



Scientific and technical journal

**ROCKET-SPACE
DEVICE ENGINEERING
AND INFORMATION
SYSTEMS**

Volume 5. Issue 1. 2018



Scientific and technical journal

“Rocket-Space Device Engineering and Information Systems”

Vol. 5. No. 1. 2018

Founder:
Joint Stock Company “Russian Space Systems”

Advisory Council

Chair:

Tyulin A.E., Cand. Sci. (Eng.), Corresponding Member of Russian Academy of Missile and Artillery Sciences, Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Deputy Chairmen:

Ezhov S.A., Dr. Sci. (Eng.), Prof., Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Romanov A.A., Dr. Sci. (Eng.), Prof., Corresponding Member of International Academy of Astronautics, Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Nesterov E.A., Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Members of the Advisory Council:

Artemyev V.Yu., Joint Stock Company “Scientific and Production Association of Measurement Equipment”, Moscow, Russian Federation

Baturin Yu.M., Doctor of Law, Prof., Corresponding Member, Russian Academy of Sciences, S.I. Vavilov Institute for the History of Science and Technology of Russian Academy of Sciences, Moscow, Russian Federation

Blinov A.V., Cand. Sci. (Eng.), Corresponding Member of Russian Engineering Academy, Joint-Stock Company “Research Institute of physical measurements”, Penza, Russian Federation

Bugaev A.S., Dr. Sci. (Phys.-Math.), Prof., Academician, Russian Academy of Sciences, Kotelnikov Institute of Radio Engineering and Electronics of RAS, Moscow, Russian Federation

Zhantayev Zh.Sh., Dr. Sci. (Phys.-Math.), Academician of Kazakhstan National Academy of Natural Sciences, Joint-Stock Company “National Center of Space Research and Technology”, Almaty, Republic of Kazakhstan

Zhmur V.V., Dr. Sci. (Phys.-Math.), Prof., Moscow Institute of Physics and Technology, Moscow, Russian Federation

Kolachevsky N.N., Dr. Sci. (Phys.-Math.), Prof., Corresponding Member, Russian Academy of Sciences, Lebedev Physical Institute of the Russian Academy of Sciences, Moscow, Russian Federation

Kuleshov A.P., Dr. Sci. (Eng.), Prof., Academician, Russian Academy of Sciences, Skolkovo Institute of Science and Technology, Moscow, Russian Federation

Nosenko Yu.I., Dr. Sci. (Eng.), Prof., Joint Stock Company “Research Institute of Precision Instruments”, Moscow, Russian Federation

Perminov A.N., Dr. Sci. (Eng.), Prof., Member of International Academy of Astronautics, Russian Engineering Academy, Russian Academy of Cosmonautics named after K.E. Tsiolkovsky, Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Petrukovich A.A., Dr. Sci. (Phys.-Math.), Prof., Corresponding Member, Russian Academy of Sciences, Space Research Institute of the Russian Academy of Sciences, Moscow, Russian Federation

Rainer Sandau, Dr. Sci. (Eng.), Adjunct Professor, International Academy of Astronautics, Berlin, Germany

Stupak G.G., Dr. Sci. (Eng.), Prof., Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Chebotaev A.S., Dr. Sci. (Eng.), Prof., Stock Company “Special research bureau of Moscow power engineering institute”, Moscow, Russian Federation

Chernyavsky G.M., Dr. Sci. (Eng.), Prof., Corresponding Member, Russian Academy of Sciences, Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Chetyrkin A.N., branch of Joint Stock Company “United Rocket and Space Corporation” – “Institute of Space Device Engineering”, Moscow, Russian Federation

The publication frequency is four issues per year.

The journal is included into the Russian Science Citation Index.

The journal is included into the List of peer-reviewed scientific publications approved by the Higher Attestation Commission (VAK RF).

The opinions expressed by authors of the papers do not necessarily those of the editors.

ISSN 2409-0239

DOI 10.30894/issn2409-0239.2018.5.1

The subscription number of the journal in the united catalogue

“The Russian Press” is 94086.

Editorial Board

Editor-in-Chief:

Romanov A.A., Dr. Sci. (Eng.), Prof., Corresponding Member of International Academy of Astronautics, Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Deputy Editor-in-Chief:

Fedotov S.A., Cand. Sci. (Eng.), Senior Researcher, Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Members of the Editorial Board:

Alekseyev O.A., Dr. Sci. (Eng.), Prof., Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Alybin V.G., Dr. Sci. (Eng.), Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Akhmedov D.Sh., Dr. Sci. (Eng.), Corresponding Member of National Engineering Academy of the Republic of Kazakhstan, SLLP “Institute of Space Systems and Technologies”, Almaty, Republic of Kazakhstan

Betanov V.V., Dr. Sci. (Eng.), Prof., Corresponding Member of Russian Academy of Missile and Artillery Sciences, Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Vasilkov A.P., Ph. Doctor in Physics and Mathematics, Science Systems and Applications Inc., Lanham, Maryland, the USA

Vatutin V.M., Dr. Sci. (Eng.), Prof., Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Danilin N.S., Dr. Sci. (Eng.), Prof., Academician of Russian and International Engineering Academies, Russian Academy of Cosmonautics named after K.E. Tsiolkovsky, Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Zhodzhishsky A.I., Dr. Sci. (Eng.), Academician of Russian Academy of Cosmonautics named after K.E. Tsiolkovsky, Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Zhukov A.A., Dr. Sci. (Eng.), Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Moroz A.P., Dr. Sci. (Eng.), Joint Stock Company “Scientific and Production Association of Measurement Equipment”, Moscow, Russian Federation

Pobedonostsev V.A., Dr. Sci. (Eng.), branch of Joint Stock Company “United Rocket and Space Corporation” – “Institute of Space Device Engineering”, Moscow, Russian Federation

Povalyayev A.A., Dr. Sci. (Eng.), Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Rimskaya O.N., Cand. Sci. (Econ.), Assoc. Prof., Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Romanov A.A., Dr. Sci. (Eng.), Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Sviridov K.N., Dr. Sci. (Eng.), Prof., Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Selivanov A.S., Dr. Sci. (Eng.), Prof., Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Strelnikov S.V., Dr. Sci. (Eng.), Joint Stock Company “Scientific Production Association Orion”, Krasnoznamensk, Russian Federation

Sychev A.P., Cand. Sci. (Eng.), Joint Stock Company “Research Institute of Precision Instruments”, Moscow, Russian Federation

Tokarev A.S. (Tech. Sec.), Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Tuzikov A.V., Dr. Sci. (Phys.-Math.), Prof., Correspondent Member of the National Academy of Sciences of Belarus, The State Scientific Institution “The United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus

Yazeryan G.G., Cand. Sci. (Eng.), Joint Stock Company “Russian Space Systems”, Moscow, Russian Federation

Joint Stock Company “Russian Space Systems”,
ul. Aviamotornaya 53, Moscow, 111250 Russia
Tel. +7 (495) 673-96-29, www.russianspacesystems.ru
e-mail: rks-304@mail.rue-mail; journal@spacecorp.ru

© Joint Stock Company “Russian Space Systems”
© FIZMATLIT



Moscow
FIZMATLIT®
2018

Contents

Vol. 5, Iss. 1, 2018

Space Navigation Systems and Devices. Radiolocation and Radio Navigation

- Determination of the Relative Position of Objects by the First Phase Measurement Differences of One Epoch
A.I. Zhodzishskiy, O.V. Nesterov, A.S. Bukin 2
- Promising Quantum-Optical Technologies for Satellite Navigation Challenges
N.N. Kolachevsky, K.Yu. Khabarova, I.V. Zalivako, I.A. Semerikov, A.S. Borisenko, I.V. Sherstov, S.N. Bagaev, A.A. Lugovoy, O.N. Prudnikov, A.V. Taichenachev, S.V. Chepurov 12
- Technique for Hardware Residuals Evaluation of GNSS-Measurements of Pseudorange in Phase of the Carrier of the Onboard Navigation System of LEO Spacecraft
A.P. Fursov, A.A. Fursov, V.S. Vdovin, A.V. Zaichikov, Z.A. Pozyaeva 25
-

Aerospace Methods for Earth Remote Sensing

- Onboard Image Processing by Inverse Filtering
V.E. Kvitka, V.D. Blinov 35
-

Radio Engineering and Space Communication

- Features of Calculation and Design of High-Speed Radio Links for Earth Remote Sensing Spacecraft
A.N. Ershov, V.V. Berezkin, S.V. Petrov, A.V. Petrov, D.A. Pochivalin 44
- The Research of Exciter Frequency Characteristics of a Quad-Band Antenna Based on a Corrugated Horn
D.D. Gabriel'ean, V.I. Demchenko, A.E. Korovkin, D.Ya. Razdorkin, A.V. Shipulin, Yu.I. Poltavets 50
- The Calculation Methodology for the Energetic Reserve of the Radio Link Spacecraft-Station
G.A. Erokhin, V.I. Mandel, Yu. A. Nesterkin, A.P. Strukov 56
- Calculation of Antenna Gain via 3D Radiation Pattern and Estimation of Their Mutual Influence
V.I. Tolkachev, O.G. Pikalov, S.V. Pan'chev, I.G. Novikov 65
-

Systems Analysis, Spacecraft Control, Data Processing, and Telemetry Systems

- Sampling Theorem and Practical Usage of Integer Functions for Signal Representation on the Receiving Side
A.I. Troshchenkov 71
- Modernization of Algorithms for Generating Correlated Data When Integrating Diverse Reception Methods and Noiseless Decoding Methods
V.L. Vorontsov 76
-

Solid-State Electronics, Radio Electronic Components, Micro- and Nanoelectronics, Quantum Effect Devices

- A Method of Detecting of Internal Defects of CMOS-Microcircuits
I.Yu. Bulaev 83

Determination of the Relative Position of Objects by the First Phase Measurement Differences of One Epoch

A.I. Zhodzishskiy, *Dr. Sci. (Engineering), Prof., ntsmou@rniikp.ru*

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

O.V. Nesterov, *postgraduate student, ntsmou@rniikp.ru*

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

A.S. Bukin, *postgraduate student, ntsmou@rniikp.ru*

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

Abstract. A new method is considered that allows determination of the relative position of objects (the vector of the baseline) within a millimeter error by the fractional parts of the first differences in the phase measurements of one epoch. It is shown that the unknown coordinates of the end of the baseline vector correspond to the basic minimum of the reduced quadratic function. An algorithm for searching for local minima has been developed, as well as two approaches to selection of the main minimum: decision-making by the threshold and decision-making by the absolute minimum. An algorithm for computing the baseline vector is given and probabilistic and time characteristics are given for its implementation for the case of sharing single-frequency (L1) range of GLONASS and GPS measurements.

The method is called the "RSS method" by the name of the patent holder (JSC «Russian Space Systems»). A classification of known methods for resolving the ambiguity of phase measurements is presented, which includes the RSS method. The RSS method is a geometric method, in which the search for spatial coordinates of the end of the baseline vector is performed in a geocentric coordinate system with the elimination of the unknown integer number of phase cycles. The method is insensitive to the loss of the count of the phase cycles of the received signals.

Keywords: global navigation satellite systems, relative positioning, phase measurements, disambiguation.

Introduction

Global navigation satellite systems (GNSS) are increasingly used in solving various tasks for both military and civilian purposes. The most accuracy when using GNSS can be achieved when determining the relative position of objects. Relative position of objects with high accuracy is required in geodesy, during construction, monitoring of displacements of engineering structures and the earth's surface, unmanned control of aircraft and ground vehicles, etc. [1]. It is also used for spatial orientation of moving objects and mechanisms.

The relative (mutual) position of objects can be defined by the vector \vec{L} of the baseline, the origin of which is at the point 1 with the coordinates $\{x_1, y_1, z_1\}$, and the end is at point 2 with the coordinates $\{x_2, y_2, z_2\}$: $\vec{L}(x_2 - x_1; y_2 - y_1; z_2 - z_1)$. In this case, the user does not need the absolute coordinates of the points, or they are known with an allowable error.

When using GNSS, the vector \vec{L} is determined from the differences in the measurements of navigation receivers installed at points 1 and 2. Code measurements make it possible to provide meter accuracy, and using phase measurement differences opens the possibility of determining the relative position of objects with a millimeter error. In the latter case, the first and second phase measurement differences are usually used.

The main problem in the processing of phase measurements is their ambiguity associated with the cyclic nature of the phase. A large amount of scientific works has been devoted to the problem of resolving phase ambiguity for the tasks of positioning objects using GNSS. The monograph [2] based on [3] and [4] contains a classification of methods for resolving the ambiguity of phase measurements. This classification is shown in Fig. 1, where references are also made to the literature in which the methods are described.

The method considered in this article will be called, for brevity, "the RSS method" by the name of the organization (JSC "Russian Space Systems"), which received a patent for it [12]. It is a geometric method in which the search for spatial coordinates of the end of the vector of the baseline is performed in the geocentric coordinate system. A distinctive feature of the RSS method is the elimination of the unknown whole cycles in phase measurements and the determination of the coordinates of the end of the baseline vector along fractional parts of these measurements at each epoch at the rate of

measurement. This gives it a number of significant advantages. In particular, it becomes insensitive to the loss of counting of whole phase cycles. The RSS method, instead of filtering phase measurements (ambiguous in nature), performs filtering (smoothing) of the calculated coordinates. This can be very effective for dynamic objects with a relatively weak power potential of the radio link or in conditions of poor electromagnetic environment.

Formulation of the problem

We assume that at points 1 and 2, at moments $t_{r,1}$ and $t_{r,2}$, the navigation signals of two satellite constellations, such as GPS and GLONASS in the L_1 frequency band, are received. As a result, at the output of receivers for one epoch we will have n measurements of pseudorange by the code of the pseudo-random sequence and the complete pseudo phase of the carrier frequency of the navigation signal: ρ_1^j and Φ_1^j for the first receiver and ρ_2^j and Φ_2^j for the second receiver. Here and below, the indices $j = 1, \dots, n_{GPS}$ will refer to the GPS satellites, and the indices $j = n_{GPS} + 1, \dots, n$ - to the GLONASS satellites.

The measured pseudorange values are related to the true ranges R_1^j and R_2^j by the relations:

$$\rho_1^j = R_1^j + cT_1^j + \xi_1^j; \rho_2^j = R_2^j + cT_2^j + \xi_2^j, j = 1, \dots, n,$$

where

$$R_1^j = \sqrt{(x^j - x_1)^2 + (y^j - y_1)^2 + (z^j - z_1)^2}, \quad (1)$$

$$R_2^j = \sqrt{(x^j - x_2)^2 + (y^j - y_2)^2 + (z^j - z_2)^2}, \quad (2)$$

$\{x^j, y^j, z^j\}$ - is the coordinates of the j -th satellite,

c is the speed of light,

T_1^j and T_2^j are the displacements of the time scales of the first and second receivers from the system time scale of the satellite constellation: for GPS $T_1^j = T_{GPS1}$ and $T_2^j = T_{GPS2}$, for GLONASS $T_1^j = T_{GL1}$ and $T_2^j = T_{GL2}$,

ξ_1^j, ξ_2^j - the total errors in the measurement of pseudoranges.

In the geocentric coordinate system, using the method of least squares, the approximate values of the absolute coordinates of the receiving antennas are calculated as $\{x_1^0, y_1^0, z_1^0\}$ and $\{x_2^0, y_2^0, z_2^0\}$, as well as discrepancies in the time scales of the receivers: for GPS - T_{GPS1}^0 and T_{GPS2}^0 , and for GLONASS, - T_{GL1}^0 and

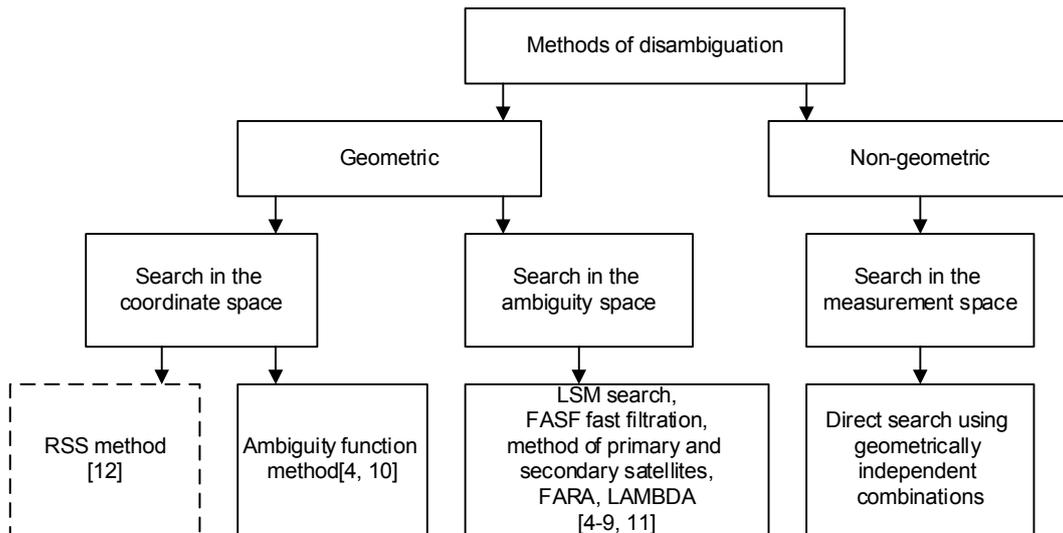


Fig. 1. Classification of methods for resolving the ambiguity of phase measurements

$T_{GL_2}^0$. The absolute coordinates found are used to refine corrections to the signal delay in the atmosphere.

When solving the navigation problem of relative positioning, the results of measurements in receivers should refer to the same time point, i.e. $t_{r1} = t_{r2} = t_r$. Then the difference of the radiation of the corresponding signals on the j -th satellite preceding this moment of time will be equal to $\Delta t_{rad}^j = \frac{1}{c}(R_2^j - R_1^j) + \Delta t_{1,2}$,

where $\Delta t_{1,2} = T_2 - T_1$ is the divergence of the receiver time scales. In this case, in calculating $(R_2^j - R_1^j)$, both the movement of the satellites along the orbit during the time Δt_{rad}^j , and the difference in the motion of the receivers due to the rotation of the Earth should be taken into account [13].

Expressions for the first differences of pseudo-phase measurements (in phase cycles), taking into account corrections for the delay of signals in the troposphere, ionosphere, known hardware delays, including calibration corrections for GLONASS interlater delay, etc. have the following form:

$$\Phi_{1,2}^j = \frac{R_2^j}{c} f_2^j - \frac{R_1^j}{c} f_1^j + f_0^j \Delta t_{1,2}^j - \psi_{1,2}^j + N_{1,2}^j + \xi_{1,2}^j, \quad (3)$$

$j = 1, \dots, n,$

where R_1^j, R_2^j are the true distances between the phase center of the transmitting antenna of the j -th satellite at the moment of radiation of the navigation signal and the phase centers of the first and second receiving antennas

at the moments of receiving this signal (via the satellite constellation),

f_1^j and f_2^j are the frequencies of signals received by receivers 1 and 2 of the j -th satellite (taking into account the Doppler frequency shift),

f_0^j is the nominal frequency of the signal emitted by the GPS satellites, or the frequency of the GLONASS satellite zero-letter signal,

$\psi_{1,2}^j$ is the difference of initial phases in synthesizers of reference signals of receivers in terms of carrier frequency of GPS and zero letter of GLONASS frequency (for GPS - $\psi_{1,2}^j = (\psi_{1,2})_{GPS}$, for GLONASS - $\psi_{1,2}^j = (\psi_{1,2})_{GL}$),

$N_{1,2}^j$ are unknown integers equal to the differences of the integers of the phases of the signals of the reference oscillators in the counters of the total phase measurement of receivers 1 and 2, which determine their initial state at the time of measurement,

$\xi_{1,2}^j$ is the difference in the total errors of pseudo-phase measurements in receivers due to multipath errors, noise, uncompensated delays in the atmosphere, and so on.

The divergence of the receiver time scales will be: in the GPS paths $\Delta t_{1,2}^j = T_{GPS_1}^0 - T_{GPS_2}^0$, and $-\Delta t_{1,2}^j = T_{GL_1}^0 - T_{GL_2}^0$ in the GLONASS paths.

If we assume that the origin of the required vector \vec{L}^* of the baseline is at the reference point $\{x_1^0, y_1^0, z_1^0\}$, then the coordinates of its end $\{x_2^*, y_2^*, z_2^*\}$ lie in the region Q with center at the point $\{x_2^0, y_2^0, z_2^0\}$, the size

Expressing the phase in meters (in wavelengths), system (3) can be represented as:

$$\tilde{\Phi}_{1,2}^j = (\hat{R}_2^j - R_1^{j,0}) - g_{\text{GPS}}^j \lambda_{\text{GPS}} (M_{\text{GPS}}^j + \eta_{\text{GPS}}) - g_{\text{GL}}^j \lambda_{\text{GL}} (M_{\text{GL}}^j + \eta_{\text{GL}}) + \xi_{1,2}^j, \quad j = 1, \dots, n, \quad (6)$$

where $\tilde{\Phi}_{1,2}^j = \frac{c}{f_2^j} \Phi_2^j - \frac{c}{f_1^j} \Phi_1^j$ is the difference of the pseudophases,

$g_{\text{GPS}}^j, g_{\text{GL}}^j$ are the coefficients determining the belonging of the j -th equation to the GPS or GLONASS satellite constellation, namely

$$g_{\text{GPS}}^j = \begin{cases} 1 & \text{for } j = 1, \dots, n_{\text{GPS}} \\ 0 & \text{for } j = n_{\text{GPS}} + 1, \dots, n \end{cases}, \quad g_{\text{GL}}^j = \begin{cases} 0 & \text{for } j = 1, \dots, n_{\text{GPS}} \\ 1 & \text{for } j = n_{\text{GPS}} + 1, \dots, n \end{cases}$$

$$\begin{cases} 0 & \text{for } j = 1, \dots, n_{\text{GPS}} \\ 1 & \text{for } j = n_{\text{GPS}} + 1, \dots, n \end{cases}$$

$M_{\text{GPS}}^j, M_{\text{GL}}^j$ are unknown integers,

η_{GPS} and η_{GL} are the unknown fractional parts of the difference of the initial phases on the carrier frequency of GPS and the zero frequency letter of GLONASS,

λ_{GPS} is the wavelength of the GPS carrier signal,

λ_{GL} is the wavelength of the zero letter carrier signal of GLONASS.

Note that the non-linearity of equation (6) with respect to the unknown displacements $\{D_x, D_y, D_z\}$ occurring in R_2^j (5), is determined only by the first term equal to the difference in distances due to the spatial diversity of the receiver antennas. The second (for GPS) and the third (for GLONASS) terms are linear with respect to the unknowns $\{M^j \text{ and } \eta\}$.

In the vicinity of the point $\{x_2^0, y_2^0, z_2^0\}$, the expressions for estimating the range (5) can be linearized:

$$\hat{R}_2^j = R_2^{j,0} + A_x^j D_x + A_y^j D_y + A_z^j D_z, \quad j = 1, \dots, n, \quad (7)$$

where $R_2^{j,0}$ is calculated by (2) with $x_2 = x_2^0, y_2 = y_2^0, z_2 = z_2^0$, and

$A_x^j = \frac{\partial R_2^j}{\partial x_2}, A_y^j = \frac{\partial R_2^j}{\partial y_2}, A_z^j = \frac{\partial R_2^j}{\partial z_2}$ are the values inverse to the direction cosines from the point $\{x_2^0, y_2^0, z_2^0\}$ to the j -th satellite.

For fixed values of the displacements $\{D_x, D_y, D_z\}$, expression (7) allows one to find estimates of the distances \hat{R}_2^j . Substituting \hat{R}_2^j in (6), we obtain the corresponding estimates for the difference of the pseudo-phases $\tilde{\Phi}_{1,2}^j$, which allow us to form residuals $(\tilde{\Phi}_{1,2}^j - \hat{\Phi}_{1,2}^j)$ and write the quadratic function

$$A(D_x, D_y, D_z) = \sum_{j=1}^n [\tilde{\Phi}_{1,2}^j - \hat{\Phi}_{1,2}^j]^2. \quad (8)$$

Here $[x]$ means the operation of pointing off the decimal places of the x (expressed in wavelengths) of less than half the wavelength.

To explain the possibility of finding displacements $\{D_x, D_y, D_z\}$ using only fractional parts of the difference of pseudophases $\tilde{\Phi}_{1,2}^j$, we assume that their values and parameters $\eta_{\text{GPS}}, \eta_{\text{GL}}$ are known and equal to $\{\tilde{D}_x^*, \tilde{D}_y^*, \tilde{D}_z^*\}$ and $\eta_{\text{GPS}}^*, \eta_{\text{GL}}^*$ (The required point has the coordinates $\{x_2^*, y_2^*, z_2^*\}$). For $M_1^j = M_2^j = 0$, substituting η_{GPS}^* and η_{GL}^* in (6) and sorting with a small step (for example, $0,01\lambda$) all values $\{D_x, D_y, D_z\}$ in the region Q , including $\{\tilde{D}_x^*, \tilde{D}_y^*, \tilde{D}_z^*\}$ (see Figure 2), we can construct the function (8). The function (8) turns out to be multimodal, and the coordinates of its main minimum will correspond to the required displacements $\{D_x, D_y, D_z\}$. Thus, the problem of determining the baseline vector from the fractional values of the first differences of pseudo-phase measurements reduces to the problem of finding the displacement coordinates $\{D_x, D_y, D_z\}$, which minimize the quadratic function (8).

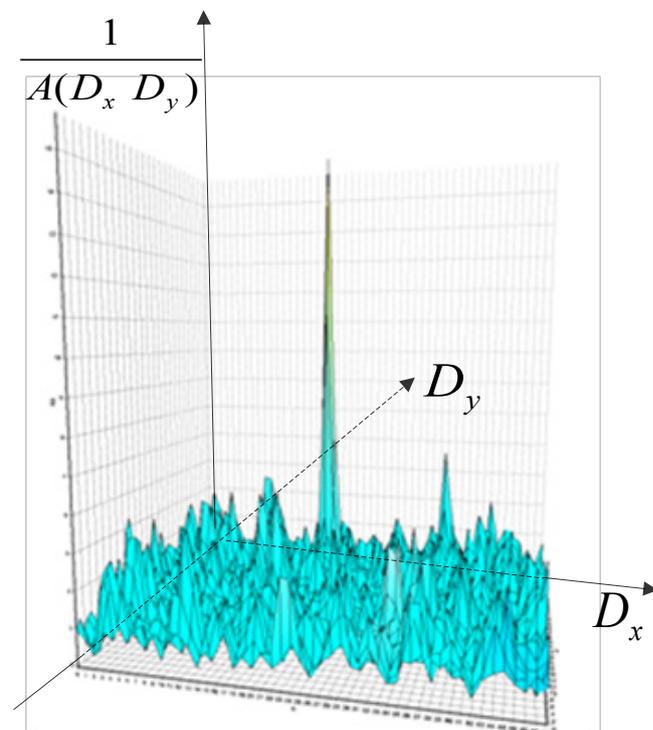


Fig. 3. A fragment of the multimodal function $\frac{1}{A(D_x, D_y)}$

In Fig. 3 for the two-dimensional case $\{x, y\}$, the function $\frac{1}{A(D_x, D_y)}$ is shown for clarity, the coordinates of the maxima of which coincide with the coordinates of the minima of $A(D_x, D_y)$. The function is constructed with the number of simultaneously visible satellites $n = 14$ and moderate values $\xi_{1,2}^j$ (not exceeding 0.1λ , that is 2 cm).

Algorithm for calculating the baseline

The patent [12] proposes a method and a device that allows one to determine the desired displacements $\{D_x, D_y, D_z\}$ from the first differences of pseudo-phase measurements of one epoch. To this end, in the region Q parallel to the coordinate axes we draw planes with step S , starting from $\{x_2^0, y_2^0, z_2^0\}$, the points of intersection of which form nodes with the coordinates $\{m_x S, m_y S, m_z S\}$, where m_x, m_y, m_z are the node numbers along the x, y, z axes (see Figure 2 for the two-dimensional case).

It will be shown below how, by analyzing the values of the function (8) in the vicinity of these nodes, one can find the coordinates of its local minima and determine the fundamental minimum. In the simplest case, when signals of wavelength λ are used and there are no measurement errors, the search step, coinciding with the grid spacing S , can be selected from the condition that spheres with radius λ and centers at the nodes fill the Q region without voids. It can be shown that in this case

$$S = \frac{\lambda}{\sqrt{3}}.$$

Below are the calculations of the probability of correct resolution of the ambiguity as a function of the step S for different values of the measurement error.

The fundamental minimum of the function (8) lies in the vicinity of one (or several) of the nodes, appearing, as a rule, displaced from it by some amount $\{D_x, D_y, D_z\}$.

For clarification, we refer to Fig. 3, which illustrates the dependence of the residuals $\tilde{\Phi}_{12}^j - \tilde{\Phi}_{12}^j(D_x)$ for several satellites on the displacement D_x , provided that along the other axes the required displacements are found and are equal to D_y^*, D_z^* . The origin ($D_x = 0$) in the figure corresponds to the point $\{x_2^0, y_2^0, z_2^0\}$, in the vicinity of which the linearization of the range estimate \tilde{R}_2^j (7), and the scale for both axes is chosen to be the same.

The residuals for all satellites are straight lines, the slope of which is determined by the geometric factor and can not exceed 45° . If there is no error in the measurements, all the lines intersect at the point $D_x = D_x^*$. The presence of measurement errors transforms the intersection of the residuals into a region, the dimensions of which depend on the errors in the measurement of the first differences in the pseudo-phase measurements.

If the destabilizing factors (atmosphere, calibration, reception paths, etc.) are correctly taken into account, these errors are mainly determined by the phase multipath. In practice, the coordinates of the reference point $\{x_1, y_1, z_1\}$ are chosen assuming the clear sky and the small multipath. Therefore, the multipath in the first differences is more dependent on the reception of the signal at point 2. Note that in the case of equality of the reflected signal to the direct (100% multipath), the error in determining the phase of the signal corresponds to $\pm 45^\circ$, i.e. $\pm \frac{\pi}{4}$.

For the coordinates of the l -th node, equal to

$$D_x^l = m_x^l S, D_y^l = m_y^l S, D_z^l = m_z^l S, \quad (9)$$

the phase estimates (6) with allowance for (7) for $M_{GPS}^j = M_{GL}^j = 0$ and $\eta_{GPS} = \eta_{GL} = 0$ will be

$$(\tilde{\Phi}_{12}^j)^{l,r=0} = \tilde{R}_2^{j,0} + A_x^j m_x^l S + A_y^j m_y^l S + A_z^j m_z^l S - R_1^{j,0}, \quad j = 1, \dots, n. \quad (10)$$

The values of the deviations $\{d_x^{l,r=1}, d_y^{l,r=1}, d_z^{l,r=1}\}$ of the local minimum from the coordinates of the node (9), as well as the values $\eta_{GPS}^{l,r=1}$ and $\eta_{GL}^{l,r=1}$ can be found by solving the system of equations using the method of the least squares:

$$[(\tilde{\Phi}_{12}^j)^{l,r=1}] = [(\tilde{\Phi}_{12}^j - (\tilde{\Phi}_{12}^j)^{l,r=0})], j = 1, \dots, n. \quad (11)$$

Here

$$(\tilde{\Phi}_{1,2}^j)^{l,r=1} = \tilde{R}_2^{j,0} + A_x^j(m_x^l S + d_x^{l,r=1}) + A_y^j(m_y^l S + d_y^{l,r=1}) + A_z^j(m_z^l S + d_z^{l,r=1}) - R_1^{j,0} - g_{GPS}^j \lambda_{GPS} \eta_{GPS}^{l,r=1} - g_{\Gamma\pi}^j \lambda_{\Gamma\pi} \eta_{\Gamma\pi}^{l,r=1}, \quad j = 1, \dots, n \quad (12)$$

the phase estimate (6) for the coordinates of the point

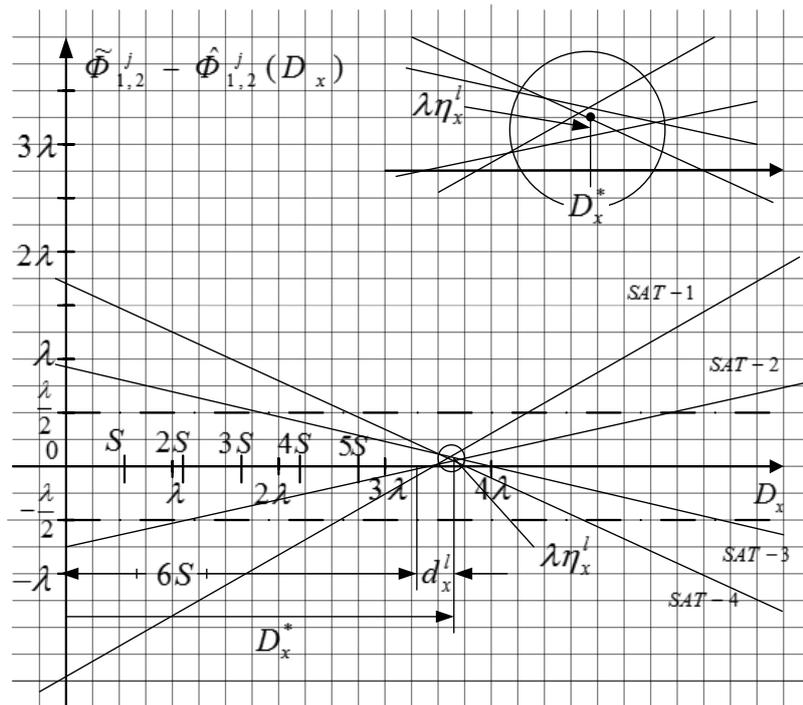


Fig. 4. Example of the dependence of the residuals on the displacement (for $D_x^* = 6S + d_x^l, D_y = D_y^*, D_z = D_z^*$)

$$\begin{aligned} D_x^{l,r=1} &= m_x^l S + d_x^{l,r=1}, & D_y^{l,r=1} &= m_y^l S + d_y^{l,r=1}, \\ D_z^{l,r=1} &= m_z^l S + d_z^{l,r=1}. \end{aligned} \quad (13)$$

Substituting the values found for $d_x^{l,r=1}, d_y^{l,r=1}, d_z^{l,r=1}, \eta_{\text{GPS}}^{l,r=1}, \eta_{\text{GL}}^{l,r=1}$ in (12), we obtain the estimate of the phase $(\tilde{\Phi}_{1,2}^j)^{l,r=1}$ on the first pass ($r = 1$) of searching for a local minimum in the vicinity of the l -th node.

To clarify the position of this minimum, we perform the second pass ($r = 2$) for the coordinates

$$\begin{aligned} D_x^{l,r=2} &= m_x^l S + d_x^{l,r=1} + d_x^{l,r=2}, \\ D_y^{l,r=2} &= m_y^l S + d_y^{l,r=1} + d_y^{l,r=2}, \\ D_z^{l,r=2} &= m_z^l S + d_z^{l,r=1} + d_z^{l,r=2}, \end{aligned} \quad (14)$$

where $d_x^{l,r=2}, d_y^{l,r=2}, d_z^{l,r=2}$ are the required corrections.

The above corrections, as well as the parameters $\eta_{\text{GPS}}^{l,r=2}, \eta_{\text{GL}}^{l,r=2}$ are found in the similar way as it was done on the first pass ($r = 1$), from the solution of the system of linear equations

$$\begin{aligned} [(\tilde{\Phi}_{1,2}^j)^{l,r=2}] &= [(\tilde{\Phi}_{1,2}^j - (\tilde{\Phi}_{1,2}^j)^{l,r=1})], \\ j &= 1, \dots, n. \end{aligned} \quad (15)$$

Let us check that the values of the corrections found are sufficiently small, for example,

$$\begin{aligned} d_x^{l,r=2} &< 10^{-4} \text{ m}, & d_y^{l,r=2} &< 10^{-4} \text{ m}, \\ d_z^{l,r=2} &< 10^{-4} \text{ m}. \end{aligned}$$

If all the conditions are satisfied, then we will consider the search for the coordinate shifts of the second receiving antenna at the l -th step complete, if not, then proceed to the calculation of the next ($r = 3$) correction.

(Note: As practice shows, if the number of passes in the calculation of corrections exceeds 3, then the calculation continues to be unsuitable due to the presence of one or more anomalous phase measurements. The measurements in this case should be rejected).

We denote the corrections to the shifts of the coordinates found at the l -th step as D_x^l, D_y^l, D_z^l , and the parameters $\eta_{\text{GPS}}^l, \eta_{\text{GL}}^l$, and introduce the residual vector as follows:

$$\nabla^{j,l} = [\tilde{\Phi}_{1,2}^j - (\tilde{\Phi}_{1,2}^j)^l] - \lambda_{\text{GPS}} \eta_{\text{GPS}}^l,$$

$j = 1, \dots, n_{GPS}$ – for the GPS measurements,

$$\nabla^{j,l} = [\tilde{\Phi}_{1,2}^j - (\tilde{\Phi}_{1,2}^j)^l] - \lambda_{GL} \eta_{GL}^l,$$

$j = n_{GPS} + 1, \dots, n$ – for the GLONASS measurements.

The values of the residuals $\nabla^{j,l}$ depend only on the spatial separation of the receiving antennas and do not depend on the divergence of the receiver time scales, including the fractional values of the difference of the initial phases $\lambda_{GPS} \eta_{GPS}^l$ and $\lambda_{GL} \eta_{GL}^l$ (see Fig. 4).

Let us calculate the value of the quadratic function (8) at the l -th step

$$A^l(D_x^l, D_y^l, D_z^l) = \sum_{j=1}^n (\nabla^{j,l})^2$$

and compare it with the a priori given threshold α . If $\sqrt{A^l} \leq \alpha$, then we assume that the fundamental minimum of the function (8) coincides with the local minimum found at the l -th step, and take the coordinates of the second receiving antenna found at this step of the search for the required, otherwise we continue to the $(l + 1)$ -th step. If for all search steps $\sqrt{A^l} > \alpha$, then for the basic minimum of the function (8) we take the coordinates defined at the search step $l = l^*$, for which $\sqrt{A^l}$ is minimal.

The first case corresponds to the decision on the threshold, the second is on the absolute minimum.

We calculate the coordinates of the vector of the basis line for the coordinates of the vector of the basis line found at the l^* -th step of the search for coordinates of the second receiving antenna: $L_x = x_2^0 + D_x^l - x_1^0$, $L_y = y_2^0 + D_y^l - y_1^0$, $L_z = z_2^0 + D_z^l - z_1^0$, which determine the mutual position of the objects.

If necessary, the value of M^j , corresponding to an integer number of cycles (wavelengths) in the pseudo-phase difference for the j -th satellite, can be determined from expression (6), substituting into it the found values $R_2^j - R_1^{j,0}$, and η_{GPS} or η_{GL} .

Probabilistic characteristics of the RSS method

In Fig. 5 for the total number of GPS satellites and GLONASS satellites equal to 17, the probability dependences P_{cr} of the correct resolution of the ambiguity (finding the basic minimum of the function (8)) are plotted as a function of the search step S for different values of the standard mean-square error $\frac{\sigma_\xi \sigma_\zeta}{\lambda \lambda}$ for the search area Q in the form of a sphere with a radius of 2 m and making a decision on the absolute minimum.

