NASA TECHNICAL TRANSLATION

NASA TT F-785

NASA TT F-785

29

(NASA-TT-F-785) EQUIPMENT FOR SPACE RESEARCH (Kanner (Leo) Associates) 417 p HC $10.50 CSCL 22B

N75-10988
THRU
N75-11016
Unclas
H1/18 53889

EQUIPMENT FOR SPACE RESEARCH

by Yu K. Khodarev, B. N. Rodionov, V. V. Andreyanov, N. N. Krupenio, L. I. Shatrovskiy, Yu. M. Shtar'kov, P. Ye. El'yasberg, and V. S. Etkin

Nauka Press
Moscow, 1972

NOV 19, RECEIVED NASA STI FACIL INPUT BRA

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • OCTOBER 1974

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TT F-785		

4. Title and Subtitle	5. Report Date
EQUIPMENT FOR SPACE RESEARCH	October 1974
	6. Performing Organization Code
7. Author(s) Yu. K. Khodarev, B. N. Rodionov, V. V. Andreyanov, N. N. Krupenio, L. I. Shatrovskiy, Yu. M. Shtar'kov, P. Ye. El'yasberg and V. S. Etkin	8. Performing Organization Report No.
	10. Work Unit No.
9. Performing Organization Name and Address Leo Kanner Associates P. O. Box 5187 Redwood City, California 94004	11. Contract or Grant No. NASw-2481
	13. Type of Report and Period Covered Translation
12. Sponsoring Agency Name and Address NASA, Code KSS-1 Washington, D. C. 20546	14. Sponsoring Agency Code

15. Supplementary Notes Translation of "Apparatura dlya kosmicheskikh Issledovaniy," Moscow, Nauka Press, 1972, pp 1-263.

16. Abstract Studies of interplanetary probe and landing equipment are given, covering radio and TV systems, orientation and computer systems, data synchronization and compression techniques and automatic follow-up and photogrammetric equipment. The state of this technology is discussed and future development are visualized.

Individual items are abstracted in this issue.

17. Key Words (Selected by Author(s))	18. Distribution Statement
	Unclassified. Unlimited STAR Category 14

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	416	$8.50

*For sale by the National Technical Information Service, Springfield, Virginia 22151

PRECEDING PAGE BLANK NOT FILMED

TABLE OF CONTENTS

Determination of Landing Point and Orientation of a Spacecraft on Mars, G. A. Mersov 1

Low-Noise Parametric Amplifiers for Radio Astronomy Observations at 18-21 cm Wavelength, B. Z. Kanevskiy, V. M. Veselov, I. A. Strukov, V. S. Etkin 23

Problems of the Design of Low-Noise Input Devices, V. M. Manokhin, Yu. A. Nemlikher, I. A. Strukov, Ye. A. Sharkov 28

Photogrammetric Method of Determination of Deflection of Light Beams by Porthole Glass Under Space Flight Conditions, Ya. L. Ziman 36

Development of the Design of a Between-the-Lens Band-Type Photographic Shutter, A. S. Dorofeyev 42

Method of Radar Studies of the Moon and the Planets from Spacecraft, N. N. Krupenio 48

Multiple-Slit Shadow Device for Sensing the Direction to a Shining Object, A. V. Bruns 65

Computers and Prospects for On-Board Spacecraft Electronic Equipment, V. V. Andreyanov 88

The Effectiveness of Correcting Codes in Reception in the Whole in Additive Normal White Noise, Yu. M. Shtar'kov 97

Selection of an Optimal Plan for Construction of a Spacecraft Antenna System, B. A. Prigoda, M. B. Faynshteyn, V. A. Gur'yev, A. S. Markin 107

Study of the Redundancy of Telemetry Information of the Luna-9 and Luna-13 Automatic Lunar Probes, A. V. Kantor, T. A. Tolmadzheva 120

Digital Computer Modeling of the Process of Reduction of Redundancy in Multichannel Telemetry Information by the Method of Adaptive Discretization with Associative Sorting, T. A. Tolmadzheva, A. V. Kantor, L. V. Rozhkovskiy 136

Circularly Polarized Antennas with Controlled Radiation Patterns, B. A. Prigoda, V. A. Gur'yev, N. S. Ul'yanin, N. N. Kalugin 163

Reduction in Redundancy of Multichannel Telemetric Information by the Method of Adaptive Discretization with Associative Sorting, A. V. Kantor, V. G. Timonin, Yu. S. Azarova 168

The Question of the Possibility of Using Associative Computer Devices (ACD) for Parallel Adaptive Discretization of Multichannel Telemetry Information, A. V. Kantor, S. M. Perevertkin, T. S. Shcherbakova 195

High Frequency Switching Devices, B. A. Prigoda, Ye. A. Vlasov, A. S. Markin, E. I. Yakovlev 217

SHF Filters in Spacecraft Antenna-Feeder Systems, Ye. A. Vlasov, B. A. Prigoda 222

Checking the Parameters of AFS in Spacecraft in Flight, Ye. A. Vlasov, B. A. Prigoda, A. S. Markin 228

The Accuracy of Automatic Tracking, V. V. Kastrov 235

Communications by Light with Long-Range Spacecraft, M. Z. Novikov 246

Data Compression, V. F. Babkin, A. B. Kryukov, Yu. M. Shtar'kov 264

Calculation of the Effectiveness of a Gamma Telescope, Yu. A. Rylov 319

The Problem of Controlling the Directivity of Low-Directivity Antennas on Spacecraft in Flight, V. I. Mashkov, B. A. Prigoda 335

Some Problems of the Design of Highly Directional Spacecraft Antennas, B. A. Prigoda 339

Potential Capabilities for Compression of Information of Certain Data Processing Systems, Yu. K. Khodarev, V. P. Yevdokimov, V. M. Pokras 353

Statistical Analysis of Information from Long-Range Spacecraft, Yu. K. Khodarev, V. P. Yevdokimov, V. M. Pokras 368

A Priori Estimate of the Quality of a Data Compression System Based on Statistical Characteristics of the Sensors Used, Yu. K. Khodarev, V. P. Yevdokimov, V. M. Pokras 379

Methods of Synchronization of Telemetry Frames Using Distributed and Combined Frame Pauses, Yu. M. Shtar'kov 387

Abstracts of Articles **404**

N75 10989

DETERMINATION OF THE LANDING POINT AND ORIENTATION OF A SPACECRAFT ON MARS

G. A. Mersov

1. Basic Relationships

The landing point of a spacecraft on the surface of Mars is determined by two coordinates: the areographic latitude ϕ_a and the areographic longitude λ_a. By definition, the areographic latitude ϕ_a is the angle between the plane of the equator of Mars and a perpendicular to the surface of the normal ellipsoid at the landing point. In this work, latitude ϕ_a shall refer to angle ϕ between a plumb line at the landing point and the plane perpendicular to the axis of rotation of Mars relative to the sun. /3*

The angle between the axis of rotation of Mars relative to the stars and the axis of rotation relative to the sun is about 2'. This quantity limits the accuracy of determination of latitude using this method. In place of the areographic longitude λ_a, we shall use the ephemeral time of the upper culmination of the sun τ at the landing point, which unambiguously determines longitude λ_a.

To determine the orientation of the spacecraft, let us introduce the coupled system of coordinates $OX_1Y_1Z_1$, stationary relative to the spacecraft, and the topocentric system OXYZ, the origin of which coincides with the origin of the coupled system; axis OZ is directed along a plumb line upward, axis OX -- along the meridian of the landing point in the direction of the equator.

Suppose we measure the following quantities in the coupled system of coordinates $OX_1Y_1Z_1$ at moment in time T: the height of the sun over the instrument horizon (plane X_1OY_1) f; the instrumental azimuth of the sun l; angles α and β, defining the direction of the plumb line in the coupled system of coordinates so that where $\alpha = 0$ and $\beta = 0$, axis OZ_1 coincides with axis OZ.

Since axis OZ coincides with the plumb line, angles α and β define the 3×5 matrix $B(\alpha, \beta)$ of the orthogonal transform of the coupled system of coordinates causing axis OZ_1 to coincide with axis OZ. Suppose in this case axis OX_1 forms angle ψ with axis OX, reading from axis OX to axis OY.

---

*Numbers in the margin indicate pagination in the foreign text.

Angle ψ describes the position of the spacecraft relative to the directions on the planet and can be defined jointly with the latitude ϕ and ephemeral time of culmination of the sun τ. We represent by $\bar{r}_{\odot_1}$ the unit vector of the direction to the sun in the coupled system of coordinates, by $\bar{r}$ the same vector in the topocentric system of coordinates; then, due to the definition of matrix $B(\alpha, \beta)$ and of angle ψ we can write

$$\bar{r}_{\odot} = \left\| \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\| B(\alpha, \beta)\, \bar{r}_{\odot_1} \tag{1.1}$$

Vector $\bar{r}_{\odot_1}$ is defined by the measured values of f and l

$$\bar{r}_{\odot_1} = \left\| \begin{bmatrix} \cos f & \cos l \\ \cos f & \sin l \\ & \sin f \end{bmatrix} \right\| \tag{1.2}$$

Vector $\bar{r}_{\odot}$ is expressed through the height of the sun over the local horizon h and the astronomical azimuth of the sun A /4

$$\bar{r}_{\odot} = \begin{bmatrix} \cos h & \cos A \\ \cos h & \sin A \\ & \sin h \end{bmatrix}. \tag{1.3}$$

Substituting expressions (1.2) and (1.3) into (1.1) and converting, we produce

$$\begin{bmatrix} \cos h & \cos A_1 \\ \cos h & \sin A_1 \\ & \sin h \end{bmatrix} = B(\alpha, \beta) \begin{bmatrix} \cos f & \cos l \\ \cos f & \sin l \\ & \sin f \end{bmatrix}, \tag{1.4}$$

where $A_1 = A - \psi$.

To establish the relationship of quantities h and A_1 with ϕ and τ at the moment of measurement T, we use the known equations for the parallax triangle [1]:

$$\sin h = \sin \varphi \sin \delta + \cos \varphi \cos \delta \cos t, \tag{1.5}$$

$$\sin A = \frac{\cos \delta \sin t}{\cos h}, \tag{1.6}$$

where δ is the declination of the sun to the equator of Mars at moment T, t is the hour angle of the sun at the same moment T.

If we ignore the change in the angular rotation rate of Mars relative to the sun over the time interval of measurements, which is several days, then

$$t = \omega_{\odot} (T - \tau),$$

where ω is the mean rotation rate of Mars relative to the sun.

Going over to measurement of the time in angular quantities (ω = 1), we can write

$$t = T - \tau. \tag{1.7}$$

We can estimate the values of ϕ, τ and ψ using equations (1.5) and (1.6) if the spacecraft makes no random motions during the entire time interval, i.e., if ψ = const.

If the spacecraft moves, the change in angle ψ can be determined by measuring the values of α and β, since there may be kinematic relations such as

$$F(\alpha, \beta, \psi) = 0,$$

caused by the contact of the spacecraft with the surface of Mars. Due to the indefinite nature of these relations, it is expedient at first to use only equation (1.5) to process the results of measurements of h(T) to determine ϕ and τ, then to use equation (1.6) and the calculated value of $A_1(T)$ to determine the value of ψ at various fixed moments in time.

Thus, we have n measurements of the value of h at moment T_i (i = 1, 2, ..., n) to estimate the values of ϕ and τ. Let us calculate the Gram determinant for this system of measured and estimated quantities: /5

$$\left|\frac{h_1 h_2 \ldots h_n}{\varphi, \tau}\right| \equiv \begin{vmatrix} \sum_{i=1}^{n}\left(\frac{\partial h_i}{\partial \varphi}\right)^2 & \sum_{i=1}^{n}\left(\frac{\partial h_i}{\partial \varphi}\right)\left(\frac{\partial h_i}{\partial \tau}\right) \\ \sum_{i=1}^{n}\left(\frac{\partial h_i}{\partial \varphi}\right)\left(\frac{\partial h_i}{\partial \tau}\right) & \sum_{i=1}^{n}\left(\frac{\partial h_i}{\partial \tau}\right)^2 \end{vmatrix} =$$

$$= \sum_{i=1}^{n}\left(\frac{\partial h_i}{\partial \varphi}\right)^2 \sum_{i=1}^{n}\left(\frac{\partial h_i}{\partial \tau}\right)^2 - \left(\sum_{i=1}^{n}\frac{\partial h_i}{\partial \varphi}\,\frac{\partial h_i}{\partial \tau}\right)^2, \tag{1.8}$$

where $\frac{\partial h_i}{\partial \varphi} \equiv \frac{\partial h(T_i)}{\partial \varphi}$.

According to the Bunyakovskiy inequality, expression (1.8) is equal to zero only if

$$\frac{\partial h_i}{\partial \varphi} = k\,\frac{\partial h_i}{\partial \tau} \quad (i = 1, 2, \ldots, n;\ k = \text{const}). \tag{1.9}$$

Let us calculate the values of the partial derivatives:

$$\frac{\partial h}{\partial \varphi} = \frac{\cos\varphi \sin\delta - \sin\varphi \cos\delta \cos(T-\tau)}{\cos h}, \tag{1.10}$$

$$\frac{\partial h}{\partial \tau} = \frac{\cos\varphi \cos\delta \sin(T-\tau)}{\cos h}. \tag{1.11}$$

Consequently, where $n \geq 2$, the ratio of partial derivatives, equal to

$$\frac{\partial h/\partial \varphi}{\partial h/\partial \tau} = \frac{\tan\delta}{\sin(T-\tau)} - \tan\phi \cot(T - t),$$

is a constant quantity only where $\phi = 0$ and $\delta = 0$.

Thus, where $n \geq 2$ and $\phi \neq 0$ or $\delta \neq 0$, the values of ϕ and τ can be estimated.

Where $\delta = 0$ and $\phi \neq 0$, derivatives (1.10) and (1.11) depend only on the modulus of ϕ and, consequently, only the modulus of the latitude can be determined by processing measurements of the height of the sun. The sign of the latitude in this case can be determined from the sign of the derivative (increment) of the instrumental longitude of the sun.

Let us go over from the values of ϕ and τ to the values of ξ and η, representing the linear deviation of the landing point from the calculated landing point (ϕ_0, τ_0) on the surface of the planet in the directions of the local parallel and meridian respectively:

$$\begin{aligned} \xi &= R \sin(\tau - \tau_0) \cos \varphi_0, \\ \eta &= R \sin(\varphi - \varphi_0). \end{aligned} \tag{1.12}$$

where R is the radius of Mars.

Since where $\tau = \tau_0$ and $\phi = \phi_0$

$$\begin{aligned} \frac{\partial \varphi}{\partial \eta} &= \frac{1}{R}, & \frac{\partial \varphi}{\partial \xi} &= 0, \\ \frac{\partial \tau}{\partial \eta} &= 0, & \frac{\partial \tau}{\partial \xi} &= \frac{1}{R \cos \varphi}, \end{aligned}$$

then

$$\begin{aligned} \frac{\partial h_i}{\partial \xi} &= \frac{1}{R} \frac{\partial h_i}{\partial \tau} = \frac{a_i}{R}; \\ \frac{\partial h_i}{\partial \eta} &= \frac{1}{R \cos \varphi} \frac{\partial h_i}{\partial \varphi} = \frac{b_i}{R} \quad (i = 1, 2, \ldots, n), \end{aligned} \tag{1.13}$$

where we represent $a_i \equiv \frac{1}{\cos \varphi} \frac{\partial h_i}{\partial \tau}$, $b_i \equiv \frac{\partial h_i}{\partial \varphi}$.

It follows from equations (1.10) and (1.11) that /6

$$a(t) = \frac{\cos\delta \sin t}{\cos h}, \tag{1.14}$$

$$b(t) = \frac{\cos\varphi \sin\delta - \sin\varphi \cos\delta \cos t}{\cos h}. \tag{1.15}$$

Equations (1.5), (1.14) and (1.15) yield the identity

$$a^2(t) + b^2(t) = 1. \tag{1.16}$$

Let us produce two more formulas, which will be used later.

If we substitute the value of cos (t) from equation (1.5) into equation (1.15), we produce

$$b(h) = \frac{\sin\delta - \sin\varphi \sin h}{\cos\varphi \cos h}. \tag{1.17}$$

Let us differentiate expression (1.17) with respect to h:

$$\frac{db}{dh} = \frac{\sin\delta \sin h - \sin\varphi}{\cos\varphi \cos^3 h}. \tag{1.18}$$

2. Equal Point-for-Point Measurements

Suppose n measurements of the height of the sun h_i at moments T_i ($i = 1, 2, \ldots, n$) are performed with errors which are independent random quantities with normal distribution and dispersion D_h.

As we process the measurements of h_i by the least squares method, the correlation matrix of the estimated values of ξ and η will be equal to

$$K_{\xi\eta} = D_h (A^{2n} A^{n2})^{-1}, \tag{2.1}$$

where A^{2n} is a $2 \times n$ matrix, and interchanging indices designates transposition. According to equation (1.13), matrix A^{2n} is equal to [2]

$$A^{2n} = \frac{1}{R}\left|\begin{bmatrix} a_1 a_2 \ldots a_n \\ b_1 b_2 \ldots b_n \end{bmatrix}\right|. \tag{2.2}$$

Substituting expression (2.2) into equation (2.1), we produce

$$K_{\xi\eta} = \frac{D_h R^2}{\Delta}\left|\begin{bmatrix} \sum_{i=1}^{n} b_i^2 & -\sum_{i=1}^{n} a_i b_i \\ -\sum_{i=1}^{n} a_i b_i & \sum_{i=1}^{n} a_i^2 \end{bmatrix}\right|, \tag{2.3}$$

where

$$\Delta = \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2 - \left(\sum_{i=1}^{n} a_i b_i\right)^2. \tag{2.4}$$

We can write the following expression for the greatest eigenvalue $D_{\rho max}$ of matrix $K_{\xi\eta}$:

$$D_{\rho max} = D_h R^2 \frac{n + \sqrt{n^2 - 4\Delta}}{2\Delta}, \tag{2.5}$$

since identity (1.16) indicates that

$$\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2 = n. \tag{2.6}$$

The value of $D_{\rho max}$ determines the greatest dispersion of the deviation in plane $\xi o \eta$ of the calculated landing point from the true landing point. Consequently, the moments of measurement T must be selected so as to minimize the value of $D_{\rho max}$ with a constant number of measurements n. /7

It follows from equation (2.5) that $\frac{dD_{\rho\,max}}{d\Delta} < 0$; therefore, the minimum of $D_{\rho max}$ corresponds to the maximum of Δ (2.4).

It follows from equation (2.6) that the greatest possible value of Δ is obtained where

$$\sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} b_i^2 = \frac{n}{2}, \tag{2.7}$$

$$\sum_{i=1}^{n} a_i b_i = 0. \tag{2.8}$$

The value of $D_{\rho max}$ takes on its least possible value in this case

$$D_{\rho\,opt} = D_h R^2 \frac{2}{n}. \tag{2.9}$$

Let us check the compatibility of equations (2.7) and (2.8) with equations (1.14) and (1.15) when the condition of visibility of the sun $h \geq 0$ is fulfilled at each moment of measurement T_i.

Since the value of a(t) is an odd function, while b(t) is an even function of t, condition (2.8) is fulfilled with any set of moments of measurement T_i which is symmetrical relative to the moment of culmination of the sun τ.

Let us now show that condition (2.7) is also possible. We note preliminarily that, due to the equivalence of cases (ϕ, δ) and ($-\phi$, $-\delta$), we can assume $\delta \geq 0$. Let us construct our program of measurements as follows.

a) If $\phi < 0$, then k/2 measurements h_i will be performed at the moment of sunrise, k/2 -- at the moment of sunset and n - k measurements h_i at the moment of culmination of the sun.

It follows from equations (1.17) that where h = 0, the value of b(h) is

$$b = \frac{\sin\delta}{\cos\varphi},$$

but it follows from equations (1.5) and (1.7) that with the solar hour angle t = 0, b = 1. Then for $\sum_{i=1}^{n} b_i^2$ we can write

$$\sum_{i=1}^{n} b_i^2 = k\left(\frac{\sin\delta}{\cos\varphi}\right)^2 + n - k.$$

Since $\sum_{i=1}^{n} b_i^2$ should be equal to n/2 (2.7), we produce the following expression for the total number of measurements at sunrise and sunset

$$k_{\text{opt}} = \frac{n}{2}\,\frac{\cos^2\varphi}{\cos^2\varphi - \sin^2\delta}. \tag{2.10}$$

By definition, $k \leq n$. Consequently, construction of the optimal program is possible if $\cos^2\phi \geq 2\sin^2\delta$. In this case, the value of $D_{\rho max}$ will be equal to $D_{\rho opt}$ (2.9).

If $\cos^2\phi < 2\sin^2\delta$, the optimal value of k = n, i.e., half of the measurements should be performed at sunrise, half at sunset. In this case, $\sum_{i=1}^{n} b_i^2$ is equal to /8

$$\sum_{i=1}^{n} b_i^2 = n\,\frac{\sin^2\delta}{\cos^2\varphi},$$

and the value of $D_{\rho max}$

$$D_{\rho\,\max} = \frac{D_h R^2}{n}\,\frac{\cos^2\varphi}{\sin^2\varphi_{\max} - \sin^2\varphi}, \tag{2.11}$$

where $\phi_{max} = \delta - \pi/2$ is the latitude of the onset of the polar night.

b) Let us study the case where $\phi \geq 0$, but ϕ and δ are not simultaneously equal to zero.

The program of measurements is constructed as follows. k/2 measurements are performed with the solar hour angle $t_1 = -6^h$, k/2 -- with the hour angle $t_3 = 6^h$; the remaining n - k measurements are performed as before at the moment of culmination of the sun.

From equations (1.15) where $t = \pm 6^h$ ($\pm 90°$ in degrees) we produce

$$b = \frac{\cos\varphi \sin\delta}{\sqrt{1 - \sin^2\varphi \sin^2\delta}},$$

while equations (1.15) and (1.5) show that where $t = 0$, $b^2 = 1$. Consequently, $\sum_{i=1}^{n} b_i^2$ is equal to

$$\sum_{i=1}^{n} b_i^2 = k \frac{\cos^2\varphi \sin^2\delta}{1 - \sin^2\varphi \sin^2\delta} + n - k. \tag{2.12}$$

Setting $\sum_{i=1}^{n} b_i^2$ equal to the value of n/2, we see that to optimize the program of measurements, the value of k should be made equal to

$$k = \frac{n}{2} \frac{1 - \sin^2\varphi \sin^2\delta}{\cos^2\delta}. \tag{2.13}$$

We can show that where $\delta \leq \pi/4$, the value of $k \leq n$ with any value of ϕ. Since the declination of the sun to the equator of Mars is not over 25°, where $\phi \geq 0$ the construction of an optimal program is always possible. The value of $D_{\rho max}$ in this case is also equal to $D_{\rho opt}$ (2.9).

c) In the case where $\phi = \delta = 0$, the value of b(t) is almost always equal to zero. Only where t = 0 is b = 1. In this case, we can only estimate the declination from the longitude λ_a.

