was conducted in two directions:

- 1) search for a "good" basis at which the probability of matrix degeneration would be insignificant;
- 2) search for an interpolation method that would not require the solution of a system of equations.

From the author's point of view, the second direction was the most successful. Namely the application for interpolation by the Lagrange polynomial. As you know, the interpolation in this case is carried out according to the following formula:

$$L(t) = \sum_{n=0}^N x_n l_n(t) , \tag{4}$$

where  $L(t)$  is the result of the interpolation at the point  $t$ ;  $x_n$  are values interpolated signal in interpolation nodes  $t_n$ ;  $l_n(t)$  is the polynomial defined by the formula:

$$l_n(t) = \prod_{k=0, k \neq n}^N \frac{t-t_k}{t_n-t_k} = \frac{t-t_0}{t_n-t_0} \cdot \frac{t-t_1}{t_n-t_1} \cdot \frac{t-t_2}{t_n-t_2} \dots \cdot \frac{t-t_N}{t_n-t_N} . \tag{5}$$

Obviously, this interpolation method always converges. In this method, it is possible to avoid such an operation as division. To do this, it is recommended to calculate all values  $\frac{1}{t_n-t_k}$  in advance and store them in the processor memory. Given that in the formula (5) all possible values of  $t$  are also known (the signal is sampled at a strictly defined time). It is possible in some cases to precompute even the entire polynomial  $l_n(t)$ . Then the formula (4) is transformed into a banal discrete convolution.

Finally, we answer the question of how the accuracy of the interpolation is affected by the quantization of the interpolated signal. Indeed, the Lagrange polynomial is constructed not on an ideal function, but on a pre-quantized one. It follows from the simulation results that the quantization error corresponds approximately to the error introduced into the ADC signal. This is because the polynomial  $l_n(t)$  is independent of the analog-to-digital transformation and is therefore considered accurate. The error occurs only when calculating the formula (4). But since this formula is essentially an ordinary digital filter, this error has the same character as the usual quantization of the coefficients of the filter.

## 5 Conclusions

This article describes a method based on the property of low entropy of current and voltage in the traction networks of Railways, which significantly increases the dynamic range of the meter when using standard ADC. This can significantly reduce the cost of the meter and



improve its accuracy. Two methods are considered - using the solution of the system of equations and the interpolation Lagrange polynomial.

## References

1. V.A. Gapanovich et al. (MISIS, Moscow, 2012)
  2. J.C.M. Acevedo, *Joint Rail Conference. – American Society of Mechanical Engineers*, 181-187 (2011)
  3. Y.V. Bocharnikov et al., *IET Electr. Power App.*, **1(5)**, 675-682 (2007)
  4. R. Barrero et al., *IEEE Veh. Technol. Mag.*, **3(3)**, 26-36 (2008)
  5. S. Gritsutenko, Thesis for the degree of candidate of technical Sciences, 154 (2007)
  6. N. Korovkin, S. Gritsutenko, *Izvestiya RAN. Energetika*, **2**, 73-86 (2017)
  7. N. Korovkin, S. Gritsutenko, *2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)* (St. Petersburg, Russia, 2017)
  8. N. Korovkin, S. Gritsutenko, *Izvestiya RAN. Energetika*, **2**, 21 (2017)
  9. S.N. Chizhma, S. Gritsutenko et al., Patent for useful model No. 88157, IPC G01R17/02
  10. L.R. Rabiner, B. Gold (Prentice-Hall, Inc., Englewood Cliffs, NJ, 1975)
  11. R.G. Lyons (Pearson Education, India, 2011)
  12. E. Biberdorf, S. Gritsutenko, K. Firсанov, *Proceedings of IEEE East-West Design & Test Symposium (EWDTS'09)* (Kazan, Russia, 2017)
  13. E.A. Biberdorf, S.S. Gritsutenko, K.A. Firсанov, *Omsk scientific Bulletin* **№(90)**. A series of "Devices and technology", **2 (90)**, 200-202 (2010)
  14. E.A. Biberdorf, S.S. Gritsutenko, K.A. Firсанov, *Questions of radio electronics*, **3**, 116-125 (2010)
  15. J.I. Jackson et al., *IEEE Trans. Med. Imaging*, **10(3)** 473-478 (1991)
  16. M. Unser, *Proceedings of the IEEE*. 88(LIB-ARTICLE-2000-002), 569–587 (2000)
  17. J.H. Wilkinson, Clarendon Press, Oxford, **87**, 570-644 (1965)
  18. W. Givens, *Oak Ridge National Laboratory. Report ORNL*, **1574**, 107 (1954)
  19. S.K. Godunov et al. (Science: Sib. Department, Novosibirsk, 1988)
  20. E.A. Biberdorf, N.I. Popova (SB RAS, Novosibirsk, 2006)
  21. L. Liokumovich, K. Muravyov, Skliarov, N. Ushakov, *Appl. Opt.*, **57 (25)**, 7127-7134. (2018)
  22. V.V. Dudelev, A.R. Akhmatkhanov, K.K. Soboleva, S.H. Abdulrazak, V.E. Bugrov, V.Ya. Shur, G.S. Sokolovskii, *Proc. - Int. Conf. Laser Opt.*, ICLO 8435259 (St. Petersburg, Russia, 2018)
  23. V.V. Dudelev, K.A. Fedorova, D.V. Chistyakov, K.K. Soboleva, V.E. Bugrov, E.U. Rafailov, G.S. Sokolovskii, *Proc. - Int. Conf. Laser Opt.*, ICLO 8435470 (St. Petersburg, Russia, 2018)
  24. D.O. Budanov, D.V. Morozov, M.M. Pilipko, *J. Commun. Technol. El.*, **62(4)**, 421-431 (2017)
  25. A. Bolotin, V. Bakayev, S. Vazhenin, *J. Physic. Educ. Sport*, **16** (2016)
- M.M. Pilipko, St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunication and Control Systems, **4(176)**, 95-101 (2013)