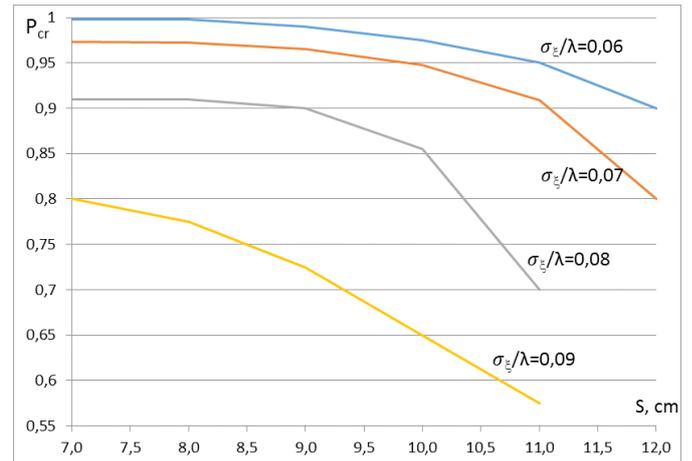


Fig. 5. The probability of correct resolution of ambiguity as a function of the search step S for different values of $\frac{\sigma_\xi \sigma_\zeta}{\lambda \lambda}$. The search area is a sphere with a radius of 2 m

Table 1. Computing costs for the implementation of the search algorithm

1 $S, \text{ cm}$	2 Maximum number of steps	3 Number of operations		4 Calculation time, ms 1 GHz processor	
		brute force enumeration	optimized enumeration	without using SSE	using SSE2
7	185192	123523064	55554600	111	28
8	132650	88477550	39795000	80	20
9	91124	60779708	27337200	54	14
10	68920	45969640	20676000	40	10
11	50652	33784884	15195600	32	8
12	35936	23969312	10780800	21	5.5

In Table 1, for the same conditions, the maximum number of search steps, the number of operations (addition, multiplication) for direct and optimized searches, and the calculation time on a 1 GHz processor for optimized enumeration (with and without using the SSE technology) are specified for the same conditions.

In this case, the optimization consisted in performing the full search procedure in two steps: first for measurements by GPS satellites, and then (in a reduced area) for all measurements, including GLONASS.

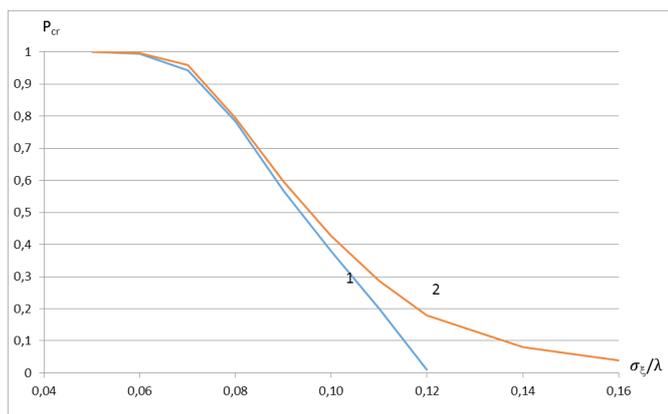


Fig. 6. Comparison of the probability of correct resolution of ambiguity when deciding on the threshold $\alpha = 1.5$ cm (curve 1) and the absolute minimum (curve 2).

The graphs in Fig. 6 illustrate differences in the values of P_{cr} when making a decision on the threshold and on the absolute minimum, depending on $\frac{\sigma_{\epsilon}}{\lambda}$. The graphs are plotted for the search area Q in the form of a sphere with a radius of 20 cm and $S = 7$ cm. The relative small search area is characteristic for the tracking mode, when it is possible to predict the change in the relative position of objects from the results of previous measurements.

The area and strategy of the search depend on the quality of the initial measurements, the dynamics of the objects and the requirements for the results. The choice and optimization of them are beyond the scope of this article.

References

- Zhodzishskiy A.I., Berezentsev M.M., Nesterov O.V. Perspektivy ispol'zovaniya v Rossii global'nykh navigatsionnykh sputnikovykh sistem grazhdanskimi potrebitelyami [Current Issues of GNSS Civilian Usage in Russia]. Raketno-kosmicheskoe priborostroenie i informatsionnye sistemy [Rocket-Space Device Engineering and Information Systems]. 2016, Vol. 3, No. 4. pp. 5–15. (in Russian)
- Antonovich K.M. *Ispol'zovanie sputnikovykh navigatsionnykh sistem v geodezii* [Use of satellite navigation systems in geodesy]. Vol. 2. Sibirskaya gosudarstvennaya geodezicheskaya akademiya, Moscow, Kart-geotsentr, 2006. (in Russian)
- Hatch R, Euler H-J Comparison of several AROF kinematic techniques. *Proceedings of ION GPS-94, 7th International Technical Meeting of the Satellite Division of the Institute of Navigation*, Salt Lake City, Utah, September 20–23, 1994, part 1: pp. 363–370.
- Hofmann-Wellenhof, Lichtenegger, Wasle. *GPS, Galileo, Glonass*. 2008.
- Chen D. Development of a fast ambiguity search filtering (FASF) method for GPS carrier phase ambiguity resolution. *Reports of the Department of Geomatics Engineering of the University of Calgary*, 1994, vol 20071 (94.20071.dschen.pdf).
- Frei E. GPS – fast ambiguity resolution approach “FARA”: theory and application. *Paper presented at XX General Assembly of the IUGG, IAG-Symposium GM 1/4*, Vienna, August 11–24, 1991.
- Frei E., Schubernigg M. GPS surveying techniques using the “fast ambiguity resolution approach (FARA)”. *Paper presented at the 34th Australian Surveyors Congress and the 18th National Surveying Conference* at Cairns, Australia, May 23–29, 1992.
- Teunissen P.J.G., Jonkman N.F., Tiberius C.C.J.M. Weighting GPS dual frequency observations: bearing the cross of cross-correlation. *GPS Solutions*, 2(2): 28–37, 1998.
- Hatch R. Instantaneous ambiguity resolution. In: Schwarz K.P., Lachapelle G. (eds): *Kinematic systems in geodesy, surveying, and remote sensing*. Springer, New York Berlin Heidelberg Tokyo: 299–308, 1990. [Mueller II (ed): IAG Symposia Proceedings, vol 107].
- Leick A. *GPS Satellite Surveying*, 3rd edition, 2004.
- Povalyaev A.A. *Sputnikovye radionavigatsionnye sistemy: vremya, pokazaniya chasov, formirovanie izmereniy i opredelenie koordinat* [Satellite radio navigation systems: time, hours, the formation

- of measurements and the determination of coordinates]. Moscow, Radiotekhnika, 2008, 324 p.
12. Zhodzishskiy A.I., Nesterov O.V. *Sposob opredeleniya vzaimnogo polozheniya ob"ektov po signalam global'nykh navigatsionnykh sputnikovykh sistem* [The method for determining the mutual position of objects by signals of global navigation satellite systems]. Patent for Invention № 2624268. JSC "Russian Space Systems" (Application № 2016135147 from 30.08.16.). (in Russian)
 13. GLONASS. *Interfeysnyy kontrol'nyy dokument. Navigatsionnyy radiosignal v diapazonakh L1 i L2* [GLONASS. Interface control document. Navigational radio signal in the L1 and L2 bands.]. Ed. 5.1, RNII KP, Moscow, 2008. (in Russian)

Promising Quantum-Optical Technologies for Satellite Navigation Challenges

N.N. Kolachevsky, *Corresponding Member of the Russian Academy of Sciences,*

Dr. Sci. (Phys.-Math.), kolachevsky@lebedev.ru

P.N. Lebedev Physical Institute of Russian Academy of Sciences, Moscow 119991, Russian Federation

K.Yu. Khabarova, *Cand. Sci. (Phys.-Math.), kseniakhabarova@gmail.com*

P.N. Lebedev Physical Institute of Russian Academy of Sciences, Moscow 119991, Russian Federation

I.V. Zalivako, *zalikes@yandex.ru*

P.N. Lebedev Physical Institute of Russian Academy of Sciences, Moscow 119991, Russian Federation

I.A. Semerikov, *ilia179@mail.ru*

P.N. Lebedev Physical Institute of Russian Academy of Sciences, Moscow 119991, Russian Federation

A.S. Borisenko, *alexander.borisenko@yandex.ru*

P.N. Lebedev Physical Institute of Russian Academy of Sciences, Moscow 119991, Russian Federation

I.V. Sherstov, *i.sherstov@skoltech.ru*

Skolkovo Institute of Science and Technology, Moscow 121205, Russian Federation

S.N. Bagaev, *Academician of the Russian Academy of Sciences, Dr. Sci. (Phys.-Math.), bagayev@laser.nsc.ru*

Institute of Laser Physics, Russian Academy of Sciences, Novosibirsk 630090, Russian Federation

A.A. Lugovoy, *lugovoy@laser.nsc.ru*

Institute of Laser Physics, Russian Academy of Sciences, Novosibirsk 630090, Russian Federation

O.N. Prudnikov, *Dr. Sci. (Phys.-Math.), oleg.nsu@gmail.com*

Institute of Laser Physics, Russian Academy of Sciences, Novosibirsk 630090, Russian Federation

A.V. Taichenachev, *Corresponding Member of the Russian Academy of Sciences,*

Dr. Sci. (Phys.-Math.), taich.alex@gmail.com

Institute of Laser Physics, Russian Academy of Sciences, Novosibirsk 630090, Russian Federation

S.V. Chepurov, *Cand. Sci. (Phys.-Math.), svc04@ngs.ru*

Institute of Laser Physics, Russian Academy of Sciences, Novosibirsk 630090, Russian Federation

Abstract. Accuracy of navigation and positioning provided by signals of global navigation satellite systems is largely determined by the frequency standards installed on board of satellites. In recent years, there has been a rapid development of new quantum-optical technologies using compact and frequency-stable laser systems, femtosecond frequency generators, ultracold atoms and ions. Optical methods of information reading off and processing provided by an atomic system were developed. This resulted in significant reduction of a relative frequency instability of ground-based frequency standards down to 18th decimal digit. A number of successful European suborbital launches demonstrated a possibility of application of some of the technologies in the space segment. The paper provides a brief overview of the latest achievements in this field and possibility of its development in Russia.

Keywords: laser cooling, ultrastable atoms and ions, ion Paul trap, femtosecond synthesizers of optical frequencies, stabilized laser, optical clock

1. Introduction

Global navigation satellite systems (GNSS) have become an integral part of our life: it is difficult to imagine sea and air navigation, traffic, and cargo transportation without them. The demand in exact positioning and synchronization is constantly growing, covering agriculture, forest and mining industry, data transmission, and fundamental science. The main objective of development of any national GNSS system is the increase in accuracy and reliability of positioning of the consumer.

One of the key factors defining the quality of functioning of GNSS is the accuracy and stability of a time signal formed by the onboard synchronizer (OS). Thus, at a daily relative error of frequency of the onboard standard in 1×10^{-14} , the corresponding error of positioning along the direction on the satellite will make about 0.25 m. There is, however, a set of the factors influencing the characteristics of the signal delivered to the consumer [1] and leading to decrease in accuracy. To achieve target characteristics and obtain in the long term decimeter accuracy, within the Federal Target Program "GLONASS of 2012–2020", a wide complex of problems touching all segments of the system including the onboard frequency standard (OFS) is being solved. Today OFS of the GLONASS system is based on microwave standards using a bunch of atoms of caesium and on a rubidium cell.

As it is noted in the report [2], one of the important problems of development of the GLONASS system is decrease in daily relative instability of OFS to 5×10^{-15} by 2020, and in the long term up to 1×10^{-15} . Achievement of these indicators employing the existing caesium or rubidium standards is actually impossible in view of a number of fundamental restrictions.

In the short term as a part of OFS, passive hydrogen mazer tests are planned [3]. Passive hydrogen mazers with a daily instability of frequency $< 1 \times 10^{-14}$ have proved themselves onboard GALILEO satellites [4] and in ground tests in Russia [5]. In its turn, an active hydrogen mazer has been successfully functioning more than 6 years onboard the Spektr-R spacecraft [6]. Considering a significant progress in developments of hydrogen mazers including for space applications (thus, within the Millimetron project, active hydrogen standards with instability $< 1 \times 10^{-15}$ are created), their use when developing promising navigation satellite systems is a relevant task.

Other important direction is creation of the onboard optical frequency reference (OOFR). Transition from a microwave ($f = 10^{10} - 10^{11}$ Hz) to optical ($f = 10^{14} - 10^{15}$ Hz) range of frequencies leads to an essential, on several orders, increase in relative stability of the oscillator, since it is defined by its *Q-factor*: $Q=f/\delta f$ (f is the transition frequency, and δf is the resonance width). Atoms in the optical range have a number of metrological (clock) transitions with a natural spectral width much less than 1 Hz. Methods of laser cooling [7] and capture of atoms [8] and ions into traps [9] allow one to solve two important problems. First, interaction time with the exciting field can be increased up to several seconds that provides the Fourier-limited spectral width of resonance up to $\delta f = 1$ Hz without increase in physical sizes of a system. Secondly, due to localization of a cold ion on the sizes much less than the length of light, linear Doppler effect and effect of return are nullified and also actually a full isolation from undesirable external fields and collisions is provided.

Today a relative instability and an error of frequency of optical standards of frequency at the level of units of the 18th sign after a comma both on the neutral atoms taken into optical lattices [10] and on single ions is demonstrated [11]. We will note that such values are shown only when checking standards directly in the optical range in extremely stable external laboratory conditions. To illustrate the last, it is enough to estimate the Doppler effect, which arises at extension, for example, of a metal platform with a characteristic size of 1 m, on which optical elements are placed. If change of temperature makes only 1 degree per hour, then the contribution of the Doppler effect when reflecting from one mirror will be up to 4×10^{-17} . It is also necessary to consider gravitational red shift of frequency ($1 \times 10^{-16}/\text{m}$) and a set of other systematic effects having a significant effect at such level of accuracy [12]. It is obvious that reduction of the size of a system and toughening of service conditions leads to decrease in characteristics. Nevertheless, for the transported sample about 1 m³, a daily instability of frequency at the level of 10^{-17} – 10^{-16} is quite achievable and is already shown by a number of laboratories [13, 14]. We will note that this indicator approximately much surpasses the instability of the best samples of active hydrogen mazers and commercially available microwave clocks of a fountain type [15].

Of course, there is a question of a possibility to transfer perspective technologies of photonics on board

the spacecraft. It is interesting to note that if about ten years ago such ideas belonged to the section of “science fiction”, then today a number of the successful launches, which have shown functionality, at least, regarding quantum and optical technologies including the femtosecond synthesizers of optical frequencies (FSOF) [16], stabilized lasers [17], and also the systems for a deep laser cooling and Bose condensation of rubidium atoms onboard the spacecraft is already carried out [18]. In fact, the essential difficulty arising during creation of any optical standard of frequency is a large number (usually about 10) various laser systems, which frequency of radiation has to be tuned on the line of transitions in an atom. It is required as for cooling and capture of atoms and ions, and for control of internal quantum states. Under such conditions, a spectral width of radiation of laser systems has to be 0.01–1 MHz with a similar accuracy of tuning. Support of operability of lasers is a difficult task; the number of failures often does not ensure a reliable functioning of optical frequency standards even on an interval of several hours.

Emergence of erbium and ytterbium fiber-optical lasers with diode pumping (continuous and femtosecond) [19, 20] and also a wide line of semiconductor lasers (Fabry–Pérot, with the distributed feedback, on quantum holes, quantum and cascade, etc.) blocking a very wide spectral range from infrared to ultraviolet has significantly increased compactness and reliability of laser systems. New reliable schemes of stabilization of frequency of semiconductor lasers on the external resonator [21] are developed that allows using them in the conditions of strong external indignations without change of the wavelength of generation. There was a break in methods of stabilization of frequency of the so-called “clock lasers” interrogating metrological transition in the laser. Compact, reliable systems based on the external ultrastable Fabry–Pérot interferometer providing relative instability of frequency of the laser at the level of 10^{-16} – 10^{-15} per 1 sec (that corresponds to the subhertz spectral width of the line) are created [22, 23].

One of the fundamental factors, which have provided real horizons of applying optical clocks as OOFR was creation of FSOF based on the fiber-optical femtosecond laser with a passive mode synchronization [24]. In the papers [25, 26] it has been shown that FSOF allows one to transform optical frequency to radio-frequency range making only an insignificant contribution to a relative instability of frequency at the level of units of the 19th

sign. An opportunity to use high characteristics of stability of a clock laser connected on the frequency to an optical resonance in an atom, in the range available to the consumer (1–10 GHz) has opened. Progress of the production technology of FSOF in Europe (Menlo Systems company) enabled one to carry out two successful suborbital launches with a compact (22 kg) FSOF onboard in 2015 and 2017 [16]. It is possible to consider that the issue of an onboard FSOF is solved at the basic level and further only efforts to increase the reliability and compactness of the system are required. In Russia, similar works are conducted by the Avesta company.

The last question, which it would be desirable to mention in Introduction, belongs to the choice of atomic or ionic system for a perspective OS. In spite of the fact that optical standards on neutral atoms in optical lattices show several somehow best characteristics of stability due to a big (up to 10^5) amount of the interrogated atoms [10] in comparison with standards on single ions [11], the last are preferable. First, the depth of potential of an ionic trap is several eV (several tens of thousands kelvin) that allows one to hold long, up to months, a single ion in a trap [27]. The main mechanism of losses are collisions with background gas in the vacuum chamber, leading to a recharge and loss of an ion. Secondly, the design of an ionic trap is much more compact and does not demand delicate adjustments of optical bunches. In the third, a lower stability of ionic clock on short times of averaging (in comparison with the clocks on neutral atoms) is not considered a restriction at designing of OS for the navigation satellite. Disadvantages are a high sensitivity to electric fields (delicate control of potentials on electrodes is required), a rather low level of a signal of a luminescence from a single ion and inaccessibility of ultraviolet transitions in some ions [28, 29].

The essential motivating factor is creation of transported optical clocks on an ion Ca^+ of 0.5 m^3 showing a relative instability in 10^{-16} per one day [14]. Moreover, a cooperation of the German institutes and companies (PTB, Toptica, Menlo Systems) has begun development of the transported clocks on a single ion of Yb^+ [30]. The similar project has been started in France [31].

In 2017, the Ministry of Education and Science of the Russian Federation supported the project 14.610.21.0010 “Development of the generator of ultrastable reference signals of frequency on cold ions of ytterbium to increase by times the accuracy of geopositioning, space navigation, and formation of new segments of mass demand in the

frequency signal of the frequency standard. A short-term stability of such standard at the same time is defined by a highly stable resonator stabilizing SICT or FSOE, and a long-term one is determined by the frequency of clock transition of an ion which relative instability can reach units of the 17th sign.

2.1. Paul trap

The fact that ions have an electric charge, other than zero, considerably facilitates their capture and localization. Interaction with an electric field permits one to hold ions by means of radio-frequency fields in so-called Paul traps [9]. The trap represents a combination of electrodes to which constant and radio-frequency potentials are attached (with the frequency ranging from 1 to 100 MHz). These electrodes create variable non-uniform potential close to quadrypolar. Thus, in three-dimensional Paul traps, which are usually in the foundation of optical ionic standards of frequency, the configuration of electrodes has an axial symmetry and the potential created by them in the center of a trap is close to the one described by a formula (1):

$$\Phi(r, z, t) = \frac{U_{dc} + V_{ac} \cos(\omega t)}{r_0^2 + 2z_0^2} (r^2 - 2z^2). \quad (1)$$

where U_{dc} is the constant component of the potential; V_{ac} is the amplitude of the variable component; ω is the circular frequency of the variable potential; r_0 and z_0 are the typical dimensional parameters of the potential determined by the certain geometry of electrodes. Fig. 2 depicts the configuration of electrodes, which enables one to provide the potential close to the one set by a formula (1), and at the same time providing good optical access to the center of a trap that is important for implementation of an effective laser cooling, manipulation of a quantum condition of an ion, and also reading off its state is presented. Here, a variable potential with an amplitude $V_{ac} \approx 250$ V is put to two electrodes-edges located along a trap axis, and hollow cylindrical electrodes surrounding them are grounded.

The movement of a charged particle in the potential set by a formula (1) is described by Mathieu equations. The analysis of these equations of the movement shows that at certain values of amplitude V_{ac} and frequency of tension ω on the electrodes of the trap, the ion keeps close to the center of the trap [32]. Its movement time can be presented as a superposition of a rather slow (secular)

movement of an ion in a harmonious pseudopotential close to the point where an amplitude of fluctuations of an electric field becomes zero and fast small fluctuations of an ion with the frequency of the field of a trap, which are called a micromovement. Since the micromovement is the necessary harmonic oscillations of a particle in the field of a trap, its amplitude is proportional to the amplitude of fluctuations of the holding field in this point of a trajectory of the secular movement of an ion.

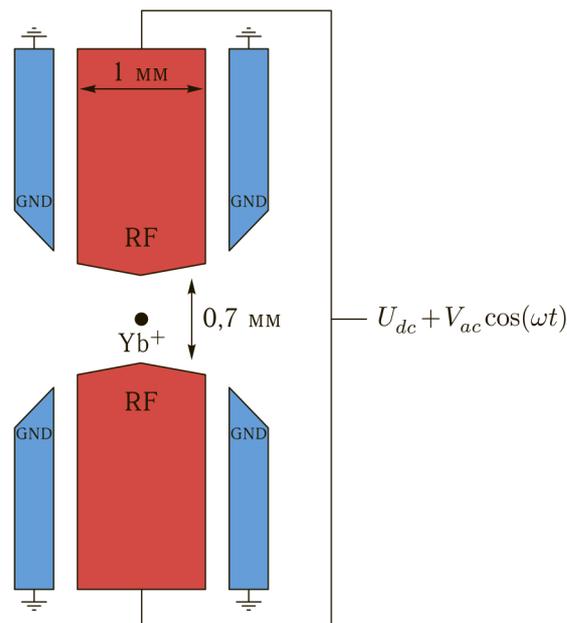


Fig. 2. A sketch of a 3D Paul trap. A variable potential is put to the electrodes-edges defined by RF, and cylindrical electrodes (GND) located concentrically around radio-frequency are grounded.

Important features of ionic traps are a big depth of a potential hole (up to several eV) that allows one to take into them ions even at the room temperature without preliminary cooling and also the fact that an ion is localized close to the point where the amplitude of the holding field becomes zero. The last leads to the fact that the frequency of clock transition of a particle is poorly subject to shift because of the Stark effect and also to minimization of shift because of the Doppler effect of the second order connected with the existence of the micromovement. The size of the area of localization of single ions in such traps in the presence of laser cooling is several tens of nanometers that leads to realization of the Lamb Dicke regime [33], this means to a full suppression of the Doppler effect of the first order. In addition, the effect of return, which in case of free atoms leads to splitting and asymmetry of the line of clock transition, is suppressed.

It is important to note that an effective suppression of the Stark and Doppler shifts of the second order is reached only in case the potential of a trap is close to the one described by a formula (1). If in the field of capture, for example, there are parasitic static electric fields, then the minimum of pseudopotential will be shifted to the area with a nonzero amplitude of an electric field that will lead to a considerable strengthening of the influence of the effects described above on the frequency of clock transition. For this reason, the design of traps also usually provides the existence of several compensating electrodes, which allow one to eliminate parasitic fields. Methods of search of optimum parameters of the compensating fields are in detail described, for example, in [34].

Loading of an ion in a trap can be carried out by means of shock ionization of atoms of ytterbium by an electron beam or by photoionization directly in the field of the capture of a trap. For this purpose, for several seconds the atomic oven in the form of the tube filled with metal ytterbium and supplied with the electric heater, which creates an atomic bunch passing through the center of a trap and also an ionization source (a laser of photoionization or an electron beam) is turned on. Photoionization is more preferable, since it provides loading only of ions of this element and isotope and also as unlike an electron beam, its using does not lead to emergence of parasitic charges near a trap.

Time of life of ions in a trap due to the high level of a vacuum ($<10^{-10}$ mbar) and a deep depth of potential is usually several days. At the same time, the main channel of losses is exchange of an ytterbium ion charge with atoms of background gas at collisions. In case of loss of an ion that can be revealed on lack of fluorescence of a particle under the influence of the cooling radiation during several cycles of cooling, the ion is repeatedly loaded into a trap that takes about a second and has no significant effect on characteristics of stability.

2.2. Laser system

The OOFR major element are the laser sources providing radiation for cooling of an ion, photoionization, manipulation of a quantum condition of a particle and initiation of clock transition. These sources have to be constantly tuned to the corresponding lines of transitions in an ion, stable and compact. Today the most stable and reliable sources are fiber lasers with the distributed feedback [35]. Fiber lasers generate only in the narrow

range of lengths of waves (1530–1560 nanometers for the lasers alloyed by ions of erbium [37]; 1030-1050 nanometers for the lasers alloyed by ions of ytterbium. Semiconductor lasers with the distributed feedback [38] (for example, made by Polyus, 780 nanometers and 850 nanometers) and with vertical resonators [39] also have a high stability of frequency. Unfortunately, the majority of the listed types of lasers generate in the ranges of lengths of the waves (taking into account the second harmonica) differing from necessary for quantum manipulations with an ytterbium ion in a trap (369 nanometers, 398 nanometers, 739 nanometers, 760 nanometers, 871 nanometers, and 935 nanometers).

Perspective laser sources including for use in space applications are semiconductor lasers with an external resonator. They have a wide range of reorganization of a wavelength, high stability of frequency, and narrowness of a range of radiation. Their compactness, simplicity of a design and operation have made them the main laser sources in spectroscopic laboratories in the world today, and laser diodes available today block a considerable part of the near infrared, visible and near ultraviolet area of a range. The design of commercially available diode lasers has high reliability [40].

The majority of the semiconductor lasers used today with the external resonator are manufactured either according to the Littrov scheme [41], or according to the Littman scheme [42]. In both schemes, the resonator of the laser is formed by the back reflecting surface of the diode and the reflecting element, and the discrimination of lengths of waves providing single-frequency generation by a diffraction lattice. At the same time, reorganization of the laser is carried out by the change of the length of the resonator and the angle of diffraction of a lattice. The main lack of these schemes is a high sensitivity of the laser to an adjustment of the reflecting element.

It is more preferable to use in OOFR the scheme offered in [43] and represented in Fig. 3, where the reflector of the cat's eye type is used, and the interferential filter is employed for discrimination of lengths of waves. A peculiarity of a reflector of the cat's eye type which is implemented by means of a lens and a mirror located in its focal plane is reflection of a bunch precisely in the opposite direction. It reduces sensitivity of the scheme to vibrations and changes of temperature. Tuning of frequency to an atomic resonance is carried out by change of current of the laser, temperature, and turn of the interferential filter by means of a piezoactuator.

A pilot study of lasers of this kind confirms their high stability. Power of radiation is 5–50 MW depending on the wavelength, and the spectral width of the line is about 1 MHz (depends on a resonator length).

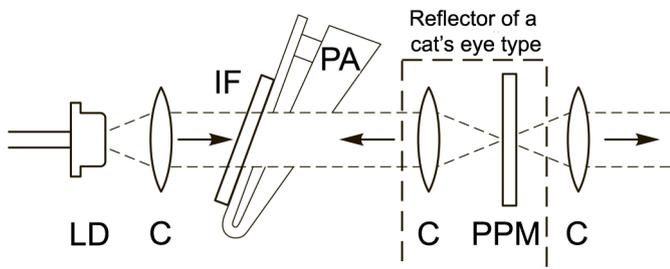


Fig. 3. The scheme of the semiconductor laser with the external resonator based on the interferential filter (IF) and the reflector of a cat's eye type. Here LD is the laser diode, C is the collimator, PA is the piezoactuator, PPM is the partially passing mirror.

After passing of a number of optical elements (insulators, optical-acoustic modulators, polarizing optics), radiation is brought to optical fiber and goes to VOS. Such scheme reduces sensitivity of a design to adjustments and vibrations.

in a metastable $^2D_{3/2}$ state and drop out of a cooling cycle. To return the studied ion to a cycle of cooling, the excitement into $^3[3/2]_{1/2}$ - state is made. The laser of 760 nanometers serves for hashing $^2F_{7/2}$ - and $^1[3/2]_{1/2}$ - of levels. This communication is necessary for a fast return of an ion to an initial state in case of transition to metastable level $^2F_{7/2}$ -. In view of existence of superthin splitting, modulation of frequency of radiation, which is implemented based on fiber electro-optical modulators (FEOM), is necessary. Stabilization of frequency of CSQS lasers can be made both by means of FSOE, and by means of MIW.

SPhDC consists of three lasers: two with the wavelength of 369.5 nanometers and one with the wavelength of 398.9 nanometers. Two lasers of 369.5 nanometers are used for Doppler cooling and preparation of an initial condition of the studied ion. The frequency of these lasers differs by 14.75 GHz for an effective excitement two superthin components of the main state. The laser of 398.9 nanometers is used for ytterbium photoionization. All lasers in the SPhDC system are equipped with the fiber optical-acoustic modulators (FOAM) for a possibility of a fast inclusion and switching off radiation. The frequency of SPhDC lasers is stabilized by means of MIW or FSOE.

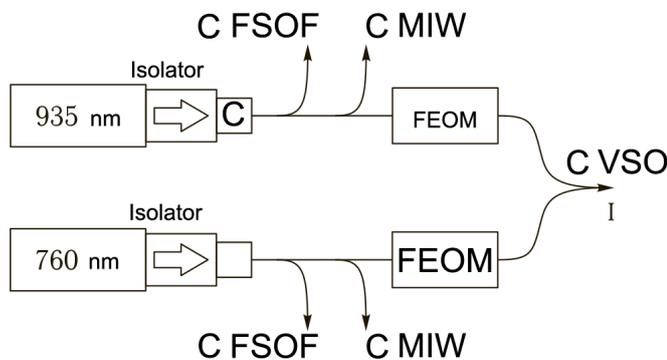


Fig. 4. The laser control system of quantum states (CSQS) based on diode lasers with the external resonator (Fig. 3). C is the collimator; FEOM is the fiber electro-optical modulator; FSOF is the femtosecond generator of optical frequencies; continuous thick lines are the optical fibers; MIW is the measuring instrument of wavelengths based on the Fizeau interferometer.

CSQS includes two lasers with the lengths of waves of 935 nanometers and 760 nanometers (Fig. 4). The laser of 935 nanometers serves for hashing of density of population $^2D_{3/2}$ - and $^3[3/2]_{1/2}$ - of levels. When applying of the cooling radiation, an ion can break up

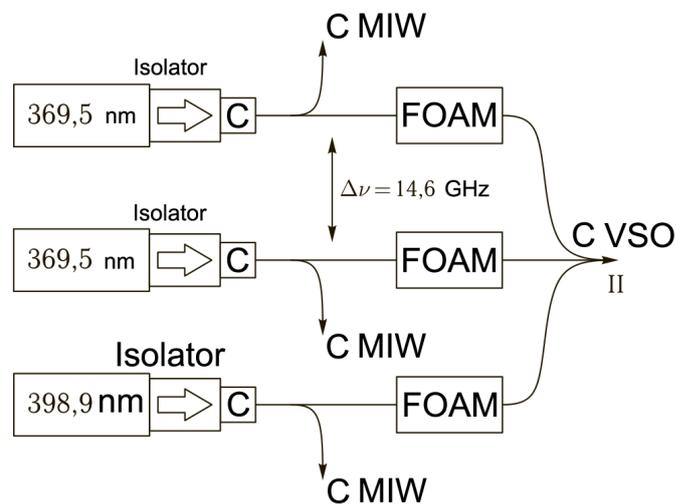


Fig. 5. A laser system of photoionization and Doppler cooling (SPhDC). Radiation is carried out by means of diode lasers with the external resonator in the scheme with the interferential filter (Fig. 3). C is the collimator; FOAM is the fiber optical-acoustic modulator; continuous thick lines are the optical fibers; MIW is the measuring instrument of wavelengths based on the Fizeau interferometer.

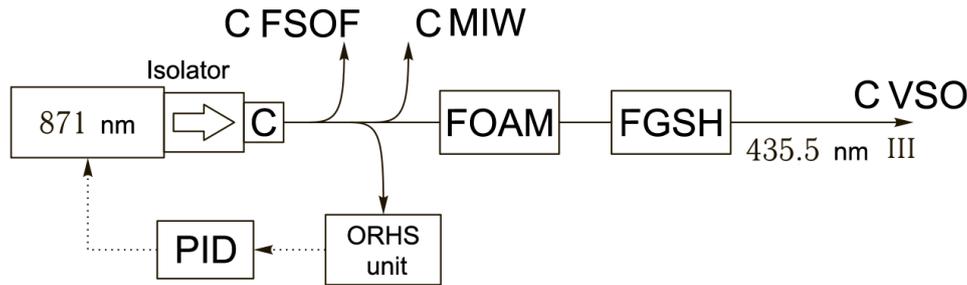


Fig. 6. A system of interrogation of clock transition (SICT). Radiation is performed by means of a diode laser with the external resonator (Fig. 3). C is the collimator; FOAM is the fiber optical-acoustic modulator; FGSH is the fiber generator of the second harmonica; ORHS is the optical resonator of high stability; PID is the proportional-integrated-differential controller; continuous thick lines are the optical fibers; dashed lines are the electric signals; MIW is the measuring instrument of wavelengths of based on the Fizeau interferometer.

In the system of interrogation of clock transition (SICT, Fig. 6), the radiation of the diode laser on the wavelength of 871 nanometer passes, at first, through the fiber optical-acoustic modulator for a possibility of reorganization of the frequency with a 100 MHz band, and then through the fiber generator of the second harmonica based on a periodically polarized nonlinear crystal [44] where the frequency of laser radiation doubles. Then on optical fiber radiation is brought to VOS. A part of radiation on the wavelength of 871 nanometer is split off for calibration of a wavemeter and stabilization of the frequency of repetition of FSOE.

2.3. Clock laser and reading off information from a single ion

For stabilization of the SICT laser on short times (1–10 sec), the ultrastable optical Fabry–Pérot interferometer, which consists of the resonator body (RB) and two mirrors with a dielectric covering fixed at end faces of a working body by the method of optical contact, is used. Mirrors and a body of the resonator are made of the glass of a special grade of with a zero coefficient of thermal expansion (Ultra Low Expansion – ULE). The optical resonator is placed in the vacuum chamber for reduction of the influence of fluctuations of the index of refraction and ambient temperature.

To stabilize the frequency of the laser relative to own frequency of the resonator, the Pound–Drever–Hall (PDH) technique is used [45]. The schematic diagram of such binding is represented in Fig. 7. Radiation of the stabilized laser on fiber is delivered to the electron-optical modulator and is modulated in phase then is reflected from ORHS and the reflected signal is detected by the

photo diode. At the same time, the complex coefficient of reflection of radiation from the resonator depends on frequency. The signal from a photo diode collides with a modulation signal. The signal of beats gains dependence on a laser frequency, where at the coincidence of the frequency of the laser to ORHS resonance frequency the signal is equal to zero.

It should be noted that by means of such equipment, it is possible to stabilize the frequency of both diode and fiber lasers [22, 46]. In case of stabilization of the frequency of the fiber laser, after it one usually installs an optical-acoustic modulator, and a servo-signal for compensation of the noise of a phase of the laser goes to it as the frequency of such lasers unlike semiconductor ones cannot be quickly modulated by means of pump current. The target relative instability of frequency of SICT is $2 \cdot 10^{-15}$ (1 sec) that it is enough for a stable interrogation of an ion and is confirmed by a series of experiments [47, 48].

To read off information on initiation of clock transition in ions, the method of quantum jumps (Fig. 8) is usually used. It is that after a cycle of initiation of clock transition (Fig. 8, I), an ion is irradiated with the cooling radiation on the wavelength of 369.5 nanometers. At this moment, registration of a signal of fluorescence of an ion by means of the photoelectronic multiplier begins. If during initiation of clock transition the ion has passed into a state $^2D_{3/2}$, then an ion will not reradiate the photons of 369.5 nanometers (Fig. 8, IIa). If initiation of clock transition is not happened, the signal of fluorescence of an ion will be observed (Fig. 8, IIb). This scheme allows one to carry out stabilization of frequency of the interviewing laser of clock transition on transition frequency in an ion.

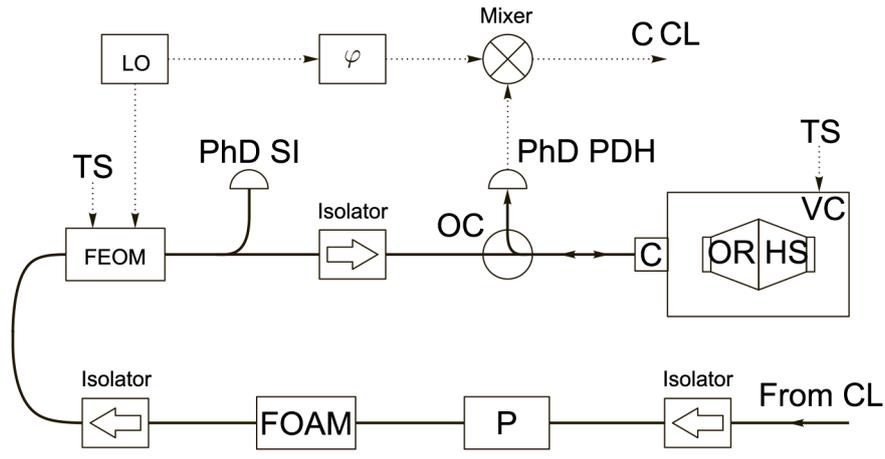


Fig. 7. The knot of the optical resonator (OR) of high stability. CL is the clock laser; P is the fiber polarizer; FOAM is the fiber optical-acoustic modulator; FEOM is the fiber electro-optical modulator; TS is the loop of temperature stabilization; PhD SI is the photo diode of intensity stabilization; OC is the fiber optical circulator; PhD PDH is the photo diode in the scheme of the Pound–Drever–Hall (PDH); C is the collimator; ORHS is the optical resonator of high stability; VC is the vacuum chamber; LO is the local oscillator; φ is the phase shifter; continuous thick lines are the optical fibers.

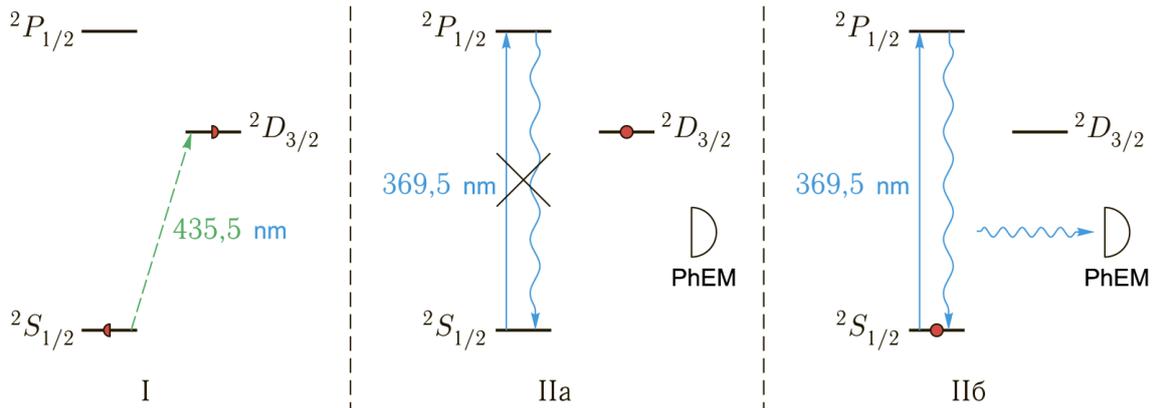


Fig. 8. Reading off quantum information from an ion. I is the initiation of clock transition; IIa is the lack of fluorescence upon transition of an ion to a state 2D3/2; IIб is the detecting of fluorescence in case of not initiation of clock transition. PhEM is the photoelectronic multiplier.

2.4. Femtosecond synthesizer of optical frequencies

The most important element of the optical standard of frequency is FSOF. Clock transition in an ion and mode of the ultrastable resonator regarding which the SICT laser is stabilized from stability of the standard in the optical range of frequencies. At the same time, practical applications demand a highly stable and exact reference point of frequency in the radio-frequency range. FSOF carries out transfer of accuracy and stability of the standard from optical range in radio-frequency.