3. Systematic Errors

Let us now study the case when, in addition to the unknown errors with dispersion D_h, there are systematic errors with dispersion $p^2 D_h$ and correlation factor equal to one. In this case, the correlation matrix of errors in measurements n of quantity h_i can be written as follows:

$$K_{hh} = D_h \begin{bmatrix} 1+p^2 & p^2 & \dots & p^2 \\ p^2 & 1+p^2 & \dots & p^2 \\ \dots & \dots & \dots & \dots \\ p^2 & p^2 & \dots & 1+p^2 \end{bmatrix}. \tag{3.1}$$

In processing the results of measurements h_i by the method of least squares with a weight matrix equal to /9

$$K = \frac{1}{D_h} \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}, \tag{3.2}$$

we will estimate the systematic error ζ along with the values of ξ and η. Then, the correlation matrix of the quantities ξ, η, ζ will be equal to

$$K_{\xi\eta\zeta} = \left(\begin{bmatrix} A^{2n} \\ I^{1n} \end{bmatrix} [A^{n2}\, I^{n1}] \right)^{-1} \begin{bmatrix} A^{2n} \\ I^{1n} \end{bmatrix} K_{hh} [A^{n2}\, I^{n1}] \times \left(\begin{bmatrix} A^{2n} \\ I^{1n} \end{bmatrix} [A^{n2}\, I^{n1}] \right)^{-1}. \tag{3.3}$$

Partitioned matrix $\begin{bmatrix} A^{2n} \\ I^{1n} \end{bmatrix}$ consists of matrix A^{2n}, having the same sense as in the preceding section [equation (2.2)], plus the $1 \times n$ row vector I^{1n}, the elements of which are derivatives of the systematic error ζ with respect to h.

Since $\frac{\partial h}{\partial \zeta} = 1$,

$$I^{1n} = [1\ 1 \dots 1],$$

$$\begin{bmatrix} A^{2n} \\ I^{1n} \end{bmatrix} = \frac{1}{R} \begin{bmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \\ R & R & & R \end{bmatrix}. \tag{3.4}$$

If the moments of measurement are located symmetrically relative to the moment of culmination of the sun τ, then as in the preceding chapter $\sum_{i=1}^{n} a_i = 0$, $\sum_{i=1}^{n} a_i b_i = 0$. Consequently:

$$\begin{bmatrix} A^{2n} \\ I^{1n} \end{bmatrix} [A^{n2}\, I^{n1}] = \frac{1}{R} \begin{bmatrix} \sum_{i=1}^{n} a_i^2 & 0 & 0 \\ 0 & \sum_{i=1}^{n} b_i^2 & R \sum_{i=1}^{n} b_i \\ 0 & R \sum_{i=1}^{n} b_i & R^2 n \end{bmatrix}, \tag{3.5}$$

$$\begin{bmatrix} A^{2n} \\ I^{1n} \end{bmatrix} K_{hh} [A^{n2}\, I^{n1}] = \\ = \frac{D_h}{R^2} \begin{bmatrix} \sum_{i=4}^{n} a_i^2 & 0 & 0 \\ 0 & \sum_{i=1}^{n} b_i^2 + p^2 \left(\sum_{i=1}^{n} b_i \right)^2 & R \sum_{i=1}^{n} b_i (1 + p^2 n) \\ 0 & R \sum_{i=1}^{n} b_i (1 + p^2 n) & R^2 n (1 + p^2 n) \end{bmatrix}. \tag{3.6}$$

Let us now invert matrix (3.5) /10

$$\left(\begin{bmatrix} A^{2n} \\ I^{1n} \end{bmatrix} [A^{n2}\, I^{n1}] \right)^{-1} = R^2 \begin{bmatrix} \frac{1}{\sum_{i=1}^{n} a_i^2} & 0 & 0 \\ 0 & \frac{n}{\Delta} & \frac{-\sum_{i=1}^{n} b_i}{R\Delta} \\ 0 & \frac{-\sum_{i=1}^{n} b_i}{R\Delta} & \frac{\sum_{i=1}^{n} b_i^2}{R^2 \Delta} \end{bmatrix}, \tag{3.7}$$

where

$$\Delta = n\sum_{i=1}^{n} b_i^2 - \left(\sum_{i=1}^{n} b_i\right)^2. \tag{3.8}$$

Substituting expressions (3.6) and (3.7) into (3.3) and multiplying the matrices, we produce

$$K_{\xi\eta\zeta} = D_h R^2 \begin{bmatrix} \dfrac{1}{\sum\limits_{i=1}^{n} a_i^2} & 0 & 0 \\ 0 & \dfrac{n}{\Delta} & \dfrac{-\sum\limits_{i=1}^{n} b_i}{R\Delta} \\ 0 & \dfrac{-\sum\limits_{i=1}^{n} b_i}{R\Delta} & \dfrac{p^2\Delta + \sum\limits_{i=1}^{n} b_i^2}{R^2\Delta} \end{bmatrix}, \tag{3.9}$$

where, as before, Δ is defined by equation (3.8).

Thus, we produce the following expressions for the dispersions of errors in quantities ξ and η:

$$D_\xi = \frac{D_h R^2}{\sum\limits_{i=1}^{n} a_i^2}, \tag{3.10}$$

$$D_\eta = \frac{D_h R^2}{\sum\limits_{i=1}^{n} b_i^2 - \dfrac{\left(\sum\limits_{i=1}^{n} b_i\right)^2}{n}}. \tag{3.11}$$

As follows from formulas (3.10) and (3.11), the dispersion of the systematic error $p^2 D_h$ does not influence dispersions D_ξ and D_η. Of course, with large values of p^2, a methodological error arises, resulting from the nonlinearity of the dependence of h on the values of ξ and η. It follows from equation (3.9)

that the dispersion of the estimate of the systematic error, equal to

$$D_\zeta = D_h p^2 \left(1 + \frac{\sum_{i=1}^{n} b_i^2}{p^2 \left(n \sum_{i=1}^{n} b_i^2 - \left(\sum_{i=1}^{n} b_i \right)^2 \right)} \right), \tag{3.12}$$

exceeds the dispersion of the systematic error, $D_h p^2$, approaching it as the number of measurements n increases.

Formulas (3.10) and (3.11) for dispersions D_ξ and D_η were produced on the assumption that the measurements were processed by the method of least squares with weight matrix (3.2). If the dispersion of the systematic error were known, measurements h_i could be processed by the method of maximum likelihood with a weight matrix the inverse of K_{hh} (3.1). In this case, the correlation matrix of estimates of ξ and η, $K_{\xi\eta max}$, would be defined by the equation /11

$$K_{\xi\eta\,\max}^{-1} = A^{2n} K_{hh}^{-1} A^{n2}. \tag{3.13}$$

Since matrix K_{hh}^{-1} is equal to

$$K_{hh}^{-1} = \frac{1}{D_h} \begin{bmatrix} 1 - \frac{p^2}{1+np^2} & -\frac{p^2}{1+np^2} & \cdots & -\frac{p^2}{1+np^2} \\ -\frac{p^2}{1+np^2} & 1 - \frac{p^2}{1+np^2} & \cdots & -\frac{p^2}{1+np^2} \\ \cdots & \cdots & \cdots & \cdots \\ -\frac{p^2}{1+np^2} & -\frac{p^2}{1+np^2} & \cdots & 1 - \frac{p^2}{1+np^2} \end{bmatrix}, \tag{3.14}$$

by substituting (3.14) into (3.13) and considering that $\sum_{i=1}^{n} a_i = 0$, $\sum_{i=1}^{n} a_i b_i = 0$, we produce

$$K_{\xi\eta\,\max}^{-1} = \frac{1}{D_h R^2} \begin{bmatrix} \sum_{i=1}^{n} a_i^2 & 0 \\ 0 & \sum_{i=1}^{n} b_i^2 - \frac{p^2}{1+np^2} \left(\sum_{i=1}^{n} b_i \right)^2 \end{bmatrix}. \tag{3.15}$$

Consequently, we produce the following formulas for dispersions $D_{\xi max}$ and $D_{\eta max}$:

$$D_{\xi\,\max} = \frac{D_h R^2}{\sum_{i=1}^{n} a_i^2}, \tag{3.16}$$

$$D_{\eta\,\max} = \frac{D_h R^2}{\sum_{i=1}^{n} b_i^2 - \frac{p^2}{1+np^2}\left(\sum_{i=1}^{n} b_i\right)^2}. \tag{3.17}$$

Comparing formulas (3.16) and (3.17) with (3.10) and (3.11), we see that the dispersions of estimates of the value of ξ produced coincide, and dispersion $D_{\eta max}$ is always less than D_η, where

$$\lim_{np^2\to\infty} D_{\eta\,\max} = D_\eta.$$

Consequently, with a large number of measurements or a large relative value of dispersion of systematic error p^2, both methods yield identical results.

Before going over to construction of the optimal program of measurements, let us introduce the relative values of dispersion $\bar{D}_\xi$ and $\bar{D}_\eta$, equal to the ratio of D_ξ and D_η to the dispersion of estimates of ξ and η produced for the case of equally accurate measurements with the optimal program D_{opt} (2.9):

$$\bar{D}_\xi = \frac{D_\xi}{D_{opt}} = \frac{1}{2\bar{a}^2}, \tag{3.18}$$

$$\bar{D}_\eta = \frac{D_\eta}{D_{opt}} = \frac{1}{2(\bar{b}^2 - (\bar{b})^2)}, \tag{3.19}$$

where $\bar{a}^2 = \frac{\sum_{i=1}^{n} a_i^2}{n}$, $\bar{b}^2 = \frac{\sum_{i=1}^{n} b_i^2}{n}$, $\bar{b} = \frac{\sum_{i=1}^{n} b_i}{n}$, while (1.16) indicates that /12

$$\bar{a}^2 + \bar{b}^2 = 1. \tag{3.20}$$

The optimal program of measurements h_i should minimize the greater of the quantities $\overline{D}_\xi$ and $\overline{D}_\eta$ with constant number of measurements n:

$$\overline{D}_{\text{opt}} = \min_{T_1, T_2, \ldots, T_n} \max\{\overline{D}_\xi, \overline{D}_\eta\}. \tag{3.21}$$

The optimal program of measurements will be sought among the programs consisting of the minimum number of measurement sessions for which it is still possible to estimate the values of ξ and η. Since the measurement sessions are placed symmetrically relative to the moment of culmination, the minimum number of sessions is equal to four. Suppose in the first session where $t = t_1$ and in the fourth session where $t = -t_1$ we perform k/a measurements in each session, while in the second session where $t = t_2$ and the third where $t = -t_2$, (n - k)/2 measurements h_i are performed in each session.

Since b(t) is an even function of t,

$$\bar{b} = \frac{\sum_{i=1}^{n} b_i}{n} = (1 - \bar{k}) b_2 + \bar{k} b_1, \tag{3.22}$$

$$\bar{b}^2 = \frac{\sum_{i=1}^{n} b_i^2}{n} = (1 - \bar{k}) b_2^2 + \bar{k} b_1^2, \tag{3.23}$$

where $\bar{k} = k/n$.

Substituting expressions (3.22) and (3.23) into (3.18) and (3.19), we produce

$$\overline{D}_\xi = \frac{1}{2(1 - (1 - \bar{k}) b_2^2 - \bar{k} b_1^2)}, \tag{3.24}$$

$$\overline{D}_\eta = \frac{1}{2\{[(1 - \bar{k}) b_2^2 + \bar{k} b_1^2] - [(1 - \bar{k}) b_2 + \bar{k} b_1]^2\}}. \tag{3.25}$$

The value of $\bar{D}_\xi$ changes monotonically as a function of $\bar{k}$ and take on its maximum and minimum values at the boundaries of the inteval $0 \leq \bar{k} \leq 1$. The value of $\bar{D}_\eta$ increases without limit as $\bar{k}$ approaches the boundaries of the interval, and takes on its minimum value where $k = 1/2$, the minimum value being equal to

$$\bar{D}_{\eta\,\min} = \frac{2}{(b_1 - b_2)^2}. \tag{3.26}$$

It follows from (3.26) that the value of $D_{\eta\min}$ decreases with an increase in modulus $|b_1 - b_2|$. Following equations (3.24), (3.25) and (3.26), we can suggest the following method for constructing the optimal program of measurements.

We select moments t_1 and t_2 so as to maximize the value of $|b_1 - b_2|$. We then compare the value of $\bar{D}_{\eta\min}$ produced with the value of $\bar{D}_\xi$ where $\bar{k} = 1/2$.

If $\bar{D}_{\eta\min} \geq \bar{D}_\xi(1/2)$, then

$$\bar{D}_{\text{opt}} = \bar{D}_{\eta\,\min} \tag{3.27}$$

and the moments t_1 and t_2 define the optimal program of measurements, and the optimal value $\bar{k}_{\text{opt}} = 1/2$.

If $\bar{D}_{\eta\,\min}, \bar{D} < (1/2)$, the optimal value of $\bar{D}_{\text{opt}}$ is determined by the equation

$$\bar{D}_{\text{opt}} = \min\{D_\eta(\bar{k}_1),\ D_\eta(\bar{k}_2)\}, \tag{3.28}$$

where k_1 and k_2 are determined by the equation

$$D_\xi(\bar{k}) = D_\eta(\bar{k}), \tag{3.29}$$

and the moments t_1 and t_2 must be selected so as to minimize the value of $\bar{D}_{\text{opt}}$, while the optimal value $\bar{k}_{\text{opt}}$ will be equal to $\bar{k}_1$ or $\bar{k}_2$, whichever corresponds to the smaller value of $\bar{D}_\eta$ from (3.28).

Let us now study various cases of the relationship of the latitude of the landing point ϕ and the declination of the sun to the equator δ.

a) Suppose $\phi > \delta > 0$. Since the derivative $ab/ah < 0$ where $h > 0$ (1.18), the value of $b(t)$ decreases monotonically with increasing h and reaches its minimum at the moment of culmination $b(0) = -1$.

Based on equation (3.26), we perform our first session at the moment of sunrise, our second at the moment of culmination.

Since where $h = 0$, $b = \sin\delta/\cos\phi$, from (3.26)

$$\overline{D}_{\eta\,\min} = \frac{2}{\left(1+\frac{\sin\delta}{\cos\varphi}\right)^2},$$

and from (3.24)

$$\overline{D}_{\xi}\left(\frac{1}{2}\right) = \frac{1}{1-\frac{\sin^2\delta}{\cos^2\varphi}}.$$

It follows from these equations that $\overline{D}_{\eta\min} \geq \overline{D}_{\xi}(1/2)$ where $\cos\phi \geq 3\sin\delta$. Based on (3.27), we produce

$$\overline{D}_{\text{opt}} = \frac{2\cos^2\varphi}{(\cos\varphi+\sin\delta)^2}, \quad \overline{k}_{\text{opt}} = \frac{1}{2}. \tag{3.30}$$

If $2\sin\delta \leq \cos\phi < 3\sin\delta$, we can find the values of $\overline{k}_{\text{opt}}$ and $\overline{D}_{\text{opt}}$ from equations (3.28) and (3.29) (the second session, as before, is performed at the moment of culmination):

$$\overline{k}_{\text{opt}} = \frac{2b_1}{1+b_1}, \tag{3.31}$$

$$\overline{D}_{\text{opt}} = \frac{1}{4b_1(1-b_1)}. \tag{3.32}$$

It follows from (3.32) that where $b_1 = 1/2$, the value of $\overline{D}_{\text{opt}}$ reaches its minimum. Consequently, the first session should be

held at sunrise if the value of b(h) where h = 0 is less than 1/2, and at the moment defined by the equation $b(t_1) = 1/2$ if b(h) is greater than 1/2 where h = 0. Since where h = 0, b = sin δ/cos φ, where 2 sin δ ≤ cos φ < 3 sin δ:

$$\overline{D}_{opt} = \frac{\cos^2 \varphi}{4 \sin \delta (\cos \varphi - \sin \delta)},$$
$$\overline{k}_{opt} = \frac{2 \sin \delta}{\sin \delta + \cos \varphi}, \tag{3.33}$$

while where cos φ < 2 sin δ:

$$\overline{D}_{opt} = 1, \ \overline{k}_{opt} = {}^2/_3. \tag{3.34}$$

b) Let us now study the case where δ > φ > 0. At the moment /14 of culmination t = 0, b(t) = 1. Where sin h = sin φ/sin δ, the value of b(h) takes on its minimum value, equal to

$$b_{min} = \frac{\sqrt{\sin^2 \delta - \sin^2 \varphi}}{\cos \varphi}.$$

Following (3.26), we hold our first session at the moment when $b = b_{min}$, the second session at the moment of culmination. We then produce the following expression for $\overline{D}_{\eta min}$:

$$\overline{D}_{\eta\, min} = \frac{2}{\left(1 - \frac{\sqrt{\sin^2 \delta - \sin^2 \varphi}}{\cos \varphi}\right)^2}.$$

Substituting the values of $b_1 = b_{min}$, $b_2 = 1$ and $\overline{k} = 1/2$ into (3.24), we produce an expression for $\overline{D}_\xi(1/2)$:

$$\overline{D}_\xi ({}^1/_2) = \frac{\cos^2 \varphi}{\cos^2 \delta}.$$

We can show that the value of $\overline{D}_\xi(1/2)$ is less than $\overline{D}_{\eta min}$ where δ ≤ π/4. Since the declination of the sun to the equator of Mars is always less than π/4, according to (3.27)

$$\overline{D}_{\text{opt}} = \frac{2}{\left(1 - \frac{\sqrt{\sin^2\delta - \sin^2\varphi}}{\cos\varphi}\right)^2}, \quad \overline{k}_{\text{opt}} = \frac{1}{2}. \tag{3.35}$$

c) Let us study the case where $\phi < 0$. At the moment of culmination, the value of b(h) reaches its maximum, equal to 1, reaching its minimum value at sunrise, equal as before to

$$b = \frac{\sin\delta}{\cos\varphi}.$$

Consequently, from (3.26) and (3.24) we produce

$$\overline{D}_{\eta\,\min} = \frac{2\cos^2\varphi}{(\cos\varphi - \sin\delta)^2},$$
$$\overline{D}_{\xi}\left(\frac{1}{2}\right) = \frac{\cos^2\varphi}{\cos^2\varphi - \sin^2\delta}.$$

Since $\overline{D}_{\xi}(1/2) < \overline{D}_{\eta\min}$,

$$\overline{D}_{\text{opt}} = \frac{2\cos^2\varphi}{(\cos\varphi - \sin\delta)^2}, \quad \overline{k}_{\text{opt}} = \frac{1}{2}. \tag{3.36}$$

d) Let us study the last case, $\phi = \delta$. At the moment of culmination $b(t) = 0$. Since the derivative $db/dh < 0$, the maximum value of b(h), reached at moment h - 0, is $b(h) = \tan\delta$. Consequently,

$$\overline{D}_{\eta\,\min} = 2/\tan^2\delta$$
$$\overline{D}_{\xi}\left(\frac{1}{2}\right) = \frac{1}{2 - \tan^2\delta}$$

Since where $\tan\delta < 2/\sqrt{3}$, $\overline{D}_{\eta\min} > \overline{D}_{\xi}(1/2)$,

$$\overline{D}_{opt} = \frac{2}{\tan^2 \delta}, \quad \overline{k}_{opt} = 1/2 \tag{3.37}$$

Discussion of Results /15

With equally accurate measurements and with the minimum number of measurements, the dispersion of estimates of ϕ, τ and ψ is approximately half the dispersion of errors in measuring h.

The decrease in the length of a day caused by placement of the landing point closer to the boundary of the area of polar night begins to influence the dispersion of estimates of ϕ, τ and ψ when the value of $\cos \phi$ becomes less than $\sqrt{2} \sin \delta$ ($\delta > 0$, $\phi < 0$).

When systematic errors appear, the dispersion of estimates of the quantities ϕ, τ and ψ remains the same if the landing point is located in an area defined by the inequality $\cos \phi \leq 2 \sin \delta$ ($\delta > 0$, $\phi > 0$). This area includes the area of polar day $\cos \phi \leq \sin \delta$.

If we do not analyze the case of landing in the area of polar night or in an area adjacent to it, the worst area is that limited by the parallels $\phi = \delta$ and $\phi = -\delta$. The dispersion of estimates of ϕ, τ and ψ in this case is approximately four times greater than the dispersion of errors of measurement of h.

REFERENCES

1. Chebotarev, G. A., Analiticheskiye i chislennyye metody nebesnoy mekhaniki [Analytic and Numerical Methods of Celestial Mechanics], Nauka Press, 1965.
2. Linnik, Yu. V., Metod naimen'shikh kvadratov i osnovy teorii obrabotki nablyudeniy [The Method of Least Squares and Principles of the Theory of Processing of Observations], Fizmatgiz Press, 1958.

N75 10990

LOW-NOISE PARAMETRIC AMPLIFIERS FOR RADIO ASTRONOMY OBSERVATIONS AT 18-21 cm WAVELENGTH

B. Z. Kanevskiy, V. M. Veselov, I. A. Strukov and V. S. Etkin

One method of increasing the sensitivity of radio astronomy apparatus is the use of low-noise parametric amplifiers (PA) for high-frequency preamplification. Quantum paramagnetic amplifiers provide a lower noise temperature for radiometer input devices. However, the difficulty involved in using these devices frequently makes the use of SHF PA as preamplifiers more desirable.

An input high frequency section including a modulator, two-stage parametric amplifier, mixer and intermediate frequency preamplifier (IFPA) was developed to be used in the performance of observations and spectral analysis of radiation at wavelength λ = 18 cm, in order to increase the sensitivity of the radiometer.

The low frequency portion of the radiometer was developed by a team led by M. I. Pashchenko (State Astronomical Institute imeni P. K. Shternberg). We present below the principal characteristics of the elements included in the SHF section of the radiometer.

Two-Stage Parametric Amplifier

The expected noise temperature of the entire input section, including antenna noise and feeder noise from the coaxial-waveguide adapter (CWA) of the radiator to the input of the amplifier T_{in} = 50°K. It was therefore decided to use the PA in a mode with forced negative bias. With the PA operating in this mode, it is possible to achieve the minimum possible noise temperature with an uncooled PA. /16

Balanced parametric amplifiers (BPA) were used as the SHF signal amplifiers in this section. The difference between these amplifiers and ordinary types of parametric amplifiers is the fact that BPA use two identical varactor diodes in the amplifier circuit.

A BPA is a two-loop parametric amplifier. The no-load frequency loop consists of two diodes. These diodes are connected into the circuit so that the capacitance of the p-n junction in the diodes is modulated in opposition (Figure 1). As a result, the voltage at the no-load frequency $U(\omega_2)$ relative to

the common terminals of the diodes (aa) is equal to zero. Thus, oscillations at frequency ω_2 can occur only in the loop formed by the diodes. Connection of any external impedance has no influence on the parameters of the no-load loop. In other words, the BPA, as a result of the use of the two diodes, automatically decouples the signal circuit from oscillations at frequency ω_2. This allows this amplifier to achieve significantly wider pass bands than those of single-diode parametric amplifiers. The internal noise level of a BPA is the same as a single-diode PA. The important advantage of the BPA is that production of the amplifier circuit does not require the use of a special filter to achieve decoupling between the signal and no-load frequency loops. This greatly simplifies the design of the amplifier and eliminates additional losses in the filters at the no-load frequency. The high-frequency section of the radiometer uses two BPA stages, based on strip lines. The equivalent circuit of one amplifier is shown in Figure 2.

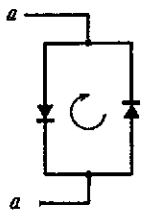


Figure 1. Connection of Diodes into Circuit

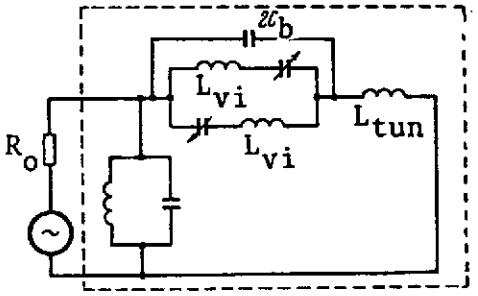


Figure 2. Equivalent Circuit of Amplifier: R_o, oscillator resistance; L_{vi}, inductance of input of one diode of varactor pair; $2C_b$, housing capacitance of balanced pair; L_{tun}, tuned inductance of signal loop

The BPA used special veractor diodes, manufactured for use in printed circuits. Two such diodes were placed on the upper and lower BPA plates and formed a balanced pair of diodes. The signal loop was tuned to fixed frequency ω_1 by means of a stub connected in series with the diodes. This stub -- a section of shorted line $l \leq \lambda/8$ in length -- acted as inductance L_{tun} of the signal loop. The pumping power was fed to the diodes through a waveguide fastened to the top of the amplifier body. A three-core transformer was used to tune the circuit.

The signal circuit is decoupled from the pumping power using a filter consisting of an open line sector $l = \lambda_p/4$ in length,

connected to the points of attachment of the diodes. Both negative-type amplifiers are connected to a wye circulator, also consisting of flat lines and mounted on the same body. The amplifier is thus made in quasi-integrated form. The basic parameters of the balanced parametric amplifier are calculated using the following formulas: /17

Relative pass band

$$2\xi|_{pb} = \frac{2K_0}{Q_M (K_0^2-1)(K_0^2-2)^{1/2}}\left[1+\frac{1}{q^2}+\left(\frac{K_0-1}{K_0+1}\right)^2 q^2\right]^{-1/2},$$

where

$$Q_M = \frac{1}{m}\sqrt{\frac{\omega_1}{\omega_2}}; \quad q = \frac{\omega_K}{\omega_2}\sqrt{\frac{\omega_1}{\omega_2}},$$

m is half the percentage modulation of the p-n junction;

the relative noise temperature

$$t_a = \left(\frac{K_0^2-1}{K_0^2}\right)\frac{\omega_1}{\omega_2}\left(\frac{\omega_2^2}{\omega_K^2}+1\right)$$

(K_0 is the gain of the amplifier at the central frequency; ω_k is the critical frequency of the diode). The source of pumping power for the two stages of the BPA is a single 3-cm waveband klystron. The power of this klystron is sufficient to achieve a gain for the two stages $K_{01}^2 \cdot K_{02}^2 \geq 25$ db. In order to assure stable operation of the amplifiers, a high-stability VS-29 klystron power supply is used and the amplifiers are thermostated at +40°C. Also, to increase decoupling from the input of the amplifier, the section uses a valve based on a Y-circulator (valve parameters: $P_d = 0.3$ db; $P_r \geq 20$ db). The diodes used in the BPA have the following parameters: $C_{-8v} = 0.25$ pf; $\tau_{-8v} = 0.7$ nsec; $C_k = 0.08$ pf; $L_{iv} = 1$ nh.

The basic parameters of the two-stage BPA are:

Input frequency f_c	1667 mHz
Pumping oscillator frequency f_p	9100 mHz
Gain of two stages	25 db
Noise temperature of amplifier T_a	60°K
Losses in circulator branch	0.3 db

Pumping power consumption for both stages	40 mw
Sensitivity of gain to pumping power drift $\Delta K_{01}^2/\Delta P$	10

The input element of the SHF section is a modulator based on coaxial lines. The active elements in the modulator are type 1A5 switching diodes.

TABLE 1

Section Elements	Noise Contribution, T °K
Modulator	15
Two-stage BPA + Input Valve ..	120
Later Stages	6
Total Noise Temperature of SHF Section of Radiometer	141

The basic parameters of the modulator are: direct losses 0.2 db; inverse losses > 20 db; pass band ≈ 60 mHz. After the SHF amplifier, the signal is sent to the mixer and IFPA. The strip mixer uses type R-15 diodes, connected in a balanced circuit. The balanced design of the mixer provides 50 db decoupling of the signal and beat-frequency circuits. The radiometric noise factor of the mixer F = 3 with an IFPA noise factor $F_1 \leq 1.7$. Table 1 presents the noise parameters of the elements of the SHF-receiver section of the radiometer. /18

Experimental Results

The receiver section was laboratory tested in the radiometer. Figure 3 presents a recording of the fluctuation signal. The measurements were performed with a time constant of $\tau = 2$ sec and $\Delta f = 2 \cdot 10^7$ Hz. With a noise temperature of the entire system of 440°K (antenna equivalent $T_n = 300°K$), a fluctuation sensitivity of about 0.2° was achieved.

Studies of the long-term stability of the SHF section have shown that with a gain of the two stages of the PA $K_{01}^2 \cdot K_{02}^2 = 25$ db the drift of the gain over one hour was about 3%.

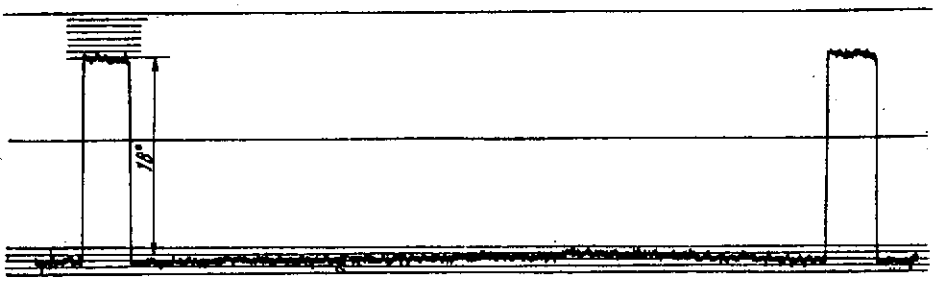


Figure 3. Recording of Fluctuation Signal

An overall view of the SHF receiver section of the radiometer is presented in Figure 4.

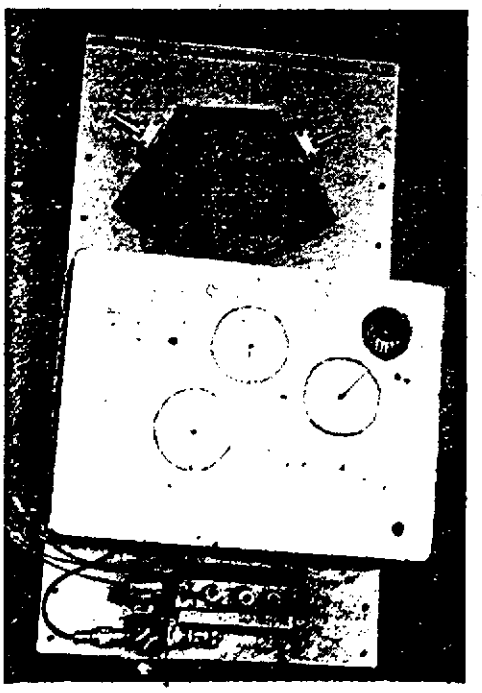


Figure 4. Overall View of SHF Receiver Section of Radiometer (Top View)

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

N75 10991

PROBLEMS OF THE DESIGN OF LOW-NOISE INPUT DEVICES

V. M. Manokhin, Yu. A. Nemlikher,
I. A. Strukov and Ye. A. Sharkov

One pressing problem in the creation of receivers for space communications is the development of low-noise input devices. The most promising low-noise input devices are parametric amplifiers, allowing effective noise temperatures of below 50°K to be achieved by cooling the parametric diode or the entire device to the temperature of liquid nitrogen, 77°K. This work presents an analysis of the requirements placed on the elements of parametric centimeter waveband amplifiers for the achievement of minimal noise temperatures. One possible design of an SHF receiver head is described and its basic experimental parameters are presented, confirming the possibility of satisfying the requirements. /19

Review of Possible Versions of Design of SHF Input Heads

The basic, determining parameter of a receiver device is its noise temperature, related to the input. Due to the relatively high frequencies of signal received, in order to achieve a low noise temperature, when modern diodes are used, a parametric amplifier must have a difference frequency in the range of 24 to 30 GHz.

We will limit our study to the selection of difference frequencies above 14 GHz. Let us discuss first of all the case of selection of a difference frequency of 14 GHz, since this difference frequency was selected in the initial version of our receiver head.

A low noise input amplifier might be a two-stage parametric amplifier. In another case, the second stage might be a tunnel amplifier.

Figure 1 shows the effective noise temperature of the receiver head (the stage following the low-noise amplifier has a noise factor of 13) as a function of losses in the circulator branch. Here and in the following, the signal frequency will be taken in the 4 GHz band.

Curves 1 and 2 correspond to a two-stage parametric amplifier, curve 1 being for the case of connection of one circulator

to the input of the amplifier, curve 2 -- for the case of connection of an additional circulator to the input of the amplifier to increase decoupling in case of a change in antenna impedence. This comment concerns all of the curves analyzed below. Curves 3 and 4 relate to the case of use of a second tunnel amplifier stage with a noise factor F = 6 db.

It is not difficult to see that the connection of a second tunnel amplifier stage results in an increase of the noise temperature of the entire device by approximately 40°K. The use of a two-stage parametric amplifier allows us to produce an effective noise temperature of the receiver device $T_n \approx 140°K$ with losses of 0.2 db in the circulator branch.

An even greater deterioration of the noise parameters of the system results from the use of a tunnel amplifier if the parametric amplifier is cooled (together with the circulator) to 77°K.

The dependence of effective noise temperature on losses in the circulator branch is illustrated by Figure 2. As we can see from Figure 2, the effective noise temperature of the receiver head (dotted line) is approximately 100°K, with losses in the branch of 0.3 db.

What we have said indicates that a tunnel amplifier is not suitable for use in the second stage of a low-noise receiver device.

A cooled parametric amplifier can be constructed according to either one of two circuits, differing in that in the one case only the crystal is cooled, while in the other case the entire device is cooled together with the circulators. The first case corresponds to the dot-dash lines, the second case -- to the dotted lines. These circuits are equivalent if the circulators used in the first case have losses of 0.1 db in the branch, while those used in the second case have losses of 0.2 to 0.3 db; both circuits allow achievement of effective noise temperatures in the receiver device of less than 50°K.

Modern varactor diodes have critical frequencies in the 100 GHz range. Therefore, in order to reduce the noise temperature of the receiver device, a pumping frequency of 70-100 GHz should be used.

Selection of such a high pumping frequency brings up a number of difficulties; therefore, let us determine what noise parameters could be expected from an amplifier with no-load frequency of 30 GHz. We will assume that the varactor diodes have a critical frequency of 60 GHz.

Figure 3 shows the noise temperature of the PA as a function of circulator losses under these assumptions.

Lines a relate to an uncooled parametric amplifier. With a difference frequency of 30 Hz, it is possible to produce a noise temperature of the system $T_n \approx 70°K$, using waveguide circulators with losses of 0.1 db. Cooling of the device (lines b, c) can achieve noise temperature on the order of 35°K.

Thus, transition to a difference frequency of 30GHz is desirable only if a noise temperature of the system of below 30°K is required or if the noise temperature of the antenna is very low, so that a gain of 20°K is appreciable.

Let us now describe the results of development, investigation and assembly of a low-noise SHF receiver head, designed for space communications. Our task did not include creation of a model with the maximum pass band width. /21

Development and Study of a Low-Noise Semiconductor Two-Loop Parametric Amplifier

A low-noise semiconductor parametric amplifier (SPA) using germanium parametric diodes was developed to provide a low noise factor for a receiver operating in the 4 GHz band. High frequency pumping ($F_p \approx 19$ GHz) was used in the SPA model to achieve low noise temperatures [1-3]. Therefore, in designing the model, the distribution of the structure of the diode was considered, since the dimensions of the body of the diode are comparable to the wave-length of the pumping oscillations and of the difference frequency.

A flat waveguide design was used. The signal circuit consisted of a strip line, the pumping circuit -- of a waveguide, the no-load circuit was a section of heterogeneous coaxial line formed by the structure of the diode and the body of the model.

The model is a T-shaped intersection between the strip line and the pumping waveguide. The diode is located at the point of intersection. The pumping waveguide is supercritical for the idler frequency oscillations, in order to decouple the pumping and no-load frequency circuits.

The pumping circuit consists of the pumping oscillator (a klystron), the decoupling diode, a variable attenuator, and an impedence transformer. A resonant wave meter is connected into the pumping circuit through a directional coupler, in order to monitor the pumping frequency.

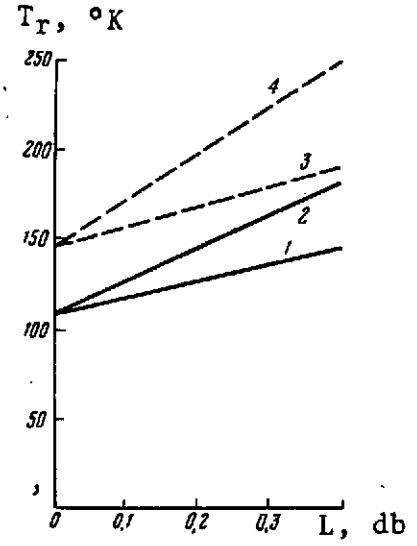


Figure 1. Effective Noise Temperature of Input Head as a Function of Losses in Input Circulator Branch

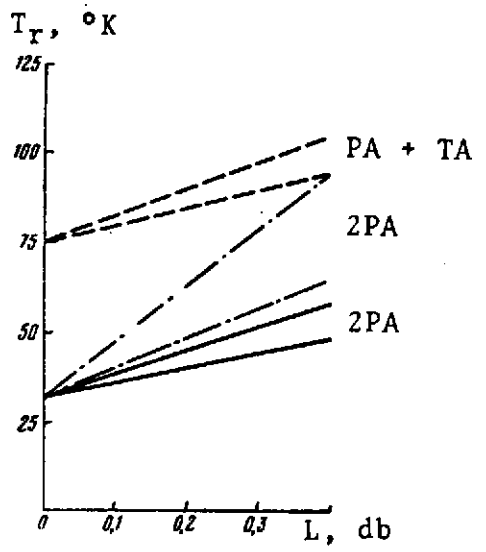


Figure 2. Effective Noise Temperature of SHF Head as a Function of Losses in Circulator for Various Cases of Cooling of Receiver System. PA, Parametric Amplifier; TA, Tunnel Amplifier

The signal circuit consists of a complex-structure strip line, connected between the diode and a 50-ohm flat Y circulator. The circulator consists of a strip line, and has losses of 0.3 db and decoupling of at least 25 db in the frequency band.

The strip line has the following structure: a line with a wave impedence Z_0 = 50 ohms, a quarter-wave transformer to match the diode to the signal generator; a half-wave (signal frequency) line is connected between the quarter-wave transformer and the diode. The last section of the line is placed in a narrow section of the body, /22 necessary to force the circuit of the strip line beyond the limit for the waveguide type of pumping oscillations.

Filters are connected into this section of the line to decouple the signal, no-load and pumping frequency circuits. The filters are resonators consisting of quarter-wave open stubs with quarter-wave couplings [4]. The distance of the filter from the diode is selected so as to provide the no-load condition in the plane of the diode for oscillations at the no-load frequency.

A coaxial stub is connected after the diode. The stub is inductive at the signal frequency, capacitive at the idler frequency. The overall equivalent circuit of the flat system is presented on Figure 4.

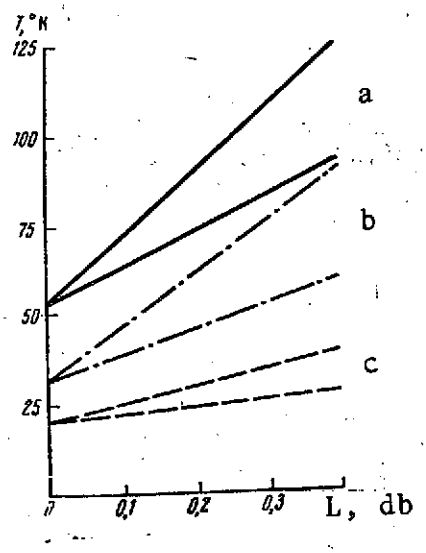


Figure 3. Effective Noise Temperature of SHF Head as a Function of Losses in Circulator Branch for Various Cases of Selection of No-Load and Cutoff Frequencies of Diode. f_i, No-Load Frequency; f_{cr}, Critical Frequency; a, f_i = 30 Hz, f_{cr} = 60 Hz; b, f_i = 60 Hz, f_{cr} = 100 Hz; c, f_i = 30 Hz, f_{cr} = 60 Hz

With the usual flat line teflon body height of 4 mm, the width of the strip becomes comparable to the wavelength and the line is not suitable for use [5]. Therefore, in order to allow flat devices to be used in the 15-20 GHz frequency range, the flat wave body height is 2 mm.

The model constructed used germanium parametric diodes with capacitances C_{-3v} = 0.22-0.25 pf with a negative bias on the order of 1.7-2 v. The parametric diode operates in a zero-current mode.

The tuning of the elements and investigation of the model of the two-loop parametric amplifier were performed using a type R2-18 sweep generator and a type P5-16 measuring receiver (radiometer).

The following amplifier parameters were achieved: power gain K^2 = 15-20 db, amplifier pass band Δf = 35-40 mHz, noise temperature, corrected to input flange of circulator, 150°K.

The sensitivity of the amplitude-frequency response was measured in relationship to pumping frequency drift $S_f = \left(\frac{\Delta K}{K} / \frac{\Delta f}{f}\right)$ and changes in pumping power $S_P = \left(\frac{\Delta K}{K} / \frac{\Delta P}{P}\right)$. With a gain of $K^2 \approx 13$ db, the measured values of sensitivity were $S_f \approx 2 \cdot 10^3$ and $S_p \approx 15$.

The balanced mixer consisted of a 3 db bridge using balanced and unbalanced flat transmission lines. The signal frequency voltage was fed through a 50-ohm terminal plug to an unbalanced flat line, the oscillator was connected to the diodes through a flat balanced line. The combination of balanced and unbalanced flat lines allows the production of bridge circuits with branch separation of at least 50 db. /23

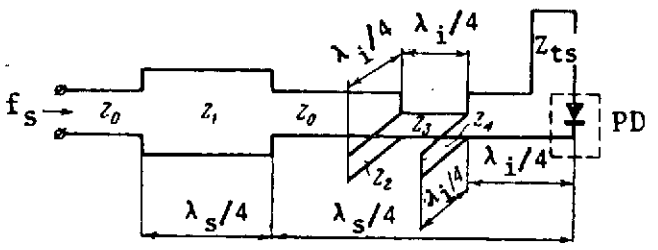


Figure 4. Equivalent Circuit of Strip System of Parametric Amplifier. PD, Parametric Diode; λ_s, λ_i, Wavelengths of Signal and No-Load Frequencies in Coaxial Line; Z_0, Z_1, Z_2, Z_3, Z_4, Wave Impedences; Z_{ts}, Wave Impedence of Tuning Stub; f_s, Frequency of Signal

The mixer uses pellet-type inverted tunnel diodes.

The use of inverted tunnel diodes in the balanced mixer allowed the oscillator power to be reduced by approximately an order of magnitude and eliminated the transformer the signal circuit.

The following parameters were achieved in the 4 GHz band: transmission factor on the order of 0.3; signal-band noise factor F = 5, oscillator power consumption less than 1 mw.

Adjustment of the high frequency section and measurement of noise characteristics of the SHF input head were performed using a sweep generator with an indicator and a type P5-16 measuring receiver.

The noise temperature of the entire receiver system was measured by the constant level method using a calibrated attenuator at the intermediate frequency.

The following characteristics were achieved for the entire SHF receiver head, a block diagram of which is shown in Figure 5: high frequency section power gain $K^2 \approx 28$-35 db, pass band at 1 db level $\Delta f \approx 25$-35 mHz, noise temperature, corrected to input flange of circulator $T_n \approx 200°K$. The experimental study confirmed the possibility of satisfying all requirements.