To create of OOFR based on ions of ytterbium, it is supposed to use FSOF based on the femtosecond fiber laser with a passive synchronization of modes with the supercontinuum generator. The range of such laser source stretches within 1000–2000 nanometers. A feature of such system is that the range consists of a set of equidistant longitudinal laser modes, which frequencies are set by the formula $f_n = f_{\text{CEO}} + n \times f_{\text{REP}}$ where f_{CEO} is the frequency of a shift of a wave package relative to the carrier, f_{REP} is the frequency of repetition of impulses of the laser (f_{CEO} and f_{REP} are usually in the range of 50–250 MHz), n is the number of a mode [49]. In the range of wavelengths

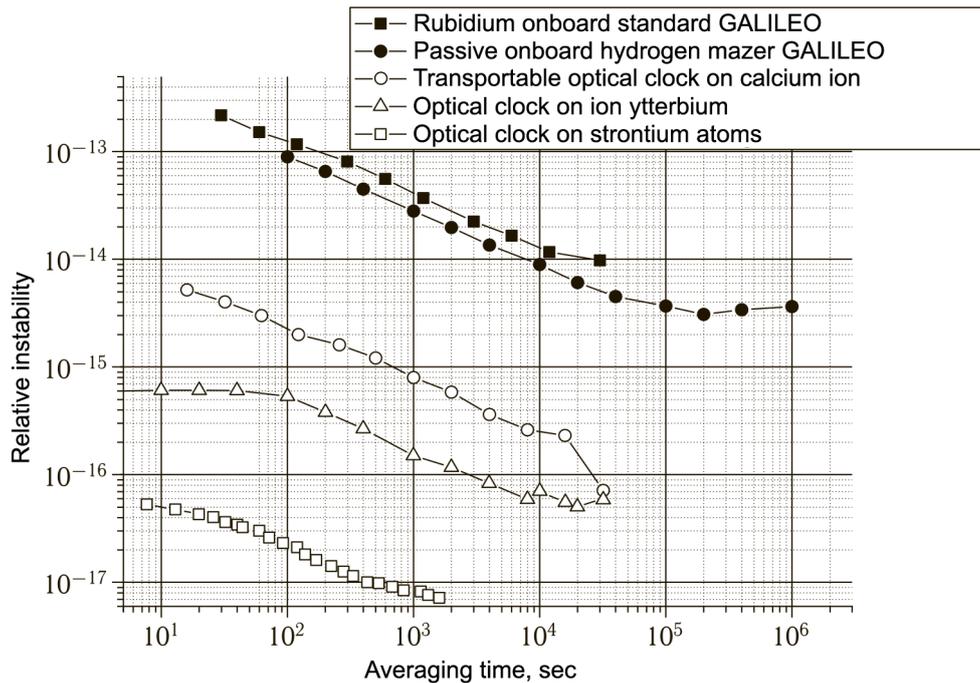


Fig. 9. Relative instability (the Allan variance) of various atomic standards of frequency depending on averaging time: the rubidium onboard standard used on the GALILEO [56] satellites, a passive onboard hydrogen maser for the GALILEO [56] satellites, the transported optical clock on a single ion 40Ca^+ [14], optical clock on a single ion 171Yb^+ [59], optical clock based on the atoms 87Sr in an optical lattice [10].

less than 1000 nanometers, transformation to the second harmonica with the use of nonlinear crystals of PPLN (periodically poled lithium niobate crystal) are usually applied to employ FSO.

Thus, FSO carries out transformation of frequencies from radio-frequency range in optical and vice versa without loss of characteristics of stability [50]. The onboard realization of FSO was performed by the Menlo-Systems company and successfully tested onboard the suborbital device in 2016 [16]. In Russia, the Avesta company is engaged in production of fiber FSO; FSO laboratory systems are also created in LPI RAS and Institute of Laser Physics, Siberian Branch of the Russian Academy of Sciences.

Creation of FSO with f_{CEO} frequency (the no offset scheme, $f_{\text{CEO}} = 0$) expelled from structure of a range is perspective. It is reached by nonlinear process of the generation of differential frequencies (GDF) between two sites of initial FSO [51–53]. Such scheme facilitates stabilization of FSO relative to optical reference points of frequency and transfer of their stability and accuracy in the radio frequency range [54].

During creation of ytterbium OOF, it is planned to use a fiber FSO that by means of nonlinear transformations provides output radiation on the wavelengths near 871 nanometers (the SICT oscillator frequency) and 1560 nanometers. Stabilization of FSO on short times will be carried out by means of a highly stable fiber laser radiating on the wavelength of 1560 nanometers. A long-term stability of FSO is provided by binding to the frequency of clock transition in an ion by means of SICT. FSO will possess a radio-frequency output of 1 GHz.

3. Conclusion

The revolutionary changes, which have happened in the last decade in the field of synthesis of highly stable signals of frequency, have led to an unconditional priority role of quantum and optical technologies on the horizon of the next 10 years both in land and in onboard systems [55]. Today optical clock surpass caesium fountains in indicators concerning instability and an error by order of magnitude (Fig. 9) and continue to evolve towards the increase in compactness, reliability, and improvement of characteristics.

Undoubtedly, creation of a prototype of a reliable and compact OOFR device integrating the listed technologies and modules is the most difficult scientific and technical task. Apart from a scientific component, its realization in onboard option requires development of a line of technologies of photonics (solid-state radiators, fiber-optical components, electron-optical elements of technology of drawing coverings). The submitted project opens an opportunity to integrate the existing scientific and technological reserve and on its basis to make the objective roadmap of development of this direction in Russia.

4. Acknowledgements

The paper is written with assistance of a grant MOH 2017-14-582-0001.

References

- Hofmann-Wellenhof B., Lichtenegger H., Wasle E. GNSS – Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and more. Springer–Wien–NewYork, 2008, pp. 109–138.
- Available at: <http://gpsworld.com/directions-2017-new-glonass-capabilities-for-users/>
- Medvedev S., Belyaev A., Myasnikov A., Pavlenko Yu., Sakharov B., Smirnov P., Storozhev E., Tyulyakov A. Passivnyy vodorodnyy mazer - chasy budushchego pokoleniya sputnikov GLONASS [Passive hydrogen maser-clock of the future generation of GLONASS satellites]. METROLOGIYA VREMENI I PROSTRANSTVA, Doklady 6-go Mezhdunarodnogo simpoziuma, 2013 [METROLOGY OF TIME AND SPACE, Proceedings of the 6th International Symposium], pp. 279–282. (in Russian)
- Steigenberger P., Montenbruck O. Galileo status: orbits, clocks, and positioning. GPS Solutions. No. 21(2), 2017, pp. 319–331.
- Demidov N.A., Belyaev A.A., Sakharov B.A. Perspektivnyy bortovoy vodorodnyy standart chastoty kosmicheskogo primeneniya [Perspective onboard hydrogen standard of frequency of space applications]. VIII Mezhdunarodnyy simpozium “Metrologiya vremeni i prostranstva”, 14–16 sentyabrya 2016 g., Sankt-Peterburg [Proceedings of VIII International Symposium “Metrology of Time and Space”, September 14–16, 2016, St. Petersburg]. (in Russian)
- Utkin A.G., Belyaev A.A., Pavlenko Yu.K. Kvantovyy vodorodnyy generator bortovogo kosmicheskogo radioteleskopa “RADIOASTRON” (konstruktsiya i rezul'taty ispytaniy) [Quantum hydrogen generator of the onboard space radio telescope “RADIOASTRON” (design and test results)]. METROLOGIYA VREMENI I PROSTRANSTVA, Materialy 6-go Mezhdunarodnogo simpoziuma [Proceedings of METROLOGY OF TIME AND SPACE, the VI International Symposium], 2012, pp. 85–87. (in Russian)
- Lett P.D., Phillips W.D., Rolston S.L., Tanner C.E., Watts R.N., Westbrook C.I. Optical Molasses. Journal of the Optical Society of America. 1989, No. 6(11), pp. 2084–2107.
- Jessen P.S., Deutsch I.H. Optical lattices. Advances in Atomic, Molecular, and Optical Physics, 1996, No. 37, pp. 95–138.
- Werth G., Gheorghie V.N., Major F.G. Charged Particle Traps II, Applications. Springer, 2009.
- Nicholson T.L. et al. Systematic evaluation of an atomic clock at 2×10^{-18} total uncertainty. Nature communications, 2015, Vol. 6.
- Huntemann N. et al. Single-ion atomic clock with 3×10^{-18} systematic uncertainty. Physical review letters, 2016, Vol. 116, No. 6, 063001 p.
- Ludlow A.D., Boyd M.M., Ye J. Optical atomic clocks. REVIEWS OF MODERN PHYSICS. 2015, No. 87, pp. 637–701.
- Vogt S., Häfner S., Grotti J., Koller S., Al-Masoudi A., Sterr U., Lisdat C. A transportable optical lattice clock. Journal of Physics: Conference Series, 2016, No. 723, 012020 p.
- Cao, J., Zhang, P., Shang, J., Cui, K., Yuan, J., Chao, S. & Huang, X. A transportable 40Ca^+ single-ion clock with 7.7×10^{-17} systematic uncertainty. arXiv preprint, 2016, arXiv:1607.03731.
- Available at: <https://www.muquans.com/index.php/products/mclock>
- Lezius M. et al. Space-borne frequency comb metrology. Optica, 2016, Vol. 3, No. 12, pp. 1381–1387.
- Schkolnik V., Hellmig, O., Wenzlawski A., Grosse J., Kohfeldt A., Döringshoff K. & Krutzik M. A compact and robust diode laser system for atom interferometry on a sounding rocket. Applied Physics B, 2016, No. 122(8), 217 p.

18. Lachmann M. et al. Creating the first Bose-Einstein Condensate in Space. *Bulletin of the American Physical Society*, 2017.
19. Jeong Y. et al. Ytterbium-doped large-core fiber laser with 1.36 kW continuous-wave output power. *Optics express*, 2004, No.12(25), pp. 6088–6092.
20. Jang H. et al. Polarization maintaining linear cavity Er-doped fiber femtosecond laser. *Laser Physics Letters*, 2015, No. 12(10), 105102 p.
21. Leibrandt D.R., Thorpe M.J., Notcutt M., Drullinger R.E., Rosenband T. & Bergquist, J.C. Spherical reference cavities for frequency stabilization of lasers in non-laboratory environments. *Optics express*, 2011, No. 19(4), pp. 3471–3482.
22. Alnis J. et al. Subhertz linewidth diode lasers by stabilization to vibrationally and thermally compensated ultralow-expansion glass Fabry-Pérot cavities. *Physical Review A*, 2008, No. 77(5), 053809 p.
23. Legero T. et al. 1.5 μm Lasers with Sub10 mHz Linewidth. *CLEO: Science and Innovations*. Optical Society of America, 2017, SW1J. 1 p.
24. Washburn B.R. et al. Phase-locked, erbium-fiber-laser-based frequency comb in the near infrared. *Optics letters*, 2004, Vol. 29, No. 3, pp. 250–252.
25. Ma L.S., Bi Z., Bartels A., Robertsson L., Zucco M., Windeler R.S. & Diddams, S.A. Optical frequency synthesis and comparison with uncertainty at the 10^{-19} level. *Science*, 2004, No. 303(5665), pp. 1843–1845.
26. Wang B., Gao C., Chen W.L., Miao J., Zhu X., Bai Y. & Wang, L.J. Precise and continuous time and frequency synchronisation at the 5×10^{-19} accuracy level. *Scientific reports*, 2012, No. 2.
27. Huang Y. et al. A comparison of two 40Ca^+ single-ion optical frequency standards at the 5×10^{-17} level and an evaluation of systematic shifts. *Applied Physics B*, 2017, Vol. 123, No. 5, 166 p.
28. Chou C. et al. Frequency comparison of two high-accuracy Al^+ optical clocks. *Physical review letters*, 2010, Vol. 104, No. 7, 070802 p.
29. Zhang J. et al. Direct Laser Cooling Al Ion Optical Clocks. *Chinese Physics Letters*, 2017, Vol. 34, No. 5, 050601 p.
30. Available at: <http://www.opticlock.de/en/info/>
31. Clement Lacroute et al. Compact Yb^+ optical atomic clock project: design principle and current status. *J. Phys. Conf. Series*, 2016, Vol. 723, 012025 p.
32. Leibfried D., Blatt R., Monroe C., Wineland D. Quantum dynamics of single trapped ions. *Review of Modern Physics*. 2003, No.75(1), p. 281–324.
33. Dicke R.H. The effect of collisions upon the Doppler width of spectral lines. *Phys. Rev.*, 1953, No. 89, pp. 472–473.
34. Berkeland D.J., Miller J.D., Bergquist J.C., Itano W.M., Wineland D.J. Minimization of ion micromotion in a Paul trap. *Journal of Applied Physics*, 1998, No. 83(10), pp. 5025–5033.
35. Qi H., Song Z., Guo J., Ni J., Wang C., Peng G. Narrow-linewidth distributed feedback fiber laser with MOPA. *Chinese Optics Letter*, No. 13(S2), 21404 p.
36. Available at: https://www.rp-photonics.com/erbium_doped_fiber_amplifiers.html
37. Available at: https://www.rp-photonics.com/ytterbium_doped_gain_media.html
38. Zhuravleva O.V., Ivanov A.V., Leonovich A.I., Kurnosov V.D., Kurnosov K.V., Chernov R.V., Shishkov V.V., Pleshanov S.A. Odnochastotnyy lazer s podstroykoy dliny volny dlya nakachki tsezievykh standartov chastoty [Single-frequency laser with wavelength tuning for cesium frequency standards]. *Kvantovaya elektronika [Quantum electronics]*, 2006, No. 36(8), pp. 741–744. (in Russian)
39. Serkland D.K., Keeler G.A., Geib K.M., Peake G.M. Narrow linewidth VCSELs for high-resolution spectroscopy. *Proc. SPIE 7229, Vertical-Cavity Surface-Emitting Lasers XIII*, 2009, 722907 p.
40. Raab S., Neuhaus R., Falke S., Nölleke C., Stuhler J., Kaenders W. 53 Years Tunable Semiconductor Laser-Past, Present and Future. *Conference on Lasers and Electro-Optics, OSA Technical Digest (online)*, Optical Society of America, 2017, paper AM3A.2.
41. Hawthorn C. J., Weber K. P., Scholten R. E. Littrow configuration tunable external cavity diode laser with fixed direction output beam. *Review of scientific instruments*, 2001, No. 72(12), pp. 4477–4479.
42. Liu K., Littman M.G. Novel geometry for single-mode scanning of tunable lasers. *Opt. Lett.*, 1981, No. 6(3), 117 p.
43. Zorabedian P., Trutna W.R. Interference-filter-tuned, alignment-stabilized, semiconductor external-cavity laser. *Optics letters*, 1988, No. 13(10), pp. 826–828.
44. Iwai M., Yoshino T., Yamaguchi S., Imaeda M., Pavel N., Shoji I. & Taira T. High-power blue generation from a periodically poled MgO: LiNbO_3 ridge-type waveguide by frequency doubling of a diode end-pumped $\text{Nd: Y}_3\text{Al}_5\text{O}_{12}$ laser. *Applied Physics Letters*, 2003, No. 83 (18), pp. 3659–3661.

45. Drever R.W.P., Hall J.L., Kowalski F.V., Hough J., Ford G.M., Munley A.J. & Ward H. Laser phase and frequency stabilization using an optical resonator. *Applied Physics B*, 1983, No. 31(2), pp. 97–105.
46. Kessler T., Hagemann C., Grebing C., Legero T., Sterr U., Riehle F. & Ye J. A sub-40-mHz-linewidth laser based on a silicon single-crystal optical cavity. *Nature Photonics*, 2012, No. 6(10), pp. 687–692.
47. Golovizin A.A., Kalganova E.S., Sukachev D.D., Vishnyakova G.A., Semerikov I.A., Soshenko V.V., Tregubov D.O., Akimov A.V., Kolachevsky N.N., Khabarova K.Yu., Sorokin V.N. Detection of the clock transition ($1.14 \mu\text{m}$) in ultra-cold thulium atoms. *Quantum Electronics*, No. 45(5), pp. 482–485.
48. Sukachev D., Fedorov S., Tolstikhina I., Tregubov D., Kalganova E., Vishnyakova G., Golovizin A., Kolachevsky N., Khabarova K., Sorokin V. Inner-shell magnetic dipole transition in Tm atoms: A candidate for optical lattice clocks. *Physical Review A*, No. 94(2), pp. 1–13.
49. Reichert J., Holzwarth R., Udem T., Hänsch T. W. Measuring the Frequency of Light with Mode-Locked Lasers. *Optics Communications*, 1999, No. 172, pp. 59–68.
50. Xie X., Bouchand R., Nicolodi D., Giunta M., Hänsel W., Lezius M., Joshi A., Datta S., Alexandre C., Lours M., Tremblin P.-A., Santarelli G., Holzwarth R., Le Coq Y. Photonic microwave signals with zeptosecond-level absolute timing noise. *Nature Photonics*, 2016, No. 11(1), pp. 44–47.
51. Puppe T., Sell A., Kliese R., Hoghooghi N., Zach A., Kaenders W. Characterization of a DFG comb showing quadratic scaling of the phase noise with frequency. *Optics Letters*, 2016, No. 41(8), pp. 1877–1880.
52. Gubin M.A., Kireev A.N., Konyashchenko A.V., Kryukov P.G., Tausenev A.V., Tyurikov D.A., Shelkovnikov A.S. Realizatsiya kompaknykh opticheskikh chasov [Implementation of compact optical clock. *Kvantovaya Elektronika*, 2008, No. 38(7). pp. 613–614. (in Russian)
53. Zimmermann M., Gohle C., Holzwarth R., Udem T. & Hänsch T.W. Optical clockwork with an offset-free difference-frequency comb: accuracy of sum- and difference-frequency generation. *Optics Letters*, 2004, No. 29(3), 310 p.
54. Kireev A.N., Tausenev A.V., Tyurikov D.A., Shelkovnikov A.S., Shepelev D.V., Konyashchenko A.V. & Gubin M.A. Femtosecond optical-to-microwave frequency divider with a relative instability of $10^{-4} - 10^{-16}$ ($\tau = 1 - 100 \text{ s}$). *Quantum Electronics*, 2016, No. 46(12), pp. 1139–1141.
55. Arias E.F., Combrinck L., Gabor P., Hohenkerk C., Seidelmann P.K., New Technologies and the Future of Timekeeping. *The Science of Time 2016: Time in Astronomy & Society, Past, Present and Future*, 2017, Springer International Publishing, pp. 379–389.
56. Rochat P., Droz F., Wang Q., Froidevaux S. Atomic clocks and timing systems in global navigation satellite systems. *Proceedings of 2012 European Navigation Conference*, 2012, pp. 25–27.
57. Huntemann N. High-accuracy optical clock based on the octupole transition in Yb+171. PhD thesis at Gottfried Wilhelm Leibniz Universität, Hannover, 2014.

Technique for Hardware Residuals Evaluation of GNSS-Measurements of Pseudorange in Phase of the Carrier of the Onboard Navigation System of LEO Spacecraft

A.P. Fursov, *Cand. Sci. (Engineering)*, contact@spacecorp.ru

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

A.A. Fursov, contact@spacecorp.ru

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

V.S. Vdovin, *Cand. Sci. (Engineering)*, contact@spacecorp.ru

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

A.V. Zaichikov, contact@spacecorp.ru

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

Z.A. Pozyaeva, contact@spacecorp.ru

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

Abstract. When using the onboard satellite navigation equipment of LEO spacecraft, there arises a problem of getting an adequate evaluation of hardware errors of the equipment of satellite navigation by real results of navigation observations without a reference orbit or measurements. This article proposes a technique of a hardware error evaluation of pseudorange measurements of the carrier frequency phase received by the onboard equipment of satellite navigation installed on LEO spacecraft. A hardware error of the mentioned measurements is determined indirectly by the so-called double differences of the initial measurements, which make it possible to rule out the overwhelming portion of systematic errors.

The results of applying this technique on two particular selections (S1, S2) of measurements with durations of 5 hours each for one of the Russian spacecraft similar to the spacecraft of the JASON space system are presented. An evaluation of phase measurement accuracy was performed for radio signals in the frequency ranges L1, L2 of GLONASS navigation spacecraft.

Keywords: navigation systems, navigation measurements, evaluation of measurement accuracy, LEO spacecraft, hardware errors, double differences

Introduction

An assessment of errors of results of navigation measurements subject to random and unknown systematic errors in the absence of control measurements, at least, much more exact, is generally impossible or significantly becomes complicated. Under certain conditions, if it is possible to exclude systematic errors of measurements and find an adequate representation of the measured function, an assessment of a random (“noise-like”) component of an error of results of measurements as a stationary process can be defined.

The offered technique is developed to assess a hardware error of results of measurements of a carrier phase received by the equipment of satellite navigation of LEO spacecraft using GNSS and GPS SC. Applying this technique and technology of its realization is considered in relation to one of domestic LEO spacecraft (further – GS SC).

The mentioned technique can be used for assessment of a hardware error of GNSS-measurements of a carrier phase in the equipment of satellite navigation of LEO spacecraft of a different purpose.

1. Technique to estimate errors of measurements of pseudorange by a carrier phase

Definition of dispersion of errors of any undisplaced measurements subject to random errors, generally comes down to definition of a residual vector of measurements which components are the differences of results of measurements and their true calculated values (the measured function). If the measured function is known and casual errors of measurements have a limited dispersion and are not displaced, then an undisplaced assessment of dispersion is the square of a norm of a residual vector divided into dimension of this vector minus the quantity of degrees of freedom of a residual vector.

In practice, results of measurements of a carrier phase do not meet the conditions mentioned above because of presence of the shifts caused by the influence of a set of various technical factors and effects associated with conditions of signal distribution.

The positive result can be received when using instead of initial results of measurements of a carrier

phase of differential measurements as a result of linear transformations of initial measurements, in particular, of the so-called double differences (DD) of a carrier phase [1].

The double differences calculated on simultaneous onboard measurements of a carrier phase of LEO spacecraft and ground measurements of a carrier phase (from ground measuring stations), have almost no systematic errors caused by the influence of the onboard equipment.

The essence of the technique offered in this paper consists in receiving an assessment of a hardware error of initial measurements of a carrier phase indirectly – according to evaluation of errors of DD of a carrier phase. If the assumption is admissible that initial phase measurements are equally accurate, then casual errors of differential measurements, as a rule, represent the errors typical of a stationary process. Further, if dispersion of errors for DD of a carrier phase is defined, then dispersion for any private component of its measurements of a carrier phase can be calculated taking into account the weight fraction considering the used combinations of initial measurements in the form of an amplification factor of the corresponding measurement in differential measurement. As a result, if δ_r is the assessment of an RMS error for DD of a carrier phase [1], then δ is the assessment of an RMS error of the components of its measurements of a carrier phase is obtained by division of δ_r into the amplification factor γ .

Let us suggest that the measured function \mathbf{u} and \mathbf{n} of the fixed values of this function $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ of the final interval of a selection are interconnected by the following ratios: $\mathbf{u} = \mathbf{f}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, $\mathbf{u} = \mathbf{u}^* + \boldsymbol{\varepsilon}_u$, $x_i = x_i^* + \varepsilon_i$, where \mathbf{u}^* , x_i^* are the true values of the variables, $\boldsymbol{\varepsilon}_u$, ε_i are the random unbiased errors.

Taking into account the chosen assumptions

$$\varepsilon_u^2 = \sum_{i=1}^n \left(\frac{\partial \mathbf{f}}{\partial x_i} \right)^2 \varepsilon_i^2$$

and initial phase measurements, pseudorange on the i -th frequency are the following:

$$L_i = D + I_i + \sum_j \Delta D_j + \varepsilon_i$$

where L_i is the result of measurement of a carrier phase on the i -th frequency; D is the true range between the satellite and receiver; I_i is the correction connected with signal distribution in the Earth ionosphere; $\sum_j \Delta D_j$ is the sum of other corrections that are not examined in

detail in this case; $\boldsymbol{\varepsilon}_i$ is the random noises of a signal on the i -th frequency.

In the expression for \mathbf{L}_i , the value $\sum_j \Delta D_j$ is the sum of systematic and slowly changing errors of phase measurements, which are eliminated when forming DD. An unremovable part of these errors is eliminated at the optimal approximation of the measurement results in the way when a dominating error of polynomial residual are only hardware errors.

An ionosphere-free combination of phase measurements on the frequencies f_1 and f_2 [1] is the following:

$$\mathbf{L}_{IF} = \frac{f_1^2}{f_1^2 - f_2^2} \mathbf{L}_1 - \frac{f_2^2}{f_1^2 - f_2^2} \mathbf{L}_2 = \mathbf{D} + \sum_j \Delta D_j^* + \boldsymbol{\varepsilon}_{IF}$$

where the $\boldsymbol{\varepsilon}_{IF}$ value is calculated buy the formula:

$$\boldsymbol{\varepsilon}_{IF} = \sqrt{\left(\frac{f_1^2}{f_1^2 - f_2^2}\right)^2 \boldsymbol{\varepsilon}_1^2 + \left(\frac{f_2^2}{f_1^2 - f_2^2}\right)^2 \boldsymbol{\varepsilon}_2^2} = \frac{\sqrt{f_1^4 \boldsymbol{\varepsilon}_1^2 + f_2^4 \boldsymbol{\varepsilon}_2^2}}{f_1^2 - f_2^2}.$$

Considering that $\frac{f_1}{f_2} = \gamma$, we obtain the formula:

$$\boldsymbol{\varepsilon}_{IF} = \frac{\sqrt{\gamma^4 \boldsymbol{\varepsilon}_1^2 + \boldsymbol{\varepsilon}_2^2}}{\gamma^2 - 1}.$$

For a geometrically-free combination of phase measurements on the frequencies f_1 and f_2 , there is a following ratio:

$$\mathbf{L}_{GF} = \mathbf{L}_1 - \mathbf{L}_2 = \mathbf{I}_{GF} + \lambda_1 \mathbf{n}_1 - \lambda_2 \mathbf{n}_2 + \boldsymbol{\varepsilon}_{GF}$$

$$\text{where } \boldsymbol{\varepsilon}_{GF} = \sqrt{\boldsymbol{\varepsilon}_1^2 + \boldsymbol{\varepsilon}_2^2}.$$

A double difference of ionosphere-free combinations of phase measurements can be presented this way:

$$L_{IF} = (\nabla \Delta L_{IF11} - L_{IF12}) - (L_{IF21} - L_{IF22}) = (D_{11} - D_{12}) - (D_{21} - D_{22}) + \sum_j \Delta D_j^* + \boldsymbol{\varepsilon}_{\nabla \Delta}$$

where

$$\boldsymbol{\varepsilon}_{\nabla \Delta} = \sqrt{\boldsymbol{\varepsilon}_{IF}^2 + \boldsymbol{\varepsilon}_{IF}^2 + \boldsymbol{\varepsilon}_{IF}^2 + \boldsymbol{\varepsilon}_{IF}^2} = 2\boldsymbol{\varepsilon}_{IF} = 2 \frac{\sqrt{\gamma^4 \boldsymbol{\varepsilon}_1^2 + \boldsymbol{\varepsilon}_2^2}}{\gamma^2 - 1}.$$

Let us consider two cases:

1. A random noise of phase measurements on the f_1 and f_2 frequencies is the same by the value, i.e., $\boldsymbol{\varepsilon}_1 = \boldsymbol{\varepsilon}_2$, then for $\boldsymbol{\varepsilon}_{\nabla \Delta}$ and $\boldsymbol{\varepsilon}_{GF}$ we receive:

$$\boldsymbol{\varepsilon}_{\nabla \Delta} = 2 \frac{\sqrt{\gamma^4 + 1}}{\gamma^2 - 1} \boldsymbol{\varepsilon}_1$$

$$\boldsymbol{\varepsilon}_{GF} = \sqrt{2} \boldsymbol{\varepsilon}_1.$$

Taking into account that for GLONASS navigation SC $\gamma = 9/7$ [1], we receive:

$$\boldsymbol{\varepsilon}_{\nabla \Delta} \approx 5.917, \quad \boldsymbol{\varepsilon}_{GF} \approx 1.414 \boldsymbol{\varepsilon}_1.$$

2. A random noise of phase measurements is directly proportional to the wavelength of the signal $\boldsymbol{\varepsilon}_i = \alpha \lambda_i$, i.e., $\boldsymbol{\varepsilon}_2 / \boldsymbol{\varepsilon}_1 = \gamma$, then for $\boldsymbol{\varepsilon}_{\nabla \Delta}$ and $\boldsymbol{\varepsilon}_{GF}$ we receive:

$$\boldsymbol{\varepsilon}_{\nabla \Delta} = 2\gamma \frac{\sqrt{\gamma^2 + 1}}{\gamma^2 - 1} \boldsymbol{\varepsilon}_1$$

$$\boldsymbol{\varepsilon}_{GF} = \sqrt{1 + \gamma^2} \boldsymbol{\varepsilon}_1.$$

Considering that for GLONASS navigation SC $\gamma = 9/7$ [1], we receive:

$$\boldsymbol{\varepsilon}_{\nabla \Delta} \approx 6.413 \boldsymbol{\varepsilon}_1, \quad \boldsymbol{\varepsilon}_{GF} \approx 1.629 \boldsymbol{\varepsilon}_1.$$

Amplification factors of a random noise for different combinations of phase measurements are given in Table 1.

Table 1. The amplification factors k_1 and k_2 of a random noise for different combinations of phase carrier in relation to the noise at the L1 frequency.

	k_1 (at $\boldsymbol{\varepsilon}_1 = \boldsymbol{\varepsilon}_2$)	k_2 (at $\boldsymbol{\varepsilon}_2 / \boldsymbol{\varepsilon}_1 = \gamma$)
$\boldsymbol{\varepsilon}_1$	1	1
$\boldsymbol{\varepsilon}_2$	1	1.286
$\boldsymbol{\varepsilon}_{GF}$	1.414	1.629
$\boldsymbol{\varepsilon}_{IF}$	2.958	3.206
$\nabla \Delta \boldsymbol{\varepsilon}_{IF}$	5.917	6.413

The results show that the estimation of an RMS error of carrier phase measurements can be calculated implicitly based on the estimation of an RMS error of the carrier phase measurements for the DD of the carrier phase measurements $\nabla \Delta \mathbf{L}_{IF}$. Further, taking into account Table 1, the estimation of an RMS error of the carrier phase measurements is used to evaluate an RMS error for initial carrier phase measurements (as a component of a corresponding difference) considering an amplification factor.

In particular, according to Table 1, the assessment of an RMS error of a random pseudonoise error of initial carrier phase measurements (as a random stationary process) is about 6 times less than an RMS error of DD of a carrier phase from ionosphere-free linear combinations of initial measurements.

Further, errors of DD of a carrier phase are considered as a residual vector which components are the differences of the mentioned DD of a carrier phase and the optimum polynomial representations [2] corresponding to them as an adequate representation of the measured function. It is supposed that results of measurements and their double differences are presented in groups – sessions of measurements. The session of measurements is an arranged sequence of measurements on the time interval of H_s duration in number of not less than 30 [2–4, 6] with an approximately same 10-second (for the equipment of satellite navigation GS SC) pace on time.

The choice of an optimum order of a polynomial (optimum polynomial) of approximation of the results of measurements on an interval of H_s can be carried out by the following criteria:

1. M – on control of a methodical error of approximation of calculated values of the measurements calculated in a priori known coordinate and high-speed parameters of the movement (CHSPM) of LEO spacecraft. Generally, CHSPM are calculated by integration of a system of differential equations of the SC movement at the known statistical assessment of initial conditions (IC) – CHSPM on the fixed (initial) timepoint. It is supposed that the mentioned statistical assessment of IC is calculated in the ground control complex (GCC) of LEO spacecraft on trajectory measurements by a method of the smallest squares or is a priori known. An optimum polynomial by this criterion is the polynomial of the minimum order at which the module of a methodical error of approximation of settlement measurements should not exceed the control size $M (dR)$ set in advance;

2. C – on control of speed of change of an RMS error of approximation of the results of measurements from a polynomial order. An optimum polynomial is a polynomial of the minimum i -th order at which the distinction of percentage of the module of a difference of an RMS error for adjacent i -th and $(i + 1)$ -th polynomials in relation to an RMS error of the i -th polynomial does not exceed the set level $C (dR)$.

A polynomial residual is a difference between the initial measurement and the calculated value corresponding to it calculated with the use of an optimum polynomial.

A repeated testing of procedures of definition of optimum polynomials with use of real measurements of the equipment of satellite navigation of GS SC has shown a high level of the coincidence of optimum orders of

polynomials by criteria of M and C at $M (dR) = 12–18$ mm and $C (dR) = 20\%$ and equivalence of the results received by these criteria.

2. An assessment of errors of phase measurements of the equipment of satellite navigation

2.1. Conditions and estimation of an RMS error measurements

As an example, employing the offered technique of an assessment of hardware errors of phase measurements (on a carrier phase) has been carried out in relation to GNSS-measurements of the equipment of satellite navigation of GS SC in the posteriori mode on a specialized automated working place of Joint Stock Company “Russian Space Systems” applying the information files containing results of navigation definitions and navigation (trajectory) information, in particular, carrier phase measurements.

To evaluate hardware errors of phase measurements, the average of statistical estimates of a mean square error of measurements and dispersion of an RMS error on selections were used.

Calculation of an RMS error was performed on measurements of a carrier phase of the equipment of satellite navigation and from the ground IGS measuring stations (in the RINEX files format) on private selections of S1 and S2.

Calculation of a required RMS error measurements of phase carrier as dot estimates was carried out taking into account the normative documents [6] and recommendations stated in [2,3,5]. The following parameters were calculated on each selection:

1. An RMS error for each session of DD of a carrier phase

$$\nabla\delta_j = \sqrt{\frac{\sum_{i=1}^n (r_i - r_p)^2}{n - N}}$$

where r_i and r_p is the double difference from compiling initial simultaneous carrier phase measurements and a calculated value corresponding to this difference, respectively, n is the number of double differences in the session, N is the order of an optimum polynomial, $j = 1, 2, \dots, k$ is the serial number of the session of second differences, k is the general number of sessions;

2. An average RMS error on all sessions of DD of a carrier phase:

$$\nabla m_s = \frac{\sum_{j=1}^k \nabla \delta_j}{k}.$$

3. Hence, dispersion and the RMS error for ∇m_s is:

$$D_s = \frac{\sum_{j=1}^k (\delta_j - \nabla m_s)^2}{k-1},$$

$$\nabla \delta_s = \sqrt{\nabla D_s};$$

4. A confidence span for DD of a carrier phase is:

$$\nabla L = (\nabla a_1, \nabla a_2)$$

where $\nabla a_1 = \nabla m_s - t_\beta \nabla \delta_s$, $\nabla a_2 = \nabla m_s + t_\beta \nabla \delta_s$ determine the limits of the ∇L interval, where with the p probability there is a true value for ∇m_s is equal with the probability $\beta = 1 - p$ – outside this interval, t_β defines the shift of the limits of the interval from its center in the units of an RMS error for the chosen p (according to the Student's distribution). Further, it is accepted that $p = 0.9973$, hence $\beta = 1 - p = 0.0027$. At $p = 0.9973$, $t_\beta = 3$.

5. The evaluation of the finding of ∇m_s outside the limits of a confidence span is

$$\beta^* = \frac{k_n}{k}$$

where k_n is the number of sessions with m_s outside a confidence span.

6. An RMS error of the carrier phase measurements considering an amplification factor:

$$= \delta_s / k_i, i = 1, 2.$$

2.2 The results of the evaluation of the errors of the phase measurements of the equipment of satellite navigation

The tables (1, 2) given below give the results of statistical valuations of the errors of carrier phase measurements obtained using DD of a carrier phase of the equipment of satellite navigation for the baseline AJAC-GSC1 (AJAC is a ground measuring station IGS, GSC1 is the equipment of satellite navigation of GS SC).

To determine initial statistical evaluations, there were used two groups made of 35 and 36 sessions of DD of a carrier phase in the S1 and S2 selections, respectively (in each session not less than 30 measurements). The two groups of sessions are presented in Figs. 1, 2 in the form of graphs of residuals of the mentioned DD of a carrier

phase from the optimum polynomial representation (in mm).

Tables 1 and 2 use the following symbols:

No. is the serial number of the session of DD of a carrier phase;

RXX is the number of the working point of a navigation spacecraft of GLONASS XX;

k_1, k_2 is the minimum and maximum values, respectively, of an amplification factor of the RMS error of DD of a carrier phase in relation to the RMS error of initial measurements of a carrier phase, $k_1 = \nabla \Delta \epsilon_{IF}$ (at $\epsilon_1 = \epsilon_2 = 5.917$), $k_2 = \nabla \Delta \epsilon_{IF}$ (at $\gamma = \epsilon_2 / \epsilon_1 = 6.413$);

$\nabla L, \nabla L_i$ is the confidence span, respectively, for DD of a carrier phase and initial measurements of a carrier phase at $t_\beta = 3$ (t_β defines the number of corresponding RMS errors, which should be decreased and increased by ∇m_s) to make a true value of ∇m_s present in the confidence span with the probability of $\beta = 0.9973$).

The statistical evaluations of the results of processing of the sessions of DD of a carrier phase of the S1 selection:

1. An average RMS error on the total sessions of DD: $\nabla m_s = 8.56$ mm.

2. An RMS error for ∇m_s (for an average RMS error of all sessions): $\nabla \delta_s = 1.88$ mm.

3. A confidence span for DD: $\nabla L = (2.92, 14.20)$ mm.

4. A confidence span for initial phase measurements of pseudorange (calculated by the ∇L value considering amplification factors):

$$\nabla L_i = (2.92 / k_i, 14.20 / k_i) \text{ mm,}$$

$$\nabla L_1 = (0.49, 2.40) \text{ mm,}$$

$$\nabla L_2 = (0.46, 2.21) \text{ mm.}$$

5. β^* is the evaluation of the presence frequency of ∇m_s (an average RMS error) beyond the limits of a confidence span: $\beta^* = 0$.

6. Average RMS errors of phase measurements of pseudorange (calculated by the ∇m_s value considering amplification factors k_1, k_2): $\nabla m_{sk1} = 1.45$ mm, $\nabla m_{sk2} = 1.33$ mm.

7. An RMS error for ∇m_{ski} of phase measurements considering an amplification factor (calculated by the $\nabla \delta_s$ value considering the amplification factors k_1, k_2): $\delta_{sk1} = 0.32$ mm, $\delta_{sk2} = 0.29$ mm.

The statistical evaluations of the results of processing of the sessions of DD of a carrier phase of the S2 selection:

Table 2. Statistical assessments of measurement errors of a carrier phase of the equipment of satellite navigation of GS SC of the S1 residual.