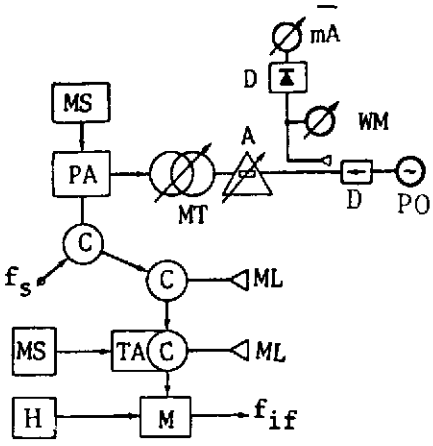


Figure 5. Overall Block Diagram of SHF Receiver Head. PA, Parametric Amplifier; TA, Tunnel Amplifier; C, Circulator; MS, Mixing Section; PO, Pumping Oscillator; H, Heterodyne; M, Mixer; MT, Matching Transformer; A, Attenuator; D, Detector Section; WM, Wave Meter; D, Diode; ML, Matched Load; f_{if}, Intermediate Frequency; f_s, Signal Frequency; mA, Milliammeter

REFERENCES

1. Etkin, V. S., Ye. M. Gershenzon, Parametricheskiye sistemy na poluprovodnikovykh diodakh [Parametric Systems Using Semiconductor diodes], Moscow, Sov. Radio Press, 1964.
2. Vasil'yev, V. N., et al., Regenerativnyye poluprovodnikovyye parametricheskiye usiliteli [Regenerative Semiconductor Parametric Amplifiers], Moscow, Sov. Radio Press, 1965.
3. Blackwell, L. A., K. L. Kotsebu, Parametricheskiye usiliteli na poluprovodnikovykh diodakh [Parametric Amplifiers based on Semiconductor Diodes], Moscow, Mir Press, 1964.
4. Schiffman, B. M., C. S. Matthaei, Trans. IEEE, 1964, MTT-12, No 1.
5. Sushkevich [Editor], Pechatnyye skhemy santimetrovogo diapazona [Centimeter Wave Band Printed Circuits], Foreign Literature Press, 1956.

N75 10992

PHOTOGRAMMETRIC METHOD OF DETERMINATION OF DEFLECTION OF LIGHT BEAMS BY PORTHOLE GLASS UNDER SPACE FLIGHT CONDITIONS

Ya. L. Ziman

Many of the observations and surveys performed in space must be made through the glass portholes of hermetically sealed spacecraft. The portholes through which high-precision navigational, astronomic, topographic and other measurements are performed should not change the direction of light rays passing through them. Even if this requirement is met as the porthole is manufactured, we do not know exactly what happens in space, when the porthole is subjected to various and variable mechanical and temperature effects. /24

The deflections of light rays caused by porthole glass can be found if the angles γ' between the directions to the stars are measured through the porthole in flight and compared with the true values of these angles γ.

The angle between the directions to stars, the equatorial coordinates of which are α_1, δ_1 and α_2, δ_2, is

$$\gamma_{12} = \arccos [\sin \delta_1 \sin \delta_2 + \cos \delta_1 \cos \delta_2 \cos (\alpha_1 - \alpha_2)]. \quad (1)$$

A measuring camera can be suggested as a tool for determination of these angles through a porthole. The camera allows the direction to large numbers of stars to be determined at fixed moments in time, to determine the changes in direction of light beams throughout the entire field of vision of the camera simultaneously. The system of photograph coordinates determined by such a camera is illustrated by Figure 1.

After measuring the coordinates x_1, y_1 and x_2, y_2 of images m and n of the two stars M and N on photograph P, we find γ'_{12} using the following formulas:

$$\begin{aligned} \gamma'_{12} &= \arccos \frac{x'_1 x'_2 + y'_1 y'_2 + f^2}{r_1 r_2}, \\ x' &= x + \Delta x, \\ y' &= y + \Delta y, \\ r_i &= \sqrt{x'^2_i + y'^2_i + f^2}, \end{aligned} \quad (2)$$

where f is the focal length of the camera; Δx, Δy are the corrections to the measured coordinates of the stars, considering the deformation of the photographic material, photogrammetric distortion, refraction of the ray at the boundary between space and the atmosphere in the spacecraft where the camera is installed, etc.

The distortion of the angle $\Delta\gamma$ introduced by the porthole will obviously be

$$\Delta\gamma = \gamma - \gamma'. \tag{3}$$

Consequently, if the measuring camera is installed on a spacecraft and photographs of stars are produced through the porthole first on the surface of the Earth, then during various stages of flight, photogrammetric processing of these photographs can determine how the porthole changes the direction of the rays passing through it under various conditions.

The problem will be most completely solved if we determine not the change in angles $\Delta\gamma$, but rather the changes in rectangular coordinates of the images of the stars on the photograph caused by the porthole. /25

A star photograph produced by a measuring camera can be used to determine its spatial orientation in the equatorial system of coordinates [1, 2]. This orientation is determined by matrix D of direction cosines d_{ij} of the coordinate axes of the photograph

$$D = \begin{pmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{pmatrix} \tag{4}$$

If we know the elements of matrix D, we can use the equatorial coordinates α and δ of the stars to calculate angles λ°, μ° and ν° between the directions to the stars and the coordinate axes x, y and z of the system of the photograph (see Figure 1).

$$\begin{aligned} \bar{l} &= \cos\delta\cos\alpha, \\ \bar{m} &= \cos\delta\sin\alpha, \\ \bar{n} &= \sin\delta, \end{aligned} \qquad \begin{pmatrix} l^\circ \\ m^\circ \\ n^\circ \end{pmatrix} = D \begin{pmatrix} \bar{l} \\ \bar{m} \\ \bar{n} \end{pmatrix} \qquad \begin{aligned} \lambda^\circ &= \arccos l^\circ, \\ \mu^\circ &= \arccos m^\circ, \\ \nu^\circ &= \arccos n^\circ, \end{aligned} \tag{5}$$

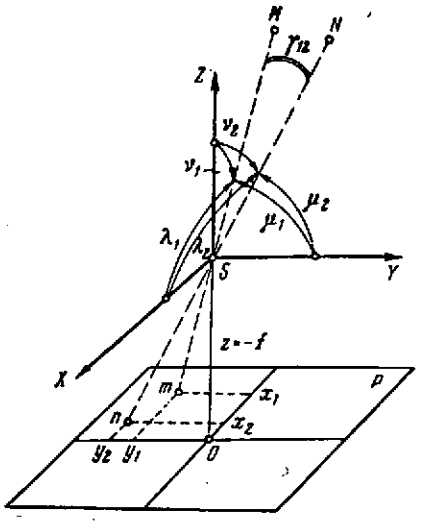


Figure 1. System of Coordinates of Photograph

where $\bar{l}$, $\bar{m}$, $\bar{n}$ and l°, m°, n° are the direction cosines of the stars in the equatorial and photograph system of coordinates respectively.

The values of λ, μ and ν can be calculated from the coordinates x', y', f of the images of the stars on the photograph [2]:

$$\lambda' = \arccos\frac{x'}{r},$$
$$\mu' = \arccos\frac{y'}{r},$$
$$\nu' = \arccos\frac{f}{r}. \qquad (6)$$

If the porthole does not change the direction of the rays passing through it, angles λ°, μ°, ν° and λ', μ', ν' for the same stars will be equal. Otherwise, their difference

$$\Delta\lambda = \lambda^\circ - \lambda',$$
$$\Delta\mu = \mu^\circ - \mu',$$
$$\Delta\nu = \nu^\circ - \nu' \qquad (7)$$

determines the distortions in the directions of projecting rays caused by the porthole. /26

Knowing the direction cosines l°, m°, n°, we can calculate the coordinates x°, y° of star images on the photograph free of all distortions, including those introduced by the porthole:

$$x^\circ = f\frac{l^\circ}{n^\circ},$$
$$y^\circ = f\frac{m^\circ}{n^\circ}. \qquad (8)$$

The differences (x° - x') and (y° - y') in calculated and true coordinates of star images can be used to find the changes

in photogrammetric parameters of the camera caused by the porthole -- changes in the coordinates of its main point and focal length [3]. These values can be looked upon as the three components Δx, Δy, Δz of displacement of the optical center of projection in the system of coordinates of the photograph.

Photography of stars during space flight should be performed with oriented spacecraft. Otherwise, due to the high rates of displacement of stars in the field of vision of the camera, their images will not be produced on the photographs. It is best to perform this task from satellites oriented according to the local vertical and the absolute velocity vector. If a star camera is installed on an oriented satellite so that its optical axis is in a plane passing through these reference directions, the stars in the field of vision of the camera will move at an angular rate of $\omega = 0.01$ deg/sec, approximately equal to the rate of motion of the satellite in its orbit. In this case, repeated exposure of one photograph at predetermined time intervals can produce several images of each star, thus increasing the number of directions, the distortions of which will be known.

The accuracy of determination of deflections of light rays by this method can be approximately characterized by the quantity of M, equal to

$$M = 2 \cdot 206265'' \cdot \frac{\mu}{f}, \tag{9}$$

where μ is the mean square error of the abcissas and ordinates of the images of the stars on the photograph. With various focal lengths of the cameras and $\mu = 0.005$ mm, the value of M will be:

f, mm	100	250	400	750
M, seconds of arc	21	8	5	3

We can see from these data what the focal length of the camera must be in order to find the deflection of light rays by the glass of the porthole with the required accuracy.

In selecting a camera lens, we should consider not only the photogrammetric, but also the photometric characteristics of the star images produced. In particular, it is desirable to obtain the greatest possible lens aperture diameter, since this increases the stellar magnitude of the stars photographed and, consequently, the number of stars imaged.

In order to be sure what types of lenses and cameras can be used, let us present as an example the results of photography of stars from the Earth by a camera with the following parameters: focal length of lens 21 cm, aperture diameter 8.4 cm, film size 13 × 18 cm, resolving capacity 30-10 lines/mm. As the camera was rotated at $\omega = 0.01$ deg/sec, images of stars were produced on type 10 N panchromatic aerial film with exposures of 0.1 and 2.0 sec, showing stars of stellar magnitude 5 and 6.5 respectively. /27
Magnitude 5 stars numbered about 25 on the photograph, magnitude 6.5 -- about 80.

It is expedient to compare the photometric characteristics of images of stars produced through a porthole under space flight conditions as well. This comparison may reveal changes in optical characteristics of the porthole. However, this question was not studied by the author and goes beyond the scope of the present article.

REFERENCES

1. Boguslavskaya, Ye. Ya., Fotograficheskaya astrometriya [Photographic Astrometry], Gostekhizdat Press, 1947.
2. Ziman, Ya. L., Opredeleniye oriyentatsii kosmicheskikh apparatov po snimkam zvezd. [Determination of the Orientation of Spacecraft from Star Photographs], Kosmicheskiye issledovaniya, 1969, No. 2.
3. Ziman, Ya. L., Opredeleniye fotogrammetricheskikh parametrov fotokamer po snimkam zvezd. [Determination of the Photogrammetric Parameters of Cameras from Star Photographs], Astron. Zh., 1969, No. 3.

N75 10993

DEVELOPMENT OF THE DESIGN OF A BETWEEN-THE-LENS BAND-TYPE PHOTOGRAPHIC SHUTTER

A. S. Dorofeyev

Modern spacecraft carry large quantities of navigational and scientific equipment. The necessity of increasing the number of scientific experiments and expanding the range of scientific apparatus on board a spacecraft sets forth increasingly rigid requirements for reduction of weight and size of devices. One means of miniaturization of photographic equipment is the use of small cameras with wide angle, short focus lenses. The photographs produced by these cameras have better measurement properties if a between-the-lens shutter is used. Wide angle lenses frequenly have such small spacings between elements that the use of between-the-lens shutters of iris type is impossible. An example of one such lens is the Russar-44bm, the space between elements of which is not over 0.5 mm.

A band-type central between-the-lens shutter [1] is the only central shutter which can be used in these cases. This shutter consists of two thin bands with slots located in the space between the elements of the lens. During operation the slots move in opposite directions and meet at the optical axis of the lens, forming a light aperture. After the slots pass, the aperture is once more closed by the two layers of the band.

The author has designed such a shutter [2] for a number of wide angle Russar and Rodina lenses, using a carbon steel band about 45 μm thick in all cases, allowing the shutter to be fit into a space between elements of 0.3 mm.

In designing aerial photography cameras with central band shutters, the recommendations of [1] as to optical efficiency of the shutter and shape of the slot in the band were accepted. Since the hexagonal slot shape was found to be difficult to manufacture, a rectangular form was used, as the form with the highest optical efficiency [3]. The slots were made in the carbon steel band by punching. The contours punched out must not have burrs, in order to assure normal operation of the shutter, placing high requirements on the accuracy of contact between the contours of the punch and die; therefore, the rectan- /28
gular shape was technologically easier to achieve. The range of shutter speeds was 1/100-1/300 sec, which could be expanded in the direction of increasing the exposure by attaching additional inertial masses to the bands.

For a number of reasons, including primarily the low relative apertures of superwide-angle Russar lenses, the shutter speeds had to be extended to 1/60-1/50 sec. The formula relating the actual exposure to the shutter parameters is

$$t_\phi = \sqrt{J_c \phi_m / M_{max} \eta} \left[\arcsin\left(\frac{x_1 + e}{r_4 \varphi_m} - 1\right) - \arcsin\left(\frac{x_1}{r_4 \varphi_m} - 1\right)\right],$$

where J_c is the reduced moment of inertia of the mechanism moving the bands during the exposure; the reference axis is the axis of the spring with radius r_4; M_{max} is the greatest force of the shutter spring for the production of exposure t_ϕ; ϕ_m is the total twisting angle of the spring for t_ϕ; η is the mechanical efficiency of the band-movement mechanism; x_1 is the path of the band from the moment of beginning of motion to the moment of beginning of exposure; r_4 is the radius of the spring shaft; e is the length of the slot in the band (here and below, all symbols used are the same as in [1]).

Analysis of this formula indicates that we can increase the exposure time by the following methods:

1) increasing the corrected moment of inertia J_c by attachment of additional inertial masses;

2) decreasing the force of the spring M_{max} while simultaneously decreasing the twisting angle of the spring ϕ_m (incomplete cocking);

3) increasing the length of the slot in the strip e while simultaneously increasing the optical efficiency of the shutter.

All three methods were tried. The first method required additional equipment to absorb the energy of the inertial masses -- shock absorbers, which complicated the design.

Incomplete cocking of the spring required no changes in the design, but did not produce stable results as the temperature was changed and led to failures as time passed due to oxidation and thickening of the lubricant.

The most favorable method was the third method -- increasing the length of the slot in the band. In addition to the increase in the actual exposure, the effective exposure also increased, since the length of the slot is related to the optical efficiency by the formula

$$e = \frac{D(1 - \eta_{opto})}{1 - \eta_{opt}}$$

where D is the diameter of the effective lens aperture; η_{opto} is the optical efficiency of opening of the shutter; η_{opt} is the optical efficiency of the shutter.

For the shutter operating with the Russar-25IM lens (D = 2.8 mm), in place of the initial value of e = 6 mm, we increase the length of the slot to e = 12 mm without significantly changing the kinematics of the mechanism. This increased the value of η_{opt} = 0.8 to η_{opt} = 0.9 while retaining the rectangular slot shape (η_{opto} = 0.576).

However, the rectangular slot form, the best from the standpoint of achievement of the highest optical efficiency of the shutter, is at the same time defective as concerns operational qualities. The right angles form points of concentration of the stresses formed during punching of the slot. Heat treatment of the band in order to relieve the stresses does not achieve the desired effect, since the band loses its elastic properties during tempering and losses to mechanical hysteresis increase. Later hardening once more produces stresses in the same dangerous areas. Practice has shown that the overwhelming /29
majority of band failures occur at the right angles of the slot, by propagation of cracks from this point. A clear indication of this is illustrated in Figure 1, showing the stresses in a band of transparent material (celluloid) with a rectangular slot. The photograph was produced by the method of crossed polarization filters. The light area shows the danger zone. Figure 2 confirms the absence of such zones around slots cut with circular contours.

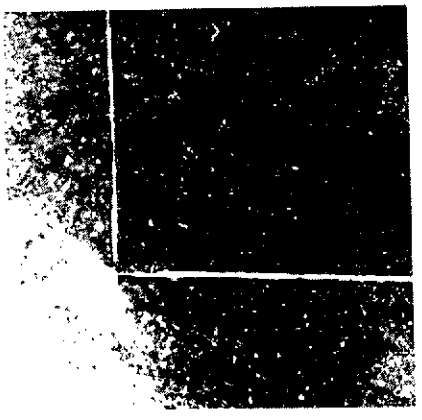


Figure 1. Rectangular Slot in Shutter Band

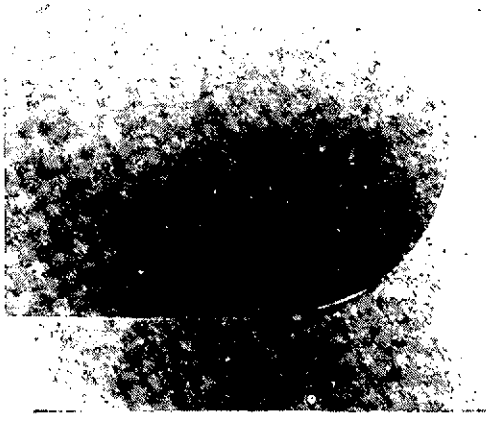


Figure 2. Rounded Slot in Shutter Band

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

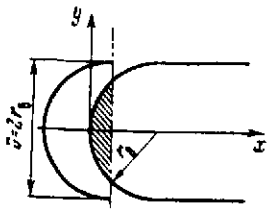


Figure 3. Diagram of Opening of Effective Shutter Aperture

Let us analyze the optical opening efficiency for this shape of slot. In order to deal with less combersome expressions, let us solve the problem of defining η_{opto} for half of the light aperture. Let us assume that the radius of curvature r_B of the slot and the radius of the effective aperture $r = D/2$ of the lens are the same. Then

$$\eta_{opto} = \frac{\int_0^{r_B} S_{(x)} dx}{\pi r_B^2 \int_0^{r_B} dx},$$

where S(x), the shaded portion of Figure 3, is the instantaneous value of the open area of the effective lens aperture.

In polar coordinates, the area of the segment is written as

$$S_{(x)} = 2\int y dx = 2\int \sqrt{2r_B x - x^2}\, dx =$$
$$= 2\left[\frac{(x - r_B)}{2}\sqrt{2r_B x - x^2} + \frac{r_B^2}{2}\arcsin\frac{x - r_B}{r_B} + C\right],$$

where C is the integration constant, the value of which can be found from the condition $S_{(x)} = 0$ where $x = 0$: /30

$$C = -\frac{3\pi r_B^2}{4}.$$

Then

$$\eta_{opto} = \frac{\int_0^{r_B}\left[(x - r_B)\sqrt{2r_B x - x^2} + r_B^2 \arcsin\left(\frac{x - r_B}{r_B}\right) - \frac{3}{2}\pi r_B^2\right] dx}{\pi r_B^2 \int_0^{r_B} dx} = 0{,}424.$$

Let us define how the value of $\eta_{opto} = 0.424$ influences the value of η_{opt}. In order to conserve the initial value of $\eta_{opt} = 0.8$, the necessary slot length

$$e_2 = D(1 - \eta_{opto})/(1 - \eta_{opt}) = 2.8(1 - 0.424)/0.2 = 8 \text{ mm}$$

in place of the earlier value of e = 6 mm.

If we retain $e_1 = 12$ mm, we calculate

$$\eta_{opt} = 1 - (1 - \eta_{opto}/e_1)D = 1 - (0.576/12)2.8 = 0.865.$$

As we can see, the influence of slot shape on optical efficiency is not great enough to give preference to the rectangular slot shape, which increases optical efficiency by only 0.035, at the expense of the danger of tearing the band. The best shape from the standpoint of manufacture of punch dies is a shape limited by arcs with a radius equal to half the slot width, always somewhat greater than the diameter of the effective aperture.

In this case, the condition $r_B = D/2$ is not fulfilled, but since the difference $r_B - r \leq 0.5$ mm, the influence of this error can be ignored.

REFERENCES

1. Rodionov, B. N., N. P. Zakaznov, "Central Band-Type Aerial Photography Shutters," Trudy MIIGAiK, No. 39, 1960.
2. Rodionov, B. N., "The AFAMIIGAiK Small-Format Aerial Camera," Trudy MIIGAiK, No. 45, 1961.
3. Mel'nikov, A. A., Teoriya fotozatvorov [The Theory of Photographic Shutters], Moscow-Leningrad, Gostransizdat Press, 1937.

N75 10994

METHODS OF RADAR STUDIES OF THE MOON AND THE PLANETS FROM SPACECRAFT

N. N. Krupenio

Radar is one method which can be used to study the physical properties and characteristics of the moon and the planets.

Radar studies from the Earth have measured the mean distance to the moon and its mean radius with great accuracy, and have defined thevalue of the astronomical unit, the direction and angular rate of rotation of certain planets.

Radar studies from the Earth have succeeded in measuring such characteristics of reradiation (reflection and scattering) of the moon and planets as the effective scattering area and the back scattering pattern for a number of wavelengths. These characteristics have been used to estimate the values of the reflection factor, dielectric permeability and surface layer material density of these planets and to estimate the mean square angles of inclination of the surfaces in the scale of several wavelengths. /31

I. Methods of Radar Studies

Radar studies of the moon and planets from the Earth require high energy potentials. Therefore, the further a planet is from the Earth, the more modest the results of radar studies of the planet. For example, whereas the moon has been studied from 1946 at twenty different wavelengths, the first radar echoes from Venus were received only in 1961, and only ten wavelengths have been used to date. Radar studies of the more remote planets -- Mars, Jupiter and Mercury -- can be performed only by the most powerful radars on the Earth: the radar of the Center for Long-range Space Communications in the USSR, the Cornell University radar in the USA and the Goldstone Tracking Station radar of the Jet Propulsion Laboratory in the USA.

Two of these radars have the following primary characteristics:

Arecibo: $P_{lim.\ cont} = 150$ kw; $A_{ef} = 36 \cdot 10^3$ m^2; $T_{noise} = 200°K$; $\lambda = 69.8$ cm; Antenna -- spherical, 305 m diameter

Goldstone: $P_{lim.\ cont.} = 100$ kw; $A_{ef} = 1.6 \cdot 10^3$; $T_{noise\ \Sigma} = 27°K$; $\lambda = 12.5$ cm; Antenna -- steerable parabola, 64 m diameter

If we consider only the range and radius of the planets, with identical reradiation characteristics and with the planets at the minimum distance from the Earth, to achieve the identical signal/noise ratio, the energy potential of the radar must have the following excess in comparison to a radar for lunar studies: for radar studies of Venus $\geq$ 59 db, of Mars $\geq$ 80 db, of Mercury $\geq$ $\geq$ 89 db, of Jupiter $\geq$ 92 db, of Saturn $\geq$ 109 db, of Uranus $\geq$ $\geq$ 133 db, of Neptune $\geq$ 140 db, of Pluto $\geq$ 156 db.

Together with the necessity of providing these tremendous energy potentials, surface radar stations for planetary studies have limitations as to usable wavelengths.

Molecular absorption of radio waves in the atmosphere of the Earth limits this range at the bottom end to wavelengths of 0.8-1.5 cm, while the terrestrial ionosphere limits the top end to wavelengths of 15-20 m. However, the millimeter and sub-millimeter wavebands, capable of yielding information on the properties of the first five to twenty cm of the surface layer of a planet, cannot be used, preventing us from obtaining information necessary to determine the characteristics needed for landing systems and systems to carry spacecraft across the surfaces of the planets.

Terrestrial radar has insufficient surface resolving capacity when reradiation characteristics are measured. However, the developing technology of celestial navigation near the moon and planets requires that the knowledge of reradiation characteristics be an order of magnitude higher than the most precise modern terrestrial radars can provide.

Radar studies from Earth provide more reliable information concerning regions lying in the center of the disc of a planet, less reliable information concerning regions near the limb, both due to the poorer signal/noise ratio (20-50 db lower), and due to the greater difficulty in determining the dimensions of a reradiating area. /32

The development of space technology has allowed radar studies to be conducted from spacecraft located near the moon. However, these studies will be performed in the near future from spacecraft launched toward Mars and Venus as well.

Radar studies from a spacecraft have a number of advantages over terrestrial radar studies:

1) better resolution of the surface is comparatively easy to

obtain (radar studies from the Earth provide resolutions of over 1 km, from spacecraft -- a few dozens of meters);

2) a broader range of wavelengths can be used, including the millimeter and submillimeter bands;

3) lower energy potential is required.

However, provision of the comparatively lower energy potential required by an on-board radar on a spacecraft encounters the power limitations of on-board power supply and on-board apparatus weight limitations. For example, placement of electronic apparatus with a mean radiated power of a few hundred watts on a spacecraft (particularly a long-range spacecraft) is a very difficult technical task. Therefore, studies of the remote planets require better methods of independent radar study.

Great difficulties arise in the production of spacecraft orientation systems when it is necessary to assure precise aiming of directional antennas. However, these difficulties are compensated by the greater information production of radar studies from spacecraft in comparison to terrestrial radar methods.

The methods of spacecraft radar studies can be divided into two groups:

1) autonomous methods, in which the radiation, reception and processing of signals are all performed on board the spacecraft;

2) non-autonomous methods, when these operations are performed partially on board the spacecraft, partially on the ground (at the research center).

Processing of reradiated signals is usually very cumbersome, requiring complex automatic devices or extensive processing time, which makes it simpler to perform at the research center than on board the spacecraft. Another significant factor in selecting the placement of the receiving or transmitting equipment on board the spacecraft is the possibility of providing the required energy potential.

Let us study methods of radar research from spacecraft used in studying the moon and the planets from artificial satellites and space probes landed on the surface.

A. Autonomous Methods of Radar Study from Spacecraft

1. Measurement of Specific Effective Scattering Area with Radio Altimeter Installed on Landed or Orbital Spacecraft

During measurements, the following parameters, determining the potential of the radio line between the instrument and the

surface, are recorded and transmitted by telemetry: 1) radiated power of the transmitter; 2) power of the signal reradiated by the surface at the input of the receiver; 3) flying altitude.

For these same moments in time, the mutual placement of the spacecraft relative to the surface, deflection of the axis of the radiation pattern from the local vertical, coordinates of the point beneath the satellite and of the point where the antenna radiation pattern intersects the surface of the planet are also determined.

All of this is done using a system of external trajectory measurements and the orientation system of the spacecraft.

The power of the reradiated signal at the input of the receiver, if the size of the surface struck by the radiation is determined by the antenna radiation pattern rather than by the transmitter pulse length, is calculated using the formula /33

$$P_r = \frac{P_{tr} G_0^2 \lambda^2}{(4\pi)^3 H^2} \int_0^{2\pi} d\varphi \int_0^{\pi/2} \sigma_0(\theta) F^2 [\arccos(\cos\theta \cos\theta_0 + \sin\theta \sin\theta_0 \cos\varphi)] d\theta. \quad (1)$$

If the size of the area simultaneously irradiated is decreased by shortening the transmitter pulse length, the coefficient $F(\rho)$, which accounts for this effect, is introduced following the integral in formula (1). The method of calculation of coefficient $F(\rho)$ can be found in [4].

In formula (1):

$$\sigma_0(\theta) \simeq \sigma_{0N} B(\theta) \text{ where } \alpha < \theta;$$

$F[\ ] = F(\alpha)$ is the radiation pattern of the antenna; $B(\theta)$ is the back scattering pattern (BSP) of the surface; ϕ is the angle in the horizontal plane; α is the angle relative to the direction of maximum radiation of the antenna; θ is the incident angle; θ_0 is the deflection of the axis of the antenna radiation pattern from the vertical; G_0 is the maximum gain of the antenna; H is the flying altitude; σ_{0N} is the specific effective scattering area of the surface with normal incidence of the radiated beam; σ_0 is the specific effective surface scattering area with incident angle θ.

During these measurements, we determine the value of σ_0. However, if we assume that the back scattering pattern does not differ greatly from the average pattern for the entire planet (which can be determined experimentally by radar studies from the Earth or in many cases analytically), calculations using formula (1) can determine the value of σ_N. As the experience of performance of such measurements on the Luna-9, Luna-13 and Surveyor-1 spacecraft has indicated, the accuracy of determination of the specific effective area (considering errors in the measurement of energy characteristics of the radio altimeter, flying altitude and position of the antenna radiation pattern relative to the surface) for highly directional antennas is on the order of ±3.0-4 db. The accuracy of measurements on the Luna-16 and Luna-17 spacecraft was increased to ±2 db.

The value of σ_0 determined in the process of measurement is averaged over the area irradiated by the radio altimeter, and the accuracy of its determination will depend on the width of the antenna radiation pattern, inclination angle of the antenna and width of the back scattering pattern of the surface.

If the spacecraft carries a highly directional scanning antenna, it is possible to measure the back scattering pattern along with σ_{ON} by rather rapid scanning of the antenna in comparison to variation of the properties of reradiation of the surface along the flight path of the spacecraft.

2. Measurement of Specific Effective Scattering Area Using a Radar with an Antenna with a Synthesized Aperture

These radar systems create high resolving capacity along the trajectory of the spacecraft by recording and coherent addition of signals reradiated by the surface over a sector of the flight trajectory of the spacecraft equal in length to the length of a large (synthesized) antenna.

High resolution in the direction perpendicular to the line of flight at the spacecraft can be achieved by using extremely short transmitter pulses.

To assure good resolution in this direction, the area studied is selected somewhat to one side of the path over which the satellite flies, since the range resolution is inversely proportional to the sine of the incident angle in the direction perpendicular to the line of flight. /34

Coherent addition over a path several hundreds of meters long requires high accuracy of compensation for the Doppler frequency shift over the sector observed from the actual

radiation pattern, plus high phase stability in the high frequency transmitting and receiving circuits.

In order to compensate destabilizing factors sufficiently to allow a resolution of 10 m on the surface, a rather complex stabilization and control system must be used in the spacecraft, including:

1) a radio altimeter measuring both flying altitude and rate of change of altitude;
2) a Doppler drift angle, lateral and longitudinal flight velocity component meter;
3) an inertial system measuring the lateral and longitudinal accelerations of the spacecraft;
4) a gyroscopic system measuring the angles and angular velocities of the spacecraft relative to its center of gravity.

The power of the reradiated signal at the input of the receiver of such a radar set can be determined by the following formula:

$$P_s = \frac{P_{tr} G^2 \lambda^2 \Delta R \Delta L F_{rep} T_a}{(4\pi)^3 R^4 S_1 S_2 S_3} \sigma_0, \qquad (2)$$

where G is the power gain of the actual antenna; σ_0 is the specific effective area of the smallest surface element which can be resolved, with dimensions $\Delta R \times \Delta L (m)^2$; F_{rep} is the transmitter pulse repetition frequency; T_a is the coherent addition time; R is the distance to the surface element resolved; S_1 are the signal losses in the transmitting and receiving circuits of the radar; S_2 are the signal losses during processing; S_3 are the signal losses due to inaccuracy of orientation of the system along a perpendicular to the line of flight of the spacecraft.

This method allows us to measure the specific effective scattering area with high resolution and thereby produce a radar map of a surface sector parallel to the line of flight of the spacecraft.

The method of measurement of surface reradiation characteristics using a radar with a synthesized aperture has not been used as yet in planetary studies, although, judging from published reports, a plan of this type was analyzed for installation on the EROS satellite [1].

The characteristics of this radar are as follows:

1) wavelength 1.82 cm;

2) width of radiation pattern in vertical and horizontal planes 10°, with reception of signals of two orthogonal polarizations;
3) surface resolution 30.5 m;
4) inspection of a strip of 10 × 600 km from an altitude of 320 km;
5) synthesized aperture diameter 15 m;
6) weight of solid-state equipment 74 kg.

This weight does not include the apparatus for transmission of radar information to the ground station.

This article [1] does not mention the radiated power or power consumption from the on-board power supplies. Calculations show that the power which would be consumed by the radar would be several kilowatts, quite within the capabilities of many modern spacecraft.

Their high weight, great power consumption and complex required stabilization and control systems apparently make it impossible to use radars with synthesized apertures to study the surface properties of the moon and planets from spacecraft, although this type of system is quite feasible for satellite studies of the Earth. /35

B. Non-Autonomous Methods of Radar Studies from Spacecraft

1. Pulse Radiation of On-Board Transmitter with Reception and Processing of Signals on the Earth

If a spacecraft carries an antenna with a broad radiation pattern, both the direct signal of the transmitter and the signal reradiated by the surface of the planet will be simultaneously received and recorded on the Earth. Time-based separation of these signals is easy if the pulse length radiated by the transformer $\tau < H/c$. Recordings made on the Earth can be analyzed to determine the dependence of the ratio of power reradiated by the surface of the planet to power of the direct signal on the delay of each probe pulse. The spacecraft orientation and stabilization system can be used to determine which areas of the planet's surface reradiate each signal pulse.

Since the radiation pattern of a spacecraft antenna is generally far from isotropic, it is necessary either to measure the antenna radiation pattern or to cause the spacecraft to rotate about one of its axes and average the results of measurement of reradiated and direct signals over the time of one rotation of the spacecraft in order to consider the influence of the radiation pattern on the results of measurement. The set of

dependences of the relative reradiated signal power on range produced by averaging can be subjected to simple trigonometric transformations using the data on the mutual position of the spacecraft, planet surface and ground station to determine the incident angles (θ) and observation angles (Φ) for each range.

Then, by grouping all measurements as to incident angle with fixed observation angle or observation angle with fixed incident angle, we can calculate the back scattering characteristics

$$i(\theta, \Phi) = \frac{P_{tr}(\theta, \Phi)}{P_{tr\ max}} \tag{3}$$

where θ = const or Φ = const. When $\theta = \Phi$, we can calculate the back scattering pattern with a resolution determined by the length of the probe pulses:

$$B(\theta) = \frac{P_{tr}(\theta)}{P_{tr\ max}} \quad \text{where } \theta = \Phi. \tag{4}$$

Comparing the powers of the direct and reradiated signals, we can determine the albedo and specific effective scattering area. This last characteristic is defined as

$$\sigma_{0N} \simeq \frac{\int_{H/c}^{H/c+\tau} P_{trd}(t)\,dt}{P_{tr}} \frac{C_r}{C_H} \cdot 2 \frac{H^2}{ac\tau i(\Phi_H)}, \tag{5}$$

where C_H, C_r are the directive gains of the on-board antenna in the direction of the local vertical and the ground station respectively; a is the radius of the planet; H is the flying altitude; Φ_H is the angle of observation for a point immediately beneath the satellite.

In order to study the reradiation characteristics of the surface of the moon, a brief experiment was performed from the Luna-10 lunar orbiter using the pulsed radiation of a decimeter transmitter. Synchronization pulses for the on-board transmitter /36

were transmitted from the Earth to the spacecraft. These same pulses were used as starting pulses in the devices used to record the signal received. The parameters of the pulses transmitted by the radar (length and repetition frequency), as well as the mutual placement of the spacecraft, moon and ground station were selected so as to allow time separation of the direct and reradiated signals at the ground station.

The accompanying photographs (Figures 1 and 2) show recordings of the amplitude of signals at the output of the ground station receiver during the time of one pulse repetiation cycle. The Luna-10 spacecraft turned around its axis as the experiment was performed, causing great changes in the amplitude of the direct and radiated signals, due to the significant variations in the antenna radiation pattern.

However, analysis of the distribution of the reradiated signal as to range for two successive repetition periods, for which the change in three-dimensional radiation pattern can be ignored, shows that the reradiated signal fluctuates strongly and that its duration is significant, indicating that the reradiation is not simply mirror reflection in the decimeter wave band, but rather that the surface in the area of measurements is quite rough.

2. Monochromatic Radiation of On-Board Transmitter with Reception and Processing of Signals on the Earth

As an artificial satellite orbits around the moon, both the angle between the direction of the velocity vector and the direction to the reradiating surface area, and the angle between the direction of the velocity vector and the direction to the ground station will change continually. Therefore, a radiated monochromatic signal from the on-board transmitter will have Doppler shift for the direct signal and a shifted Doppler spectrum for the reradiated signal. The mutual position of the spacecraft, planet and ground station, determined by trajectory measurements, can be used to correlate each component of the reradiated signal spectrum to a definite area on the surface of the planet. The constant Doppler frequency in the reradiated signal spectrum will correspond to the area defined by the lines of intersection of two concentric cones, the axis of which corresponds to the velocity vector of the spacecraft, with the spherical surface of the planet. The mean angle at the tip of a cone is α_k, defined by the following formula:

$$\frac{\alpha_к}{2} = \arccos \frac{V f_0}{c f_{d_к}}. \tag{6}$$

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

/37

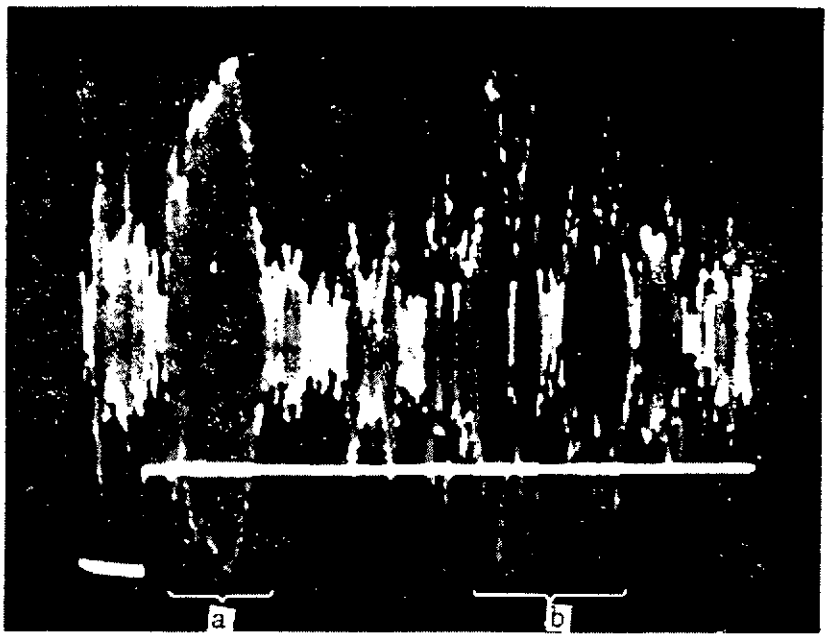


Figure 1. Recording of Amplitude of Direct (a) and Reradiated (b) Signals with Pulse Modulation of Transmitter of Luna-10 Spacecraft

The difference of the angles at the peaks of these two cones is determined by the resolution of the Doppler frequency spectrum being analyzed and is calculated by the formula

$$\delta\alpha = 2\arcsin\frac{\delta f_{d_\Sigma} c}{V f_0 \sin\alpha_\kappa}. \tag{7}$$

Here f_{d_k} is the Doppler frequency corresponding to the area of space being studied; V is the modulus of the spacecraft velocity vector; δf_{d_Σ} is the overall resolution of the system for Doppler frequencies: f_0 is the transmitter carrier frequency.

In the case of mirror reflection, the line corresponding to /38
a constant Doppler frequency on the surface degenerates to a point. Considering the finite resolution of the system for frequencies, we obtain an area of finite dimensions rather than

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

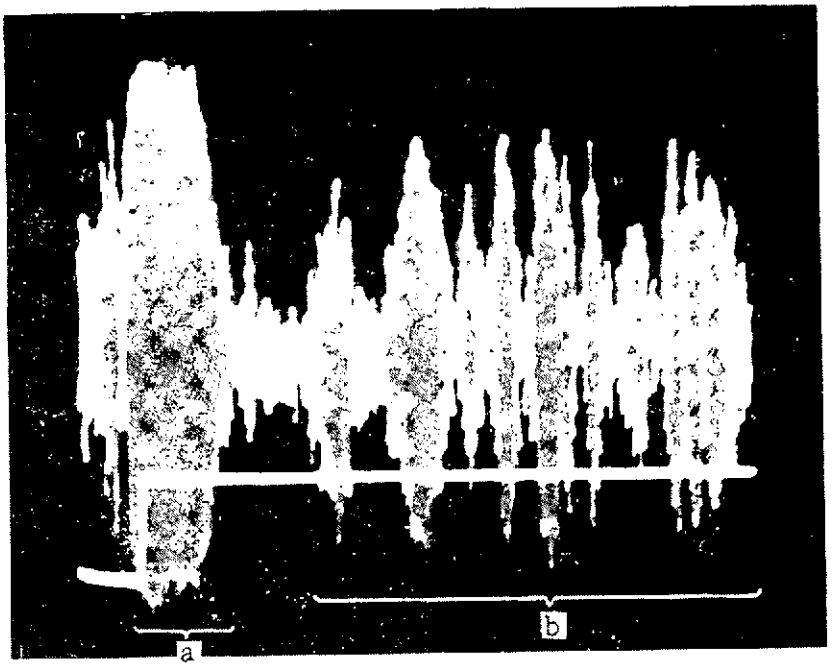


Figure 2. Recording of Amplitudes of Direct (a) and Reradiated (b) Signals with Pulse Modulation of Luna-10 Transmitter, Changing with Time

a point on the surface of the planet. The dimensions of this area and its placement can be easily determined using the formulas presented above considering the additional condition $\theta = \Phi$. Furthermore, comparison of the powers of the direct and reradiated signals, the spectra of which degenerate to a common line for the case studied above can be used to determine the reflection coefficient:

$$\rho_m(\theta_m) = \frac{P_{tr}\, G_r}{P_{per} G_{mir}\, \alpha_m}, \tag{8}$$

where G_{mir}, G_r are the radiation patterns of the on-board antenna in the direction of the mirror reflection and the ground station; α_m is the coefficient of divergence, equal to

$$\alpha_m = \frac{a \cos \theta_m}{R_m + a \cos \theta_m}, \tag{9}$$

Here a is the radius of the planet; R_m is the distance from the spacecraft to the point of mirror reflection; θ_m is the incident angle for the point of mirror reflection. The values of R_m and θ_m are related to the spacecraft flying altitude and the Earth-moon-spacecraft angle ϕ_a by the following formulas:

$$\tan (2\theta_m - \varphi_a) = 2 \tan \theta_m, \tag{10}$$

$$R_m = (a + H) \frac{\sin(\varphi_a - \theta_m)}{\sin \theta_m}. \tag{11}$$

After performing repeated measurements, we can determine the dependence of reflection factor on incident angle and then, comparing it with the theoretical dependence for various ε considering polarization, it is possible to determine the permittivity of the surface layer of the planet, ε, for the wave length for which measurements were performed.

In the case of non-mirror reradiation, analysis of the results of measurement is significantly more difficult, since the Doppler frequencies in the spectrum of the reradiated signal cannot be unambiguously correlated to an area on the surface of the planet. To determine the properties of the surface layer, the reverse problem is usually solved. Since the spectrum of the reradiated signal is determined by the unevenness of the surface, by selecting the electrical properties of the surface layer, as well as the size and distribution of irregularities on the surface, it can be arranged so that the distribution characteristics of the model correspond to the spectrum of the reradiated signal which matches the Doppler spectrum produced experimentally. This is an extremely simple task. Therefore, in practice the simplified reverse problem was solved first: the ratio of integral power of the reradiated signal to the power of the direct signal was determined as a function of the incident angle and flying altitude for various values of permittivity ε and mean square angles of inclination of the surface σ_α [2]. These theoretical dependences were then compared with the results of changes in this ratio produced on the Luna-11 and Luna-12 artificial satellites. Calculations have shown that the results of experiments at λ = 1.7 m correspond best to ε = 2.7-3.0 and σ_α = 10-12°.

A more complex experiment was performed during the flight of the Lunar Orbiter-1 satellite. During the sector of the flight trajectory near the Castner crater, the spectra of direct and reradiated signals were recorded continually [3]. The spectrogram showed several continuous lines, the Doppler frequencies of which changed with time. It was assumed that each line of the /39 spectrum corresponds to reradiation from a fixed surface sector. Processing the signal received using the data from trajectory measurements, the experimenters determined the coordinates of the "radio bright" areas. The intensity of the reradiated signal from these areas was 3 to 5 db higher than from the surrounding surface. Thus, radar mapping from a satellite was achieved. A similar experiment was performed on the Explorer 35 spacecraft with continuous radiation of the on-board transmitter at $\lambda = 2.2$ m [14].

The Luna-14 performed a similar experiment. However, processing of the "instantaneous" Doppler spectra allowed consideration of the statistics of the surface and, therefore, production of more reliable data on the dielectric permeability of the upper layer of the local regions of the moon at $\lambda = 1.7$ m [13].

II. Some Results of Measurement of Reradiation Characteristics of Local Regions of the Lunar Surface

The reradiation characteristics of local regions of the lunar surface were measured during the landings of the Luna-9, Luna-13, Luna-16, Luna-17, Surveyor-1, Surveyor-3, Surveyor-5, Surveyor-6 and Surveyor-7 spacecraft, and also by the Lunar Orbiter-1. The Luna and Surveyor spacecraft measured the specific effective scattering area as the spacecraft landed on the surface of the moon. All of these spacecraft, except for the Surveyor-7, landed in relatively smooth areas. The Surveyor-7 landed in a continental region in the ray system of Tycho crator. Processing of the signals at the input of the on-board radio altimeters and Doppler velocity meters produced dependences of the specific effective scattering area on flying altitude. These dependences are illustrated on Figure 3 (Luna-9 and Luna-13) and Figures 4, 5, 6, 7 (Surveyor-1).