No.	Navigation SC1 – Navigation SC 2	Session duration, s	Quantity of measurements in the session	$\nabla\delta$ – an RMS error (mm) of residuals of DD of a carrier phase	$\nabla\delta_{k1}$ – an RMS error (mm) of initial measurements of a carrier phase, $k_1 = 5.917$	$\nabla\delta_{k2}$ – an RMS error (mm) of initial measurements of a carrier phase, $k_2 = 6.413$
1	R20 – R21	379.999000	39	8.592000	1.452087	1.339779
2	R11 – R22	360.000000	37	7.303000	1.234240	1.138781
3	R09 – R16	869.999000	88	9.885000	1.670610	1.541400
4	R19 – R10	709.999000	72	7.247000	1.224776	1.130048
5	R10 – R01	710.001000	72	9.784000	1.653541	1.525651
6	R10 – R20	580.001000	59	8.114000	1.371303	1.265242
7	R21 – R22	380.000000	39	6.390000	1.079939	0.996414
8	R10 – R11	459.999000	47	6.049000	1.022309	0.943240
9	R10 – R20	380.000000	39	7.556000	1.276998	1.178232
10	R16 – R10	420.000000	43	9.577000	1.618557	1.493373
11	R10 – R20	379.999000	39	9.938000	1.679567	1.549665
12	R10 – R20	420.000000	43	8.594000	1.452425	1.340090
13	R11 – R21	330.000000	34	7.004000	1.183708	1.092157
14	R19 – R09	870.000000	88	8.922000	1.507859	1.391237
15	R19 – R09	579.999000	59	7.340000	1.240493	1.144550
16	R11 – R21	320.001000	33	5.698000	0.962988	0.888508
17	R19 – R10	289.999000	30	9.338000	1.578165	1.456105
18	R11 – R22	459.999000	47	6.545000	1.106135	1.020583
19	R10 – R20	320.000000	33	10.458000	1.767450	1.630750
20	R20 – R21	789.998000	80	12.370000	2.090586	1.928894
21	R18 – R19	489.998000	50	8.744000	1.477776	1.363480
22	R01 – R11	899.999000	91	11.669000	1.972114	1.819585
23	R21 – R22	589.999000	60	5.710000	0.965016	0.890379
24	R20 – R11	519.999000	53	7.874000	1.330742	1.227818
25	R09 – R10	459.999000	47	6.557000	1.108163	1.022454
26	R16 – R20	330.000000	34	10.222000	1.727565	1.593950
27	R11 – R21	379.999000	39	6.508000	1.099882	1.014814
28	R20 – R11	320.000000	33	11.749000	1.985635	1.832060
29	R20 – R01	580.001000	59	9.681000	1.636133	1.509590
30	R22 – R05	470.001000	48	9.160000	1.548082	1.428349
31	R09 – R21	519.999000	53	7.626000	1.288829	1.189147
32	R11 – R21	380.001000	39	7.618000	1.287477	1.187900
33	R08 – R10	579.999000	59	9.241000	1.561771	1.440979
34	R20 – R11	380.000000	39	10.842000	1.832347	1.690628
35	R20 – R11	300.000000	31	6.085000	1.028393	0.948854
36	R09 – R08	580.000000	59	12.090000	2.043265	1.885233
37	All sessions	494.166306	50	8.557778	1.446303	1.334442

Table 3. Statistical assessments of measurement errors of a carrier phase of the equipment of satellite navigation of GS SC of the S2 residual.

No.	Navigation SC1 – Navigation SC 2	Session duration, s	Quantity of measurements in the session	$\nabla\delta$ an RMS error (mm) of residuals of DD of a carrier phase	$\nabla\delta_{k1}$ – an RMS error (mm) of initial measurements of a carrier phase, $k1 = 5.917$	$\nabla\delta_{k2}$ – an RMS error (mm) of initial measurements of a carrier phase, $k2 = 6.413$
1	R09 – R20	539.999000	55	12.544000	2.119993	1.956027
2	R01 – R09	540.000000	55	8.341000	1.409667	1.300639
3	R21 – R02	360.000000	37	13.431000	2.269900	2.094340
4	R09 – R21	319.999000	33	5.312000	0.897752	0.828317
5	R11 – R21	400.000000	41	4.703000	0.794828	0.733354
6	R07 – R09	409.999000	42	9.624000	1.626500	1.500702
7	R20 – R11	719.999000	73	8.027000	1.356600	1.251676
8	R10 – R20	699.999000	71	6.963000	1.176779	1.085763
9	R09 – R20	440.000000	45	8.115000	1.371472	1.265398
10	R11 – R22	419.999000	43	7.549000	1.275815	1.177140
11	R08 – R10	710.001000	72	7.339000	1.240324	1.144394
12	R10 – R01	540.001000	55	7.865000	1.329221	1.226415
13	R20 – R19	440.001000	45	6.473000	1.093967	1.009356
14	R11 – R21	360.000000	37	9.088000	1.535913	1.417121
15	R19 – R11	539.999000	55	7.619000	1.287646	1.188056
16	R11 – R21	370.000000	38	6.008000	1.015379	0.936847
17	R19 – R11	719.999000	73	13.280000	2.244381	2.070794
18	R08 – R10	539.999000	55	13.180000	2.227480	2.055200
19	R11 – R21	720.000000	73	12.872000	2.175427	2.007173
20	R10 – R01	740.000000	75	7.925000	1.339361	1.235771
21	R21 – R22	720.001000	73	11.145000	1.883556	1.737876
22	R21 – R22	329.999000	34	6.997000	1.182525	1.091065
23	R20 – R19	320.000000	33	9.477000	1.601656	1.477780
24	R10 – R09	430.000000	44	10.201000	1.724016	1.590675
25	R20 – R19	540.001000	55	14.275000	2.412540	2.225947
26	R01 – R19	719.999000	73	16.344000	2.762211	2.548573
27	R10 – R20	400.000000	41	7.487000	1.265337	1.167472
28	R20 – R11	519.999000	53	7.937000	1.341389	1.237642
29	R20 – R11	399.999000	41	8.999000	1.520872	1.403243
30	R21 – R02	400.000000	41	7.357000	1.243367	1.147201
31	R10 – R01	440.001000	45	12.836000	2.169343	2.001559
32	R21 – R22	369.999000	38	5.281000	0.892513	0.823484
33	R10 – R11	329.999000	34	6.044000	1.021464	0.942461
34	R09 – R20	339.999000	35	10.275000	1.736522	1.602214
35	R10 – R11	419.999000	43	8.348000	1.410850	1.301731
36	All sessions	491.713971	50	9.121743	1.541616	1.422383

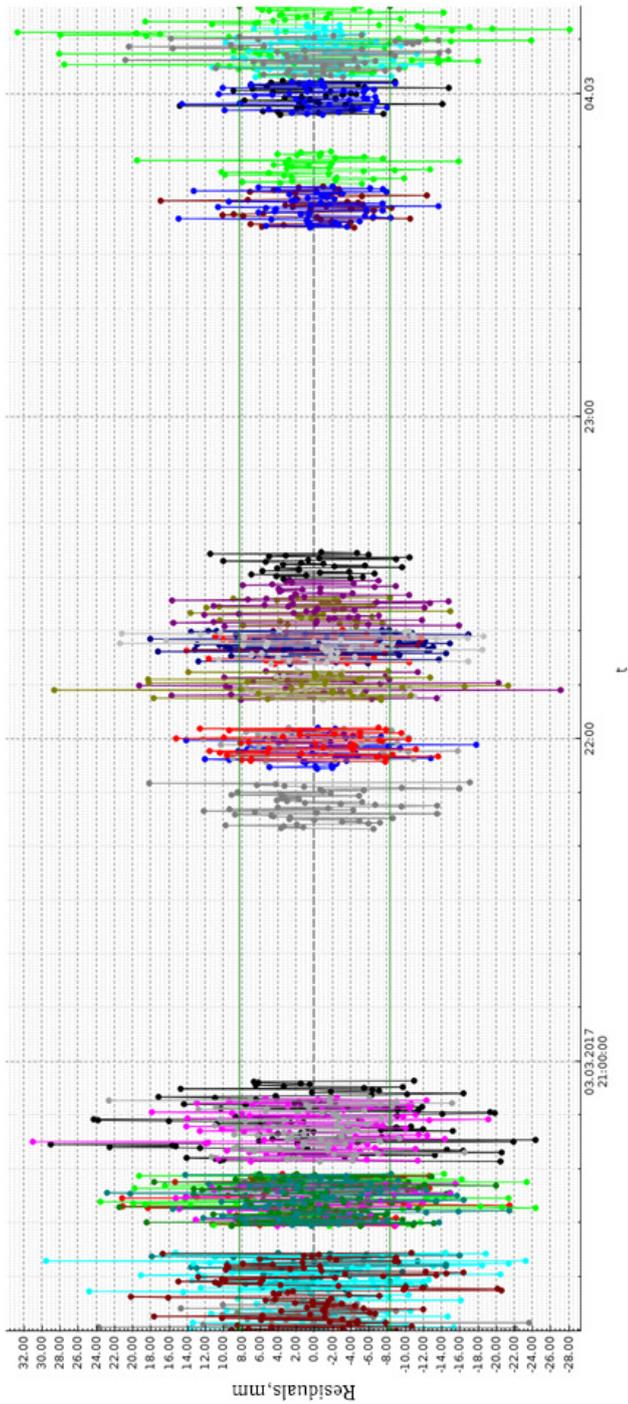


Fig. 1. A graph of residuals of DD of a carrier phase from an optimum polynomial representation of measurement sessions of the S1 selection

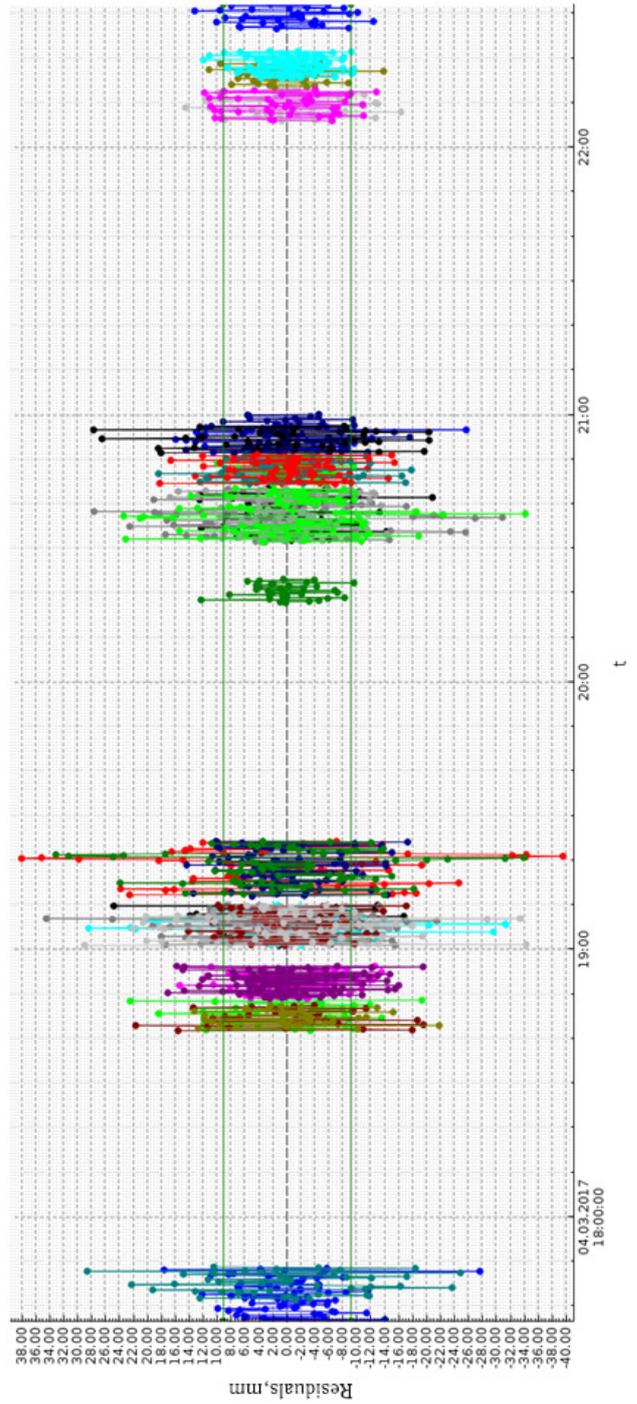


Fig. 2. A graph of residuals of DD of a carrier phase from an optimum polynomial representation of measurement sessions of the S2 selection

1. An average RMS error on total sessions of DD: $\nabla m_s = 9.12$ mm.

2. An RMS error for ∇m_s (for an average RMS error of all sessions): $\nabla \delta_s = 2.89$ mm.

3. A confidence span for DD: $\nabla L = (0.46, 17.78)$ mm.

4. A confidence span for initial phase measurements of pseudorange (calculated by the ∇L value considering amplification factors):

$\nabla L_i = (0.46 / k_i, 17.78 / k_i)$ mm, $\nabla L_1 = (0.08, 3.00)$ mm, $\nabla L_2 = (0.07, 2.77)$ mm.

5. β^* is the evaluation of the frequency of presence of ∇m_s (an average RMS error) beyond the limits of a confidence span: $\beta^* = 0$.

6. Average RMS errors for initial phase measurements of pseudorange (calculated by the ∇m_s value considering the amplification factors k_1, k_2): $\nabla m_{sk1} = 1.54$ mm, $\nabla m_{sk2} = 1.42$ mm.

7. An RMS error for ∇m_{ski} of phase measurements considering amplification factor (calculated by $\nabla \delta_s$ value considering the amplification factors k_1, k_2): $\delta_{sk1} = 0.49$ mm, $\delta_{sk2} = 0.45$ mm.

According to the offered technique, an evaluation of the precise characteristics of measurements of pseudoranges on the phase of the carrier phase received from the equipment of the satellite navigation of GS SC was carried out. The results of the evaluations are calculated on two selections having 36 (the S1 selection) and 35 (the S2 selection) sessions, respectively, of DD with at least not less than 30 measurements in each session.

The final results of the mentioned evaluations are given in Table 4.

Table 4. The evaluations of the precise characteristics of the carrier phase characteristics

Selection	max ∇m_{ski} , mm	∇L_l , MM	∇L_r , mm	β^*
S1	1.45	0.46	2.40	0
S2	1.54	0.07	3.00	0

Table 4 uses the following symbols:

$\max \nabla m_{ski}$ is the maximum value of an RMS error of initial measurements of a carrier phase for the group at the amplification factors $k_i, i=1,2$;

∇L is the confidence span of an average of a statistical evaluation of an RMS error;

p is the possibility of finding of an average of the evaluation of an RMS error in ∇L ;

β^* is the assessment of the possibility of the finding of ∇m_s beyond ∇L ;

∇L_l is the left limit of a confidence span ∇L (a minimum limit of ∇L at the amplification factors $k_i, i=1,2$ with the possibility of confidence $p = 0.9973$);

∇L_r is the right limit of the confidence span of ∇L (a minimum limit of ∇L at the amplification factors $k_i, i=1,2$ with the possibility of confidence $p = 0.9973$ is equal to "nonconfidence" – $\beta = 1 - 0.9973 = 0.0027$).

The obtained statistical characteristics in the given Tables 1–3 and the graphs in the Figs. 1–2 are the representation evaluations of hardware errors of measurements on a carrier phase for the chosen spacecraft.

Conclusion

1. The technique to evaluate a hardware error of navigation measurements of a carrier phase received in a two-frequency mode by onboard navigation systems of LEO spacecraft is offered.

2. Selective evaluations of hardware errors of navigation measurements of a carrier phase received by the equipment of satellite navigation of GS SC are received. In particular, it was determined that with the probability of $p = 0.9973$, an average value of an RMS error of measurements of a carrier phase using a navigation spacecraft GLONASS is in the range 2–3 mm, that agrees with the RMS error for other similar instruments.

3. The offered technique and technology of the evaluation of hardware errors of the carrier phase measurements received in the two-frequency mode of the equipment of satellite navigation of LEO GS SC can be used as typical of monitoring precisions of GNSS-measurements of a wide class of Russian and foreign LEO spacecraft.

4. A visual analysis of residuals of DD of a carrier phase given in Figs. 1,2 allows one to make an assumption that there is a small trend in them, the presence of which can be explained by the influence of the signal propagation medium. The conclusion can be obtained based on the results of processing the measurements on long intervals of time and is an item of further research.

References

1. Antonovich K.M. Ispol'zovanie sputnikovykh radionavigatsionnykh sistem v geodezii [Use of

- satellite radio navigation systems in geodesy]. Moscow, FSUE "Kartgeotsentr", 2005, 334 p. (in Russian)
2. Anderson T. Statisticheskiy analiz vremennykh ryadov [Statistical analysis of time series]. Moscow, Mir, 1976, 774 p. (in Russian)
 3. Venttsel' E.S. Teoriya veroyatnostey [Probability theory]. Moscow, Nauka, 1968, 576 p. (in Russian)
 4. Zhdanyuk B.F. Osnovy statisticheskoy obrabotki traektornykh izmereniy [Basics of statistical processing of trajectory measurements]. Moscow, Sov. radio, 1978, 384 p. (in Russian)
 5. Bronshteyn I.N., Semendyaev K.A. Spravochnik po matematike dlya inzhenerov i uchashchikhsya vtuzov [A handbook on mathematics for engineers and students of technical colleges]. Moscow, Nauka, 1986. (in Russian)
 6. GOST P 50779.21-2004. Statisticheskie metody. Pravila opredeleniya i metody rascheta statisticheskikh kharakteristik po vyborochnym dannym. Chast' 1. Normal'noe raspredelenie. [State Standard R 50779.21-2004. Statistical methods. Rules for the determination and methods for calculating statistical characteristics from sample data. Part 1. Normal distribution]. (in Russian)

Onboard Image Processing by Inverse Filtering

V.E. Kvitka, *postgraduate student, kva-vasja@yandex.ru*

Joint Stock Company "Space Rocket Centre "Progress" –

*Scientific and Production Enterprise "OPTEKS", Zelenograd, Moscow, Russian Federation
Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russian Federation*

V.D. Blinov, *v_blinov@mail.ru*

Joint Stock Company "Space Rocket Centre "Progress" –

Scientific and Production Enterprise "OPTEKS", Zelenograd, Moscow, Russian Federation

Abstract. Satellite images of the Earth of high resolution have blurring of small details and low contrast. This is due to the low values of the modulation transfer function of the large space-based optical Earth observation systems. The problems of image reconstruction by the inverse filtering method are considered in the article. Various ways of its implementation are suggested. The necessity of preliminary noise reduction of images before their restoration is revealed. A quantitative and qualitative analysis of the performance of the most well-known noise reduction algorithms has been performed. Adaptive algorithms of preliminary noise reduction are developed and tested. By computer simulation of images and comparison of quality criteria, the optimal algorithm of preliminary noise reduction of images is for this task is discovered. Regularization, an alternative solution to the problem of noise amplification in an image when performing inverse filtering is studied. A computer simulation of various regularization methods has been carried out. Visual analysis of the results of image modeling and comparison of quantitative criteria for their quality made it possible to determine the optimal parameters of the regularization function.

Keywords: noise reduction, inverse filtering, modulation transfer function

Problem definition

One of the properties of high-resolution Earth remote sensing spacecraft (ERS SC) is the low value of the modulation transfer function (MTF) at high spatial frequencies. Visually this is manifested in the blurring of images of small objects [1]. Let us investigate the possibility of improving the image by inverse filtering performed onboard a spacecraft.

As is well known, the modulation transfer function $MTF(\nu)$ relates the Fourier image of $F(\mathbf{A})$ of the original image \mathbf{A} and the Fourier image $F(\mathbf{B})$ of the image \mathbf{B} obtained by an optoelectronic system, which is shown by equation (1). Here and in what follows, we are talking about a two-dimensional discrete Fourier transformation (2DDFT).

$$F(\mathbf{B}) = F(\mathbf{A}) \cdot MTF(\nu). \tag{1}$$

Naturally, a suggestion arises to restore image \mathbf{A} , having image \mathbf{B} obtained by a HR SC surveying the Earth's surface. We carry this out by the formula (2):

$$\mathbf{A} = F^{-1} \left(\frac{F(\mathbf{B})}{MTF(\nu)} \right). \tag{2}$$

However, the implementation of this method, known as inverse filtering, has a great practical difficulty: for small values of $MTF(\nu)$, the noise always present in image \mathbf{B} will lead to strong distortions of the reconstructed image \mathbf{A} [2]. Accordingly, the question arises of investigating the limits of applicability of this method of image reconstruction. The key point is the need to perform the procedure not on the Earth, but on board a spacecraft.

Otherwise, there will be additional distortions associated with the operation of the video compression algorithm.

Image simulation

The procedure of the image simulation is based on the use of aerial photographs with a 12-bit brightness gradation. Since inverse filtering is considered for a high-resolution spacecraft (about 0.5 m), it is necessary that the original images have a significantly smaller pixel projection. Otherwise, a visual analysis of the results of inverse filtering will be incorrect, since the MTF of the air camera, which is significantly different from 1, will introduce its distortions. Therefore, aerial photographs with L_{aer} pixel projection of not more than 0.2 m are used for the modeling.

Let's note its main stages:

1. The pixel projection of the aerial photograph L_{aer} is calculated, after which it is rescaled to the projection of the pixel L_{sc} (scaling factor $k = L_{aer} / L_{KA}$). The results of this operation are described by the matrix \mathbf{A} .

2. The modulation transfer function (MTF) of the optoelectronic path $MTF(\nu)$ is shown in Fig. 1.

3. 2DDFT of the snapshot \mathbf{A} is performed. We denote the result as $F(\mathbf{A})$.

4. We simulate the blurring of small details of the image \mathbf{A} by multiplying the spectrum $F(\mathbf{A})$ by the MTF of the optoelectronic tract $MTF(\nu)$. The result $F(\mathbf{B}) = F(\mathbf{A}) * MTF(\nu)$ is the Fourier spectrum of the image \mathbf{B} formed in the focal plane of the telescope. Accordingly, in order to obtain the image \mathbf{B} , it is necessary to perform the reverse 2DDFT using formula (3):

$$\mathbf{B} = F^{-1}(F(\mathbf{B})) = F^{-1}(F(\mathbf{A}) \cdot MTF(\nu)). \tag{3}$$

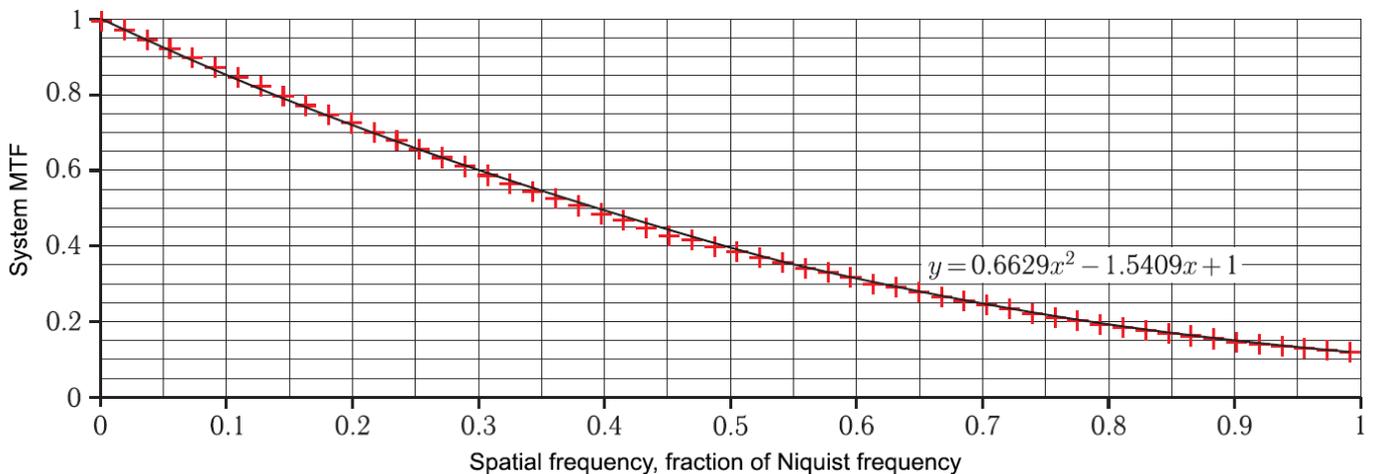


Fig. 1. Modulation transfer function of the system and its approximation.



Fig. 2 (Image A)



Fig. 3 (Image B)



Fig. 4 (noisy image C)

5. We simulate the noise of image **B** taking into account the photon noise, based on the energy calculation of the spacecraft for the corresponding height of the Sun above the horizon. Then the RMS of the noise has its own value for each pixel. We denote the result as **C**.

Here we apply a dimensionless spatial frequency normalized to the Nyquist frequency (55.56 pairs of lines per mm).

Figures 2-4 show fragments of the resampled aerial photograph **A**, the image on the focal plane of the objective **B** and the noisy image **C**, respectively, at the sun height $h_s=5^\circ$. All three images include four groups of three test objects on the upper left (high and low contrasts, periods of 2 and 4 pixels).

Study of inverse filtering

We perform inverse filtering without prior noise reduction for images **C** (h_s), noisiness of which corresponds to different elevations of the Sun over the horizon h_i . The restored image of A_{inv} is described by the formula (4):

$$A_{inv} = F^{-1} \left(\frac{F(C)}{MTF(\nu)} \right). \quad (4)$$

The result is shown in Figure 5, where enlarged fragments of the images are shown

It can be seen from Fig. 5 that a satisfactory result of inverse filtering can be obtained only at elevations of the Sun exceeding $h_s = 30^\circ$, when the noisiness of the images is low. We shall produce image C_{fl} using noise reduction and apply inverse filtering to it. The results are shown in Figure 6.

It can be seen from Fig. 6 that the best result is provided by inverse filtering of an image with noise reduction performed using the pseudo-median filter described in [3]. In this method, the pixel signal is replaced by $\text{med}(y_1, y_2, y_3)$, where y_1, y_2, y_3 are the median signal values in the rows of the 3×3 pixel window. Restoration after the median noise filtering has caused distortion of the image of the test object, and the geometric average filter blurs the details to such a degree that the inverse filtering is unable to restore the sharpness of the edges.

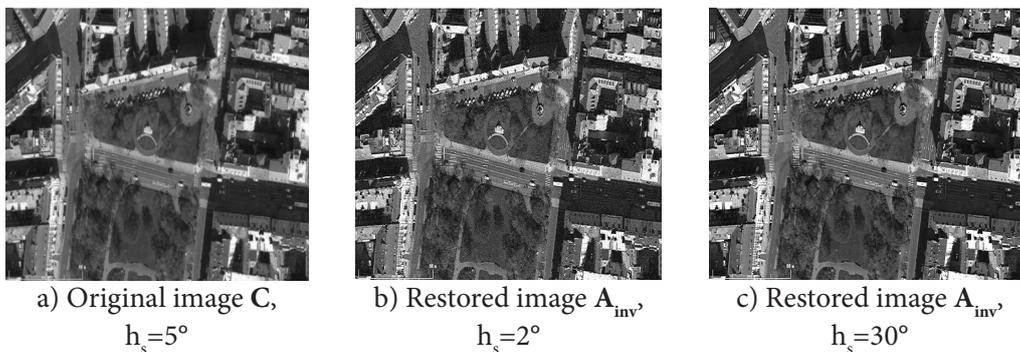


Fig. 5 Inverse filtering without prior noise reduction

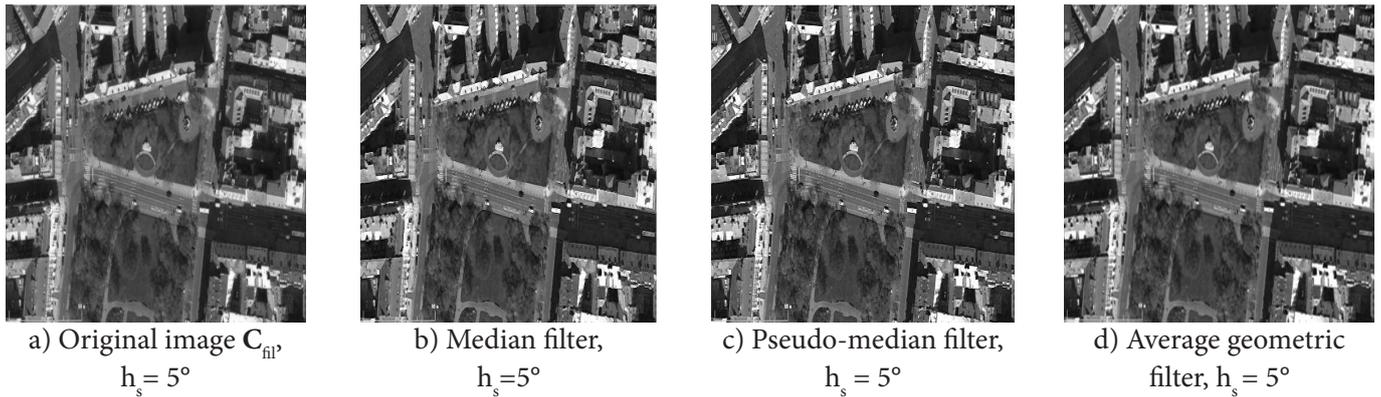


Fig. 6. Results of noise reduction and subsequent inverse filtering

Preliminary processing with adaptive filter

We investigate the inverse filtering performed on the image, which has undergone noise reduction by the adaptive algorithm. The idea of such a filter is described in [2, p. 355]. It should take into account the dependence of RMS of the noise on the signal in a pixel, and in the presence of strong luminance differences it should not smooth them, since these are not noise, but the outlines of the objects.

The following algorithm is proposed: as a sign of the presence of contours in the 3x3 pixel window **E**, we shall assume as a requirement that $RMS^2(\mathbf{E}) / noise^2(i,j) \geq 9$. That is, the actual RMS of the noise in the 3x3 pixel window is 3 times larger than the noise in pixel **C**(i, j). A flowchart of the algorithm is shown in Figure 7.

In this case, the noise RMS is calculated by the on-board device according to the formula (5):

$$noise(i, j) = \frac{4095}{P_{zap}} \cdot \sqrt{C(i, j) \cdot \frac{P_{ch}}{4095}} \tag{5}$$

The charge capacity of the photodetector P_{ch} is known in advance. However, the calculation of the squared value of the RMS of the signal in the 3x3 pixel window requires about 20 operations, which is too many for an on-board processing task.

Detecting outlines using a gradient

We shall use for noise reduction a gradient algorithm that takes into account the direction of the contours in the sliding window of 3x3 pixels. As is known, the gradient of a function of two variables $U(x, y)$ is the vector that determines the direction of the greatest change in the function. If there is a contour in the window, the vector gradient will be perpendicular to the sharp edge. Then

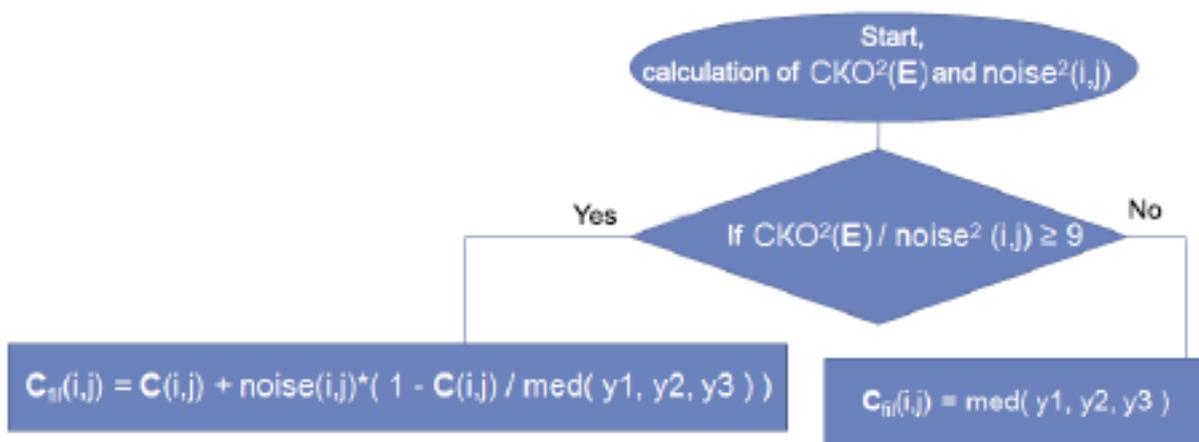


Fig. 7. Flowchart of the algorithm.

we average the brightness of the pixels exactly in the direction of the contour to prevent it from blurring. Mathematically, this is described by formula (6).

$$C_{\text{fil}}(i, j) = a \cdot C(i, j) + b \cdot \left(\sin^2 \theta \cdot \frac{C(i, j - 1) + C(i, j + 1)}{2} + \cos^2 \theta \cdot \frac{C(i - 1, j) + C(i + 1, j)}{2} \right). \quad (6)$$

In this case, the condition $a + b = 1$, which is necessary for the invariance of the average brightness of the image of a homogeneous surface, is satisfied. The angle θ , which is the angle between the gradient vector and the X axis, is determined by the formula (7):

$$\text{tg } \theta = \frac{\frac{\partial U}{\partial y}}{\frac{\partial U}{\partial x}} = \frac{C(i + 1, j) - C(i, j)}{C(i, j + 1) - C(i, j)}. \quad (7)$$

Then $\sin^2 \theta = \frac{\text{tg}^2 \theta}{1 + \text{tg}^2 \theta}$, and $\cos^2 \theta = \frac{1}{1 + \text{tg}^2 \theta}$.

These equations make it possible to avoid the time-consuming calculation of the sine and cosine by the Taylor series expansion method.

Note that the corner pixels of the 3x3 window are not involved in the calculations. To improve the transmission of non-horizontal and non-vertical contours, we introduce into the algorithm a branch that takes into account the angular pixel values of the 3x3 pixel window. A flowchart of the gradient noise reduction algorithm is shown in Figure 8.

When considering the diagonal contour $30^\circ < |\theta| < 60^\circ$, one of the branches works when the contour can be considered diagonal and at the upper part of the 3x3 window the contour is at the top, and in the right part of the window the contour is at the bottom. Similarly, the other branch works if the diagonal contour is at the bottom of the 3x3 window. The results of applying the algorithm with the values of the parameters $a = 0.5$ and $b = 0.5$ are shown in Figures 9-11.

As can be seen from Figures 9-11, at a height of the sun exceeding 10 degrees, a completely acceptable result of inverse filtering is provided.

Laplacian based noise reduction

The contours and inhomogeneities in the 3x3 window can be detected using the Laplacian operator. It is defined by the equation (8):

$$\begin{aligned} \text{Lap}(i, j) &= \left| \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} \right| = \\ &= \left| (C(i, j - 1) + C(i, j + 1) - 2 \cdot C(i, j)) + \right. \\ &\quad \left. + (C(i - 1, j) + C(i + 1, j) - 2 \cdot C(i, j)) \right|. \end{aligned} \quad (8)$$

Obviously, for a white dot on a black background, it results in $\text{Lap}(i, j) = 2 \cdot |0 + 0 - 4095| = 8190$. For an absolutely sharp horizontal white line on a black background, we have $\text{Lap}(i, j) = |0 + 0 - 4095| = 4095$.

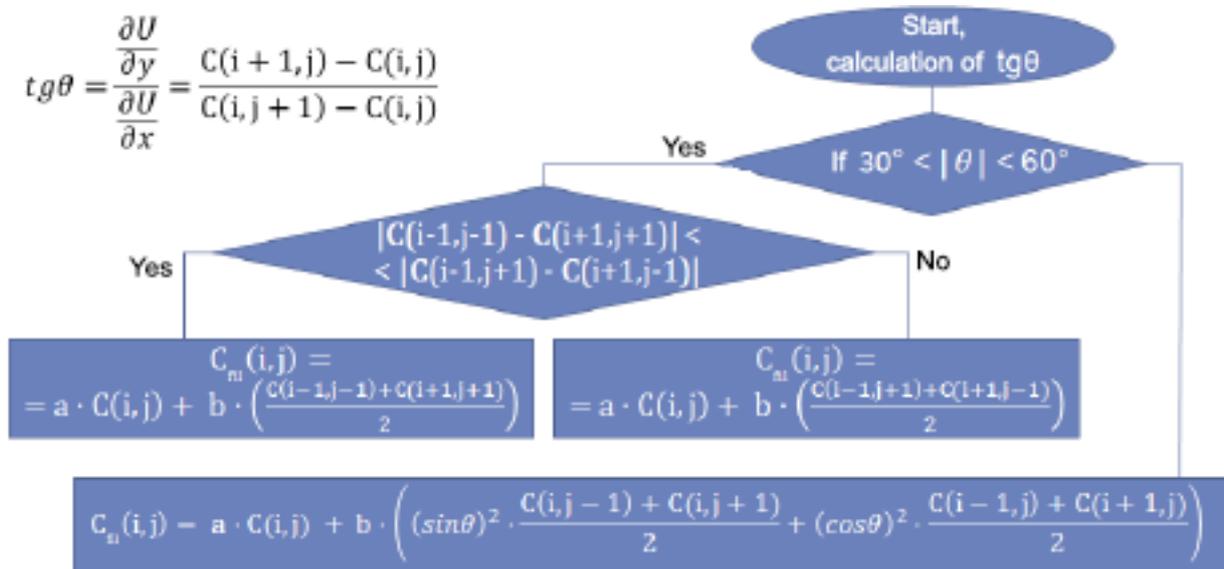


Fig. 8. Flowchart of the gradient noise reduction algorithm.



Fig. 9. $a=0,5; h_s=2^\circ$



Fig. 10. $a=0,5; h_s=5^\circ$



Fig. 11. $a=0,5; h_s=10^\circ$

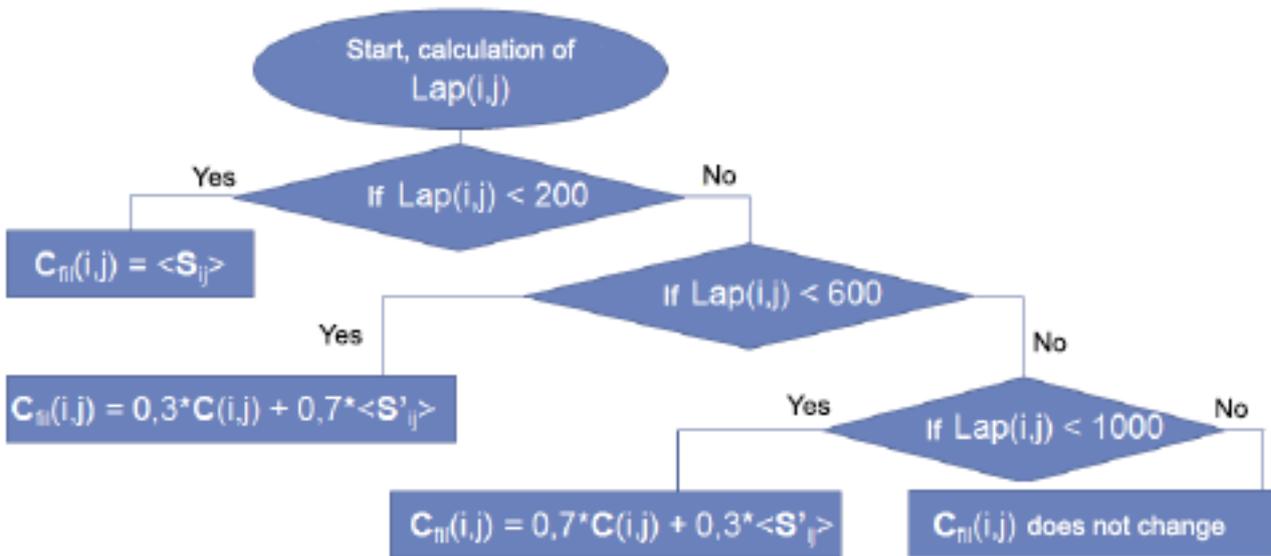


Fig. 12. Flowchart of the Laplacian noise reduction algorithm.

Consider an algorithm based on the three threshold values of the Laplacian at the current point of the image $C(i, j)$. The flowchart is shown in Fig. 12.

Here S_{ij} is a 3x3 pixel window, and S'_{ij} is the surroundings of the point (i, j) , that is, the 3x3 window without the central pixel. Accordingly, $\langle S_{ij} \rangle$ is the average value of the signal of 9 pixels, and $\langle S'_{ij} \rangle$ is the average value of 8 pixels.

Thus, we prevent the changing of the signal at points with very strong local inhomogeneities ($Lap(i, j) \geq 1000$) to avoid blurring of small details. On the contrary, for equiluminous sections, a branch of the algorithm that averages all 9 pixels of the window with the same weights is used.

We apply the described algorithm for the preliminary noise reduction performed before inverse filtering. The result of obtaining of the A_{inv} images for different elevations of the Sun is shown in Fig. 13–14.

As can be seen from Fig. 13-14, the result of image restoration using this algorithm of preliminary noise cancellation is worse than using the gradient method, since there are individual dark or light points in homogeneous areas.

Resolution regularization

In addition to the use of adaptive noise reduction algorithms, the problem of dividing the Fourier image by a small value of the MTF can be solved by regularizing the solution described in [4]. The idea of regularization is described by formula (9):

$$A_{inv} = F^{-1} \left(\frac{F(C)}{MTF(\nu) + Reg(\nu)} \right), \quad (9)$$

Where: $Reg(\nu)$ is the regularization function.

From the analysis of formula (9) we can formulate two requirements for the regularization function:



Fig. 13. $A_{inv}, h_s=2^\circ$



Fig. 14. $A_{inv}, h_s=10^\circ$



a) $\alpha = 0.01$



b) $\alpha = 0.5$



c) $\alpha = 2$

Fig. 15. Simulation of inverse filtering without prior noise cancellation

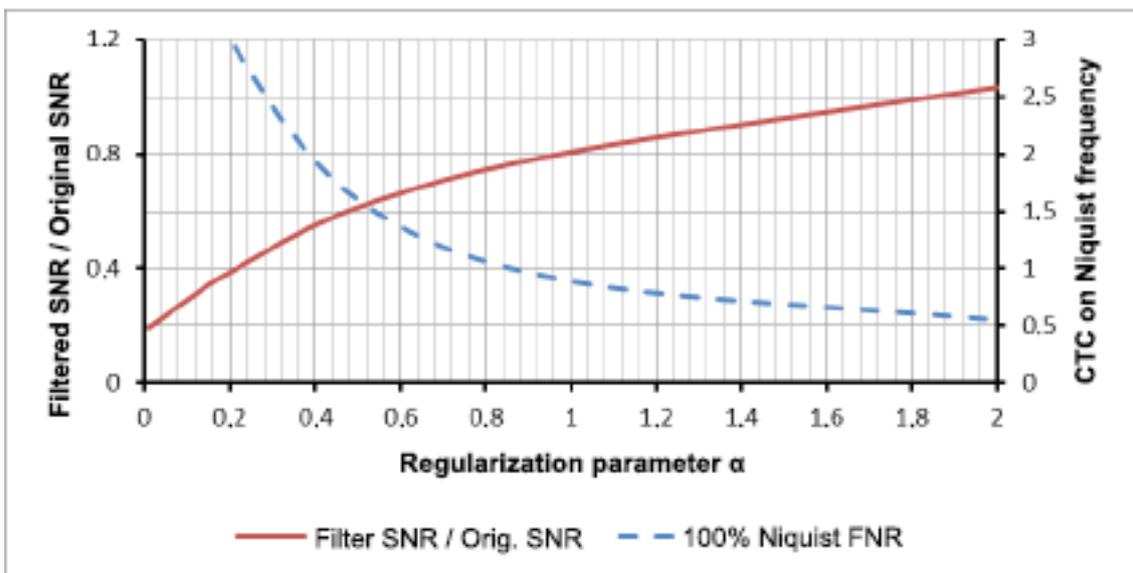


Fig. 16. Dependences of the contrast transfer coefficient (CTC) and the change in signal-to-noise ratio (SNR after filtering / SNR initial) on the coefficient α .

$Reg(0) = 0$, since at small spatial frequencies the distortions are minimal and regularization is not needed.

$Reg(1)$ must be non-zero, but not close to 1, so as not to allow large distortion of the restored image.

The use of regularization instead of the preliminary noise cancellation is of interest, since it requires fewer operations with the data array. We analyze various variants of the regularization function.

Quadratic dependence

$Reg(v) = \alpha \cdot vt^2$, where α is the experimentally chosen dimensionless coefficient. We will try to experimentally optimize the value of α . Figure 15 shows the simulation of inverse filtering for different α values without prior noise cancellation.

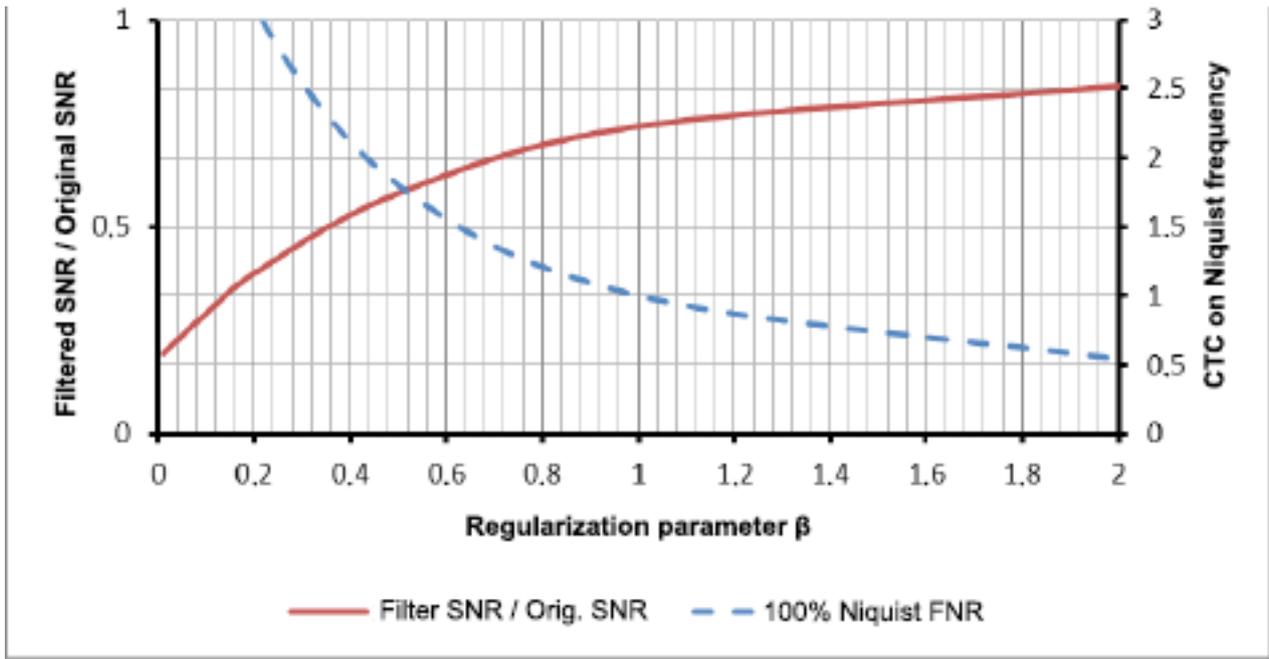


Fig. 17. Dependencies of the CTC and changes in the signal-to-noise ratio (SNR after filtration / SNR initial) from the coefficient β .

Table 1. Main characteristics of inverse filtering options

pre-noise reduction variant	SNR ₅₀		FNR	
	in the source image	on the processed image	0.5 v _N	v _N
without noise reduction	31	5.6	3	8.6
pseudo-median	31	15	2.4	-5
gradient algorithm	31	9	3	7
Laplacian algorithm	31	21	3.7	8.6
regularization $\alpha \cdot v^2$ ($\alpha = 0.8$)	31	$31 \cdot 0.7=22$	1	1
the regularization $\beta \cdot v^3$ ($\beta=1$)	31	$31 \cdot 0.75=23$	1	1

When analyzing Figures 15a-15c, it is seen that as the α increases, the signal-to-noise ratio increases, but the contrast decreases. The nature of this relationship is shown in Fig. 16. Next, we will talk about the signal-to-noise ratio given for a signal that is half of the maximum, which corresponds to an albedo of 0.5. Denote this value SNR₅₀.

Assuming that the CTC should be close to 1, like the change in SNR, we determine that the optimal value of α is in the range from 0.6 to 1.

Cubic Dependence

Consider another regularizing function: $Reg(v) = \beta \cdot v^3$, where β is the experimentally selectable dimensionless coefficient. Then the effect of the change in the denominator of fraction (3) at low frequencies will be less than for quadratic regularization. Figures 23a - 23c show the simulation of inverse filtering without noise reduction. Analogously to the quadratic regularization, the signal-to-noise ratio increases with increasing β , but

the CTC falls off. The nature of this relationship is shown in Figure 17.

Assuming that the contrast transfer coefficient should be close to 1, like the change in SNR, we find that the optimal value of β is in the range from 0.8 to 1.2.

Conclusion

For a mathematical description of the quality of inverse filtering, let us summarize in Table 1 the main characteristics of its variants. Here FNR ($0.5 \nu_N$) is the ratio of the target object image contrasts with a spatial frequency of 50% of the Nyquist frequency for the processed and original images. The negative value of the FNR corresponds to the inversion of the bars of the test object: the black bar becomes white and vice versa. A similar phenomenon can be observed in Figure 6 with pseudo-median filtration.

Let us analyze the data in Table 1 and the results of preliminary noise reduction algorithms. Optimal solutions for image preparation are the use of the gradient method of noise reduction or cubic regularization. In addition, it is possible to use these operations jointly when the regularization is performed on the Fourier image that underwent noise reduction.

References

1. Baklanov A.I. *Sistemy nablyudeniya i monitoringa: uchebnoe posobie* [Monitoring and monitoring systems: textbook]. Moscow, BINOM. Laboratoriya znaniy, 2009, 234 pp. (in Russian)
2. Gonsales R., Vuds R. *Tsifrovaya obrabotka izobrazheniy* [Digital image processing]. Moscow, Tekhnosfera, 2005, 1072 pp. (in Russian)
3. Akinshin N.S., Esikov O.V. Pseudomediannaya fil'tratsiya dlya obrabotki opticheskikh izobrazheniy [Pseudo-median filtering for processing optical images]. *Materialy XII nauchno-tekhnicheskoy konferentsii "Sistemy nablyudeniya, monitoringa i distantsionnogo zondirovaniya Zemli"* [Proceedings of the XII Scientific and Technical Conference "Systems of Earth observation, monitoring and remote sensing"]. Kaluga, Manuskript, 2015, 436 pp. (in Russian)
4. Kashkin V.B., Sukhinin A.I. *Tsifrovaya obrabotka aerokosmicheskikh izobrazheniy* [Digital processing of aerospace images]. Digital resource, a summary of lectures. Krasnoyarsk, IPK SFU, 2008. (in Russian)

Features of Calculation and Design of High-Speed Radio Links for Earth Remote Sensing Spacecraft

A.N. Ershov, *ean241@mail.ru*

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

V.V. Berezkin, *Cand. Sci. (Engineering), nkpor@spacecorp.ru*

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

S.V. Petrov, *petrov_sv@spacecorp.ru*

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

A.V. Petrov, *nkpor@spacecorp.ru*

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

D.A. Pochivalin, *nkpor@spacecorp.ru*

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

Abstract. The paper discusses the issues of selecting the parameters of the fine structure of modulated signals to optimize frequency-energy resources when designing a high-speed (hundreds and thousands Mbit/s) radio links for data transfer. Their features is application of signal-code constructs with the kinds of a high-order modulation and coding methods with high code rates.

Based on the analysis of a model of a radio link with nonlinearity, which limits a peak power in the output of the power cascade of the transmitter and a modulating signal with a nonzero value of a peak factor, the estimations of the main characteristics of a radio link are made.

It is noted that when realizing the set value of a signal-to-noise ratio, the fulfilment of the norms on providing the ratio of the values of the radiating powers in the main and adjacent channels can be limited.

The reasons of the spread in the estimations of the radio link parameters given in domestic and foreign publications on the question under consideration are analyzed.

It is marked, that the results received in the paper, are expedient to use during designing, tests, measurements, and optimization of the parameters of onboard and ground complexes of the radio lines of high-speed data transmission aimed to operate in Earth remote sensing (ERS) systems.

Keywords: high-speed radio links, modulated signals, signal-code constructs, frequency-energy resources, Nyquist filter, peak factor

The main feature of high-speed radio links (with transmission speeds from several hundred Mbit/s to several Gbit/s) is the need to use high modulation rate ($m \geq 3$) and code rate ($R \geq 0.8$), which is determined by strict requirements imposed by limitations in the availability of frequency resources. The modulation multiplicity with the value $m > 3$ determines the use of a multilevel amplitude-phase structure, since for $m > 3$ this structure, compared with a single-level (phase only structure), provides a maximum of the minimum Euclidean distance between the points of the equivalent "signal constellation".

In this case, the amplitude-phase structure implies two types of signal constellation: in the form of a rectangular lattice (QAM) and in the form of points located on a series of concentric circles (APSK).

The quantitative energy characteristics corresponding to these types of modulation are determined by two parameters:

- the minimum Euclidean distance between the points of the signal constellation;
- peak factor of the modulated signal (in a nonlinear channel with limited peak power).

The experience of design of complexes and equipment of high-speed radio links shows that these features are not always taken into account by developers of technical requirement specifications for both systems and instruments.

For the APSK, the error probability is determined by averaging the probability integral based on a search for all Euclidean distances of the signal constellation:

$$P_e \leq \frac{1}{M} \sum_{x \in X} \sum_{\substack{x' \in X \\ x' \neq x}} Q \left(\sqrt{\frac{E_s |x - x'|^2}{2N_o}} \right)$$

$E_s = |x - x'| = d'$ is the Euclidean distance ($d'_{ij} = \sqrt{(r_{p(i)}^2 + r_{q(j)}^2 - 2 \cdot r_{p(i)} \cdot r_{q(j)} \cdot \cos(\beta)_{ij})}$), β is the angle between the points of the constellation [1]).

For QAM, the error probability is determined by the following formula [2, p. 586]:

$$P_e \approx \frac{2(1 - L^{-1})}{\log_2 L} Q \left[\sqrt{\left(\frac{3 \log_2 L}{L^2 - 1} \right) \frac{2Eb}{N_o}} \right]$$

In both formulas, $Q(x) = 1/\sqrt{2\pi} \times \int_x^\infty e^{-\frac{t^2}{2}} dt$ is the probability integral, for small error probabilities

replaced by the approximation [2]: $Q(x) \approx \exp(-0,5x^2)/(x\sqrt{2\pi})$.

Peak-factor PARP (pick / average ratio power) is the ratio of the peak power P_{peak} to the average power of P_{av} :

$$PARP = P_{peak}/P_{av} = (U_{peak}/U_{rms})^2.$$

For a signal at period 0,T [6]:

$$PARP = \frac{\max_{t \in [0,T]} |s(t)|^2}{\frac{1}{T} \int_0^T |s(t)|^2 dt}.$$

Or the same in terms of discrete components [3]:

$$PARP = \frac{\max_{n=0,1,\dots,N-1} |x_n|^2}{\frac{1}{N} \sum_{n=0}^{N-1} |x_n|^2}.$$

The value of PARP in actual transmission systems depends on two parameters:

- the multiplicity of the modulation type;
- the rounding factor of the matched Nyquist filter, which determines the bandwidth of the emitted signal on air.

Table 1 lists the main parameters for several types of modulation that are currently in use or are expected to be used in the satellite systems of high-speed information transmission in the foreseeable future. The complex of these parameters is partially calculated by the above formulas and partially borrowed from these sources [3, 4, 5, 7].

Table 2 shows the PARP values of the digital filter (DF), depending on the selected rounding factor α [8].

Since the bandwidth occupied by the spectrum of the modulated signal is defined as $\Delta f \approx (1 + \alpha) \times f_c$ (where f_c is the character frequency of the signal on air), the desire to save frequency resources (choosing a smaller value of α) calls for a larger extension between the operating point of the nonlinear amplifier and the point of its saturation and, consequently, leads to losses of energy resources.

The resulting peak-factor is formed by summing the data by the peak factor, given in Table 1 and Table 2.

For example, Table 3 gives comparative data on the total PARP value for a DF with different rounding and modulation coefficients of type 16APSK and 16QAM.

According to the data in Table 3, two conclusions can be drawn:

1. Since the resulting peak factor (PARP) of APSK is less than that of QAM, this makes it possible (in the case of peak power limitation) to use the power of the output stage of the transmitter more efficiently and, for this reason, APSK is used in space transmission systems

Table 1

Modulation type	BPSK, QPSK	8PSK	16 PSK	16QAM	16APSK (4+16)	32 QAM	32 APSK (4+12+16)	64QAM	64APSK (4+12+16+32)
Peak factor [db]	0	0	0	2.56	1.065, (R=4/5), $\gamma^*=2.75$	2.3	2.06	3.68	2.62
E_{bit}/N_o [db] at $P_{er}=10^{-6}$	10.6	13.8	18	14.4	15.5	16	17.4	18.8	19.5

Note: γ^* is the ratio of the diameters of the outer and inner rings of the signal constellation for the code rate (R = 4/5).

Table 2

α	0.15	0.2	0.3	0.4	0.5
PARP [dB]	6.3	5.6	4.5	3.5	2.8

Table 3

α	0.15	0.2	0.3	0.4	0.5
PARP Σ APSK [dB]	7.36	6.66	5.56	4.66	3.86
PARP Σ QAM [dB]	8.86	8.16	7.06	6.06	5.36

in an overwhelming number of cases, despite the fact that QAM has a minimal Euclidean distance slightly larger than that of APSK.

2. These tables make it possible to estimate the required OBO (Output Back Of) value, the linearity margin from the saturation point of the PA and the corresponding change in the average output power of the transmitter and, accordingly, the efficiency.

Figure 1 shows the corresponding schemes of a real X-band transmitter with an average output power $P_{out.av} = 8$ watts.

The graph in Figure 1 shows that in this particular case, the presence of a peak factor in the modulating signal requires a margin of linearity of the power amplifier stage of about 4.6 dB.

Simultaneously with the requirement to provide the required transmitter power in order to realize the required signal-to-noise ratio, there is a requirement not to exceed the specified level of radiation in the adjacent channel.

In this case, the ACPR (Adjacent Channel Power Ratio) parameter is used as a measure of the linearity of systems operating using multi-position modulation signals, which is determined by the following relation [9]:

$$ACPR=10\lg(P_{mc}/P_{ac})$$

where R_{mc} is the average signal power in the main channel, and P_{ac} is the average signal power in the adjacent channel, as shown in Fig. 2.

A mode with a given level of radiation power in an adjacent channel is provided by a simulation or experimentally by changing the load level.

The effect of channel nonlinearity on the transmission quality can be estimated from the curves in Fig. 3, which show the dependence of the error probabilities on the energy of the radio link - $P_{er}=f(E_{bit}/N_o)$ for 16QAM and 16APSK for $\alpha = 0.35$.

The meaning of these curves is that in the case of 16APSK, the achievable average power (and, hence, E_{bit}/N_{er}) for a certain dependence of the nonlinearity of the PA is about 1 – 2 dB more. This leads to the possibility of realizing the magnitude of the error probability by about 1 – 2 orders of magnitude less than in the case of using QAM.

It should be noted that in the sources cited in the literature, there is a certain variation in the absolute and relative noise immunity of different types of signals. This fact is due to the multi-vector nature and some uncertainty in setting the initial data for the calculation.

The total ambiguity and error of calculations consists of the following components:

- errors in the computation of the probability integral and its approximations for small error probabilities;
- an ambiguous choice of the value of the rounding factor from the possible range (usually ~ 0.2 – 0.5);
- some heuristics when selecting the required value of the OBO;

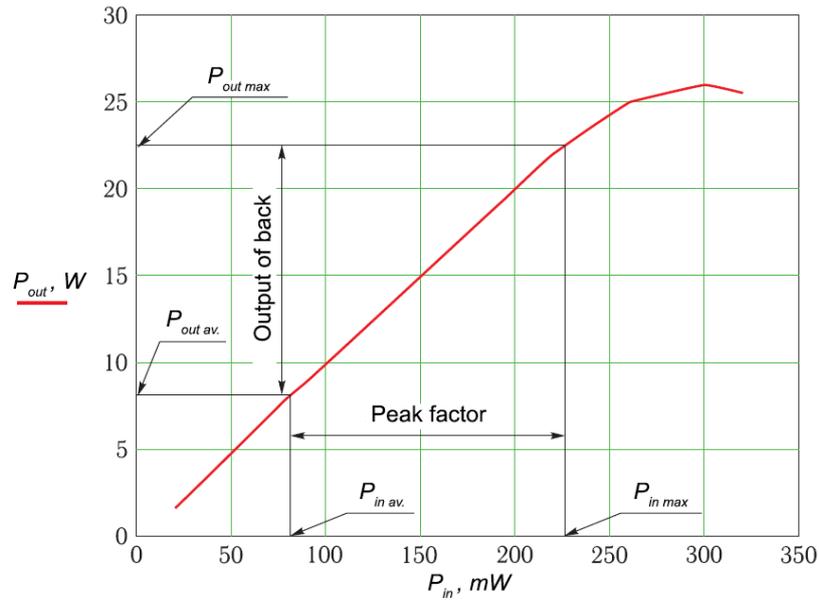


Fig. 1. Influence of the peak factor on the output parameters of the PA

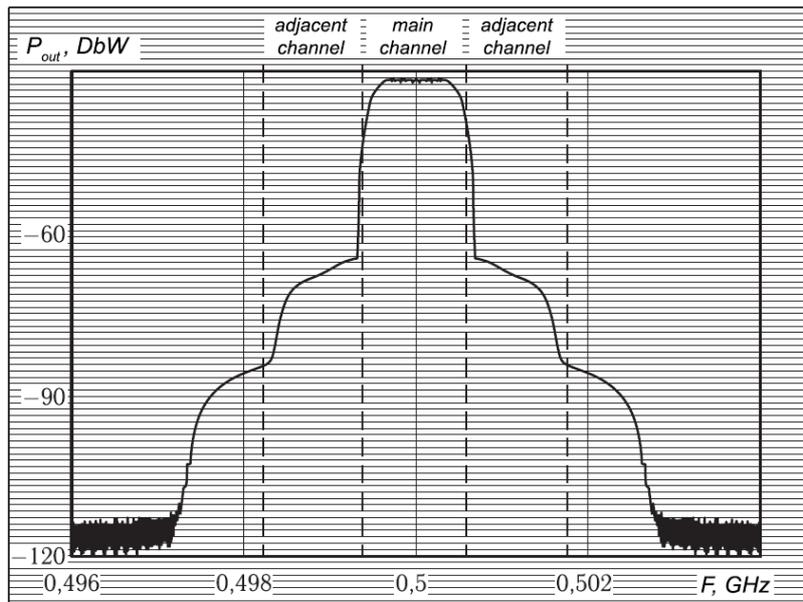


Fig. 2. Typical distortions of the signal spectrum due to the nonlinearity of the transmitter

- the use in the calculation of the value of the Euclidean distance either minimum or averaged (sometimes truncated) by the signal constellation;
- the use of various methods for approximating the parameters of the nonlinearity of output power stages (including taking into account the nonlinearity of the AGC of the transmitter);
- the use of slightly different relationships between the radii of circles of the signal constellation, which are usually chosen different for different code rates.

Conclusion

In the article the questions concerning the choice of parameters of the transmission path of the high-speed radio link are considered when using the Nyquist filter in the transmitter modulator.

It is shown that in the case of limiting the peak power of the transmitter and the requirement to save frequency resources by reducing the rounding factor and selecting an APSK of higher multiplicity, it is necessary to take

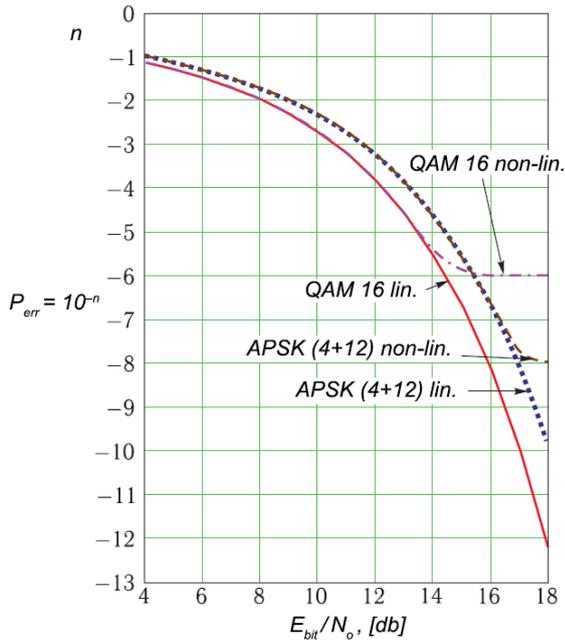


Fig. 3 Comparative dependences $P_{err} = f(E_{bit}/N_o)$ for 16QAM and 16APSK in a linear and nonlinear channel

into account the increase in the magnitude of the peak factor of the transmitter signal, which leads to a decrease in the emitted average power and the efficiency of the transmitter.

It is noted that in the process of increasing the power of the emitted signal, norms should be provided for the signal-interference ratio in the main and adjacent channels.

The reasons for the variation in radio link parameter estimates, which are presented in Russian and foreign papers, are analyzed, and corresponding quantitative estimates are made.

References

- Jordanova L., Laskov L., Dobrev D. Constellation and Mapping Optimization of APSK Modulations used in DVB-S2. *Engineering, Technology & Applied Science Research*. 2014, Vol. 4, No. 5, pp. 690–695. Available at: <https://zenodo.org/record/14687#.WXHwmS7WgeE> (accessed 30 October 2017).
- Sklar B. *Tsifrovaya svyaz'. Teoreticheskie osnovy i prakticheskoe primeneniye* [Digital Communications: Fundamentals and Applications]. Moscow, Vil'yams. 2nd edition, revised. 2007, 586 p. (translated from English).
- Chayratsami P., Thuaykaew S. *The Optimum Ring Ratio of 16-APSK in LTE Uplink over Nonlinear System*. Available at: http://www.icact.org/upload/2013/0554/20130554_Journal_B.pdf (accessed 30 October 2017).
- Gaudenzi R., Fabregas A., Martinez A. Turbo-Coded APSK Modulations Design for Satellite Broadband Communications. *International Journal of Satellite Communications and Networking*. Published online 19 May 2006 in Wiley InterScience (www.interscience.wiley.com). Available at: http://www.dtic.upf.edu/~amartinez/publications/apsk_ijsc_published.pdf (accessed 30 October 2017).
- Lottici V., Luise M., Reggiannini R. *Adaptive Nonlinear Compensation of Satellite Transponder Distortion for High-Level Data Modulations*. Available at: https://www.researchgate.net/publication/228600391-APSK_Coded_Modulation_Schemes_for_Nonlinear_Satellite_Channels_with_High_Power_and_Spectral_Efficiency (accessed 30 October 2017).
- Baldi M. A comparison between APSK and QAM in wireless tactical scenarios for land mobile systems. *Journal on Wireless Communications and Networking*. Published October 19, 2012. Available at: <http://jwcn-urasipjournals.springeropen.com/articles/10.1186/1687-1499-2012-317> (accessed 30 October 2017).
- John S. Seybold. *Crest Factors for QAM Signals*. Available at: <http://mathscinotes.com/2012/11/crest-factors-for-qam-signals/> (accessed 30 October 2017).
- RF signal processing. June 2002. Output Back-Off Requirements For Root-Raised, Cosine-Filtered Digital Signals*. Available at: <http://defenseelectronicsmag.com/site-files/defenseelectronicsmag.com/files/archive/rfdesign.com/images/archive/0602Seybold50.pdf>
- Malev A.S., Solov'ev A.M., Shutov V.D. Podkhody k optimizatsii metodov formirovaniya signalov s mnogopozitsionnoy modulyatsiey po minimumu pik-faktora [Optimization of signal's parameters with multiway modulation for the purpose of crest factor minimization]. *Teoriya i tekhnika radiosvyazi*

- [Radio communication theory and equipment]. Voronezh, OAO "Kontsern "Sozvezdie", 2012, No. 2, pp. 50–56. (in Russian)
10. Pedro E. Santacruz. *Analysis of the effects of nonlinear amplification on turbo Coding*. Thesis Presented to the Faculty of the Graduate School of The University of Texas at El Paso in Partial Fulfillment of the Requirements for the Degree of Master of Science. Available at: <http://mpc.ece.utexas.edu/users/pesantacruz/files/thesis.pdf> (accessed 30 October 2017).
 11. Ershov A.N., Petrov S.V., Berezkin V.V., Petrov A.V., Pochivalin D.A. *Metody proektirovaniya i apparatnoy realizatsii tsifrovyykh fil'trov dlya vysokoskorostnoy radiolinii v sistemakh DZZ* [Methods of Design and Hardware Implementation of Digital Filters for High-Speed Radio Link in Earth Remote Sensing Systems]. *Raketno-kosmicheskoe priborostroenie i informatsionnye sistemy* [Rocket-Space Device Engineering and Information Systems]. Moscow, FIZMATLIT, 2017, Vol. 4, No. 1, pp. 25–31. (in Russian)

The Research of Exciter Frequency Characteristics of a Quad-Band Antenna Based on a Corrugated Horn

D.D. Gabriel'ean, *Dr. Sci. (Engineering)*, rniirs@rniirs.ru

FSUE "Rostov-on-Don Research Institute of Radio Communications", Rostov-on-Don, Russian Federation

V.I. Demchenko, *Cand. Sci. (Engineering)*, rniirs@rniirs.ru

FSUE "Rostov-on-Don Research Institute of Radio Communications", Rostov-on-Don, Russian Federation

A.E. Korovkin, rniirs@rniirs.ru

FSUE "Rostov-on-Don Research Institute of Radio Communications", Rostov-on-Don, Russian Federation

D.Ya. Razdorkin, rniirs@rniirs.ru

FSUE "Rostov-on-Don Research Institute of Radio Communications", Rostov-on-Don, Russian Federation

A.V. Shipulin, rniirs@rniirs.ru

FSUE "Rostov-on-Don Research Institute of Radio Communications", Rostov-on-Don, Russian Federation

Yu.I. Poltavets, *Cand. Sci. (Engineering)*, rks0901@yandex.ru

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

Abstract. The objective of the paper is to study the frequency characteristics of the exciter based on a corrugated horn for a quad-band antenna for satellite communications. The problems of the construction of the corrugated horn as the basis for the radiating system of the quad-band reflector antenna for satellite communications with overlapping of C-, X-, Ku-, and Ka-bands are considered. The conducted analysis of the interrelation of the characteristics of the multiband reflector antenna with the characteristics of the radiating system permits defining a set of the requirements to the parameters of the corrugated horn, which is the base of a radiating system. The corrugated horn, which provides the required characteristics in C-, X-, Ku- and Ka-bands, is developed. The radiation patterns of the corrugated horn within the specified bands and the frequency dependences of the mode converter providing the interface of the horn with a circular feed waveguide are analyzed. The experimental checking proved that within C-, X-, Ku-, and Ka-bands, an average value of the VSWR does not exceed 1.12, and the maximum value makes 1.15 at the frequency.

Keywords: multiband reflector antenna, corrugated horn, performance indicators of the radiating system and feed horn

Introduction

One of the ways to boost the capacity of communication links, volume of the obtained surveillance information at signal intercept in satellite communication links (SCL) is a simultaneous usage of several frequency ranges. Since SCL apply reflector antennas, thus the work in several frequency ranges can be provided by employing either several single-band reflector antennas or one multiband reflector antenna (MBRA) being as good as single-band antennas in regard to radiation characteristics in each of frequency ranges. The last seems to be more preferable, since it allows one to reduce considerably costs when organizing and servicing communication links and data reception.

It is obvious that solving the task of MBRA creation is connected, first of all, with development of original multiband exciters. Hence, the aim of the paper is to examine frequency characteristics of the exciter based on the corrugated horn for the quad-band satellite communication antenna.

The tasks to be solved:

- To establish interrelation of the characteristics of the exciter system with the characteristics of MBRA
- To develop and select the parameters of a corrugated horn to build a radiation system of the quad-band reflector antenna
- To examine experimentally the characteristics of the radiation of the prototype of the developed horn

Interrelation of the characteristics of the exciter system with the characteristics of MBRA

A performance indicator of MBRA can be presented by the following ratio [1]:

$$P_{MBRA} = \prod_{j=1}^J \prod_{i=1}^2 \left(H_{j,i}^{(0)} K_{j,i}^{(1)} K_{j,i}^{(2)} \right)^{m_{j,i}} \quad (1)$$

where $H_{j,i}^{(0)}$ is the noise Q-factor of MBRA in the j -th frequency range when receiving the signals of the i -th polarization; $K_{j,i}^{(1)}$ is the coefficient taking into account the decrease of the signal-to-noise ratio (SNR) due to depolarization effects in the antenna and waveguide transmission line (AWTL) of the antenna; $K_{j,i}^{(2)}$ is the coefficient considering the decrease of the noise Q-factor due to the errors in beam pointing of the antenna.

A noise Q-factor is determined as follows [2]:

$$H_{j,i}^{(0)} = \frac{4\pi S}{\lambda_j^2} \cdot \eta_{j,i} \quad (2)$$

where S is the square of the main mirror; $\eta_{j,i}$ is the coefficient of using of the area (CUA) in the j -th frequency range for the signals of the i -th polarization; λ_j is the wavelength of radiation in the j -th frequency range.

The value of CUA $\eta^{(j,i)}$ allowing for the matching of the exciter with the feeder circuit in case MBRA based on [2] can be given as follows:

$$\eta_{j,i} = \eta_1^{(j,i)} \cdot \eta_2^{(j,i)} \cdot \eta_3^{(j,i)} \cdot \eta_4^{(j,i)} \cdot \eta_5^{(j,i)} \cdot \eta_6^{(j,i)} \cdot \eta_7^{(j,i)}. \quad (3)$$

Private indicators of CUA are determined with the following ratios:

- Aperture coefficient of usage

$$\eta_1^{(j,i)} = 2ctg^2\left(\frac{\theta_0}{2}\right) \cdot \frac{\left| \int_0^{\theta_0} \left(|F_E^{(j,i)}| + |F_H^{(j,i)}| \right) \cdot tg(\theta/2) d\theta \right|^2}{\int_0^{\theta_0} \left(|F_E^{(j,i)}|^2 + |F_H^{(j,i)}|^2 \right) \cdot \sin \theta d\theta} \quad (4)$$

- Coefficient due to the intercept of the share of the energy flow with the primary radiator

$$\eta_2^{(j,i)} = \frac{\int_0^{\theta_0} \left(|F_E^{(j,i)}|^2 + |F_H^{(j,i)}|^2 \right) \cdot \sin \theta d\theta}{\int_0^{\pi} \left(|F_E^{(j,i)}|^2 + |F_H^{(j,i)}|^2 \right) \cdot \sin \theta d\theta} \quad (5)$$

- Coefficient considering inequality of the phase distribution in the aperture

$$\eta_3^{(j,i)} = \frac{\left| \int_0^{\theta_0} \left(F_E^{(j,i)} + F_H^{(j,i)} \right) \cdot tg(\theta/2) d\theta \right|^2}{\left| \int_0^{\theta_0} \left(|F_E^{(j,i)}|^2 + |F_H^{(j,i)}|^2 \right) \cdot tg(\theta/2) d\theta \right|^2} \quad (6)$$

- Coefficient taking into account the transition of the share of the radiated energy into the cross-polarized component

$$\eta_4^{(j,i)} = \frac{\left| \int_0^{\theta_0} \left(|F_E^{(j,i)}| + |F_H^{(j,i)}| \right)^2 \cdot \sin \theta d\theta \right|^2}{\left| \int_0^{\theta_0} \left(|F_E^{(j,i)}|^2 + |F_H^{(j,i)}|^2 \right) \cdot \sin \theta d\theta \right|^2} \quad (7)$$

- Coefficient considering the covering of the radiator by the primary radiator

$$\eta_5^{(j,i)} = \frac{\left| \int_{\theta_B}^{\theta_0} \left(|F_E^{(j,i)}| + |F_H^{(j,i)}| \right)^2 \cdot \sin \theta d\theta \right|^2}{\left| \int_0^{\theta_0} \left(|F_E^{(j,i)}|^2 + |F_H^{(j,i)}|^2 \right) \cdot \sin \theta d\theta \right|^2} \quad (8)$$

- Coefficient taking into account the errors in radiator development

$$\eta_6^{(j,i)} = \exp \left[- \left(\frac{4\pi\mathcal{E}}{\lambda_j} \right)^2 \right] \quad (9)$$

where \mathcal{E} is the root-mean-square deviation of the surface of the radiator from the set form.

Moreover, when defining of the amplification factor of MBRA, it is necessary to take into account the matching of the radiator with the antenna external device (AED) determined by VSWR or the reflection index $\Gamma_{j,i}$ in the sectional view of the coupling of the radiator with the AED in the form of the multiplier:

$$\eta_7^{(j,i)} = 1 - |\Gamma_{j,i}|^2 \quad (10)$$

The ratios (1)–(10) define the interrelation of the MBRA efficiency with the parameters of the primary radiator.

The given ratios show that primary radiators of high-effective MBRA used in SCL should have the following characteristics in their ranges:

- Axisymmetric amplitude radiation pattern (RP) in the limits of the working corner sector
- A low level of side lobes out of the working corner sector
- A quasistationary (poorly dependent on frequency) position of the phase center
- A low level of the cross-polarized radiation
- A high level of the matching with the feed waveguide.

Development and selection of the parameters of a corrugated horn to build a radiation system of a quad-band reflector antenna

Varieties of corrugated cone horns with a feed circular waveguide have a broad application as primary radiators of MBRA. The highest efficiency of corrugated cone horns is provided by exciting the hybrid mode HE_{11}

having the least critical length of the wave and suppression of the higher modes in the required wave ranges. The last is achieved by the realization of the conditions of exciting of “speed” waves in these ranges [2].

The field of the mode HE_{11} in the aperture of the corrugated horn of the waveguide is determined by the following components:

$$\begin{aligned} E_x &= A_1 J_0 \left(\frac{\nu_{11}}{R} \cdot r \right) - A_2 J_2 \left(\frac{\nu_{11}}{R} \cdot r \right) \cdot \cos(2\varphi) \cdot \frac{X-Y}{k \cdot R}, \\ E_y &= A_2 J_2 \left(\frac{\nu_{11}}{R} \cdot r \right) \cdot \sin(2\varphi) \cdot \frac{X-Y}{k \cdot R}. \end{aligned} \quad (11)$$

where $J_0(\bullet)$, $J_2(\bullet)$ are the Bessel functions of the first class of zero and second orders, respectively; k is the wave number of a free space; A_1 , A_2 are the amplitudes of the components determined by the parameters of the horn; X and Y are the normalized impedance and conductivity of the corrugated structure; ν_{11} is the root of the Bessel function of a zero order.

On the sides of the horn at $r=R$, the normalized impedance and conductivity of the corrugated structure can be defined as follows:

$$\begin{aligned} X &= -iZ_0^{-1} E_\varphi / H_z, \\ Y &= iZ_0 H_\varphi / E_z. \end{aligned} \quad (12)$$

When the condition $X=Y=0$ is performed, the field of the wave HE_{11} in the aperture of the horn does not depend on the angle φ and $E_y=0$ that defines the lack of the cross-polarized component and axial symmetry of a radiation field. The condition called the condition of the balance of hybrid waves correspond to the specific ratio of lateral components of electric and magnetic fields.

It can be accepted that $X \approx 0$ in the corrugated field. The condition of the balance for the hybrid mode HE_{11} and minimum level of transformation into the mode EH_{12} greatly contributing into the cross-polarized radiation is achieved at the closest to the zero on the border of the corrugated surface of the conductivity horn. Thus, to fulfill the condition of the balance of hybrid waves, it is necessary to meet $Y \approx 0$. The fulfillment of this condition is achieved at the following parameters of the corrugated structure [2]:

$$t < \lambda/10, \quad 1 - t < \lambda/10$$

Fig. 1 depicts the distribution of the electric field intensity in the aperture of the corrugated horn for the ratio $\xi = A_2(X-Y)/(kA_1R)$ equal to 0, 0.1, and 0.2, respectively.