Table 1 presents the results of calculation of the specific effective scattering area with normal incidence from the data of measurements by the satellite. The results presented in this /41 table from calculation of the mean value of specific effective scattering area over the disc of the moon, taken from terrestrial radar observations, are taken from [10].

As we can see from Table 1, the values of σ_{ON} measured in the maria regions by Luna-13, Surveyor-1, Surveyor-3, Surveyor-5

and Surveyor-6 differ little from the average values of σ_{ON} for the disc of the moon at the same wavelengths. The results of measurement of σ_{ON} by Surveyor-7 in a continental region of the moon are double the average for the entire disc of the moon. The results of measurements by Luna-16 and Luna-17 yielded lower values of σ_{ON} than the average values for the disc on the moon and can be explained by some difference in the electrical properties of the surface material of the local regions of the moon and slightly higher densities of material and permittivity of the substance of the continental regions in comparison to the maria regions.

TABLE 1

Spacecraft	Wavelength, cm	Radiation	Antenna Radiation Pattern Width, °	Specific Effective Scattering Area with Normal Incidence				Region of Measurements
				Measured Value (Mean)		Average for Entire Disc		
				Value	Accuracy, db	Value	Accuracy, db	
Luna-9	3.2	Pulse	3.5	0.5	±3.5	0.64	±3	Mare
Luna-13	3.2	"	3.5	1.0	±3.5	0.64	±3	"
Luna-16	3.0	Continuous	4.0	0.2	±2.0	0.56	±3	"
Luna-17	3.0	"	4.0	0.36	±2.0	0.56	±3	"
Surveyor-1	2.25	Continuous	6	0.25	±4	0.3	±3	"
Surveyor-3	Same as for Surveyor 1			0.2	±4	Same as for Surveyor 1		"
Surveyor-5				0.16	±4			"
Surveyor-6				0.2	±4			"
Surveyor-7				0.4	±4			Continent

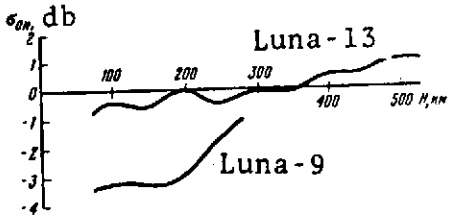


Figure 3. Specific Effective Scattering Area with Normal Incidence at λ_0 = 3.2 cm (Luna-9; Luna-13)

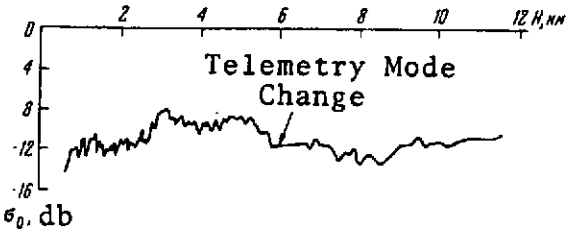


Figure 4. Specific Effective Scattering Area λ_0 = 2.25 cm (Surveyor-1, first ray, θ_0 = 25°)

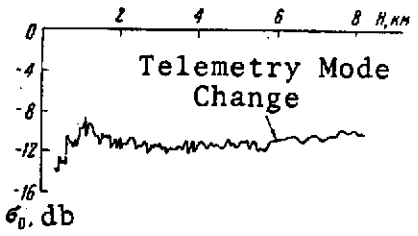


Figure 5. Specific Effective Scattering Area λ_0 = 2.25 cm (Surveyor 1, second ray, θ_0 = 25°)

The experiments on Lunar Orbiter 1 produced a radar map of the sector of the lunar surface near Castner crater (eastern edge of lunar disc) [3] at λ = 13.1 cm. The results of matching of the radar map with a photographic map of this region [11] are illustrated in Figure 8. Analysis of the radar map indicates that the radio bright areas correlate to the structure of the surface of this region: /40

1) As we can see, the inner slopes of craters and floors of craters have higher reradiation;

2) The radio bright areas do not coincide with smooth surface sectors;

3) Not all radio bright areas match with sectors of the surface satisfying the requirement of mirror reflection at the moment of measurement.

These data agree well with the results of radar mapping of individual sectors of the lunar surface from the Earth [12].

Conclusions /42

The first results of radar studies of the moon from spacecraft have shown that this method allows studies to be performed with higher surface resolution than the best measurements from the Earth. The accuracy of measurement of reradiation characteristics of the surface from spacecraft is not much less than the accuracy of radar measurements from the Earth (approximately 0.5-1 db). These qualities of radar studies of the properties of the surface from

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

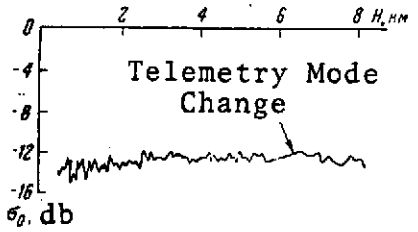


Figure 6. Specific Effective Scattering Area at λ_0 = 2.25 cm (Surveyor-1, third ray, θ_0 = 25°)

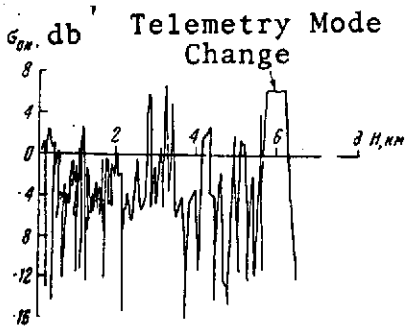


Figure 7. Specific Effective Scattering Area with Normal Incidence at λ_0 = 2.32 cm (Surveyor-1)

spacecraft, as well as the possibility of performing measurements with high resolution near the limb and even on the back side of the moon make radar studies from spacecraft one of the most informative methods of investigation of the surface of the moon and the other planets.

In conclusion, I would like to express my gratitude to A. V. Kurbatov, V. G. Timonin, G. S. Zaslavskiy and R. K. Kazakova for their help in performing the present work.

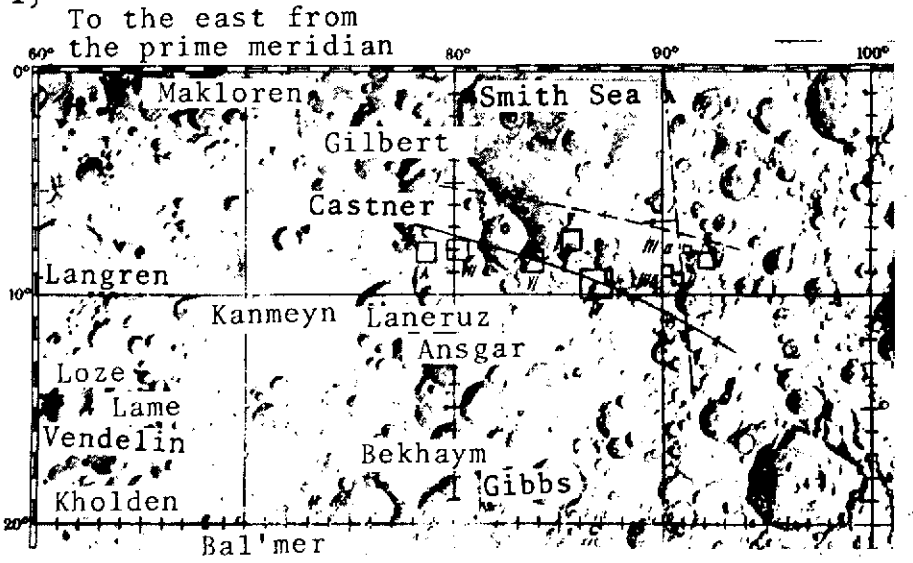


Figure 8. Radar and Photographic Map of Region of Kasner Crater (Lunar Orbiter 1, λ_0 = 13.1 cm)

REFERENCES

1. Electronics, 1967, Vol. 40, No. 2, p 47.
2. Yakovlev, O. I., et al., Kosmicheskiye issledovaniya, 1968, Vol. 6, No. 3, p 432.
3. Taylor, G. L., et al., Science, 1967, Vol. 157, No. 3785, p. 193.
4. Zonnenshtral', G. A., Yu. I. Fel'dman, Kosmicheskiye issledovaniya, 1970, Vol. 8, No. 4.
5. Krupenio, N. N., Kosmicheskiye issledovaniya, 1968, Vol. 6, No 2, p. 252.
6. Brown, W. E., J. Geoph. Res., 1963, Vol. 72, No 2. p. 791.
7. Krupenio, N. N., Kosmicheskiye issledovaniya, 1966, Vol. 6, No. 6, p. 951.
8. Jaffe, L. D., et al., J. Geophys. Res., 1968, Vol. 73, No. 12, p. 3983.
9. Jaffe, L. D., Report at Symposium on Physics of the Moon and Planets, Kiev, 1968.
10. Krupenio, N. N., Radiolokatsionnyye issledovaniya Luny [Radar Studies of the Moon], Moscow, Nauka Press, 1971.
11. Polnaya karta Luny [Complete Map of the Moon], Composed by State Astronomical Institute imeni P. K. Shternberg and USSR Topogeodetic Survice, Nauka Press, 1967.
12. Pettengill, G. H., T. W. Thompson, Icarus, 1968, No. 8, p. 3.
13. Yakovlev, O. I., S. S. Matyugov, N. N. Krupenio, K. M. Shvachkin, Report at Ninth All-Union Symposium on Radio Wave Propagation, Khar'kov, 1969.
14. Tyler, G. L., J. Geoph. Ref., 1968, Vol. 73, No. 24, p 7609.

N75 10995

MULTIPLE-SLIT SHADOW DEVICE FOR SENSING THE DIRECTION TO A SHINING OBJECT

A. V. Bruns

In developing the photoelectric spectrometer for studies of the sun from outside the Earth's atmosphere, it was required to develop a simple, highly reliable tracking system to guide the rotating mirror toward the sun so that the light reflected from the mirror would illuminate the input slit of the spectrometer.

The system developed was based on an element to detect the direction of the sun -- a multi-slit shadow direction sensor of jalousie type (JDS). The system held the direction to the sun with an accuracy of ±7', which satisfied the requirement of the job.

In order to check the calculation method and determine the maximum accuracy to be provided by a sensor of this type, a single-coordinate version was developed and tested.

I. Operating Principle

The direction sensor is based on a shadow analyzing element (SAE), used to determine the mutual placement of the sensor and light source. It consists of a set of z-cross section diaphragms in front of a photosensor. There are two SAE for each coordinate in the JDS (Figure 1). /43

In the properly directed state, when the axis of symmetry of the JDS coincides with the direction toward the light source, both of the photosensors are in the shadow (Figure 1b). In case of a mismatch in direction, one of the two photosensors will be struck by light, while the other will remain in the shadow. The photosensor struck by light depends on the direction of the mismatch (Figure 1a, c).

Under actual operating conditions, due to light scattered on the walls of the diaphragms, diffraction on the edges of the diaphragms, and also the half shadow produced due to the finite dimensions of the object being tracked, both photosensors do receive some light flux even when the device is properly aimed. Therefore, a deviation in the axis of symmetry of the JDS from the direction toward the light source causes the light flux striking one photosensor to increase while the other decreases relative

to their initial values. If the SAE are at some angle to each other, the sensor may have some initial level of illumination of the photosensors, depending on this angle.

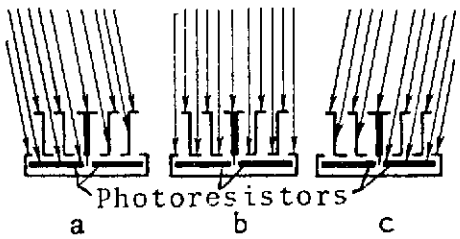


Figure 1. Diagram of Operation of Multi-slit Shadow Sensor

The use of a set of many z-shaped diaphragms in each SAE allows a comparatively steep output characteristic to be achieved, in spite of the small dimensions of the sensor: the mismatch angle is the output signal.

1. Output Characteristics of SAE for Point and Larger Light Sources

When the sensor tracks a point light source, a parallel beam of light strikes the SAE, creating light level E_0 (Figure 2a).

Light flux Φ passing through each element of the SAE with incident angle α and plate width b will be:

$$\Phi = E_0 hb \sin \alpha. \tag{1}$$

If we use h elements in the SAE packet, when the sensitivity of the photosensor is η, the output signal is

$$P = \eta\, nEhb \sin \alpha. \tag{2}$$

When the device tracks an extended light source (ELS), the surface of the photosensor will not receive a constant level of illumination. Due to the finite dimensions of the ELS, the photosensor, in addition to the shaded and illuminated areas, also has a half-shaded area. If the size of this half-shaded area is negligible in comparison to the entire illuminated surface, the output signal can be defined, as is the case for a point source, using formula (1). /44

Let us study the operation of the sensor in the half-shaded area (Figure 2b), i.e., when light reaches the field of vision from only a portion of the ELS. We will limit ourselves to analysis of the phenomena in the plane of the figure, assuming that in the perpendicular direction, the field of vision of the

SAE is significantly larger than the angular dimensions of the object being tracked.

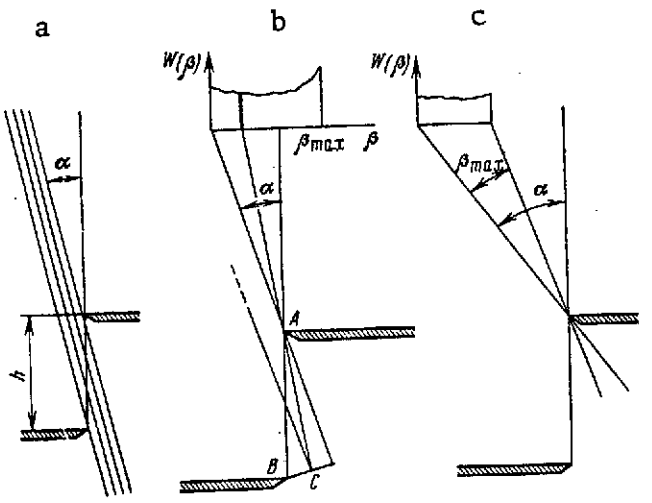


Figure 2. Operation of Shadow Analyzing Element with Various Light Sources in the Field of Vision: a, Point Source; b, Section of Extended Source; c, Entire Extended Source

Let us assume that the distribution of brightness over the objects being tracked is described by function $W(\beta)$, where β is an angular coordinate, read from the edge of the ELS (Figure 2b). Strictly speaking, $W(\beta)$ is the integral brightness of an elementary band of the source, located in the direction perpendicular to the figure and having coordinate β in the plane of the figure.

The condition of operation of the SAE in a half shadow area will be $\alpha \leq \beta_{max}$, where α is the angular distance between the edge of the ELS and the sighting line passing the edges of the upper and lower lateral elements of the diaphragm (points A and B of Figure 2b); β_{max} are the angular dimensions of the ELS. The surface of the ELs, from each edge to the point with coordinate β, creates illumination $E(\beta)$ at point C of the photosensor, equal to

$$E(\beta) = k \int_0^{\beta} W(\beta)\, d\beta. \tag{3}$$

The normalizing factor k is defined by means of the illumination E_0 received from the entire ELS surface:

$$E_0 = k \int_0^{\beta_{max}} W(\beta)\, d\beta; \quad k = \frac{E_0}{\int_0^{\beta_{max}} W(\beta)\, d\beta}. \tag{4}$$

Thus, expression (3) can be written as

$$E(\beta) = \frac{E_0}{\int_0^{\beta_{max}} W(\beta)\,d\beta} \int_0^{\beta} W(\beta)\,d\beta. \tag{5}$$

If the ELS has constant brightness over its entire surface, then /45

$$W(\beta) = kL(\beta), \tag{6}$$

where $L(\beta)$ is the transverse dimension of the ELS as a function of angle β. In this case, expression (5) becomes

$$E(\beta) = \frac{E_0}{\int_0^{\beta_{max}} L(\beta)\,d\beta} \int_0^{\beta} L(\beta)\,d\beta. \tag{7}$$

Knowing the distribution of illumination, let us determine the light flux striking the illuminated portion of the surface of the photosensor:

$$\Phi = \int_S E(\beta)\,dS = bh \int_0^{\alpha} E(\beta)\cos(\alpha - \beta)\,d\beta. \tag{8}$$

Here $dS = bh \cos(\alpha - \beta)d\beta$ is an element of the surface with illumination $E(\beta)$ (point C, Figure 2b).

If the field of view of the sensor contains only a portion of the ELS ($\alpha \leq \beta_{max}$), illumination $E(\beta)$ is determined from expressions (5) or (7), the light flux is related to angle α by the dependence

$$\Phi(\alpha) = \frac{E_0 bh}{\int_0^{\beta_{max}} L(\beta)\,d\beta} \int_0^{\alpha} \cos(\alpha - \beta) \int_0^{\beta} L(\beta)\,d\beta. \tag{9}$$

For the case when the ELS is entirely contained in the field of vision ($\alpha < \beta_{max}$, Figure 2c), the photosensor receives both the half shade and an area with constant illumination E_0. Expression (8) in this case is broken into two expressions:

$$\Phi(\alpha) = bh \int_0^{\beta_{max}} E(\beta)\cos(\alpha-\beta)\,d\beta + E_0 bh \int_{\beta_{max}}^{\alpha} \cos(\alpha-\beta)\,d\beta =$$
$$= bh \int_0^{\beta_{max}} E(\beta)\cos(\beta_{max}-\beta)\,d\beta + E_0 bh \sin(\alpha-\beta_{max}). \tag{10}$$

The first component corresponds to the light flux reaching the photosensor in the half shade area (within limits of angle $0 < \beta < \beta_{max}$). The second component corresponds to the light flux passing through the area with constant illumination.

We should note that with small angular source dimensions ($\beta_{max} << \alpha$), the first component becomes negligible and expression (10) is converted to formula (1), derived earlier for a point light source.

2. Basic Relationships when Working with the Sun

We will analyze the sun as a disc with even brightness over its surface. The distribution of brightness $W(\beta)$ is determined according to expression (6) as the dependence of the chord length (in angular measure) on its position on the disc (Figure 3):

$$L(\beta) = 2\sqrt{\left(\frac{\beta_{max}}{2}\right)^2 - \left(\frac{\beta_{max}}{2} - \beta\right)^2}$$

or, representing $\phi = \beta/\beta_{max}$, we produce

$$L(\delta) = 2\beta_{max}(\delta - \delta^2)^{1/2}. \tag{11}$$

The distribution of illumination over the half shade area can be found by substituting expression (11) into (7) and setting the corresponding integration limits: /46

$$E(\delta) = \frac{E_0 \int_0^{\delta} (\delta - \delta^2)^{1/2} d\delta}{\int_0^1 (\delta - \delta^2)^{1/2} d\delta} = \frac{E_0}{\pi} \{0{,}5\pi - \arcsin(1 - 2\delta) - $$
$$- (1 - 2\delta)[1 - (1 - 2\delta)^2]^{1/2}\}. \qquad (12)$$

To determine the light flux, we must integrate the illumination $E(\delta)$ according to expressions (9) or (10) with respect to the entire surface of the photosensor S. Considering the small angular dimensions of the sun, we can assume within the half shade $dS = hbd(\alpha - \beta)$.

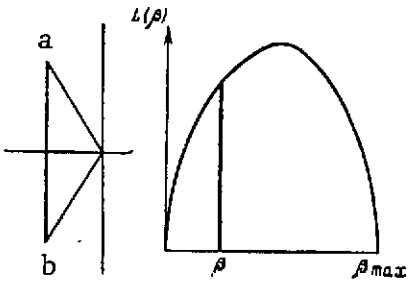


Figure 3. Determination of Solar Brightness Function

The light flux in this case (where $\alpha \leq \beta_{max}$) is equal to

$$\Phi(\delta) = E_0 bh\beta_{max} \int_0^{\delta} E(\delta)\, d\delta =$$
$$= 0.5\beta_{max} E_0 bh \left\{ \frac{2 + (1 - 2\delta)^2 [1 - (1 - 2\delta)^2]^{1/2}}{3\pi} + \right.$$
$$\left. + \frac{(1 - 2\delta)[\arcsin(1 - 2\delta) - \pi/2]}{\pi} \right\}. \qquad (13)$$

If the sun is in the center of the field of vision, i.e., $\alpha > \beta_{max}$, according to expression (10) for the light flux we have

$$\Phi(\delta) = E_0 bh \frac{\beta_{max}}{2} \left\{ 1 + \frac{2}{\beta_{max}} \sin[\beta_{max}(\delta - 1)] \right\}, \qquad (14)$$

where $\delta = \alpha/\beta_{max}$. The output signal $P(\delta)$ of one SAE with n slits and a photosensor with sensitivity η will be

$$P(\delta) = n\eta\Phi(\delta). \qquad (15)$$

The light flux in this expression is found from formulas (13) or (14), depending on the relationship between α and β_{max}. Let us represent

$$P_0 = E_0 nbh\eta \frac{\beta_{max}}{2}, \tag{16}$$

so that formula (15) can be written as

$$P(\delta) = P_0 M(\delta), \tag{17}$$

where $M(\delta)$ is the expression in the braces in formula (13) or (14).

The sun direction sensor has two SAE connected in opposition for each coordinate. Therefore, the output signal is the difference of the output signals of the two SAE:

$$P_{jds} = P_1(\delta_1) - P_2(\delta_2). \tag{18}$$

If both SAE are identically manufactured and have photosensors with identical characteristics, the expression for the output signal of the sensor becomes /47

$$P_{jds} = P_0[M(\delta_1) - M(\delta_2)]. \tag{19}$$

Since the sensor must provide rather high accuracy, its operation near the balance point is interesting. In the properly aimed state (where $P_{jds} = 0$), $M(\delta_1) = M(\delta_2) = M(\delta_H)$.

For slight deviations, we can write

$$P_{jds} = 2P_0 M'(\delta_H)\Delta\delta, \tag{20}$$

where $M(\delta_H) = dM(\delta_H)/d\delta$ is the corrected steepness of the characteristic of one SAE at point $\delta = \delta_H$; here $\Delta\delta$ is the deviation from the equilibrium point. It is clear from formula (20) that in order to produce the maximum signal at the output of the sensor with slight deviations, the SAE must be located relative to each other so that in the properly aimed state they are at the point of their characteristic with the maximum steepness.

Stability

The effects of various factors on the sensor change both the steepness and the position of the output characteristic as time passes.