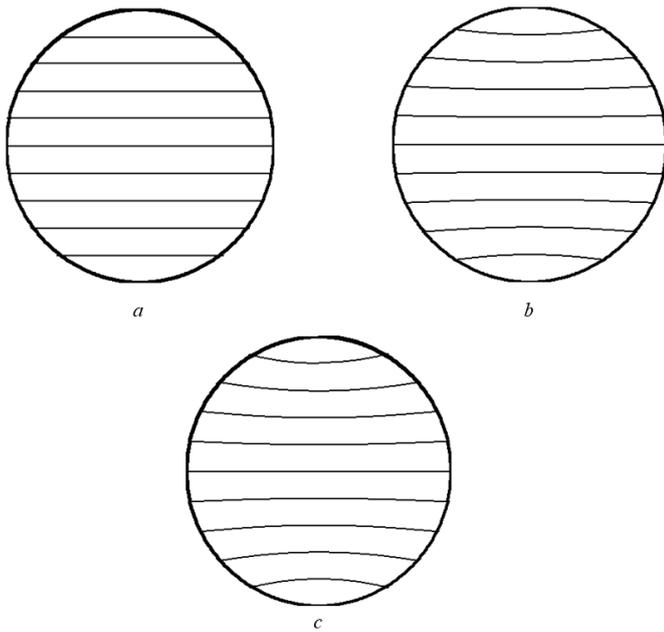


Fig. 1. The lines of electric field intensity in the aperture of the corrugated waveguide:
 a) $\xi=0$; b) $\xi=0.1$; c) $\xi=0.2$.

The conducted research has shown that the best matching for the main and higher modes and for the same level of the power selection, the minimum number of grooves in the structure, stability of the radiation pattern in the frequency band is provided when using the grooves of a trapezoidal cross-section for the horn, and a complicated cross-section for the multimode converter (Fig. 2). Moreover, such selection considering the requirements of tolerances satisfies the requirements of production effectiveness.



Fig. 2. The form of the groove in the multimode converter

The example of the offered horn feed is designed for the C , X , Ku , and Ka frequency bands. The size of a circular waveguide, which agrees with the horn, is $0.7C\{\lambda_0\}$ [3, 4]. The exciter for weight saving is made completely from aluminium alloy. The horn has the aperture $3C\{\lambda_0\}$ and length $1.8C\{\lambda_0\}$. The depth of a groove of the horn is $-0.24C\{\lambda_0\}$ with a constant period. A profile of a transition begins with the diameter $1.4C\{\lambda_0\}$ and ends with the diameter $0.9C\{\lambda_0\}$ on the length $0.8C\{\lambda_0\}$. The multimode transformer is designed to provide simplicity

of engineering implementation in the form of a set of rings of the same thickness with the selections shown in Fig. 2 and has the length $1.5C\{\lambda_0\}$.

To use as a part of a high-effective MBRA designed based on the dual-reflector scheme, the exciter ensuring the best CUA according to (2)–(9) is expediently to make from the described corrugated horn, the multimode transformer of which creates the required out-phasing of the filed in the exciter. This allows one to provide a negligible change of the width of the radiation pattern with the change in the frequency.

An experimental study of the characteristics of the radiation of the prototype of the developed horn

Experimental radiation patterns of the corrugated horn in the E - and H -planes are given in Fig. 3. The width of radiation pattern at the level -15 dB on the lower and upper frequencies of the working ranges are 58° in the E -plane and 60° in the H -plane, respectively. The given level of radiation is achieved by the highest effectiveness of the reflector system [1,4]. In X -, Ku -, and Ka - bands, on can observe a monotonous decrease in the width of radiation pattern, which makes 54° in both planes for X -band, 47° in the E -plane, 48° in the H -plane for the Ku -band, 34° in the E -plane, and 35° in the H -plane for the Ka -band. Such measurement of the width of radiation pattern ensures maximum effectiveness of the multiband exciter system according to the indicator (3). In addition to this, formed radiation patterns have almost the same width in E - and H -planes and a low level of side radiation that determines a high axial symmetry of radiation pattern and high cross-polarized radiation.

The measured VSWR of the horn exciter in the frequency bands of C , X , Ku , and Ka is given in Fig. 4. The mean value of VSWR in the frequency band does not exceed 1.12 at the maximum value 1.15 in the frequency of $C\{\lambda_0\}$.

Thus, using the considered quad-band horn permits one to achieve the highest axial symmetry of radiation pattern in the limits of the working sector of angles, low level of side lobes, high level of matching with the feed waveguide, and high coefficient of a polarizing cross coupling of orthogonal linear polarizations of the exciter that provides a high efficiency of MBRA in general.

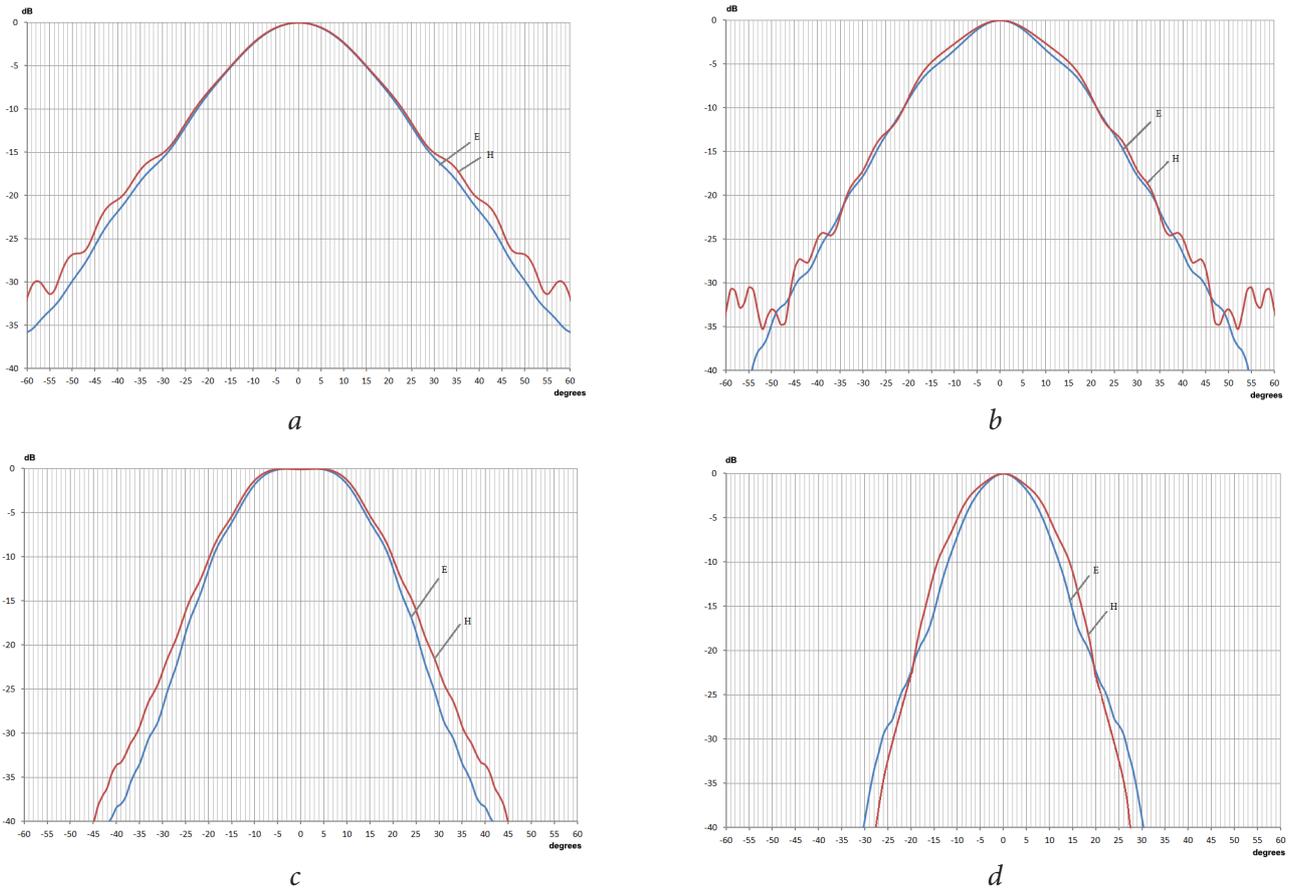


Fig. 3. Radiation patterns of the corrugated horn: a) C-band, b) X-band, c) Ku-band, d) Ka-band

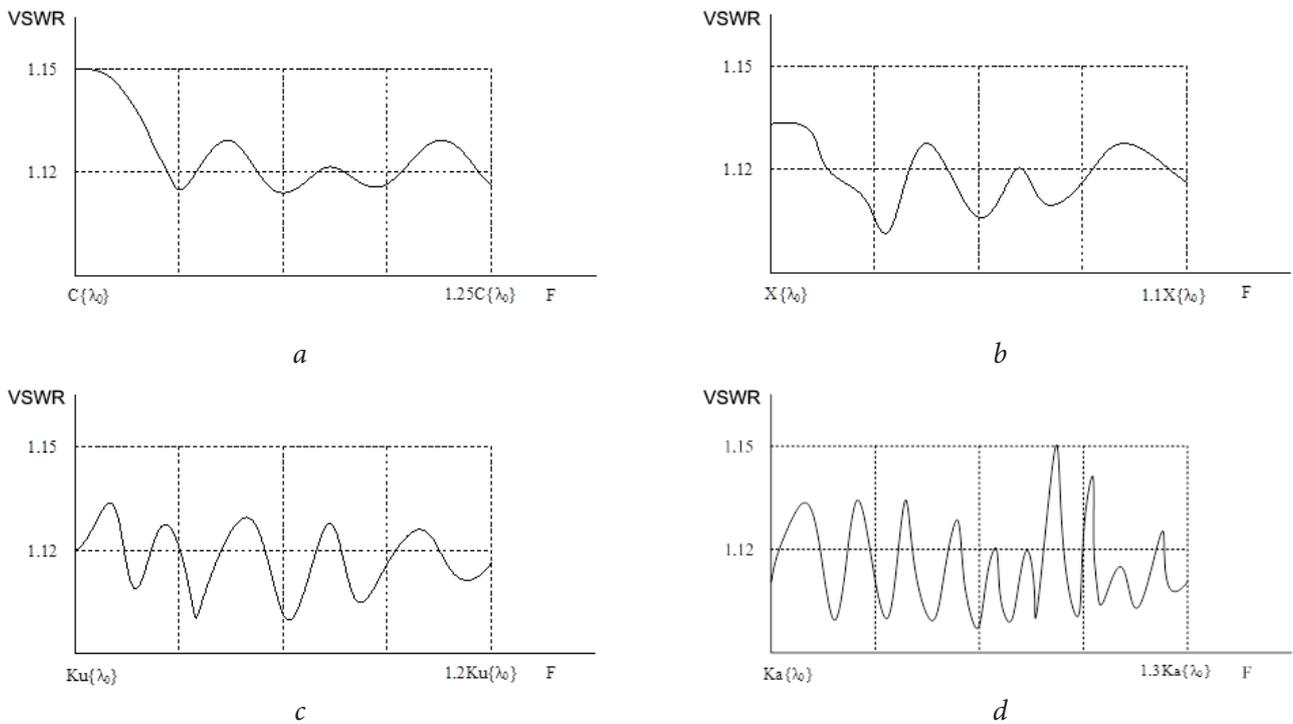


Fig. 4. Frequency dependences of VSWR of the quad-band corrugated horn: a) C-band, b) X-band, c) Ku-band, d) Ka-band

Conclusions

1. The conducted research has allowed one to establish the interrelation of characteristics of the exciter system with the characteristics of MBRA and to prove the requirements imposed to the exciter system. It is shown that the corrugated horn, when the condition of the balance of the hybrid mode and minimum level of transformation into the mode is fulfilled, to the full extent meets the requirements to the exciter system of MBRA.

2. The quad-band corrugated horn with matching to the circle waveguide containing the original multimode converter of types of waves adjoining through the radial transition to a corrugated horn is developed. Based on the mathematical simulation, the parameters of a corrugated horn and mode converter are chosen.

3. Experimental check has shown that in the bandwidth of C-, X-, Ku-, and Ka-bands, an average value of VSWR does not exceed 1.12, and the maximum value is 1.15 at the frequency. Radiation patterns of the exciter have a high axial symmetry.

References

- Gabriel'yan D.D., Demchenko V.I., Korovkin A.E., Razdorkin D.Ya. Vybór pokazateley i kriteriya effektivnosti obluchatelya mnogodiapazonnoy zerkal'noy anteny sistemy sputnikovoy svyazi [A selection of performance indicators and criterion of the exciter of the multiband reflector antenna of the satellite communication system]. IX Vserossiyskaya nauchno-tekhnicheskaya konferentsiya "Radiolokatsiya i radiosvyaz" [Collected papers of IX All-Russian scientific and technical conference "Radiolocation and radio communication], November 23–25, 2015, pp. 52–57. (in Russian)
- Clarricoats P.J.B., Olver A.D. Corrugated horns for microwave antennas. Electromagnetic waves series 18. London, UK, Peter Peregrinus Ltd., 1984.
- Gabriel'yan D.D., Demchenko V.I., Korovkin A.E., Razdorkin D.Ya., Shipulin A.V. Issledovanie chastotnykh zavisimostey komponentov vektornoy diagrammy napravlenosti gofrirovannogo rupora [A study of frequency dependences of the components of the vector radiation pattern of the corrugated horn]. Sbornik trudov V Vserossiyskoy Mikrovolnovoy konferentsii [Collected papers of V All-Russian Microwave conference], November 29 – December 1, 2017, pp. 57–61. (in Russian)
- Gabriel'yan D.D., Demchenko V.I., Korovkin A.E., Razdorkin D.Ya., Gvozdyakov Yu.A., Poltavets Yu.I. Postroenie obluchately mnogodiapazonnykh zerkal'nykh antenn sistem sputnikovoy svyazi [Building of Exciters of Multiband Mirror Antennas for Satellite Communication Systems]. Raketno-kosmicheskoe priborostroenie i informatsionnye sistemy [Rocket-Space Device Engineering and Information Systems], 2017, Vol. 4, No. 1. pp. 40–45. (in Russian)

The Calculation Methodology for the Energetic Reserve of the Radio Link Spacecraft-Station

G.A. Erokhin, *Cand. Sci. (Engineering)*, ru3dlr@mail.ru

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

V.I. Mandel, mandel.vlada@gmail.com

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

Yu.A. Nesterkin, uriy_nesterkin@mail.ru

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

A.P. Strukov, workbox69@gmail.com

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

Abstract. Nowadays different sets of calculation methods for energetic reserve are used at the factories of the rocket and space industry, meanwhile, not all the factors influencing the energetics of a radio link are taken into consideration. This causes difficulties while comparing some calculations with others and requires further recalculations, which are different from the previous ones and usually have negative results.

The article given below includes the methodology for a general calculation of the energetic reserve of a radio link transmitting the information from a spacecraft to a receiving station. Most of the methods of evaluating the effect of the atmosphere on the signal can be found in the recommendations of the International Telecommunication Union (ITU). Methods of calculation of the losses caused by the environment, which are kept in the recommendations of the ITU, are complemented by loss accounting methods resulting from the guidance errors, Faraday effect, and receiver noise temperature calculation.

Keywords: satellite communication, atmospheric attenuation, radio link "spacecraft-station", energy calculation of a radio link

Introduction

Development of space systems and space complexes at all stages of a product life cycle demands carrying out calculation of an energy margin of a radio line. At the stage of the preproject, a predesign of an energy margin of the radio line is performed; at the stage of the conceptual design, an analysis of several options of creation of the equipment with calculations of an energy margin of a radio line for each of the options is carried out; at the stage of development of design documentation, a detailed calculation of the chosen option and correction of the calculations by results of ground and experimental development is fulfilled; and at the operational phase, the analysis and confirmation of the carried-out calculations for correction and specification of the applied technique of calculations based on the received statistics is carried out. It is important to understand that the fullest calculation taking into account the influence of the Earth surface and the atmosphere already at the preproject stage will further allow avoiding corrections of requirements imposed to the onboard and ground equipment. Nowadays at the enterprises of a rocket and space industry, there is no uniform technique of an assessment of the influence of losses (atmospheric, polarizing, etc.) arising at signal distribution. The above-stated circumstances cause relevance of an objective.

Use of frequencies for the systems of a radio communication and broadcasting is strictly regulated by the International Telecommunication Union and the State Radio Frequency Commission. Requirements to the increase in the volumes of information transferred from spacecraft lead to the requirement on the increase in speed of information transfer that leads, in its turn, to the increase in a necessary frequency band, and this, taking into account a load of a current frequency plan leads to the increase in the value of a carrier frequency. Each of frequency ranges possesses the specifics of a set of losses at signal distribution, which needs to be considered at calculation. Detailed calculations are necessary not only for the signals with in advance known frequencies, such as GLONASS, AIS (tracking of vessels) and automated dependent observation-broadcasting (ADOB, tracking of aircraft), but also for again developed radio lines: high-speed radio links for transfer of target information and intersatellite radio lines.

An assessment of weakening of a signal becomes especially relevant in view of search of balance between

power decrease of radio-transmitting devices and an informational content increase. It is necessary to solve a task of the multicriteria choice of parameters of a radio link for information transfer with the set speed and the reliability satisfying the customer.

The analysis of publications allows one to draw the conclusion that there is no standard technique of an assessment of the influence of the losses (atmospheric, polarization, etc.) arising at signal distribution. A question of an assessment of the influence of the atmosphere on a signal contains most fully in the recommendations of the International Telecommunication Union.

The purpose of the present paper is to show and standardize a technique of calculations of an energy margin of radio lines representing the calculation of a radio line "spacecraft-station".

1. Calculation of an energy margin of a radio line

To ensure information transfer with a required speed and a set probability of a bit mistake, it is necessary to analyze physical processes and to carry out calculation of the parameters influencing the distribution of a radio signal in natural radio routes. Within this article, the signal transfer on the radio line "spacecraft-station" is considered.

SC moves in space on the set orbit; a reception of a signal is conducted on a receiving station, which can be located both directly on the ground and at some height over it. From parameters of SC movement, it is necessary to determine the maximum and minimum ranges between SC and the station and elevations under which SC is observed from the station.

Calculations should be conducted for all possible limiting cases: this will permit one to determine the range of changes of power flux density (PFD) in the fold of a receiving antenna forming the requirements to the dynamic range of the receiver. When calculating losses in the atmosphere, one has to analyze the altitude of station location above sea level and make a conclusion whether the Earth atmosphere influences the parameters of a radio signal.

Possibility of signal reception and availability of a radio line are defined by a positive value of an energy margin calculated as a difference between energy potential in the input of a receiving low noise amplifier (LNA) and its sensitivity. Calculation of energy margin

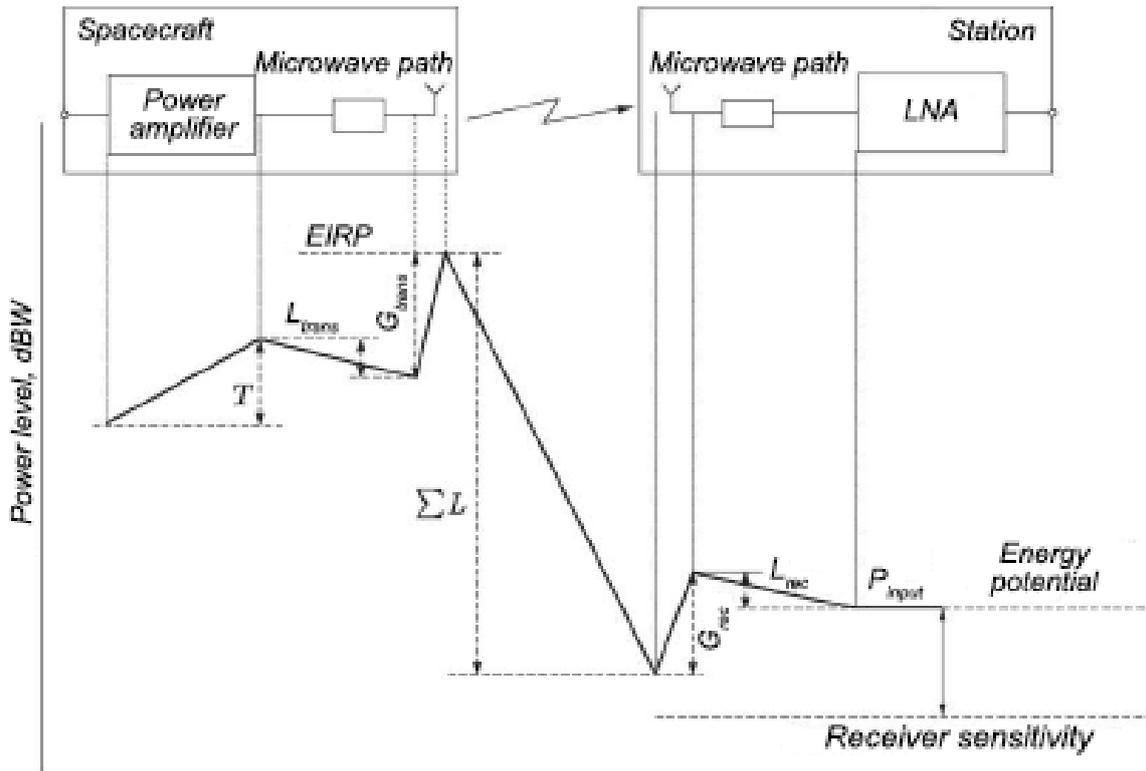


Figure. Structural scheme of the "spacecraft-station" radio link

begins with determination of the value of the equivalent isotropically radiated power (EIRP) of SC and conditions of signal distribution. The block diagram of a "spacecraft (SC)–station" radio line and graphical representation of a power level of a signal is given in the figure.

Transmission begins with SC, EIRP, which is expressed in dBW and calculated as [1]

$$EIRP = G_{trans} + T - L_{trans} \text{ (dBW)} \quad (1)$$

where G_{trans} is the gain of the transmitting antenna expressed in decibels relatively to isotropic gain, dBI; T is the value of power of a signal in the output of the power amplifier, dBW; L_{trans} is the losses in the microwave path from the power amplifier to the antenna input, dB.

During signal propagation, in the medium there takes place damping of wave oscillations in a free space caused by signal spreading when ranging from the transmitter. A signal comes to the receiving antenna of the station with the gain G_{rec} connected with the receiver through a waveguide with the losses L_{rec} . The value of signal power P_{input} in the input of a LNA is calculated by the formula:

$$P_{input} = EIRP - \Sigma L + G_{rec} - L_{rec} \text{ (dBW)} \quad (2)$$

where ΣL is the value of sum losses; G_{rec} is the gain of a receiving antenna expressed in decibels relative to isotropic gain, dBI; L_{rec} is the losses in the microwave path from the antenna output to the LNA input (all losses have dB regularity).

Comparison of signal power in the input of a LNA with the sensitivity of the R_x receiver determines the target value of an energy margin and radio link availability:

$$Z = P_{input} - R_x \text{ (dBW)}. \quad (3)$$

Sensitivity is defined by a minimum level of signal power in the input of a receiver when information reception is provided with the required speed and set probability of a bit error. In case when signal power in the LNA input is less than sensitivity of a receiver ($Z < 0$), it is not possible to ensure signal reception with the required validity. If the level of signal power in the LNA input is more than sensitivity ($Z > 0$), then reception is provided with the required validity.

The receiver sensitivity is calculated by the formula:

$$R_x = \kappa + T_{\text{эKB}} + B + C/N \text{ (dBW)}, \quad (4)$$

where $k = -228.6 \cdot 10^{-23}$ (dBW/kHz) is the Boltzmann constant; T_{eqv} is the equivalent noise temperature of a system expressed in decibels relative to the value 1 K, dBK; B is the signal bandwidth expressed in decibels relative to the value 1 Hz, dBHz; C/N is the required signal-to-noise ratio, dB.

Formula (4) shows a connection of the sensitivity and the required signal-to-noise ratio in the input of the receiver. As an example: providing of the less value of signal-to-noise ratio in the input of an LNA results in toughening of the requirements to the receiver (decrease in losses in the microwave path, decrease in noise coefficient of the receiving path, etc.).

A ratio of a signal power to noise power recalculated to the input of an LNA is expressed by the equation [1-3]

$$C/N = E_b/N_0 + R - B, \quad (5)$$

where E_b/N_0 is the required ratio of the energy of the information bit to the one-side spectral power density of noise for the set validity of reception and speed of data transfer, dB; R is the speed of data transfer expressed in decibels relative to the value 1 bit/s, dBbit/s.

Let us substitute (5) into (4):

$$R_x = k + T_{eqv} + R + E_b/N_0 \text{ (dBW)}. \quad (6)$$

2. Technique to calculate equivalent noise temperature of the system

When calculating the value of an energy margin of a radio link, it is important to determine a general power of noises, which are generated in the input of a receiver by different sources — an equivalent noise temperature. An equivalent noise temperature of a system is calculated by the formula [11]

$$T_{eqv} = T_a \cdot L_{rec} + 290 \cdot (1 - L_{rec}) + (F - 1) \cdot 290 \text{ (K)}, \quad (7)$$

where T_a is the noise temperature of an antenna, K; F is the noise coefficient of a LNA's receiver; L_{rec} is the losses in the microwave path from the antenna's output to the LNA's output.

An equivalent noise temperature for the onboard antenna can be presented by the following components:

$$T_{ant} = T_{b,atm} + T_{b,E} + 2cT_{b,space} \text{ (K)}, \quad (8)$$

where $T_{b,atm}$ is the noise temperature of the atmosphere, K; $T_{b,E}$ is the brightness temperature of the Earth, K; $T_{b,space}$ is the brightness temperature of the

prolonged space sources, K; c is the coefficient considering an average level of side and back lobes of a radiation pattern of the antenna.

An equivalent noise temperature for the ground antenna can be depicted as following:

$$T_{ant} = T_{b,space} + T_{b,atm} + cT_{Earth} + T_{n,cone} \text{ (K)}, \quad (9)$$

$T_{n,cone}$ is the noise temperature due to the influence of a blister of the antenna, K; T_{Earth} is the noise temperature of the Earth radiation, K.

Radiation of the Earth atmosphere has a thermal nature and at full extent is stipulated by signal absorption in the atmosphere. Owing to thermodynamic balance, the atmosphere radiates the same amount of energy at this frequency, which it absorbs. A noise temperature of the atmosphere is determined by the formula:

$$T_{b,atm} = T_{atm.av} \cdot (1 - 10^{-(A_r - A_c - A_g/10)}) \text{ (K)}, \quad (10)$$

where $T_{atm.av}$ is the average thermodynamic temperature of a standard atmosphere; A_r is the weakening due to hydrometeors, dB; A_g is the weakening in atmosphere gases, dB; A_c is the weakening due to a clouded sky, dB.

A value of an average thermodynamic temperature of a standard atmosphere is given in [4]. A detailed calculation of losses is presented in section 3.4.

Noises of space origin are determined, in general, by radiations of the Galaxy, Sun, and Moon. An average temperature of the noises of the Galaxy is negligibly small in the range of frequencies 6/4 GHz and more and does not surpass several degrees of Kelvin at the frequencies more than 2 GHz at any elevation. At the same time, the radiation of the Sun can completely break communication when falling onto the main lobe of the radiation pattern of the antenna. The radiation of the Moon causes less influence than the noises of the Galaxy, since its noise temperature is several degrees less than the noise temperature of the Sun.

The coefficient considering an integral level of side lobes of the radiation pattern of the antenna is determined by the formula:

$$c = \frac{1}{2} \sum_{i=1}^n \int_0^{\Omega_{sidei}} G_{sidei}(\alpha, \theta) d\Omega \bigg/ \int_0^{\Omega_{main}} G(\alpha, \theta) d\Omega, \quad (11)$$

where G_{sidei} is the coefficient of the antenna gain within the limits of back and side lobes; G is the coefficient of the antenna gain within the limits of the main lobe.

A component of the noises of the antenna from the thermal radiation of the Earth T_{Earth} at the elevations of the antenna from $5-7^\circ$ to 90° is stipulated by its reception of side and back lobes. Due to side lobes, the increase in the temperature of noises of the antenna of the ground station can be approximately estimated by the formula:

$$cT_s = 23 + 0.2(90^\circ - \theta) \text{ (K)} \quad (12)$$

where θ is the elevation of the receiving antenna, degrees.

For the onboard antenna, $T_{\text{b,Earth}} = 290$ K.

In some cases, antennas of ground stations are covered from the influence of precipitations with a radiotransparent blister. Losses of a signal and correspondent augmentation of noises are usually not big and can be almost not taken into account. However, during intensive rains, a water film appears on the surface of a blister, which is a cause of significant absorption of a signal and formation of secondary noises.

3. Losses of a radio link

At distribution of radio waves along natural radio routes, the environment influences the characteristics of a radio signal. Apart from losses in a free space, a radio wave passes through additional losses: losses at distribution through the atmosphere, losses because of the antenna pointing error, polarizing losses, losses on the antenna blister, etc.

The general losses $\sum L$ in the radio line are calculated by the formula:

$$\sum L = L_{\text{losses}} + At + L_{\text{pol}} + L_{\text{point,trans}} + L_{\text{point,rec}} + L_{\text{otl}} \text{ (dB)} \quad (13)$$

where L_{losses} are the losses in a free space;

At is the general weakening of a radio signal in the atmosphere,

L_{pol} are the polarizing losses;

$L_{\text{point,trans}} + L_{\text{point,rec}}$ are the losses caused by the errors of pointing of the transmitting and receiving antennas, dB;

L_{otl} are the other losses caused by the intersymbol interference, hindrances of the neighboring channel, the losses connected with restriction of the band (since all systems use in the filters to transfer energy in a limited or separated band, similar filtration decreases a total quantity of the transmitted energy that leads to weakening), etc., dB.

Further, calculation of all components, comprising the formula, is given (13).

3.1. Losses in a free space

Losses in a free space are calculated according to the formula (14) [1–3]:

$$L_{\text{losses}} = (4\pi \cdot (d_{\text{trans-rec}}/\lambda))^2, \quad (14)$$

where $d_{\text{trans-rec}}$ is the range of radio communication, m; λ is the wavelength, m.

3.2. Losses caused by pointing errors

The losses caused by a pointing error are generated due to inaccuracy of antenna pointing and are considered independently both for transmitting and receiving antennas. The antenna gain is calculated in the maximum of radiation pattern and decreases with a shift from it according to its characteristic. The shift, which defines the decrease of the gain, is an error of the elevation.

The losses caused by a targeting error are calculated by the formula:

$$L_{\text{point}} = 12 \cdot (\text{APE}/\text{BW})^2 \text{ (dB)}, \quad (15)$$

where APE is the error of the elevation, degrees; BW is the width of the antenna gain by the level 3 dB, degrees.

3.3. Polarizing losses

Polarizing losses arise due to the fact that the polarization of a coming wave differs from the polarization of the receiving antenna. At consideration of polarizing losses, one should consider Faraday's effect, i.e., rotation of the plane of polarization of the wave when passing through the ionosphere. At the frequencies over 2 GHz, the influence of the effect is insignificant.

Polarizing losses are calculated by the formula [3]:

$$L_{\text{pol}} = -10 \log_{10} \left(\frac{1}{2} \left[1 + \frac{4e_{\text{trans}} \cdot e_{\text{rec}}}{(1 + e_{\text{trans}}^2)(1 + e_{\text{rec}}^2)} + \frac{(1 - e_{\text{trans}}^2)(1 - e_{\text{rec}}^2) \cos(2 \cdot \text{pol})}{(1 + e_{\text{trans}}^2)(1 + e_{\text{rec}}^2)} \right] \right) \text{ (dB)}, \quad (16)$$

where $e_{\text{trans}}, e_{\text{rec}}$ is the coefficient of ellipticity of polarization of the transmitting and receiving antennas (the ratio of a small half shaft of an ellipse to the big one); pol is the type of polarization, degrees.

3.4. Weakening in the atmosphere

At distribution of radio waves in the Earth atmosphere, there is weakening of field tension of due to absorption in gases, dispersion and absorption in hydrometeors (rain, hail, snow, fog, and clouds) and also due to absorption in ionized areas. The main absorption of radio waves generates oxygen and water vapor.

At consideration of the influence of the troposphere on distribution of radio waves, it is necessary to take into account major factors: refraction of radio waves, reradiation by troposphere hydrometeors, weakening by gases and hydrometeors, and depolarization.

It is also necessary to consider such phenomenon as absorption in the ionosphere. It is caused by collisions of free charged particles with neutral molecules and atoms. In a process of such collisions, the energy that charged particles receive due to origination of an ordered speed at the influence of the electromagnetic field is transferred to neutral molecules and atoms, i.e., thermal losses take place. The absorption in the ionosphere significantly decreases at the increase in the frequency f (in inverse proportion to a frequency square) because of lag effect of charged particles and, hence, the smaller energy which is taken away from the influencing field. It becomes negligibly small at the frequencies over 100–150 MHz, i.e., it can

be not considered in those ranges of frequencies which are applied in modern communication and broadcasting systems using spacecraft.

All listed phenomena are dependent on frequency, geographical location, and elevation. The dependence of the phenomena arising in the atmosphere on frequency is given in Table 1. The type of weakening is presented in the left column; the right column gives the dependence of this phenomenon on various factors of the influence.

The general total weakening of radio signals in the atmosphere is calculated by the formula [4]:

$$At = A_g + \sqrt{(A_r + A_c)^2 + A_s^2} \text{ (dB)}. \quad (17)$$

3.4.1. Weakening in clouds

The value of weakening due to the influence of a clouded sky on inclined routes is calculated by the formula [5]:

$$A_c = (L \cdot K) / \sin(\theta) \text{ (dB)}, \quad (18)$$

where θ is the elevation; L is the statistics of a general column volume of water over Russia is taken from P.840-5; K is the coefficient of running weakening, a detailed calculation of which is given in [5].

Table 1. Weakening phenomena in the atmosphere

Type of weakening	Influencing factor
Weakening due to overcast A_o , dB [5]	At the frequencies below 10 GHz, this factor can usually be neglected. At the frequencies over 10 GHz while their increase, weakening becomes a more and more important factor, especially for small elevations
Absorption, dispersion and depolarization due to hydrometeors (water drops and ice particles in rainfall) A_r , dB [4,6]	These phenomena are especially noticeable at the frequencies over 10 GHz
Weakening of radio waves in atmospheric gases on inclined routes A_g , dB [7]	The frequency range is 1–1000 GHz
Weakening due to tropospheric twinkling A_s , dB [4,8,9]	At low elevations (10°) and at the frequencies about more than 10 GHz, tropospheric blinking can sometimes cause serious deterioration in performance data. At very small elevations (4° for the routes passing over the land, and 5° for the routes passing over water or along the coast), fading caused by multibeam distribution can be especially strong. In some places ionospheric blinking can play an important role at the frequencies lower about 6 GHz

3.4.2. Weakening in rain

Weakening of a signal in hydrometeors is caused, first, by dispersion of an electromagnetic energy by particles. Under the influence of an influencing field, each particle becomes a secondary radiator disseminating an electromagnetic energy in various directions, and, as a result, the energy share decreases, which distributing to a reception point. The second reason of weakening of field intensity in hydrometeors is nonresonance absorption in particles, which also depends on their quantity, electric properties, and frequency. This phenomenon depends on frequency, elevation, and intensity of rainfall.

The influence of rain is calculated by the formula [4]:

$$A_r = A_{0.01} \times (p/0.01)^{-(0.655+0.033 \ln(p)) - 0.045 \ln(A_{0.01}) - \beta(1-p)\sin \theta} \text{ (dB)}, \quad (19)$$

where $A_{0.01}$ is the predicted weakening value, dB; p is the percent of the time of an average year, %.

The percent of the time of an average year at which the calculation of losses in the atmosphere is carried out, for example 0.01%, means that during 99.99% of time of a year weakening on the radio line will not exceed the calculated value and availability of the radio line will be 99.99%.

The predicted value of weakening $A_{0.01}$ being exceeded during 0.01% of the time of the average time is found by the formula:

$$A_{0.01} = \gamma_r \cdot L_e \text{ (dB)}, \quad (20)$$

where L_e is the effective route length, km; γ_r is the running weakening, a detailed calculation of which is given in [4], dB/km.

An effective route length is calculated by the formula:

$$L_e = L_r \cdot v_{0.01} \text{ (km)}, \quad (21)$$

where L_r is the elevation view of the length of the inclined route, km; $v_{0.01}$ is the coefficient of fine tuning in vertical direction for 0.01% of time.

A detailed calculation of the coefficient of fine tuning in vertical direction, $v_{0.01}$ for 0.01% of time, the coefficient of weakening in horizontal direction and other parameters are given in [4].

If the height of the station above sea level is higher than the rain layer height, then the predicted weakening for any percent of time is equal to zero and further calculations are not required. The calculation of the height of the layer of rain is given in [10].

An assessment of the weakening exceeded for other percent of the time of an average year in the range from 0.001% to 5% is defined by recalculation on the value of weakening of 0.01% of the time of an average year.

3.4.3. Weakening in atmospheric gases

Weakening of a signal is caused by the absorption phenomenon in atmospheric gases and depends on the frequency, elevation, altitude above sea level, and density of water vapor.

Weakening on the route for communication systems can be calculated by division of the atmosphere into the horizontal layers defining a profile of the change of such parameters as pressure, temperature, and humidity along the route.

Full losses in atmospheric gases are calculated by the formula [7]:

$$A_g = \sum_{n=1}^k a_n \gamma_r \text{ (dB)}, \quad (22)$$

where a_n is the length of a part of the route in the layer, km.

A running attenuation γ_r is calculated by summing spectral lines by the formula:

$$\gamma_r = 0.1820 f_{\text{down}} (N''_{\text{oh}}(f_{\text{down}}) + N''_{\text{wat. vap}}(f_{\text{down}})) \text{ (dB/km)}, \quad (23)$$

where $(N''_{\text{oh}}(f_{\text{down}}), N''_{\text{wat. vap}}(f_{\text{down}}))$ is the imaginary part of the frequency dependent complex reflective capability, the calculation of which is given in [7].

To calculate the general attenuation on the satellite line it is necessary to know not only a running attenuation in each point of the line, but also the length of the route on which there is a running attenuation of such value. To determine this length, it is necessary to consider a curvature of the beam spreading above the Earth.

3.4.4. Weakening because of tropospheric blinkings

The value of tropospheric blinkings depends on the value and structure of changes of the index of a refraction, increasing with the increase of the frequency and length of the route passing through the distribution environment and decreasing in process of narrowing of the antenna radiation pattern due to averaging of its aperture.

The depth of fading is calculated by the following formula [4]:

$$A_s = a(p) \cdot \sigma \text{ (dB)}, \tag{24}$$

where σ is the standard deviation of a signal for the considered period and the route of propagation, dB; $a(p)$ is the time percent coefficient for the considered p .

A detailed calculation of the components, comprising the formula (24), is presented in [4,8,9].

3.5. Results of calculation

Basic data and results of calculation can be presented in the table form, the template of which is given below (Table 2).

Table 2

Parameter	Necessarily / Desirably for calculation
Information speed of data transmission, Mbp/s	Necessarily
Frequency band, MHz	Desirably
Required ratio $E_b/N_{0, \text{required}}$, dB	Necessarily
Frequency, GHz	Necessarily
EIRP of the transmitter, dBW	Necessarily
Gain of the transmitting antenna, dB	Desirably
Losses in a microwave path, dB	Desirably
Power of the transmitter, W	Desirably
Losses of pointing of a transmitting antenna, dB	Necessarily
Width of the gain of the transmitting antenna, degrees	Desirably
Error of a pointing angle, degrees	Desirably
Losses in a free space, dB	Necessarily
Range of communication, km	Desirably
Elevation, degrees	Necessarily
Reception altitude above the sea level, km	Necessarily
Reception width, degrees	Necessarily
Coefficient of percent of time, %	Necessarily
Intensity of rainfall, km	Necessarily
Losses in the atmosphere, dB	Necessarily
Weakening in the rain, dB	Necessarily
Weakening in atmospheric gases, dB	Necessarily
Weakening in clouds, dB	Necessarily
Weakening because of blinkings, dB	Necessarily
Diameter of the receiving antenna, m	Necessarily
Efficiency of using of the receiving antenna	Necessarily
Losses of pointing of the receiving antenna, dB	Necessarily
Width of the gain of the receiving antenna, degrees	Desirably
Error of a pointing angle, degrees	Desirably
Losses on the blister of the receiving antenna, dB	Necessarily
Q-factor (G/T)	Necessarily
Gain of the receiving antenna, dB	Desirably
Noise temperature of the system, K	Desirably
Losses in a microwave path, dB	Desirably
Stock, dB	Necessarily

Conclusion

The article presents the method of calculation of an energy margin in the radio link a point-to-point (“spacecraft– station”). All factors, influencing the value of an energy margin in the radio link are listed. It is shown from what components the value of losses in the atmosphere of Earth is produced. Analytical expressions for carrying out the corresponding calculations are given. References to ITU are provided. Formulae of calculation of an equivalent noise temperature of the receiver are also given. Thus, gaps in calculations arising due to incomplete record of the components are eliminated. These gaps can cause overestimations of the results and further corrections.