This change may result from two factors: 1) the influence of temperature and other types of deformation on the sensor, and 2) changes in the photoelectric parameters of the photosensors. Careful design, proper selection of materials and manufacturing quality control can completely eliminate the influence of the first factor, so that the drift in the characteristics of a sensor will result only from the influence of changes in the photosensor parameters. The maximum constancy of parameters is that of photosensors with the photoemissive effect. Of diode type photosensors, silicon elements have the highest stability. Various parameters and characteristics of photosensors, including questions of stability, are discussed in [1, 2].

Since the operation of the JDS is based on the balance of currents of two photosensors, its stability is determined entirely by the possible relative drift of the parameters of each SAE. Obviously, in order to increase the stability of the JDS, we must select photosensors with characteristics as similar as possible.

The instability of the parameters of photoreceptors is manifested primarily in changes of the steepness of the output characteristic. The influence of variations in dark current t when illuminated by the sun is generally slight. Therefore, changes in the output signal arising with all sorts of instabilities are proportional to the value of this signal at the working point selected and, consequently, the selection of this point determines the angular error of the JDS.

The stability of various points on the SAE characteristic can be estimated quantitatively on the basis of the ratio of the increment in the angle $\Delta\alpha$ to the corresponding relative increment in light flux $\Delta\Phi/\Phi$. Let us call this ratio the instability factor and represent it by K_C. Then

$$K_c = \frac{\Delta\alpha}{\frac{\Delta\Phi(\alpha)}{\Phi(\alpha)}}. \tag{21}$$

In the case of the use of photoreceptors with a linear light characteristic, the relative change in light flux in expression

(21) can be replaced by the relative increment in the output signal of the SAE, $\delta P(\alpha)/P(\alpha)$: /48

$$K_c = \frac{P(\alpha)\,\Delta\alpha}{P(\alpha)} \tag{22}$$

or, considering formula (17),

$$K_c = \frac{M(\delta)\,\beta_{max}}{M'(\delta)}. \tag{22a}$$

TABLE 1*

α, min	δ	M	M'	K_c, calc., min	K_c exp., min	α, min	δ	M	M'	K_c, calc., min	K_c exp., min
0	0	0	0			18	0,5625	0,2796	1,159	7,6	7,65
2	0,0625	0,0012	0,0522	0,74		20	0,625	0,3569	1,2718	8,6	8,6
4	0,125	0,0071	0,1446	1,57	6,25	22	0,6875	0,4448	1,4658	9,7	10,0
6	0,1875	0,0197	0,2598	2,43	5,6	24	0,75	0,5450	1,6108	10,8	12,0
8	0,250	0,04	0,3712	3,45	5,47	26	0,8125	0,6447	1,74	11,9	14,0
10	0,3125	0,0699	0,5342	4,2	5,56	28	0,875	0,7571	1,8554	13,2	16,0
12	0,375	0,1069	0,6848	5,0	5,6	30	0,9375	0,8762	1,9478	14,3	18,0
14	0,4375	0,1546	0,8410	5,9	6,0	32	1,0	1,0000	2,0000	16,0	20,0
16	0,5	0,2122	1,0000	6,6	6,66						

*Translator's Note: Commas represent decimal points.

It follows from expression (22) that the instability factor can also be defined as the ratio of the SAE output signal to the steepness of its characteristic at the point in question.

We can use expressions (22) and (13) to calculate the value of K_c for the calculated characteristic of the SAE. Table 1 presents the calculated values of $M(\delta)$, $M'(\delta)$ and K_c, and Figure 4 shows these dependences graphically. Here also we see the output characteristic and change in stability factor for an experimental SAE model. A curve for K_c was plotted using experimental characteristics by calculation with the formula

$$K_c = \frac{(\alpha_1 - \alpha_2)(J_1 + J_2)}{2(J_1 - J_2)}. \tag{23}$$

The calculated value of K_c decreases with decreasing β and reaches a minimum as $\alpha \to 0$, when the light flux on the photosensor approaches 0. However, due to scattered light and diffraction of the working light flux on the edge of the diaphragm, the illumination of the photosensor does not drop to zero, as follows from expression (13). Therefore, under actual conditions, the function $K_c = f(\alpha)$ has its minimum at another point. The minimum value of K_c can be used to characterize the quality of each specific set of diaphragms used in the SAE.

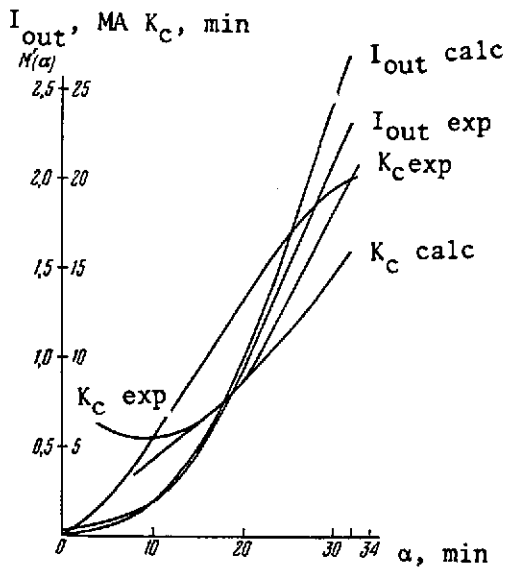


Figure 4. Calculated and Experimental Characteristics of SAE

The basic means for reducing the value of the instability factor in specific JDS designs is reduction of the scattered light level. /49 To do this, the inner surface of the z-shaped diaphragms must be specifically treated and covered with light-absorbing paint.

If we know the instability factor, we can use the known change in photosensor parameters to determine the SAE angular error resulting:

$$\Delta\alpha = \frac{\Delta P}{P} K_c. \qquad (24)$$

If the photosensor used in an SAE, with characteristics as illustrated on Figure 4, undergoes a change in sensitivity of, for example, 2%, with initial angle α_H = 16 min, this leads to an angular error of

$$\Delta\alpha = 0.02 \cdot 6.6 \simeq 0.13'$$

Selection of Initial Level of Illumination of Photosensors

The initial signal level at the output of each photosensor with JDS guidance is determined by the angle at which the SAE are

placed in relation to each other. Since various points on the characteristic are not equivalent, the operation of the sensor is determined to a great extent by proper selection of the section of the characteristic. We can note two extreme points on the characteristic of an SAE, usually limiting the area of selection of a working section: the point at which the sensor has the greatest sensitivity, and the point with the maximum stability of the characteristic. Based on expression (20), the sensitivity of JDS is proportional to steepness at the working point of the SAE output characteristic, which, according to formulas (13) and (14), reaches its maximum value where $\alpha = \beta_{max}$. The stability of each sector of the characteristic is determined by the value of K_c and increases with decreasing K_c. The working point is selected between these extreme points, depending on the specific requirements placed on the JDS.

Field of Vision

The field of vision of an SAE consisting of identical z-shaped plates is determined by the relationship between the width of the horizontal bands l_1 and l_2 and the distance between them H:

$$\alpha = \arctan \frac{l_1 + l_2}{h}, \tag{25}$$

With small angles $\tan \alpha = \alpha$, we can write the following formulas for the field of vision

$$\alpha = \frac{l_1 + l_2}{h}; \tag{26a}$$

$$\alpha_{max} = \frac{a}{nh}; \tag{26b}$$

$$\alpha_{max} = \frac{S_{wrk}}{nbh}. \tag{26c}$$

Here a is the dimension of the photosensor across the field of vision; S_{wrk} is the working surface of the photosensor.

The problem is usually reduced to attaining a sufficiently large field of vision with maximum sensitivity. Obviously, in order to increase the field of vision we must decrease h or increase $l_1 + l_2$. However, as follows from (13), a decrease in h results in a proportional decrease in the sensitivity of the sensor. On the other hand, an increase in dimension $l_1 + l_2$ worsens the utilization of the working surface of the photosensor which also decreases sensitivity in the final analysis. Thus, the expansion of the field of vision of the SAE in any case results in some reduction in sensitivity. /50

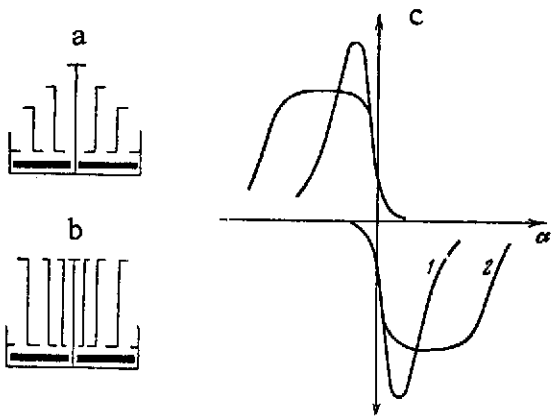


Figure 5. Methods of Expanding the Field of Vision of a Sensor (a, b) and the Change in Characteristic upon Expansion of the Field of Vision (c)

A method of expanding the field of vision without any loss in sensitivity is shown in Figure 1. The expansion of the field here is achieved by removing the upper shelves from the outer plates. The shortcoming of this method is that the light flux, passing through only one slit, cannot create a great signal at the output of the photosensor, so that some reduction in output signal will occur around the edges of the field of vision in comparison with the central portion, in which several slits operate simultaneously. If the rule of change of output signal across the field of vision is important, the sensitivity must be decreased by creating designs as shown in detail on Figure 5a, b. Here the output characteristic is produced by a gradual change either in the height of the z-shaped plates or in the length of their horizontal shells. The approximate shapes of the characteristics for the sensors corresponding to the diagrams of Figures 1 and 5a, b are shown on Figure 5c.

Calculation Method

Based on these considerations, we can recommend a procedure for calculation of JDS, the result of which should be the determination of the output characteristic of the sensor, as well as determination of its basic dimensions.

For clarity, simultaneously with the presentation of the method of calculation we will perform a test calculation of a one-coordinate JDS.

Eliminating the dependent variables from expressions (16), (20) and (26c), we produce

$$K_c \frac{\Delta P_{JDS}}{\Delta \alpha} = E_0 \eta \frac{S}{\alpha_{max}} M(\delta_H) \beta_{max}.$$

It follows from this expression that any four of the following five parameters should be fixed as the basic initial data:

1) steepness of JDS output characteristic $dP/d\alpha$;
2) instability, defined by value of coefficient K_c;
3) type of photosensor, geometric dimensions of working portion of its photoguide $S_{wrk} = ab$;
4) sensitivity of photosensor to sunlight $E_0\eta$;
5) field of vision α_{max}.

Let us take the area of the silicon photosensors as the unknown parameter in this example, consider the other four parameters known, equal to: /51

$$\frac{dP_{JDS}}{d\alpha} = 0.17 \text{ ma/ ang. min}; \; E_0\eta = 18 \text{ ma/cm}^2;$$

$$K_c = 6.5 \text{ ang. min}; \; \alpha_{max} = 80 \text{ min}.$$

The basis for construction of the characteristics of the JDS consists of the characteristics of the SAE which compose it and the initial signal level in the properly aimed state.

Let us perform the calculation by graph analysis. We can use the value of K_c (see Figure 4) to determine the initial angle α_H = 16 minutes and the value of $M'(\alpha_H)$.

Then, using expression (20), we find P_0 -- the value of the signal at the output of one SAE when $\alpha = \beta_{max}$:

$$P_0 = \frac{\Delta P_{JDS}}{\Delta \alpha} \cdot \frac{\beta_{max}}{2M'(\alpha_H)} = 0.17 \cdot \frac{32}{2} = 2.7 \text{ ma}.$$

Knowing the value of P_0 from expression (17) and using Table 1, we can construct the characteristics of the SAE.

Figure 4 shows a characteristic constructed by this method and an experimental characteristic. The reduction in the output signal rise rate in the experimental study in comparison with calculations where $\alpha > 20'$ results from the design of the SAE, namely from the small dimensions of the outer horizontal shelf -- $l_1 = 0.4$ mm. The required working area of the silicon photosensors can be determined from formula (16) considering (26c):

$$S_{wrk} = ab = 2P_0 \frac{1}{E_0\eta} \frac{\alpha_{max}}{\beta_{max}}; \quad S_{wrk} = 2 \cdot 2.7 \cdot 1/18 \cdot 2.6 = 0.78 \text{ cm}^2.$$

Assuming a coefficient of utilization of photocell area $K_{use} = 0.8$, we get

$$S_{tot} = S_{wrk}/0.8 = 1 \text{ cm}^2$$

The basic SAE dimensions are interrelated by formulas (26a, b, c). In this example we assume: $n = 8$, $b = 6$ mm, $\alpha_{max} = 0.026$. Then $h = S_{wrk}/nb\alpha_{max} = 61$ mm; $l_1 + l_2 = h\alpha = 1.6$ mm. The width of the shelf closest to the photosensor is taken from the condition $l_2 = (1.5\text{-}2)h\beta_{max}$. Fulfillment of this relationship prevents scattered light reflected from the walls less than three times from striking the photosensor.

The calculations which we have performed indicate the basic dimensions of JDS elements.

An experimental check of the correctness of the calculated data should consist first of all in determination of the value of the instability factor. If an experimental SAE model succeeded in producing the desired value of K_c with the corresponding α_H, the remaining parameters of the JDS would differ little from the calculated values.

II. JDS Designs

1. Two-Coordinate Sensor in the Tracking System of a Photoelectric Spectrometer

This version of the sensor has four photosensors and tracks the sun in two mutually perpendicular directions (Figure 6). The main design element of the sensor is a cylinder with a wall of I cross section. In four places, by cutting away portions of the shelves, the form of the cross section is reduced to a z shape. /52
The sensor consists of a set of such cylinders with different diameters but identical length. The diameters are selected so that the cylinders fit inside each other. The coaxial system thus produced has the cross section characteristic for SAE in four places.

A ring is pressed into the photosensor end of the inner cylinder. This ring, together with the opposite edge of the inner cylinder, limits the field of vision of the JDS to 15°. Light passing through the central aperture in the ring strikes a photosensor, the output signal of which can be used to judge whether the sun is present. The field of vision of this "presence sensor" is a cone with vertex angle of 14°.

The photosensors used are polycrystalline cadmium sulfide type SF-KO photoresistors.

This version of a JDS is designed for precise aiming of a diffraction spectrometer at the sun. Furthermore, the device includes four coarse aiming sensors with fields of vision of ±60°. The transition from the coarse guidance mode to the fine guidance mode is performed when a signal is received from the "presence sensor."

The control circuit for each coordinate (Figure 7) consists of a four-arm bridge made of silicon transistors PP1, PP2 and load resistors R_1 and R_3. The guidance motors are connected into the diagonals of the bridge.

In order to decrease the harmful influence of scattered light, photoresistors FS_3 and FS_4 of the coarse guidance sensor are connected so as to form a static flip-flop circuit with collector base coupling with transistors PP1 and PP2. As a result, when one of the transistors is open, the other is fully blocked, so that even if the light shining on the photoresistor connected to its input is rather strong, the guidance is not changed.

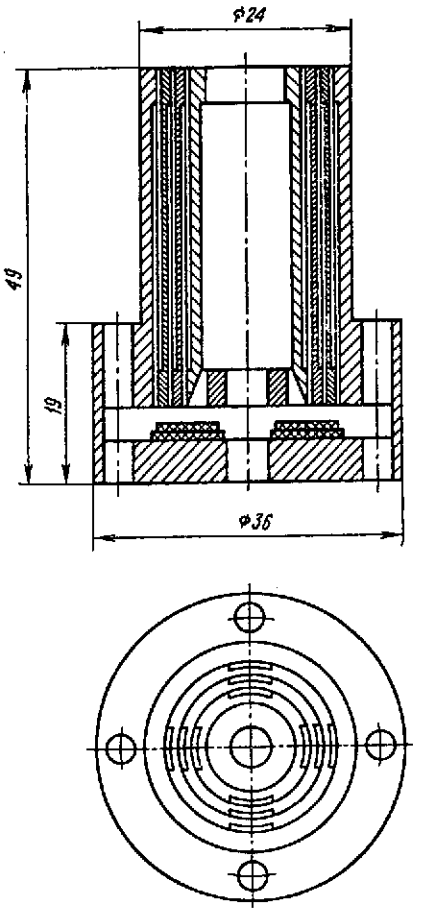


Figure 6. Design of Two-Coordinate Sensor

Photoresistors FS_1 and FS_2, the working photosensors of the JDS, are usually connected between the base of the transistor and the positive pole of the power supply. The transistors are then connected in a common-collector circuit. Initial aiming is performed using signals from the coarse guiding sensors. When the mismatch angle drops to 7°, the sun enters the field of vision of the JDS and the "presence sensor." The mode switching device operates, blocking diodes D_7 and D_6 and disconnecting the coarse guidance photoresistors from the input of the amplifier. Further aiming is performed using the fine guidance sensors. Elements R_2, C, D_1 and D_3 are included in the circuit to improve the characteristics. /53

The system as a whole includes two identical control units for the two axes and an operating mode switching circuit, consisting of a threshold device -- a Schmitt trigger, operated by the signals, and the presence sensor FS_5.

Figure 8 shows: 1) $U = f(\alpha)$ -- the voltage across the motor as a function of the mismatch angle and 2) $\omega = \psi(U)$ -- the rate of motion of the system as a function of the voltage across the motor.

2. One-Coordinate Sun Direction Sensor

The design of a second version of the sensor is based on a rectangular plate with slots in its central portion, giving it a z-shaped cross section (in contrast to the cylinder in the first version).

The JDS includes two sets of identical plates (Figure 9). Each of the sets is assembled as an individual unit by means of four bolts 8, then the two units are attached by bolts 2 to the

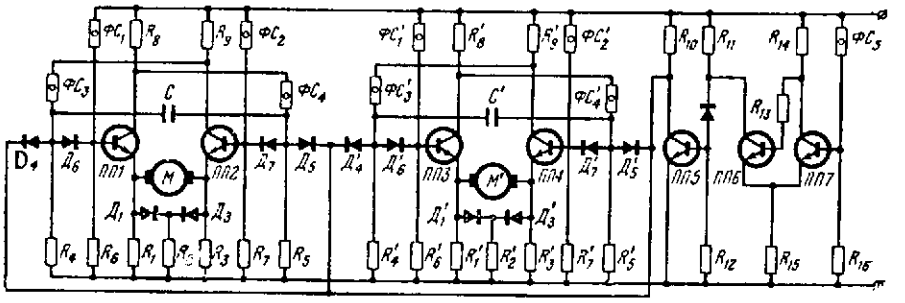


Figure 7. Control Circuit. Key: ФС = PS = Photoresistor; Д = D = Diode; ПП = PP = Transistor; R and C = elements included in the circuit to improve the characteristics

seats on part 3. The difference in the height of the upper and lower seats determines the angle between sets 4 and thereby determines the initial light flux level. Part 3 is fastened to body 1; both sets of plates can be illuminated by the sun through windows 10 in the body. The lower portion of the body, part 9, contains silicon photosensors 11. These sensors, with their high sensitivity, differ from other elements with photoconductive effect in their good stability of parameters [1]. Protective glasses 6 are fastened in front of the photoreceptors in slots cut in parts 3 and 5. The JDS is connected electrically through plug 7.

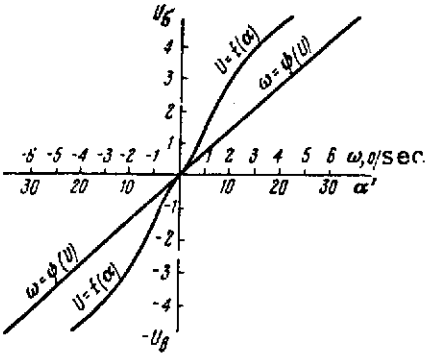


Figure 8. Characteristics of Two-Coordinate System

The values of the basic parameters are presented in Table 2.

In order to measure the characteristics, the JDS is fastened to a special rotating stage. The stage can be rotated at any angle within limits of ±3.5° relative to the base, fastened to a parallax device[1], by means of a micrometer screw. The parallax device is equipped with a clockwork drive in order to compensate for the motion of the sun /54

[1]A device allowing an instrument attached to it to be rotated around two mutually perpendicular axes, set so that one of the axes of rotation is directed toward the pole. A 15-inch telescope from the Crimean Astrophysical Observatory, Academy of Sciences, USSR, was used.

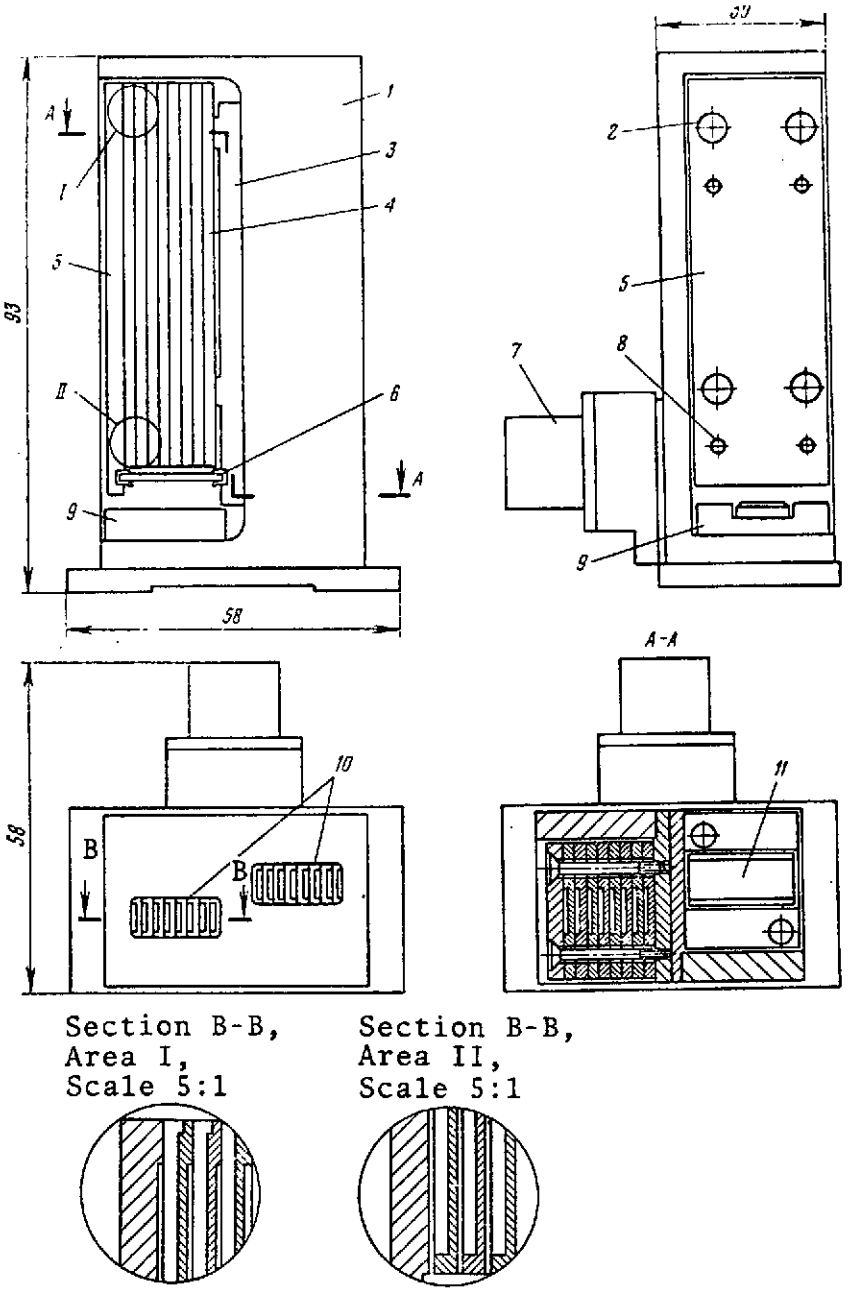


Figure 9. Design of One-Coordinate Sensor

relative to the JDS fastened on the device. Rotation of the stage with the micrometer screw can set the required mismatch /55

angle between the JDS and the direction to the sun. Each scale division of the reading device corresponds to one minute of arc.