The technique will be useful to the experts who are engaged in calculations of radio lines, and its use at the enterprises of a space-rocket industry will allow one to correlate easily the calculations and to carry out expert estimates by uniform rules.

References

1. Sklyar B. Tsifrovaya svyaz'. Teoreticheskiye osnovy i prakticheskoye primeneniye [Digital communication. Theoretical foundations and practical usage]. 2nd edition. Moscow, ID “VIL”YAMS”, 2003, 1104 p. (in Russian) (translated from English)
2. Markov G.T., Sazonov D.M. Antenny. Uchebnik dlya studentov radiotekhnicheskikh spetsial'nostey vuzov [Antennas. Manual for students of radio engineering universities]. 2nd edition, revised and enlarged. Moscow, Energiya, 1975, 523 p. (in Russian)
3. Sputnikovaya svyaz” i veshchaniye [Satellite communication and broadcasting]. Ed. Kantor L.Ya. 2nd edition. Moscow, Radio i svyaz', 1988. (in Russian)
4. ITU-R Recommendations. P.618-13. Propagation data and prediction methods required for the design of Earth-space telecommunication systems. Geneva, 2017, 27 p.
5. Recommendations ITU-R P.840-7. Attenuation due to clouds and fog. Geneva, 2017, 4 p.
6. ITU-R Recommendations. P.838-3. Specific attenuation model for rain for use in prediction methods. Geneva, 2005, 9 p.
7. ITU-R Recommendations. P.676-11. Attenuation by atmospheric gases. Geneva, 2016, 23 p.
8. ITU-R Recommendations. P.835-6. Reference standard atmospheres. Geneva, 2017, 10 p.
9. ITU-R Recommendations. P.453-13. The radio refractive index: its formula and refractivity data. Geneva, 2017, 23 p.
10. ITU-R Recommendations. P.839-4. Rain height model for prediction methods, (1992-1997-1999-2001-2013). 1 p. (in Russian)
11. Somov A.M., Kornev S.F. Sputnikovyye sistemy svyazi. Uchebnoye posobiye dlya vuzov [Satellite communication systems. Manual for universities]. Ed. Somov A.M. Moscow, Goryachaya liniya-Telekom, 2012, 244 p. (in Russian)

Calculation of Antenna Gain via 3D Radiation Pattern and Estimation of Their Mutual Influence

V.I. Tolkachev, *Cand. Sci. (Engineering)*, tolkachev_1707@mail.ru

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

O.G. Pikalov, *Cand. Sci. (Engineering)*, pikalov-og@bk.ru

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

S.V. Pan'chev, *Cand. Sci. (Engineering)*, sergey_p76@mail.ru

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

I.G. Novikov, novige@mail.ru

Federal State Budget Educational Institution "Moscow Technological University" (MIREA), Moscow, Russian Federation

Abstract. The paper considers the task to determine the gain of the optionally beamed antenna on the set direction. It is necessary to solve the task when evaluating electromagnetic compatibility of radio engineering systems involving both receiving and transmitting segments equipped with their own antennas. The main attention is focused on the treatment and analysis of the form of the antenna radiation pattern appearing as a 3D rotation figure.

The determination of the intensity of the 3D interaction of receiving and transmitting antennas by means of simulating the radiation patterns considering their mutual orientation is one of the most important task of obtaining the degree of the influence of the transmitting segment on the reception devices of the radio engineering system.

The method given in the paper presumes a simple software realization and is slated to be applied in the special software product created for the analysis of the electromagnetic situation in radio engineering systems, as well as for evaluating electromagnetic compatibility of radioelectronic facilities.

Keywords: antenna radiation pattern, antenna gain, electromagnetic compatibility

Introduction

It is known that any equipment connected with information reception or transmission by means of electromagnetic waves is supplied with the antenna in this or that kind.

A great work has been done by engineers to achieve the required parameters from the created antennas, first of all, reducing power costs for transmission of useful electromagnetic signals in the necessary direction and also improving reception of useful signals from the directions of interest. One of such parameters is the antenna gain defined as the relation of density of a stream of the radiated energy in some direction by means of the considered antenna to the stream density which would be radiated in the same direction by the isotropic antenna (that is the imagined, ideal antenna radiating with uniform intensity in all directions). The display of antenna gains in the spherical system of coordinates where the intensity of radiation is displayed by the radial coordinate, and the direction by the azimuthal one, is called the antenna radiation pattern. Certainly, that each type of the antenna, each of its dimensions depending on the used frequencies of radiation or reception, has its specific form of the radiation pattern.

Forms of radiation patterns have various appearances. They can either possess the principal lobe and side and reversed petals (that is a characteristic of narrow-beam, for example, parabolic antennas, Fig. 1) or not possess them as, for example, radiation patterns of rod antennas (Fig. 2).

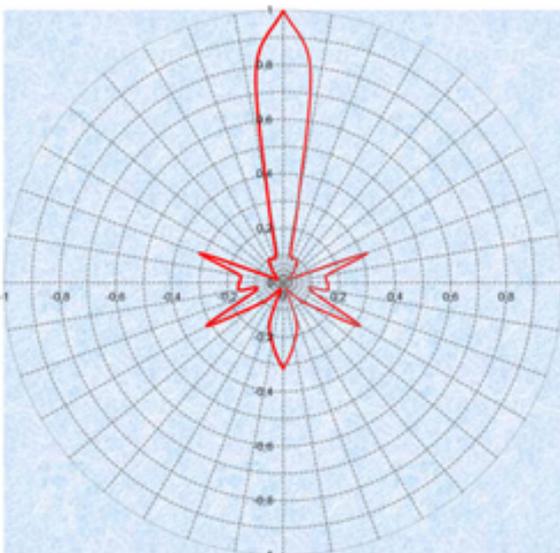


Fig. 1. Radiation pattern of a narrow-beam antenna.

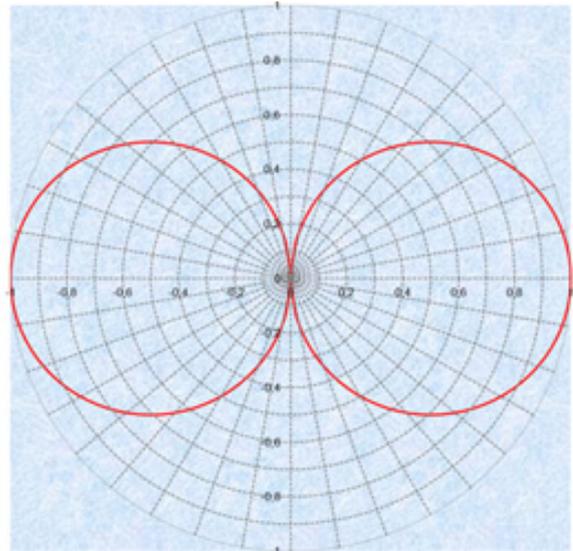


Fig. 2. Radiation pattern of a rod antenna.

When it concerns the maximum quality of signal transfer or its reception, so based on the given drawings, it is possible to understand intuitively that it is reached by orientation of the antenna so that the radial coordinate of the radiation pattern (in other words, antenna gain) was maximum in the direction of the receiver (in case of transmission) or a signal source (in case of reception).

This article offers to consider another task: **to determine the antenna gain of a known randomly directive antenna in the set direction.** The solution of such task is necessary, for instance, at the assessment of a problem of electromagnetic compatibility for the systems containing at the same time receiving and the transmitting segments equipped with their own antennas. In this case, it is meant that receiving and transmitting segments operate on the channels independent from each other. At the same time there is a need to estimate parasitic influence of the work of the transmitting segments on the reception channels. One of the components of such calculation is the assessment of mutual gain of transmitting and receiving antennas.

In the article the method of the assessment of mutual antenna gains for one pair of antennas, one of which is transmitting and another is a receiving one, will be considered. Of course, for the big system comprising several receiving and transmitting segments (for example, for a spacecraft), it will be necessary to carry out the paired analysis.

We will consider the necessary entrance data for the solution of this task:

- The coordinates of placement of transmitting and receiving antennas (the Cartesian coordinates x, y, z set in the random, but uniform system of coordinates for both antennas are meant).
- The rotation angles of axes of antennas: an azimuth and elevation (a zero azimuth is considered the OX axis and turn on an azimuth is the turn in the XOY plane. An elevation is the angle between an axis of the antenna and the XOY plane).
- The forms of radiation patterns of the considered antennas.

It is worth noticing that irrespective of a radiation pattern form, a straight line located on a zero corner of the diagram will be an axis. In the figures given above an axis is the vertical straight line passing through the center.

First, we will perform the calculation of antenna gain of any set transmitting antenna in the direction of any set receiving antenna.

Finding of an angle between the axis of the transmitting antenna and the direction on the receiving antenna will be the first step of our calculation (which in this case we simply consider a point).

The direction angle between the antennas is the angle between three points in space: $t1$ is the location of the antenna of the receiver; $t2$ is the location of the antenna of the transmitter; $t3$ is the random point on the axis of the antenna radiation pattern of the transmitter (if we look for an angle between the axis of the antenna of the transmitter and the direction on the antenna of the receiver, Fig. 3).

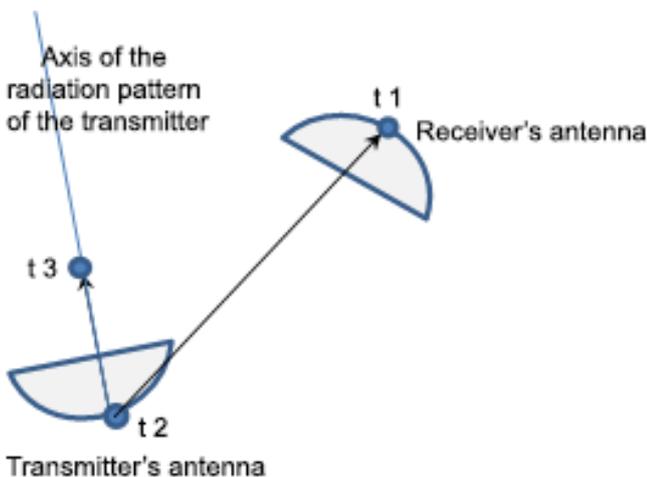


Fig. 3. A schematic image of the location of the antennas.

As it has been mentioned above, $t1$ is set by the Cartesian coordinates, $t2$ is also set by the Cartesian coordinates. The $t3$ point will be taken as a point on the axis of the radiation pattern of the transmitter remote from the point $t2$ by 1 unit (for example, 1 m, if the scale of the system of coordinates is set in meters). Thus, we know the coordinates of the points $t1$ and $t2$, and the coordinate of the point $t3$ is to be defined.

By means of simple trigonometrical transformations, we will obtain that the coordinates of the point $t3$ are in number equal to:

$$x_3 = x_2 + \cos(a) \cdot \cos(b), \tag{1}$$

$$y_3 = y_2 - \sin(a) \cdot \cos(b), \tag{2}$$

$$z_3 = z_2 + \sin(b), \tag{3}$$

where a is the azimuth of the transmitter's antenna; b is the elevation of the transmitter's antenna;

x_2, y_2, z_2 are the Cartesian coordinates of the transmitter.

Now we have the coordinates of all three points. It is possible to find an angle between them as an angle between the vectors $t2t1$ and $t2t3$.

The cosine of the angle between the vectors is as the relation of a scalar product of vectors to the product of their lengths.

The angle itself (we will call it α) is found by the formula:

$$\alpha = \arccos \left(\frac{(x_3 - x_2) \cdot (x_1 - x_2) + (y_3 - y_2) \cdot (y_1 - y_2) + (z_3 - z_2) \cdot (z_1 - z_2)}{\sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2 + (z_3 - z_2)^2} \times \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}} \right) \tag{4}$$

In the formula (4), the subtraction of the values x_2, y_2, z_2 from the coordinates of other points of interest shows a parallel transfer of the considered system in the beginning of the coordinates.

Thus, the angle between the axis of the transmitting antenna and the direction on the receiving antenna is defined. Now there is a problem to determine the gain of transmitting antenna corresponding to this angle. To do this, we need to carry out a certain analysis of the radiation pattern of the transmitting antenna.

For a mathematical task of a form of the radiation pattern, we will present it in the form of a two-dimensional table. The first field will reflect a deviation in degrees from the main axis. An interval, which needs

to be considered in the interval, is 0–180°. Here we will make a certain assumption and consider the radiation pattern symmetric in relation to the antenna axis, and consider its section perpendicular to the axis the plane, respectively, circular.

The second field of the table will reflect the relative power radiated by the antenna in the direction corresponding to value of the first field. We will accept the maximum value of such relative power to be 1 (Table 1).

Table 1. Table of the rough description of the radiation pattern.

Angle between the axis of the antenna and direction (degrees)	Relative power (times)
0	1.00
20	0.79
40	0.32
60	0.40
80	0.32
100	0.50
120	0.32
140	0.50
160	0.32
180	0.79

For illustrative purposes, these table data can be displayed on the diagram and to designate the tops of the received figure (Fig. 4).

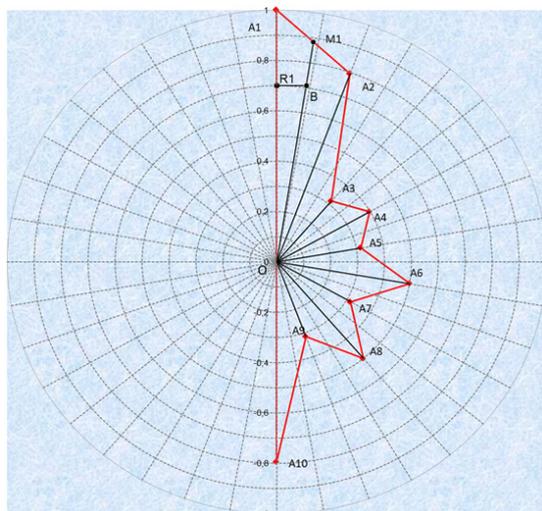


Fig. 4. A diagram display of the table data.

Now the radiation pattern set by Table 1 can be given as a body of rotation of the figure represented in Fig. 4 around the direct line OA_1 . As it is easy to see, the figure consists of the triangles ($A_1OA_2, A_2OA_3, A_3OA_4$, etc.), and for each such triangle the lengths of the sides equal to relative powers, and the angle between them are known (equal to the difference of the angles set in Table 1).

The general power of the transmitter radiated by this antenna can be interpreted as radiation pattern volume (that is the bodies of rotation of the considered figure). To find the volume of a body of rotation, we will use the second Pappus–Guldinus theorem, which says that:

The volume of the body formed by rotation of a flat figure around the axis located in the same plane and which does not cross the figure is equal to the area of the figure multiplied by the length of a circle, which radius is the distance from a rotation axis to a figure barycenter.

Thus, it is possible to draw a conclusion that the total amount of a body of rotation will represent a composition of the bodies of rotation of the triangles stated above.

Let us take the first triangle A_1OA_2 .

Its square is easily calculated by the formula:

$$S_1 = 1/2 \cdot OA_1 \cdot OA_2 \cdot \sin(A_1OA_2). \tag{5}$$

Now it is necessary to find a barycenter (center of mass) of the triangle. As geometry says, the barycenter of a triangle lies on the crossing of medians, and medians cross in 2:1 ratio.

First, the length of the side A_1A_2 should be found by the cosine theorem:

$$A_1A_2 = \sqrt{(OA_1^2 + OA_2^2 - 2 \cdot OA_1 \cdot OA_2 \cdot \cos(A_1OA_2))}. \tag{6}$$

Knowing the sides of the triangle, we can determine the length of the median leading to A_1A_2 by the formula:

$$OM_1 = 1/2 \cdot \sqrt{(OA_1^2 + OA_2^2 - A_1A_2^2)} \tag{7}$$

and then multiply this length (OM_1) by 2/3 determining the position of the barycenter (B) on the median. Now it is necessary to define the distance from the barycenter to the axis OA_1 , because for calculation of the volume it is necessary to know the length of a circle, which the barycenter traces at rotation. The radius of this circle is the length of the perpendicular lowered from the barycenter on the axis OA_1 .

Understanding that the median halves the area of the triangle and knowing the area of the triangle OA_1A_2 (S_1), we find the angle between the median and the axis OA_1 :

$$A_1OM_1 = \arcsin \left(1/2 (S_1 / (1/2 \cdot OA_1 \cdot OM_1)) \right). \quad (8)$$

Then we determine the distance from the axis OA_1 to the barycenter (Fig. 4):

$$R_1 = 2/3 \cdot OM_1 \cdot \sin(A_1OM_1). \quad (9)$$

Applying the second Pappus–Guldinus theorem, we receive that the volume of a body of rotation of the triangle A_1OA_2 is equal to:

$$V_1 = S_1 \cdot 2 \cdot \pi \cdot R_1. \quad (10)$$

By same way we can find the volumes of bodies of rotation of other triangles of a figure and then, making an addition, we receive the total amount of a body of rotation of the figure (V).

At the emission of the same power, the isotropic antenna (we will remember that the isotropic antenna is the ideal antenna radiating evenly in all directions), the radiation pattern would be as a sphere of the same volume V with a radius being equal to:

$$R_i = \sqrt[3]{3/4 \cdot V/\pi} \quad (11).$$

Thus, we geometrically have interpreted the power of isotropic radiation by the value R_i and have had an opportunity to define the gain of the antenna in the direction which angle is calculated by the formula (4). For this purpose, it is necessary to find the value of the radial parameter R according to Fig. 4 for this angle (α). It is possible to do by means of a simple linear calculation by proportions. At first, the triangle where the angle beam α falls, is determined and then the following formula is used:

$$R = (\alpha - A_1OA_n) / (A_1OA_{n+1} - A_1OA_n) \times (OA_{n+1} - OA_n) + OA_n, \quad (12)$$

where n is the number of the triangle where the angle beam α falls.

Now we have an opportunity to find the ratio of the radial parameters R and R_i and, thus, to define the gain of the transmitting antenna in the direction of the receiving antenna:

$$GER = R/R_i. \quad (13)$$

The designation GER is based on the English words *gain, emitter, and receiver* and shows the considered direction of gain (“from the transmitter to the receiver” contrary to GRE , that is the gain of the receiving antenna in the direction of the transmitter).

The gain of the receiving antenna in the direction of the transmitter is in the same way calculated. It is worth reminding that the received GER and GRE values are measured in “times”, therefore the gain of power transfer from the transmitter to the receiver can be considered as $GER \times GRE$. In certain cases, the GER and GRE values are given in decibels, and then the gain of power transfer from the transmitter to the receiver represents their sum of $GER + GRE$.

Conclusion

The advantages of the described method can be an opportunity to record a 3D radiation pattern of the antenna in the form of a circular section that is most widespread and the simplicity of its implementation by electronic computing machines. The method does not demand difficult software and is easily implemented in such common applications as, for example, Excel.

The disadvantages are the simplification of models of radiation patterns (a rotation figure). However, the method is opened for further modernizations. In future the method can be functionally expanded to the calculation of the radiation charts of the elliptic section. It is also simple to develop this method before recording the apertures of the interacting antennas. For this case, the mechanism of the integrated calculation for space angles and comparison of the received values with conic volumes, which are the segments of an isotropic sphere, will be created.

References

1. Elektromagnitnaya sovmeštmost' sistem sputnikovoy svyazi [Electromagnetic compatibility of satellite communication system]. Eds. Kantor L.Ya., Nozdrin V.V. Moscow, NIIR, 2009, 280 p. (in Russian)
2. Badalov A.L., Mikhaylov A.S. Normy na parametry elektromagnitnoy sovmeštmosti RES [Norms for electromagnetic compatibility parameters of radioelectronic facilities]. Moscow, Radio i svyaz', 1990, 272 p. (in Russian)

3. Teoriya i metody otsenki elektromagnitnoy sovmestimosti radioelektronnykh sredstv [Theory and methods of evaluating electromagnetic compatibility of radioelectronic facilities]. Ed. Feoktistov Yu.A. Moscow, Radio i svyaz', 1988, 216 p. (in Russian)
4. Upravlenie radiochastotnym spektrom i elektromagnitnaya sovmestimost' radiosistem [Radio frequency spectrum and electromagnetic compatibility of radio systems]. Ed. Bykhovskiy M.A. Moscow, Ekotrendz, 2006, 377 p. (in Russian)
5. Goncharenko I.V. Antenny KV i UKV [SB- and USB-antennas]. Moscow, Radiosoft, 2016, 744 p. (in Russian)
6. Ustroistva SVCh i anteny [Microwave devices and antennas]. Ed. Voskresenskiy. Moscow, Radiotekhnika, 2006, 378 p. (in Russian)
7. Elektromagnitnayasovmestimost' radioelektronnykh sredstv i radiokontrol'. Metody otsenki effektivnosti [Electromagnetic compatibility of radioelectronic facilities and radio control. Effectiveness evaluation methods]. Ed. Saiy P.A. Moscow, Radiotekhnika, 2015, 400 p. (in Russian)
8. Galimov G.K. Anteny radioteleskopov, sistem kosmicheskoy svyazi i RLS [Antennas of radiotelescopes, space communication systems, and radars]. Moscow, Advanced Solutions, 2013, 392 p. (in Russian)

Sampling Theorem and Practical Usage of Integer Functions for Signal Representation on the Receiving Side

A.I. Troshchenkov, *post-graduate student, as181191@mail.ru*
Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

Abstract. The article contains the provisions uniting engineering and mathematical approaches to process telemetry signals and other information signals on the receiving side. The connection between a tabular composition of the values of telemetric parameters and the functions simulating these parameters is shown. The method to evaluate the nature of the change of different types of telemetry parameters is described. A number of the tasks, the solution of which will make possible to apply the mathematical apparatus of Kotelnikov for economical coding of information as an effective alternative to the methods used today, is offered.

The pay-offs of the paper consist in clearly set of engineering tasks, innovative approaches to processing and representation of the space telemetry. In conjunction with application of modern mathematical methods of data transformation, the presented provisions allow one to realize the development of essentially new methods of information transfer.

Keywords: sampling theorem, continuous function, telemetry signal, spectral form, cutoff frequency, reading, sample, sample rate, time axis, approximating polynomial, orthogonal basis

In radio communication, the sampling theorem is known [1–3]. In the formulation of C. Shannon [4], the theorem is as follows:

Theorem. *If the $f(x)$ function does not contain components with a frequency over W Hz, then it is completely defined by the sequence of its values in the points remote at distance $1/2 W$ from each other.*

Shannon called the theorem “the sampling theorem”. Under this name it is also known in English-speaking technical literature.

Further in the paper the terms “function” and “signal” will be considered synonyms, and $f(x) = f(t)$.

Definition. The functions, which do not contain components of frequencies above the set frequencies (higher than W Hz), are called functions with a limited spectrum.

A statement formulated in the theorem is the sampling theorem in time representation. Also the sampling theorem in frequency representation is known (for example, [4]).

Definition. “Sampling” in the Shannon’s theorem is the sequence of its values in the points remote at distance $1/2 W$ from each other. Actually, based on these values, support functions which sum represents a restored form of a signal are formed.

It is known that any continuous function can be spread out on a final piece in a Fourier series that is presented in a spectral form – in the form of the sum of a number of sinusoids with multiple (enumerated) frequencies with certain amplitudes and phases. At rather smooth functions the spectrum quickly decreases (spectrum module coefficients quickly tend to zero). For representation of the “cut-up” functions, with gaps and “breaks”, the sinusoids with greater frequencies are necessary.

In engineering practice, nobody remembers that Shannon has behind “sampling” a function, and engineers do not use these support functions to restore a signal. In engineering, “sampling” is understood not as a function, but the value of a signal in the set time of “sample” of a signal (“the sequence of its values in the points remote at distance $1/2 W$ from each other”).

Definition. “Sample” is a process of determination of the current instant value of a signal in the set time point.

Definition. “Sample rate” is a frequency of “sampling” of a signal.

It is clear, that in engineering practice to restore a signal on the receiving side, it is more convenient to

use not “functions – sampling”, but “sampling” – instant values of a signal in points of “sampling” of a signal.

The class of functions with a limited spectrum is rather big. It is enough to note that spectra can be continuous and line.

Any periodic signal can be presented in the form:

$$s(t) = \frac{1}{2} \sum_{n=-\infty}^{\infty} A_n e^{jn\omega_1 t}$$

where

$$A_n = \frac{1}{2}(a_n - jb_n) = \frac{1}{2} A_n e^{-j\varphi_n}$$

is the complex amplitudes of a spectrum containing the information both about amplitude and about phase spectra.

The spectrum of a periodic signal – a line spectrum is given:

$$\dot{S}_n = \frac{1}{2} A_n T = \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} s(t) e^{-j\omega t} dt$$

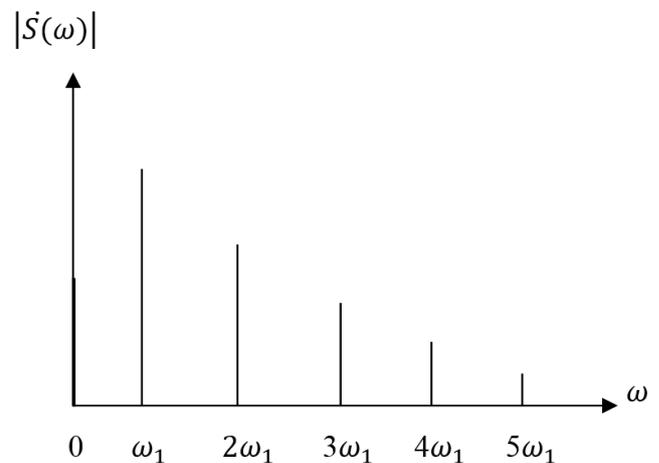


Fig. 1. The example of a line spectrum.

A **continuous** spectrum (a spectrum of a nonperiodic signal):

$$\dot{S}(\omega) = \int_{-\infty}^{\infty} s(t) e^{-j\omega t} dt$$

$\dot{S}(\omega)$ is the complex spectrum that contains the information both on the spectrum of amplitudes and the spectrum of phases.

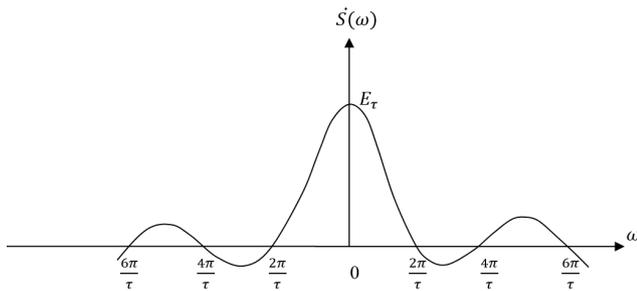


Fig. 2. The example of a continuous spectrum.

From the strict mathematical theory of a Fourier series [5], it is known that functions with a limited spectrum are representable on an infinite axis of time, and the function limited on time is representable on an infinite axis of frequencies (i.e., is not a function with a limited spectrum). In fact, from Figs. 1 and 2 given above it is clear that a periodic signal is the signal that is not limited on time has a spectrum in the form of a limited number of harmonicas. A nonperiodic signal is the signal limited on time has a spectrum, which is not limited on an axis of frequencies.

In engineering, data transmission (in television, in radio telemetry) when time of data transmission is obviously limited, theoretically we have to deal with the signals having an unlimited spectrum. At the same time, there are three engineering problems:

1. A problem of “assignment” of a cutoff frequency of a spectrum concerning which more high-frequency components can be neglected.
2. An engineering problem of approximate representation of the transmitted signal on the receiving side on its sampling (numbers) formed on the transmitting side.
3. A problem of assessment of an error of the representation.

V.A. Kotelnikov at the proof of the theorem applied an artificial mathematical method using a row [1], which was later called the Kotelnikov’s row.

In engineering, at the representation of a signal on the receiving side the mathematical apparatus of Kotelnikov (as well as Shannon’s “functions – sampling”) that he offered at the proof of the theorem is not applied.

Thus, the sampling theorem for engineers is a mathematical justification of a possibility of representation (on a limited interval of time) of the signal consisting of an infinite number of points using a final quantity of numbers – sampling.

To represent a signal on the receiving side, a well-

developed theory of approximation is used. In this theory, approximation methods using polynomials are employed. The Weierstrass theorem that he proved in the 19th century is the cornerstone of this method of approximation. One knows several methods of the proof of an approval of the Weierstrass theorem among which the most preferable for engineering is the proof provided in [7].

In relation to continuous functions of one valid variable set on a final piece $[a, b]$, the first Weierstrass theorem claims: for any continuous $f(x)$ function on $[a, b]$, the sequence of ordinary polynomials which is evenly meeting on $[a, b]$ to $f(x)$ [7] exists.

The essence of the first Weierstrass theorem tells that any continuous function in the final closed interval can be spread out in evenly meeting row which members are polynomials.

The second Weierstrass theorem gives a clearer link of decomposition of a continuous periodic function with the 2π period (p. 40, [6]):

If $F(t)$ is a continuous function with the 2π period, then whatever the number $\epsilon > 0$ is, there is the trigonometrical sum $S_n(t)$, $S_n(t) = a_0 + \sum (a_k \cos kt + b_k \sin kt)$, where summing is carried out by k measured from 1 to n :

$$[n = n(\epsilon)].$$

This sum for all t meets the inequality

$$|F(t) - S_n(t)| \leq \epsilon.$$

As it is known, polynomials relate to the class of integer functions [8].

Polynomials, an exponential function, trigonometrical functions (sine and cosine), and others belong to the integer functions. The integer functions, which are not polynomials (they are called transcendental), in many respects behave as some kind of “polynomials of an infinitely high degree”. In engineering, the approximation of functions (signals) with use of polynomials of a final degree is of interest.

From the representatives of the integer functions we will concern only research of polynomials of a final degree and trigonometrical functions.

In the introduction to Ya.I. Khurgin and V.P. Yakovlev [8] book, it is written that the sampling theorem “is the theorem of an opportunity for signal transferring with a limited spectrum in principle to use not all its values but only separate periodically chosen (“equidistant”,

author's note) values, and at the same time to restore unambiguously a signal on the receiving side on the entire time axis".

In addition, the theorem of special not equidistant "sampling", or the theorem of sampling in special points proving a possibility to use "special" sampling for signal restoration is known.

Special readings are the values of a signal taken in special points of a signal where the first, second, and other derivatives are equal to zero [10, 11], and equidistant readings are considered as a special case of not equidistant readings. At the same time to restore a signal by not equidistant readings, polynomials of a final degree are also used.

Further in V.P. Khurgin's preface to [9] we read: "However soon it was succeeded to understand: all the matter is that functions with a limited range are the whole analytical functions, and, therefore, Kotelnikov's formula is one of the possible interpolation formulas used in the theory of the integer functions". Moreover, below it is about the proof and generalization by students B.S. Tsybakov and V.P. Yakovlev of the sampling theorem "by means of the instrument of the theory of interpolating of the integer functions".

However, the analysis of literature showed that B.S. Tsybakov and V.P. Yakovlev's publications with the proof of this statement are absent, or proofs are published in the little-known edition by a small circulation.

To restore the specified gap, it is of interest to prove or disprove five statements and to specify living conditions of these statements:

1. Signals, mathematically representable on a final piece $[a, b]$ in the form of polynomials of a final degree are functions with a limited spectrum.
2. Signals, mathematically representable on a final piece $[a, b]$ in the form of trigonometrical functions (a sine and a cosine) are functions with a limited spectrum.
3. Signals, mathematically representable on a final piece $[a, b]$ in the form of a piece of an exponential function, are functions with a limited spectrum.
4. Signals, mathematically representable on a final piece $[a, b]$ in the form of integer functions are functions with a limited spectrum. In view of complexity, it is possible to leave this fourth (general-theoretical) statement without proof in view of taking into account the practical importance for engineering practice of the first three statements (theorems).
5. The continuous signal presented on the

transmitting side in the form of a polynomial of a final degree is a signal with a limited spectrum, and W depends on coefficients of a polynomial and is defined by the following expression:

$$s(t) = \frac{1}{2} \sum_{n=-\infty}^{\infty} A_n e^{jn\omega_1 t}$$

At the practical engineering approach the proof of three first private statements, living conditions of these statements, and founding is of interest to these three cases of the cutoff frequency of W and "numbers" that is analogs to numbers "sampling" in the sampling theorem. We will note that K. Shannon offered the term "sampling", and V.A. Kotelnikov used the term "numbers".

It is clear, that the statement is justified: if on a final piece $[a, b]$ a mathematical representation of a signal in the form of the integer function (polynomial, trigonometrical, and indicative) is known, thus a priori knownness of a general view of the integer function on the transmitting side, a signal can be transferred by transfer of a final quantity of the numbers representing the coefficients of this integer function.

The mathematical apparatus of decomposition of functions in orthogonal bases is known [12]. A mathematical language of pure mathematicians [12] demands essential efforts for adaptation (translation) of this language into the language clear to the practical engineers who are engaged in development of algorithms of preliminary data processing on the transmitting side and the hardware of communication realizing these algorithms.

If the orthogonal basis has a final quantity of orthogonal axes of basis of decomposition, then representation of a signal with a limited range supports the final number of members of decomposition in this basis.

A final, without the rest, number of members of representation of the function continuous on a final piece $[a, b]$, also represents the main idea of a possibility of representation of this function with the use of a final quantity of numbers.

It is possible to use a concept of the "generalized limited range" outlined in [10].

The polynomial of a final degree represents the decomposition of a signal on the degrees representing decomposition in the orthogonal basis $(1, x, x^2, x^3 \dots)$.

If the coefficients of decomposition members are arranged on the extent of decrease on the absolute value,

then the size of coefficient is an analog to the amplitude, and the smallest decomposition coefficient is a certain analog to “the boundary frequency” of W for the signals with the generalized limited range.

Conclusion

1. The statement that “functions with a limited range are integer analytical functions” [8], contains either inaccuracy or a slip. Actually, one should speak about a narrower class of the integer analytical functions – about analytical functions with the final number of members. And if it is about polynomials, then about polynomials of final degree especially as the formula (row) of Kotelnikov supports the final number of members (this private or general statement should be proved or disproved in the theorem 1).

2. A continuous function with a limited range on $[a, b]$ using sampling can be presented (is approximated) with apply of a polynomial of a final degree.

3. Representation (approximation) on $[a, b]$ of continuous signals employing a polynomial of a final degree sets tasks to evaluate an error of approximation and the choice–justification (depending on a preset value of an error of approximation) of the maximum degree of the approximating polynomial.

References

1. Kotelnikov V.A. O propusknoy sposobnosti “efira” i provoloki v elektrosvyazi [On the capacity of ether and wire in telecommunication]. Radiotekhnika [Radioengineering], 1995, No. 4–5 (double), pp. 42–55. (in Russian)
2. Shannon C. Raboty po teorii informatsii i kibernetike [Works on Information and Cybernetics]. Moscow, Izdatel'stvo inostrannoy literatury, 1963, pp. 243–332. (translated from English)
3. Jerry J.A. Teorema otschetov Shennona, ee razlichnye obobshcheniya i prilozheniya [The Shannon Sampling Theorem — Its Various Extension And Application]. TIIEER, 1977, Vol. 65, No. 11, pp. 53–89. (in Russian)
4. Goldman S. Teoriya informatsii [Information theory]. Moscow, Izdatel'stvo inostrannoy literatury, 1957, 446 p. (translated from English)
5. Akhiezer N.I. Lektsii po teorii approksimatsii [Lectures on approximation theory]. Moscow, OGIZ, GIT-TL, 1947, 324 p. (in Russian)
6. Gnedenko B.V. Vvedenie v spetsial'nost' “matematika” [Introduction to mathematics]. Moscow, Nauka, 1991. (in Russian)
7. Markushevich A.I. Tselye funktsii. 2-oe izdanie [Entire functions]. 2nd edition. Moscow, Nauka, 1975, 119 p. (in Russian)
8. Khurgin Ya.I., Yakovlev V.P. Metody teorii tselykh funktsiy v radiotekhnike, teorii svyazi i optike [Methods of the theory of entire functions in radio engineering, communication theory, and optics]. Moscow, Fizmatgiz, 1962, 220 p. (in Russian)
9. Pobedonostsev V.A. Osnovaniya informmetrii [Informetrics foundations]. Moscow, Radio i svyaz', 2000, 193 p. (in Russian)
10. Pobedonostsev V.A. Teorema o neravnootstoyashchikh otschetakh [The theorem on unequally spaced readings]. Radiotekhnicheskie tetrady. Spets. vyp. [Radio electronics booklets Special edition]. Moscow, MEI, 1995, No. 8, pp. 25–27. (in Russian)
11. Pobedonostsev V.A. Teoreticheskie voprosy izmereniya kolichestva informatsii nepreryvnykh signalov na konechnykh intervalakh [Theoretical Questions in Measuring the Amount of Information in Continuous Signals on Finite Intervals]. Raketno-kosmicheskoe priborostroenie i informatsionnye sistemy [Rocket-Space Device Engineering and Information Systems], 2014, Vol. 1, No. 2, pp. 47–58. (in Russian)
12. Razmakhnin M.K., Yakovlev V.P. Funktsii s dvoynoy ortogonal'nost'yu v radioelektronike i optike [Functions in double orthogonality in radio electronics and optics]. Moscow, “Sov. Radio”, 1971, 256 p. (in Russian)

Modernization of Algorithms for Generating Correlated Data When Integrating Diverse Reception Methods and Noiseless Decoding Methods

V.L. Vorontsov, *Cand. Sci. (Engineering)*, a762642@yandex.ru

A branch of Joint Stock Company "United Rocket and Space Corporation"

– "Institute of Space Device Engineering", Moscow, Russian Federation

Abstract. Some aspects of integration of diverse reception methods and noiseless coding (decoding) methods are considered for cases of soft decision demodulation (in particular, Viterbi decoding). Since the developed algorithms for generating correlated data A4 and A42 that adapt to the time-varying conditions of the interference situation and ensure high data reliability, are oriented toward hard decision demodulation they need to be modernized to implement the above-mentioned integration. The essence of the modernization is described, and legitimacy of the modernization is justified. The modernization is connected with the formation of reliability estimates for the symbols (in particular, symbols in the form of the four-position signal implementation) received under the unfavorable conditions that are due to systematic errors caused by channel interference, as it has been deduced from experiments.

Keywords: algorithms for correlated data generation, analog implementation of four-position (two-position) signal, hard decision demodulation, noiseless coding (decoding) methods, diverse reception methods, soft decision demodulation, symbol.

It is shown [1, 2] that diverse reception methods and noiseless coding (decoding) methods are not alternative in principle. Moreover, when they are integrated, the possibilities for reliability improvement increase. In such a case, it is necessary to observe certain rules. In particular, in the case of sequential application, first, the correlated data is obtained, and then they are decoded [1, 2]. This approach makes it possible to use rationally the integration possibilities; it is relevant in the case of applying highly efficient algorithms A_4 and A_{42} for generating correlated data [1, 2], adapting to the time-varying conditions of the interference situation, and subsequent decoding oriented toward the hard decision demodulation.

However, if the decoding is oriented toward the soft decision demodulation (in particular, Viterbi decoding), then the algorithms A_4 and A_{42} [3] need to be modernized (in case they are applied). In this case, it is necessary that the format (structure) of a datum at the software-hardware output for correlated data generation or at the decoder input with traditional soft decoding be the same. The components of such a datum are sharply delimited. It consists of an elementary datum and its estimate, namely: its high bits refer to an information datum selected from a certain received symbol, or to an analog implementation of a digital signal (refer to the elementary datum in terms of [1, 2, 3]), and low bits refer to reliability estimate of this symbol (of this analog implementation of a digital signal or of this elementary datum).

The purpose of the work is recommendations on correlated data generation for subsequent decoding oriented toward the soft decision (for) demodulation.

An example of an output three-bit datum of an eight-level soft modem for the modernized algorithm A_{42} is considered [3] (see Fig. 7.8 and explanations to it [4]). Fig. 1 is similar to Fig. 7.8 of [4]). In Table 1, e_i are values of an elementary one-bit datum of i -th diverse channel (0 and 1); the leftmost column of estimates e_i contains estimates of the most reliable elementary data «0» and «1», and rightmost column contains the least reliable ones. Values of these estimates are conditional (they may not be the same for the circuit implementation of the corresponding software and hardware). They are primarily chosen for clarity. Their substance is related to the Euclidean distance that is characteristic of the soft decoding scheme (and not to the Hamming distance, as in the case of the hard decoding scheme) [4].

Table 1. Values of the elementary one-bit data e_i and corresponding two-bit estimates

e_i	Estimates e_i			
0	00	01	10	11
1	11	10	01	00

Further generation of correlated data shall be performed by the algorithm A_4 (A_{42}), with using, as usual, only elementary data acquired from the diverse channels [1, 2]. If an elementary datum e_{cor} is selected for the correlated data block and elementary data of the same value and corresponding to the same transferred datum and the aforesaid correlated datum e_{cor} are acquired from the i_1, \dots, i_h diverse channels, then the reliability estimate of a symbol from one of the i_1, \dots, i_h channels is added to this correlated datum, which value matches the most reliable symbol (elementary datum). This addition results in generation of a datum e_{cor_est} , the structure of which is identical to the structure of a soft modem output datum (Table 2). It means that the format of the words (data) e_{cor_est} and e_{i_est} are the same (where e_{i_est} is a datum containing the reliability estimate of the received symbol from the i -th diverse channel).

Table 2. An example illustrating how correlated data e_{cor_est} are generated by using the modified algorithm A_{42}

e_{cor}	Reliability estimates of symbols for diverse channels, $i =$					e_{cor_est}
	1	2	3	4	5	
1	-	01	00	-	11	111
0	-	-	-	-	01	001
0	10	10	01	11	-	001
.....

Thus, the first row of Table 2 describes the case for which e_{cor} is 1, the best estimate of reliability is 11 (it refers to the fifth diverse channel), and the datum e_{cor_est} is 111.

The propriety of the established procedure for selection of reliability estimates for correlated elementary data is related to the entities of symbol reliability estimates (reliability estimates of elementary data) pertaining to a separate diverse channel (e.g., shown in Figure 1), as well as the peculiarities of the elementary data selection for a correlated data set when implementing the algorithm A_4 (A_{42}) [1, 2].

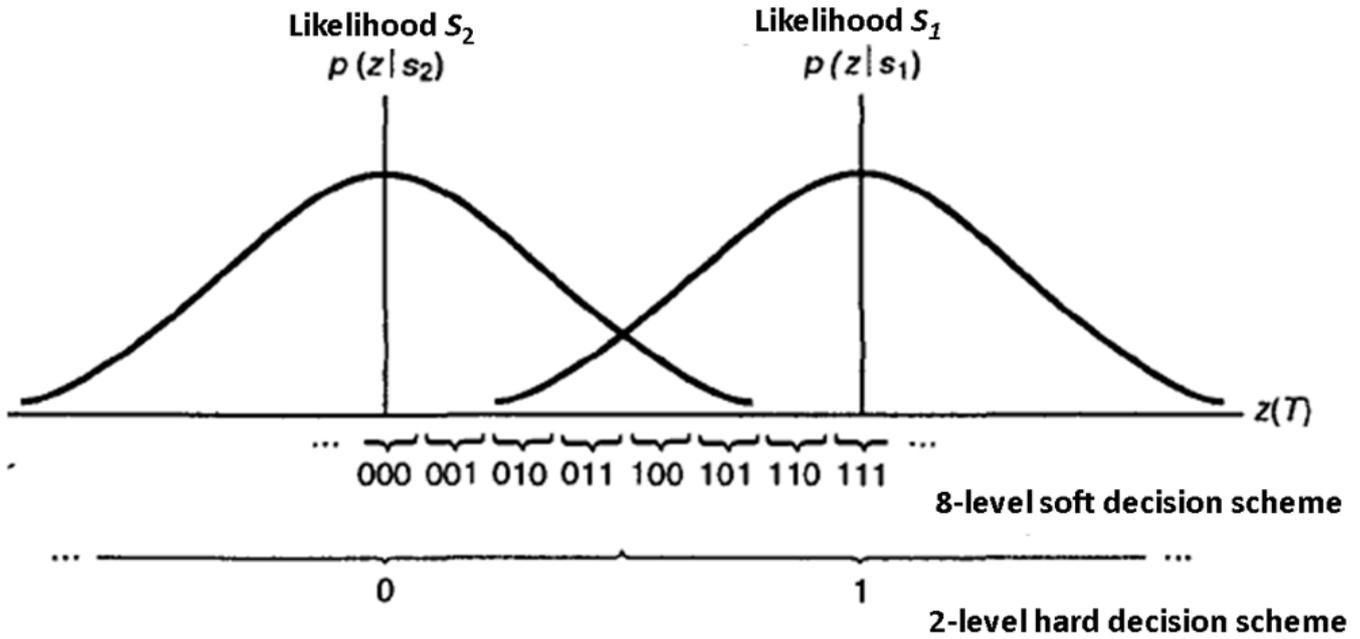


Fig. 1. Graphs explaining how the hard decision decoding scheme and the soft decision decoding scheme work

The closer the received signal value z is to any position of an ideal signal («0» or «1» are for the example in question, see Fig. 1), the less the probability of its erroneous identification is. The following statement is also true: the minimum deviations of the received signals z from their ideal values most often correspond to the most reliable elementary datum. On the other hand, the reliability provided by the algorithms A_4 and A_{42} is not worse than the reliability when using auto-selection (i.e. $P_{er_i} \geq P_{er_{cor}}$, $i = 1, 2, \dots, n$, where P_{er_i} is the probability of identification error in the i -th diverse channel, $P_{er_{cor}}$ is the probability of an erroneous elementary correlated datum) [1]. For example, if from n data blocks received from the diverse channels and corresponding to the same transmitted data block, the minimum amount of invalid elementary data (in comparison with the remaining blocks) is contained in the n -th block, but at the same time $P_{er_n} > P_{er_{cor}}$ and correlated data are obtained using the algorithms A_4 and A_{42} , then the reliability estimates of the elementary data of the n -th block should be worse than the estimates for the block of the received correlated data selected from the estimates corresponding to the elementary data of the i_1 -th, \dots , i_h -th diverse channels. In other words, it is logical that reliability estimates of the correlated elementary data obtained using the algorithms A_4 and A_{42} are no worse than when using auto-selection.

The normalization of the received signal is essential.

It is shown [5] that the normalization of the received four-position telemetry signal (Figure 2) is based on bringing the telemetry values of each diverse channel to a unified measurement scale (to nominal levels).

Table. 3 implies that in the case of the above normalization, selection of two high bits of a binary eight-bit word corresponding to an analog implementation of a four-position signal (see Figure 2) is equivalent to the use of threshold separation (equivalent to data recognition). The values of the thresholds 1, 2, and 3 are 63.5, 127.5, and 191.5 binary units respectively. Such normalization and identification provide very high reliability, if the interference component of a received signal is additive, and its mathematical expectation is zero [5]. They also provide clear separation of chunks of data related to the reliability estimates of elementary data, and elementary data themselves, which is very important.

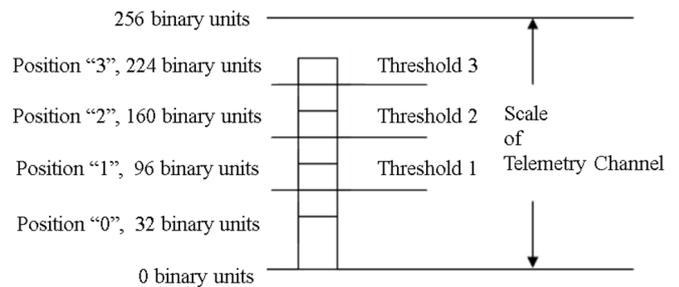


Fig. 2. A graph of a four-position signal with nominal values of position levels

Table 3. The results of binary-coded representation of the values of nominal position levels for a four-position signal

Position No.	Values of position levels, binary units	bit values							
		1	2	3	4	5	6	7	8
1	32	0	0	0	0	0	1	0	0
2	96	0	0	0	0	0	1	1	0
3	160	0	0	0	0	0	1	0	1
4	224	0	0	0	0	0	1	1	1

Information on reliability estimates is contained in the fragment b_e of the word e , from the first to the sixth bits, shown in Table. 3. They characterize deviation of a received four-position signal from its nominal level. These characteristics are the initial data for obtaining reliability estimates of elementary data in the format required for further decoding. In the context of the purpose of this work, the determination of this format is premature.

However, it is necessary pay attention to some features associated with it.

Thus, because of the asymmetry of the communication channel, interference with a negative sign and a positive sign at the positions «0» and «3», respectively, do not cause errors in the data recognition considered above (see Figure 2). In these cases, values of deviation of a received four-position signal from its nominal levels «0» and «3» (respectively from 32 and 224 binary units) are not important [6]. Because of the asymmetry (in particular, the above one) with a received degraded four-position signal, double bit errors are possible. For example, it is the case when position «2» (10) is transferred, and it is recognized as position «1» (01); and vice versa: «1» is transferred, and it is recognized as «2». To reduce the probability of such errors, Gray modulation codes are used [4]. For example, for the above four-position signal (see Figure 2) in such a case, it is expedient to determine the following correspondence between the position number and binary digits of an elementary datum: «0» - 00, «1» - 01, «2» - 11, «3» - 10. Therefore, the probability of a double bit error will significantly decrease (see the example above). We have a single bit error: the position «2» (11) is transmitted, and the position «1» (01) is recognized; and vice versa. It is shown [6] that elimination of asymmetry can be also achieved by shifting the levels

of the first and third thresholds to the corresponding nominal levels of the extreme positions. However, the data reliability degrades, which is unacceptable.

In the case considered (see Table 3), the datum b_e contains information on the reliability estimate of an elementary datum presented by two high bits – the seventh and eighth ones. However, it is possible to have the datum e_{cor_est} in the format of a one-bit elementary datum instead of a two-bit elementary datum, which is necessary for further decoding. This means that it becomes necessary to generate two words from each source word containing a two-bit elementary datum, in each of which an estimate of its reliability (a b_e -type datum) is attached to a bit of the elementary datum. It is assumed that the essence of such generation will be determined in the future.

The bit capacity of the b_e -type datum is set depending on the required accuracy of reliability estimates. To get rough estimate, you just need to drop its lower-order digits.

Some practical aspects concerning the justification of the required accuracy of the reliability estimates of the b_e -type are considered, related to the normalization of the received four-position telemetry signal (see Figure 2).

Experiments proved [5] that the quality of the normalization performed by hardware at ground receiving and recording stations (RRS) of the BRS-4 (BRS-4M, BRS-4MK) type is unacceptably low. There is a wide spread of telemetry values obtained by different RRSs involved in receiving telemetry information (TMI) from one source, compared with each other and compared with nominal levels (see, for example, the onboard calibration mean values of the position median of four-position signals $U_{med_i_mean}$, where i is a position number, available from experiments and represented in binary units in Table 4). Therefore, now at the Cosmodrome computing center (CC) the normalization is repeated when the TMI is processed by an analog implementation of an analog signal.

Table 4. Experimental data characterizing the quality of normalization performed at RRS

RRS	$U_{med_1_mean}$	$U_{med_2_mean}$	$U_{med_3_mean}$	$U_{med_4_mean}$
MK-12a	31	98	166	235
MK-166	30	92	155	220
Nominal levels	32	96	160	224

The results of the studies [5] performed at the Cosmodrome CC showed that the best possible normalization capabilities are provided when using the four-position sinusoidal signal of the onboard calibration of BITS of BRS-4 type «Scut-40», a feature of which is a priori known phase (Fig. 3). To bring the signals to nominal levels (see Fig. 2), the following transformations are necessary:

$$u_{norm} = au + b, \quad (1)$$

where u_{norm} (u) is the value of normalized (not normalized) telemetry;

a, b – coefficients.

Coefficients a and b (1) are obtained by solving a system of two equations:

$$\begin{cases} U_{norm_max} = aU_{med_4} + b, \\ U_{norm_min} = aU_{med_1} + b. \end{cases} \quad (2)$$

From (2) it follows that

$$\alpha = \frac{U_{norm_max} - U_{norm_min}}{U_{med_4} - U_{med_1}} = \frac{224 - 32}{U_{med_4} - U_{med_1}} = \frac{192}{U_{med_4} - U_{med_1}}, \quad (3)$$

$$b = U_{norm_max} - aU_{med_4} = 224 - aU_{med_4}.$$

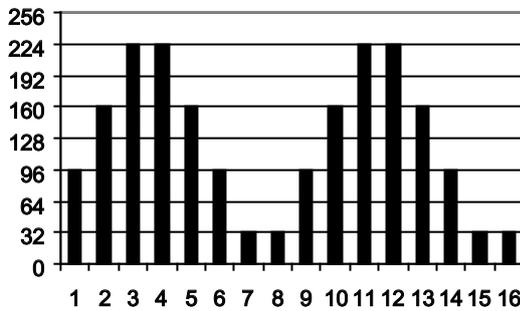


Fig. 3. A graph of a four-position sine wave signal of onboard calibration

The position medians of the onboard calibration signal (not the average values of position levels) make it possible to avoid shifting estimates of the values of the first and fourth positions when the received signal exceeds the edges of the telemetry channel scale (see Fig. 2).

Requirements Δu_{acc} for acceptable interference are established [5]:

$$|U_{nom_i} - U_{med_i}| < \Delta u_{acc}, \quad i = 1, 2, 3, 4, \quad (4)$$

где U_{nom_i} is the nominal value of the i -th position level of a four-position signal.

During the normalization process, the values of the coefficients a and b are calculated and continuously updated for each data block within time intervals of relatively stable reception of TMI (when condition (4) is satisfied). If condition (4) is not satisfied (i.e. there is relatively strong interference within the diverse channels), then the previously calculated and stored values of the coefficients a and b are used.

It is seen (see Fig. 2) that in the normalized form

$$U_{norm_med_1} = U_{norm_min} = 32, \quad (5)$$

$$U_{norm_med_4} = U_{norm_max} = 224,$$

where $U_{norm_med_1}$ ($U_{norm_med_4}$) is the normalized value of the first (fourth) position median of the onboard calibration four-position signal.

For example, note that for the TMI RRS MK-12a and MK-16b (see Table 4), the specified formula (1) will be given by:

$$u_{norm_MK-12} = 0,9412 \cdot u + 2,824,$$

$$u_{norm_MK-16} = 1,0105 \cdot u + 1,690.$$

The maximum position deviation from the nominal values is $\Delta u_{max_MK-12} = 224 - 235 = -11$ (binary units), which is approximately 5.7% of the nominal scale of the reference signal; $\Delta u_{max_MK-16} = 160 - 155 = 5$ (binary units) which is approximately 2.6% of the nominal scale of the reference signal (see (4)). In fact, these are characteristics of a systematic error; their values are relatively small, but they can significantly degrade reliability (see Table 5 below and explanations to it).

The process of normalization is very important in terms of code conversion «LS-UK», which consists in selecting two semantic bits from an analog implementation of a four-position signal, represented in the structure of LS and S4 in the form of an eight-bit binary word [5].

At the Cosmodrome CC, tests of threshold setting methods when selecting a two-digit elementary datum from an analog implementation of a four-position signal were conducted [5].

In particular, the following methods of setting thresholds have been tested.

1. Automatic setting of thresholds in Threshold Separation Units (TSU) of Digital Input Equipment (DIE) (normal operation mode of TSU of DIE, usually used in computer centers) with the training of TSU devices based on an onboard calibration four-position signal (see Fig. 3). (The name of the equipment «DIE» is assigned by its developer – IT NPO (IT Research and Production Association)).

2. Selection of the two high bits from eight-bit words corresponding to an analog implementation of a four-position signal (a method implemented with some applied computer programs providing transformations of the TMI structure of S4-type or LS-type into the structure of the UK-type).

3. Automatic setting, within the data block length, of threshold levels equal to half the sum of the medians of adjacent positions of an on-board calibration four-position signal. For example, to determine the value of the first threshold with the test on-board calibration four-position signal, the medians for the first and second positions are calculated, summed and divided by 2. The method is proposed for operation with the onboard radio telemetry system "SKUT-40" TMI containing a priori known onboard calibration four-position signal (see Fig. 3).

The first method was chosen as the basic one. During the tests a real TMI was used. Table 5 presents the results obtained with the scores E [1, 2] (E = -1 means that the test method provides *significantly* less reliability than the basic method, E = 0 means that the test method provides approximately the same reliability, and E = 1 implies *significantly* more reliability).

Table 5. The test results of threshold setting methods when applying code conversion «S4-UK»

Method No.	Quantity of data block (%), E =		
	-1	0	1
2	2	32	66
3	0	8	92

The third method provides the highest reliability, and its advantage is overwhelming (see Table 5). From the essences of the second and third methods it follows that with high-quality normalization of the received four-position signals and subsequent selection of the two high bits from the eight-bit words of the TMI with structure of S4-type or LS-type (as it is done with the second method), an effect similar to the third method is obtained. That means that they provide the relatively high reliability, as in the case of the third method. Herein, normalization is equivalent to eliminating systematic errors caused by interference (their mathematical expectation is more than zero), and the results of the experiment show that even with relatively small systematic errors that cause the position deviations of the received four-position signal from their nominal values, the reliability significantly [1, 2] deteriorates.

Thus, there are good possibilities for improving reliability by combining methods of diverse reception and noiseless coding (decoding) in cases where the decoding is oriented toward a soft decision demodulation. At the same time, modernization of the developed algorithms for generating correlated data A4 and A42, which is necessary to create conditions for the aforementioned decoding, does not require much effort. However, the successful modernization requires measures to normalize the received signals (to bring the levels of received m-position signals to their nominal values). The reliability provided by the algorithms A4 and A42, as well as the accuracy of reliability estimates of the received symbols (reliability estimates of analog implementation of the received m-position signals) required for a soft decision decoding depend on the quality of the above normalization.

References:

1. Vorontsov V.L. *Metody raznesennogo priema telemetricheskoy informatsii i usloviya ikh primeneniya v protsesse razvitiya telemetricheskogo kompleksa kosmodroma* [Methods of Diversity Receiving of Telemetry Data and Conditions of Their Application in the Development of a Telemetry Complex of a Cosmodrome], 2nd edition, revised and enlarged. Naberezhnye Chelny: Kama State Academy of Engineering and Economics, 2009, 284 p. (in Russian)
2. Correlated Data Generation, Research and Development for Space Data System Standards CCSDS 551.1-O-1, Issue 1, Orange Book, Consultative Committee for Space Data Systems, July 2015, <https://public.ccsds.org/Publications/NonEnglishVersions.aspx> (link to the Russian version of CCSDS 551.1-O-1).
3. Vorontsov, V.L. *Uluchsheniye dostovernosti dannykh putyom ispolzovaniya vozmozhnostey raznesonnogo priyoma* [Reliability Improvement through Making the Most Use of Data Diverse Reception Capabilities]. Rocket-Space Device Engineering and Information Systems, 2017, Vol. 4, Iss. 1, pp. 61-69. (in Russian)
4. Sklar B. *Digital Communications: Fundamentals and Applications*. 2nd edition, revised. Moscow: Publishing House «Williams», 2003, 1104p. (translated from English, in Russian)

5. Vorontsov V.L. O kontrole vernosti teleizmereniy bystromenyayushchikhsya parametrov pri letnykh ispytaniyakh raketno-kosmicheskoy tekhniki [On Control of Telemetry Correctness of Rapidly Changing Parameters during Flight Tests of Rocket and Space Technology]. Instruments and Systems: Monitoring, Control, and Diagnostics, 2004, № 3, pp. 49-55. (in Russian)
6. Vorontsov V.L. Ob ustraneni nesimmetrichnosti kanalov razneseniya pri vydelenii dannykh iz analogovykh realizatsiy chetyrekhpozitsionnykh signalov [On Eliminating the Asymmetry Of Diverse Channels When Data Is Extracted From Analog Implementation Of Four-Position Signals], in the collection of research papers: Research and Modeling of Rocket-Space Complexes and Their Components, Moscow, MAI, 2004, pp. 26-31. (in Russian)

A Method of Detecting of Internal Defects of CMOS-Microcircuits

I.Yu. Bulaev, bulaev.ivan@gmail.com

Joint Stock Company "Russian Space Systems", Moscow, Russian Federation

Abstract. For the acquisition of highly reliable equipment it is necessary to allow only those electronic components, which before the installation phase are fully functional and parametric. The testing center is often not able to conduct a full-fledged functional control of the electronic components products, because does not have information about internal blocks of complex functional products and how they interact with each other. The situation is complicated by the fact that the degree of integration and functionality of modern microcircuits is constantly growing. Another problem is the forecasting of further trouble-free operation of the product and the detection of internal hidden defects. At an early stage of operation, the microcircuit can contain insignificant internal hidden defects, the value of which practically does not affect the performance of the microcircuit. However, after a certain amount of time has elapsed as a result of the degradation processes of the product materials, the defect may increase and lead to irreversible failure of the product. Therefore, it is extremely important at the stage of autonomous testing of a product, not as part of equipment, to detect such internal defects. The article explains the need to search for small hidden defects in ECB products, suggests comparison of the most popular diagnostic methods, discusses the problems of using existing methods for diagnosing modern microcircuits, suggests an approach to detecting small internal defects of the microcircuit before the stage of its installation in equipment.

Keywords: defect, failure, CMOS, diagnostics, tests

Internal defects of electrical, electronic and electromechanical (EEE) parts, which include diodes, transistors, integrated microcircuits, microassemblies, passive components, etc., can sometimes not be detected by ordinary functional or parametric control. In cases where the defects are too small, they do not have a significant effect on the operation of the EEE products (hereinafter - the products) and do not lead to either the falling of electrical parameters beyond the permissible norms, or to a malfunction.

However, after a certain amount of time has elapsed as a result of the degradation processes of the product materials, the defects may increase in magnitude and lead to irreversible failure of the product. Therefore, it is extremely important at the stage of off-line testing of a product, not as a part of equipment, to detect such internal defects.

In addition, the complexity of modern microcircuits is so great that the task of performing full functional control of all internal components “in the far corners” of the product becomes practically impossible. The situation is aggravated by the fact that the tester, as a rule, does not have any information about the internal structure and circuitry implementation of the microchip.

To identify hidden defects and compensate for insufficient functional monitoring, such methods as burn-in testing are used, during which the products are influenced by various factors that accelerate the aging process of materials: voltage, temperature, etc. With such testing, within a relatively short time period (several days), it is possible to simulate a certain lifetime of the product (several years). After the test, either the potentially unreliable products fail, or the drift of their informative parameters (the difference in values measured before and after burn-in testing) will be very different from the drift of serviceable products.

However, burn-in testing is a fairly long and expensive test, requiring specialized equipment, rigging and method selection techniques. Burn-in testing methods can be replaced by diagnostic non-destructive testing (NDT) methods, which indirectly, by measuring electrical parameters (currents, voltages, time intervals), judge whether there are internal defects in the product.

Practically all methods of NDT are based on statistical processing and comparison of the measured parameters for a batch of products under the same impact. NDT applicable only for products of the same batch.

Hereafter, products containing a defect will be referred to as “unreliable”, and the products not containing a defect will be referred to as “reliable”.

NDT methods, in addition to their advantages, which include the ability to reject unreliable products at minimal cost, have a number of disadvantages.

It should be noted that not all of the above methods are applicable to any type of EEE products. Table 1 demonstrates the applicability of NDT methods for each type of EEE product.

From the table it is seen that for the NDT of microcircuits and microassemblies all the above methods are applicable, while for the control of the passive components there are the fewest methods. On the one hand, this situation seems logical; microcircuits and microassemblies are the most complex types of EEE products, and to verify their quality, it is necessary to conduct as many inspections as possible. On the other hand, from the conclusions of the commissions on the analysis of failures of devices it is evident that often the cause of the malfunction is a hidden defect in one of the passive components.

Thus, it is necessary to conduct research on the development of new, more advanced NDT techniques for passive components.

Table 2 shows the advantages and disadvantages of NDT methods.

A common disadvantage of NDT methods is that in the case of a defect in most products in the batch, there is a risk of rejecting of reliable products, because their parameters may not fall into the confidence range, while unreliable products will not be rejected at all. The main reason for this problem lies in the absence of a “reference” sample with previously unknown culling boundaries, whereby it is necessary to determine the boundaries using statistical methods.

To solve this problem, it is proposed to organize the cooperation of manufacturers of EEE products with test centers, which should consist in transferring to the testing centers of the models on which it would be possible to assess the degree of change in the parameters of the product when a defect of a certain value is introduced into one or another part of the product.

In addition, it follows from the table that a number of methods require the use of high-precision measuring equipment and sophisticated testing equipment, which is practically absent from the Russian market. The use of foreign-made equipment entails certain problems related

Table 1. Applicability of NDT methods for different types of EEE products

Type of method	Integrated microcircuits and microassemblies	Semiconductor devices	Passive components
Tightened norms	+	+	+
Static current consumption	+	-	-
Dynamic current consumption	+	-	-
Critical power supply voltage	+	-	-
Hysteresis of the parameters	+	+	+
Current-voltage analysis	+	+	-
Radiation methods	+	+	+
Low frequency noise	+	+	+
Electrostatic method	+	+	+

Table 2. Advantages and disadvantages of different NDT methods.

Type of method	Advantages	Disadvantages
Tightened norms	A method simple in the implementation that allows detecting products with anomalous characteristics	Danger of rejecting the reliable samples
Static current consumption	A method simple in the implementation that allows detecting internal parasitic resistances	It is impossible to detect small defects in products with a high degree of integration
Dynamic current consumption	A method that makes it possible to detect internal parasitic resistances and capacitances	High complexity of implementation and processing of results
Critical power supply voltage	A method simple in the implementation that allows detecting internal parasitic resistances	None
Hysteresis of the parameters	The ability to detect products with unstable construction	Requires high-precision measuring equipment and expensive test equipment
Current-voltage analysis	The possibility of detecting defects of $p-n$ -junctions	There is no possibility to diagnose $p-n$ -junctions of internal gates of microcircuits
Radiation methods	Detection of objects with reduced resistance to radiation	High complexity of implementation, risk of damage to the product
Low frequency noise	Detection of samples with an increased number of impurities, dislocations and microfissures	High-precision measuring equipment is required
Electrostatic method	Detection of objects with reduced resistance to electrostatic discharge	High complexity of implementation, risk of damage to the product

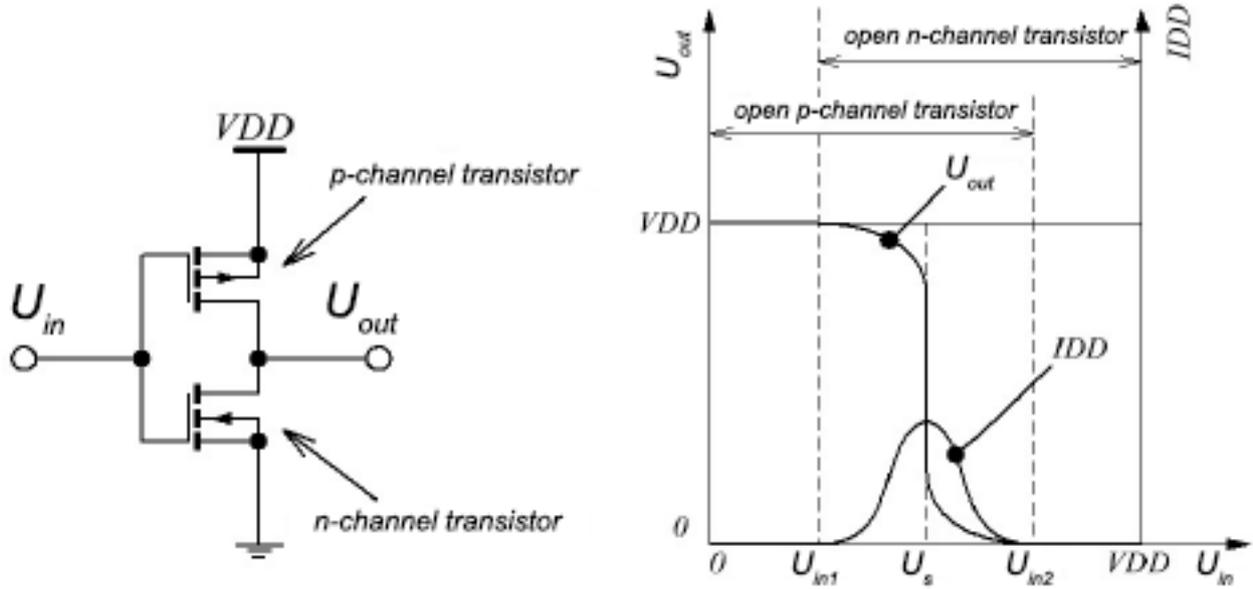


Fig. 1. CMOS inverter transfer characteristic

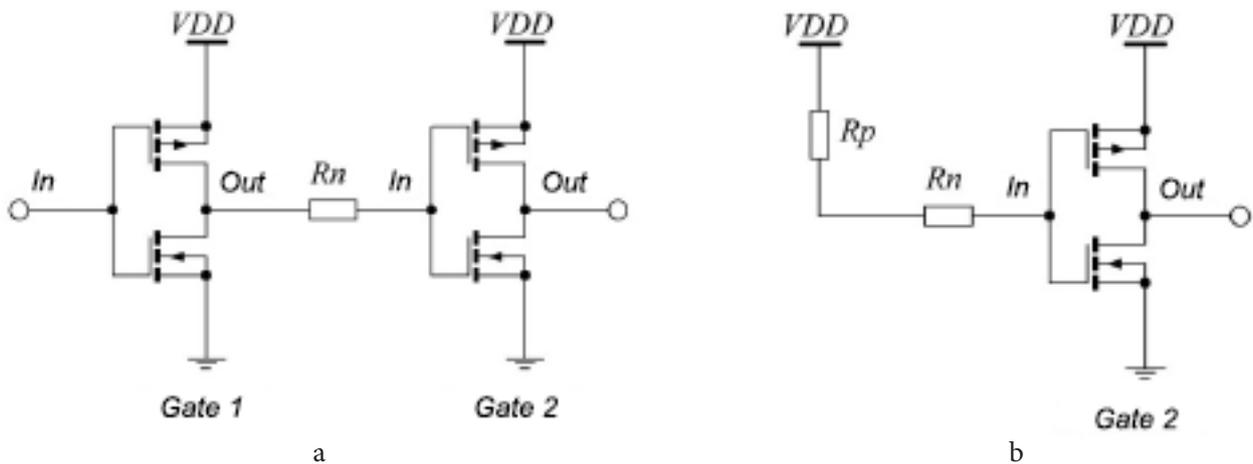


Fig. 2. The connection of the two gates through the parasitic resistance R_n (a) and the equivalent circuit with the voltage at the input of the first gate equal to 0 V (b)

to the cost, delivery times and the level of technical support. To solve this problem, it is necessary to develop own equipment that can provide the necessary level of accuracy and effects.

A major problem is that, due to the increased integration of modern EEE products and the use of fundamentally new materials, the old NDT methods show low efficiency and their use in the testing process is not economically viable. It is necessary to carry out work, involving test centers and manufacturers of EEE parts, to improve the old methods and develop new ones. The work should include mandatory testing of the effectiveness of NDT methods, for example, by performing a correlation analysis of NDT results and resource tests.

In addition, correlative analysis of NDT and burn-in testing results should be carried out. In the case of obtaining a high correlation, it is necessary to replace the long-term and expensive burn-in testing with NDT.

At the moment, the EEE manufacturers do not recognize the results of the NDT carried out at a testing center. This leads to the fact that the customer of the tests does not have the opportunity to carry out the complaint work for products rejected by the results of NDT. To solve this problem, it is proposed to approve some NDT methods at the level of state or industry standards.

The most widely used NDT methods, such as rejection for static and dynamic consumption currents,

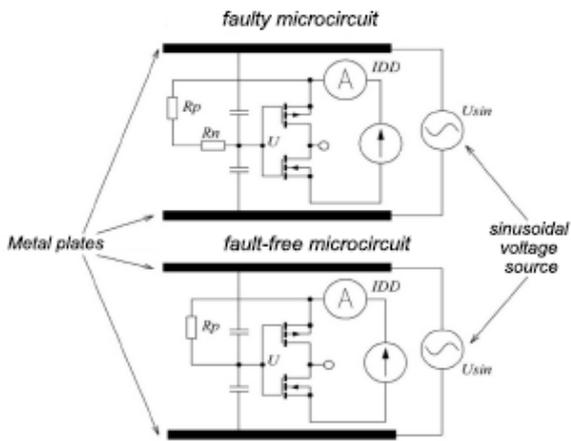


Fig. 3. Scheme of the experiment

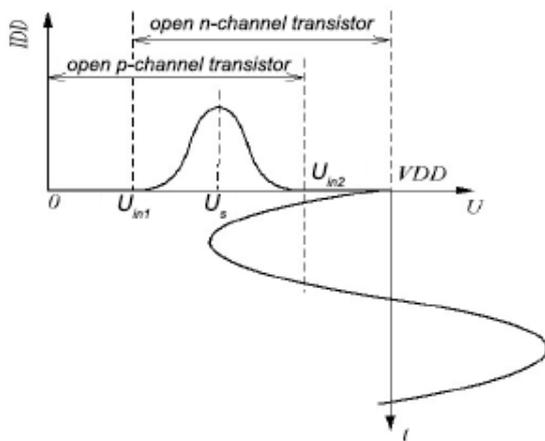
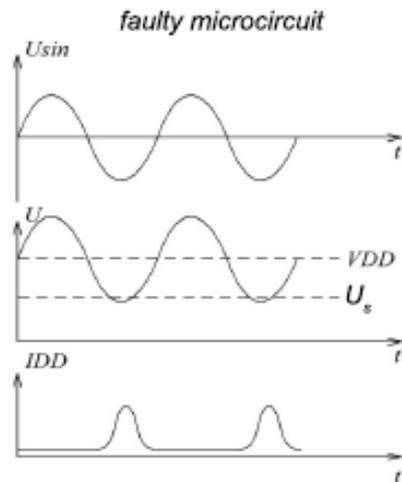


Fig. 4. The appearance of the pulse of the IDD current consumption during interference at the input of the gate U of the threshold value U_t

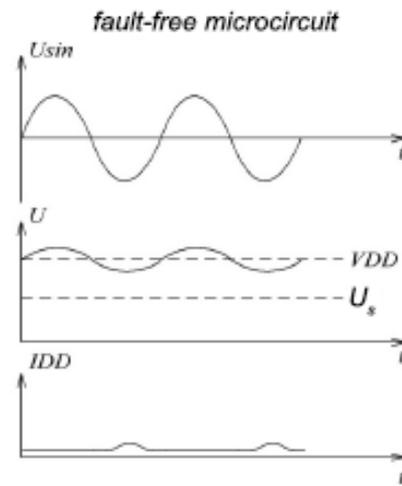


Fig. 5. The difference in the oscillograms of the current consumption of IDD in a fault free and faulty microcircuit with the same voltage level U_{sin}

critical supply voltages, etc., have one common drawback: in order to detect a defect, it must be “activated”, i.e. switching a gate containing a defect from one logical state to another. Thus, to ensure 100% test coverage, it is necessary to switch each gate of the EEE product and at the same time make an informative parameter measurement, which is difficult for modern microcircuits. Therefore, there is a need to develop NDT methods that affect the entire product in an integrated manner and detect the presence of a defect anywhere in the structure of the product, albeit without the possibility of its localization.

The most common internal defects of modern CMOS chips are low resistance (in the extreme case, a short circuit) between two points of the circuit where the resistance should be maximum, or a large resistance (in

the extreme case, an open) between two points where resistance should be minimal. In foreign literature, such defects are called “short” and “open” [1].

If the defect of the microcircuit is minor (for example, closing the output of the gate to 10 kΩ or an incomplete opening of the conductor between two gates, increasing the resistance of the circuit by 100 Ω, etc.), it may not be detected during routine functional and parametric control because of the imperfection of the tests and characteristics of control and measuring equipment.

However, when such a chip will work for some time as a part of equipment, the proportions of the defects may increase, which will lead to the failure of the chip. Thus, it is necessary to detect such defects before the installation of the chip into the equipment. The problems of detecting

defects of the “short” type were discussed in [2]. This article provides a theoretical justification for the method of searching for defects of the “open” type.

Consider the transfer characteristic of a CMOS gate (Figure 1). With the voltage at the input of the gate U_{in} from 0 V to U_{in1} , the n-channel transistor is closed, the p-channel transistor is open; there is no connection between the power supply and the ground of the chip and the I_{DD} consumption current is practically zero. A similar situation occurs when the voltage at the gate input is in the range from U_{in2} to V_{DD} , when the p-channel transistor is closed and the n-channel transistor is open. With the voltage at the gate input in the range from U_{in1} to U_{in2} , both transistors open, forming a galvanic connection between power and ground. In this case, the current consumption of the chip increases, and the maximum it will be at the voltage at the input of the gate equal to the threshold switching voltage U_s .

Defects of the “open” type in CMOS structures are divided into two main types: a gate opening and a drain or source opening. Consider the first type: the appearance of parasitic resistance R_n between the output of the first gate and the input of the second gate (Figure 2, a). Suppose that the input of the first gate is set to 0 V. Then the output of the first gate connects the input of the second gate to the power supply through the series resistance of its p-channel transistor R_p and the parasitic resistance R_n (Fig. 2b).

Thus, in comparison with a fault free microcircuit, in a defective chip, the gate input has a weaker connection with the power supply, and, therefore, is more susceptible to interference effects.

We place the microcircuit in an alternating electric field, placing it between two metal plates to which a sinusoidal voltage is applied, and we will make a measurement of the current consumption (Figure 3).

Between each point of the chip and the metal plates, a capacitance is formed, therefore, the inputs of all the gates of the chip will be interfered with, the magnitude of which can be varied by the amplitude and frequency of the sinusoidal voltage.

The level of induced interference U at the input of the gate of the faulty chip will be greater than that of the fault free one. By increasing the amplitude and frequency of the sinusoidal voltage U_{sin} , one can achieve a level of interference sufficient to open the p-channel transistor (Fig. 4).

In this case, the through consumption current will flow. For a working chip with the same sinusoidal voltage parameters, the interference level will be much lower and will not lead to gate switching (Fig. 5).

Thus, by varying the parameters of the sinusoidal signal and comparing the consumption currents of individual samples of a single batch of chips, it becomes possible to identify microcircuits with internal defects of the “open” type.

References

1. Segura J., Hawkins C.F. *CMOS electronics: How it works, how it fails*. Wiley-IEEE Press. 2004, 346 p.
2. I. Yu. Bulaev. Diagnosticheskiy nerazrushayushchiy kontrol' EKB. Variatsii metoda kriticheskogo napryazheniya pitaniya [Diagnostic non-destructive control of ECB. Variations of the method of cutoff supply voltage]. Materialy XII nauchno-tekhnicheskoy konferentsii “Tverdotel'naya elektronika. Slozhnye funktsional'nye bloki REA” [Proceedings of XII scientific and technical conference “Solid-state electronics. Complex functional units of electronic equipment”]. Moscow, OAO “NPP “Pul'sar”, 2013, pp. 256–259. (in Russian)