TABLE 2

	Symbol	Number
Number of Z-shaped plates in each SAE	n	8
Distance between horizontal shelves	h	60 mm
Length of outer horizontal shelf	l_1	0.4 mm
Length of inner horizontal shelf	l_2	1.15 mm
Width of Z-shaped plates	b	6 mm
Current produced by 1 cm^2 surface of photosensor used in sensor when illuminated by direct sunlight	ηE_0	18 ma/cm^2
Maximum stability factor	K_c	6.5 min
Value of angle α where $K_c = 5.3$		9 min

The sensor was studied for two operating modes of the silicon photosensors: the diode and short circuit modes.

Figure 10 shows the output characteristics of the individual SAE (curves 1; 2) and the entire sensor as a whole (curve 3). The circuits in which the photosensors were connected to measure these characteristics are shown in Figure 11. The resistance of resistor R was 820 ohms.

Curve set 1, 2, 3, 4 (Figure 12) shows the influence of the resistance of resistor R on the steepness of the output characteristic of the JDS, with the resistance at 0, 330, 500 and 820 ohm respectively.

The change in output characteristic as a function of initial level is shown in Figure 13. The characteristic was measured with two values of initial angle: $\alpha_H = 16'$ (curve 1) and $\alpha_H = 5'$. The mismatch of the points of intersection of the curves with the abcissa resulted from inaccuracy in repeated placement of the sensor after the change in the initial angle. It is not difficult to see that curve 1 has a rise steepness almost thrice that of curve 2. The resistance of the resistor in both cases was 1.5 kΩ.

The second group of tests was performed with the photosensors in the short-circuit mode. This mode was achieved by compensating the output voltage of the photosensor to zero by an

additional source of current. The output characteristics of one SAE (curves 2, 3) and of the entire sensor (curve 1) in this mode are given on Figure 14.

Figure 4 presents a graph of the dependence of instability factor K_c on angle α. The minimum value (K_c = 5.2) of this factor is reached where α = 9'.

In order to determine the possible error in the JDS, we must know the changes in characteristics of photosensors used in the JDS under actual operating conditions. The parameters of silicon photoresistors are most strongly influenced by radiation and changes in temperature. Unprotected silicon photoresistors are most damaged by the stream of protons with energies of 0.1-0.5 MeV. Calculations and experiments [3] have shown that protective covering of photosensors with a thickness of 7.6 g/cm^2 protects them almost completely from damage.

As was pointed out above, the accuracy of operation of JDS is determined by the drift of the parameters of photosensors of one SAE relative to the other. There is therefore interest not only in the absolute value of temperature factors for photosensors, but also in the limits with which they change.

In order to explain the possible dispersion of values of temperature factor for the short circuit current and the EMF of the silicon photosensors, the characteristics of 30 elements /56
were studied, from three different production groups. The results of the study showed that the temperature factor of the short circuit current is K_i = (0.1±0.03)%/°C, for the EMF, K_E = (0.53±0.15)%/°C.

Thus, we should give preference to the operation of photosensors in the short circuit mode.

Let us estimate the error produced by a sensor with stability factor K_c = 6.5 min when random photosensors are used in the SAE, for which the temperature coefficients are within various tolerance limits K_i = 0.07%/°C and K_i = 0.13%/°C. As the temperature changes by 75°C, the SAE will produce errors $\Delta\alpha_1$ = 0.345' and $\Delta\alpha_2$ = 0.645'. The drift of the JDS from the initial direction in this case will be $\Delta\alpha$ = 0.3'.

Figure 15 shows the initial sector of the experimental /57
characteristic of a JDS with the area of possible error $\Delta\alpha$ = ±0.3' produced if the parameters of the photosensors diverge during operation by 6%.

Using silicon photosensors selected as to sensitivity and temperature coefficient in SAE which operate together, and also

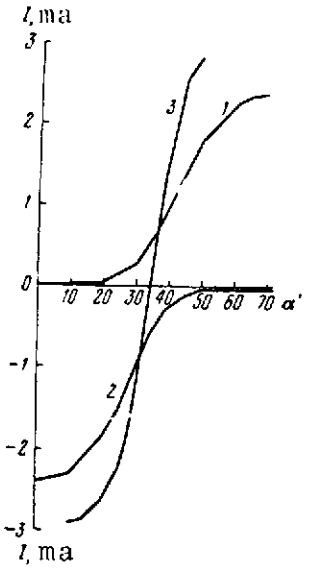


Figure 10. Characteristics of Single-Coordinate Sensor where R_H = 820 ohm

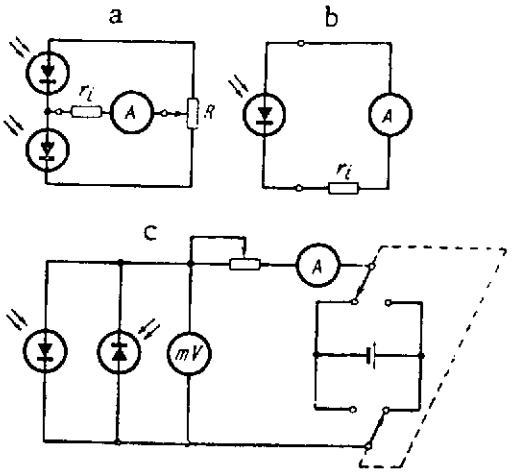


Figure 11. Diagrams of Connection of Photosensors of Sensor. a, b, With Various Load Resistances; c, When Tested in Short-Circuit Mode (R_H = 0)

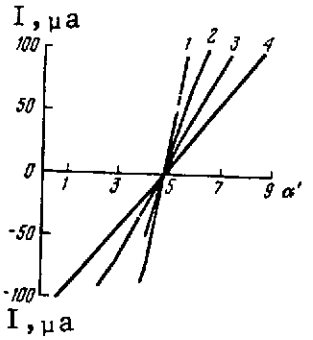


Figure 12. Output Characteristics of JDS Under the Influence of Various Resistances of Resistors

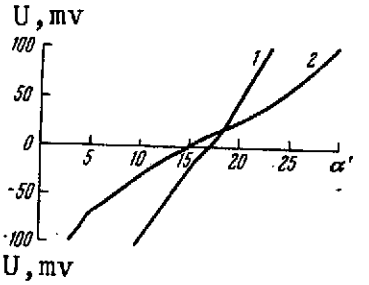


Figure 13. Characteristics of Sensor with Various Values of Angle between SAE

establishing α_H = 9', we can increase the accuracy of this version of the sensor significantly. Further increases in accuracy are possible by more efficient selection of sensor design and the use of vacuum photosensors. We should expect that with careful manufacture, we can achieve a value of K_c = 4.0 min.

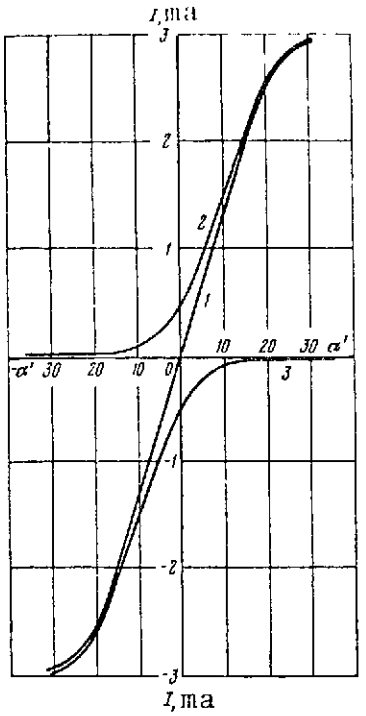


Figure 14. Characteristics of Individual SAE and of Entire Sensor where $R_H = 0$

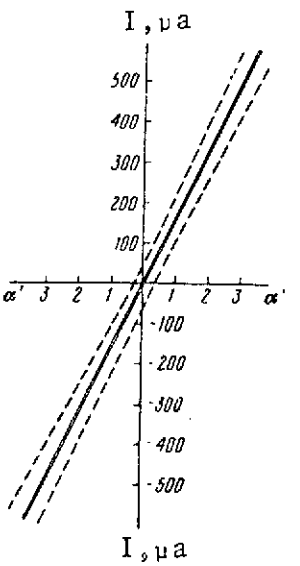


Figure 15. Characteristics of Sensor with Slight Mismatch Angles

Conclusions

1. A multiple-slit shadow sensor for determination of direction toward a shining object is described; the sensor can be used in sun orientation systems.

2. A method is given for design of JDS, allowing the basic dimensions to be determined after fixing: 1) the steepness of the output characteristics, 2) the type of photosensor, 3) the sensor instability factor and 4) the field of vision.

3. The maximum accuracy provided by the sensor is determined basically by two factors: the identity of the parameters of the photosensors and the value of instability factor K_C for a selected working point.

4. An experimental check has confirmed the correctness of the suggested calculation method. The value of the instability factor is determined as $K_C = 5.4$. This means that mutual divergence of parameters of photosensors used in the sensor by 10% leads to an angular error of 0.54'.

5. A description and results of investigation of two versions of a sensor are presented.

REFERENCES /58

1. Soboleva, N. A., A. G. Berkovskiy, N. O. Chechik, R. Ye. Yeliseyev, Fotoelektronnyye pribory [Photoelectronic Devices], Nauka Press, 1965.
2. Koltun, M. M., Golovter, Optika i spektroskopiya, 1966, Vol. 21, No. 5, p. 630.
3. Grigor'yeva, G. M., V. A. Gumennyy, L. B. Kreynin, A. P. Landsman, Kosmicheskiye issledovaniya, 1966, Vol. 4, No. 5, p. 740.

N75 10996

COMPUTERS AND PROSPECTS FOR ON-BOARD SPACECRAFT ELECTRONIC EQUIPMENT

V. V. Andreyanov

Introduction

Although discussions are still sometimes published as to whether preference in the investigation of space should be given to man or machines, the necessity of making a number of on-board electronic devices and systems independent, even in piloted spacecraft, is apparently beyond doubt. Long-range manned flights require that the functions of electronic apparatus be expanded and their characteristics improved which, in turn, require optimal control of the apparatus and individual parameters depending on changing external and internal conditions. The cosmonauts should give their primary attention not to servicing of equipment, but rather to observations and studies.

Thus, the property of self-adaptation of equipment should be developed regardless of whether the equipment is to be used in a piloted or pilotless spacecraft. Obviously, this is true not only of space communications and navigations equipment, but of on-board research and measurement equipment as well.

1. Introduction of On-Board Computers

The first decade of the space age "got by" with some exceptions without on-board computers. In most cases, on-board instruments and devices were controlled directly by radioed instructions from the Earth or by means of time programming devices (TPD) with rigid programs. Much less frequently, self control of devices was used: AGC and AFC in receivers, automatic target tracking in orientation systems, automatic switching of redundant devices. All measurements performed were transmitted to the Earth almost without preliminary selection, then later evaluated as to information content.

The development of space studies and manned flights has required that on-board computers be used for both control of technical equipment and for selection of useful information and control of the radio devices themselves.

The TPD, automatic regulation and switching, were essentially the beginning of introduction of computers to spacecraft. We will not set ourselves the task of discussing whether specialized or universal on-board computers are needed. We will study only how

telemetry and transceiving devices can be changed if digital computers are available.

The practical production of on-board machines became possible /59 due to the successes in microelectronics and the gradual increase in the role of computers in the control of spacecraft. Doubtless, navigational problems requiring complex and accurate calculations are the most important type for computers, but the capabilities of research and communications apparatus are also increased and altered when they operate in combination with computers.

In the current decade in space, the space equipment used will be planned for work in conjunction with computers. We can assume that computers will at first perform a narrow range of important tasks and will not influence the construction of many devices as, for example, in the American Apollo spacecraft, where computers are primarily used for navigational problems, or specialized small spacecraft will be planned for a limited range of studies, entirely designed to work with computers.

The problem of timely processing of the tremendous flow of measurement information produced by complex spacecraft requires careful selection of the measurements performed on board before telemetry to the Earth.

2. Prospects for Changes in On-Board Telemetry Apparatus

In order to reduce the volume of measurements transmitted without loss of information content, computers must control the selection of the most informative measurements from the total number of measurements. The following operations can be used to achieve this goal.

1. Compression of information by the most common method, namely comparison of each subsequent measurement of a parameter with the next preceding measurement, with permission to transmit given only if the comparison operation indicates a change in the measured parameter.

2. Comparison of measured quantities with desired values or combinations of values of parameters recorded in advance in the machine, for example with digital tests or samples. In this case, permission to transmit is given when a measurement result agrees with the desired specimen.

3. Determination (calculation) of the rate of change of parameters in order to control the rate of interrogation of sensors, which can be achieved either by changing the order of

interrogation (a rapidly changing parameter is interrogated more frequently) or by changing the overall interrogation cycle.

4. Determination of the mean of the current dynamic range of a measured parameter in order to increase the accuracy of measurement of parameters with slight relative changes. With a given accuracy of measurement (code word length), this operation can be performed by changing the scale of measurement.

5. Estimation of the importance of measurements (messages), giving priority to transmission of the most significant information.

One criterion of importance might be a change in the rate of change of parameters being measured (operation of emergency sensors, comparison with a desired sample, appearance of rare measurements) or criteria recorded in the machine while still on Earth.

6. Control of the radiated power (and, of course, power consumption) of telemetry transmitters as a function, for example, of the rate of information transmission, as well as control of the parameters of sensors, i.e., adaptation of instruments.

These possible functional relationships between telemetry apparatus, computers and radio transmission channels are illustrated in Figure 1. /60

Here the telemetry apparatus interrogates the sensors, converting their outputs to digital code, providing the required accuracy of measurement of parameters and differentiability (addressing) of each parameter, and is adjusted by the computer.

The computer detects changes in the measurements, compares the measurements with recorded samples, determines the rate of change of the measured parameters and the ranges of change and separates the most essential (important) information.

As a result of these operations, the computer develops telemetry apparatus control signals (controlling the rate and sequence of interrogation, selection of the scale), signals controlling the transmission of measurements into the radio channel (sequence and rate of transmission of current information and recorded information) and signals controlling the transmitter power.

It is assumed in this system that the memory device for comparison operations (operative memory) is in the computer. Since the transmission of information to the radio channel is not continuous, but rather occurs only during sessions of communication with the reception points on the ground, the storage of

information developed using the computer (main memory) is in the memory and grouping device.

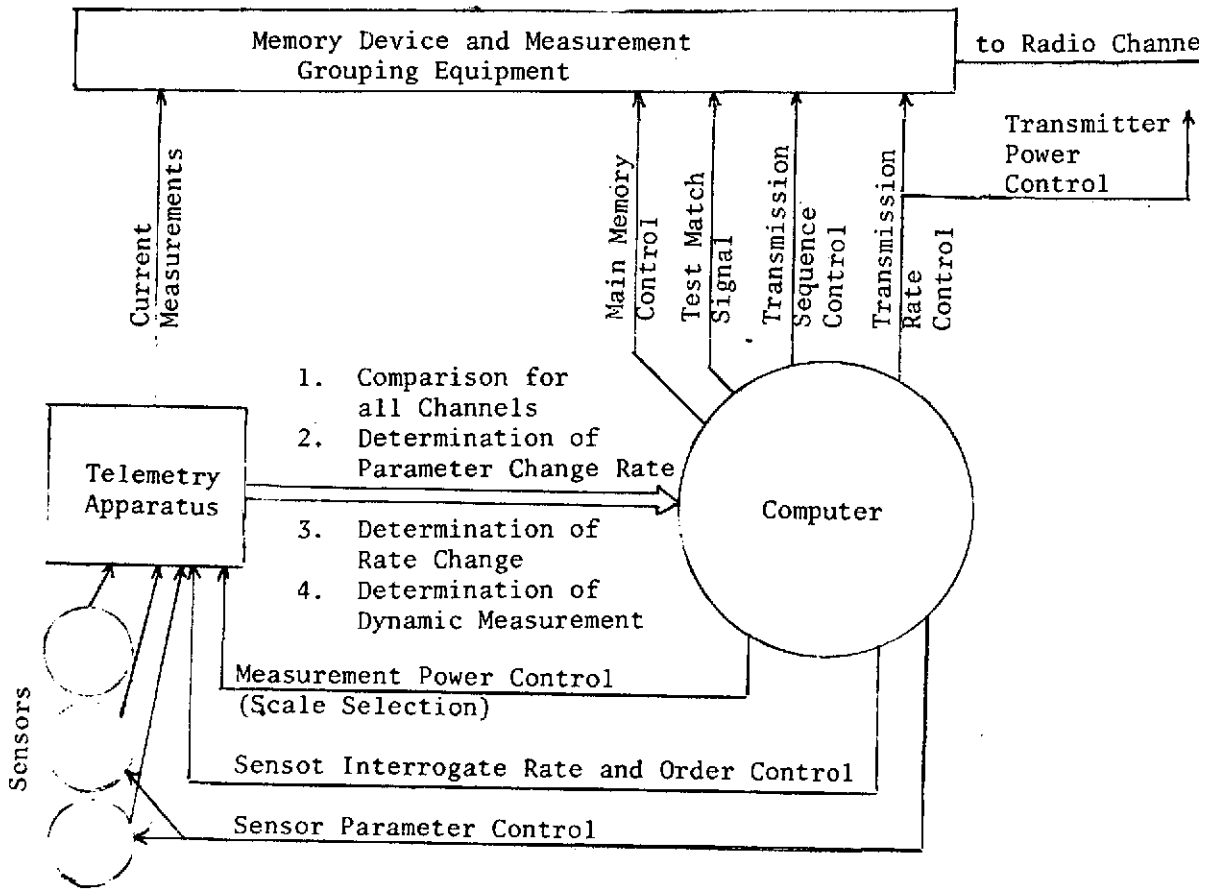


Figure 1. Diagram of Possible Interconnections Between Computer and Telemetry Apparatus

The exchange of information between the telemetry apparatus and the computer is primarily in parallel code.

The memory and grouping device contains the main memory, switching and selecting circuits. The current measurements adapted by the computer are recorded during the interval between communications sessions in the main memory. During transmission, the computer establishes the sequence of transmission of current and recorded information, giving priority to the most significant measurements, changes the rate of transmission of information and the transmitter power. Information on the energy potential of the radio channel may be fed to the computer from an on-board receiver using the same antenna as the transmitter, measuring the level of the signal transmitted from the ground.

3. Prospects for Changes in On-Board Transceivers /61

The presence of an on-board computer makes it possible to improve many of the parameters of transceivers and to control their characteristics, but requires significant changes in the planning of these devices, since it is necessary to produce discrete characteristics in place of the usual, rather smoothly changing characteristics (referring to modulation, demodulation, direction-finding, discrimination, amplitude, phase and other characteristics).

Let us study some of the possibilities for development of transceiver equipment to work together with computers.

1. The movement of the spacecraft relative to the reception point (for example, a ground station) and relative to its own axes makes it impossible to use effective directional spacecraft antennas unless orientation equipment is used.

Until recently, antennas were oriented using star or radio sensors by rotating the antennas or the spacecraft as a whole. It has been tempting to use electronically controlled antenna devices. This has become possible with the development of so-called antenna arrays, where the maximum of reception (or transmission) can be moved in space by the proper phasing of the oscillations fed to the antenna elements making up the array. In the SHF range, it is easiest to produce discrete phase shifters, consisting either of waveguide sections or varactors, or ferrites, the state of which is changed by means of SHF switches (for example, PIN diodes). It is easy to control these phase shifters by computer, allowing the computer to establish the required combination of discrete values of phase shifts and amplitudes in the circuits of the many antenna elements.

Along with electronic control of the radiation pattern of the antenna, this system allows comparatively simple changes in the total radiated power of the transmitter under computer control by disconnecting a portion of the antenna elements (together with their elementary transmitters) and redistributing the power of the remaining transmitters.

2. Another possibility is opened for demodulation of signals in receivers.

Space communications channels primarily use the method of phase modulation (PM), allowing easy multichannel transmission and near-optimal reception with relatively simple on-board transmitters. The only obstacle to the use of PM is the difficulty of linear demodulation of signals with high modulation index, since ordinary phase detectors have a sinusoidal demodulation

characteristic, leading to significant nonlinear distortions with indices of over 0.5 rad, and to ambiguity with indices over 1.5 rad.

On the transmitting end, the attainment of high modulation indices involves no complications, cince linear modulation with a low index can be done on a frequency lower than the carrier frequency, after which teh index is increased by multiplying the frequency. The computer facilitates production of a linear, or more precisely a stepped digital phase detector, capable of demodulating PM signals with indices of up to ±3 rad. For this purpose the phase of the signal being received is continually compared with the phase of a reference waveform, the value of which is computer-varied with a predetermined step, or the computer calculates the number of reference pulses between zeros of the signal and the reference waveform. Essentially, computers and microelectronics facilitate practical construction of on-board correlation detectors for the reception of signals in general, the circuit complexity of which generally results from the necessity of large numbers of multipliers and comparison /62 devices and of controlling correlation time.

If pseudonoise modulation is used, the computer memory contains specimens of the pseudonoise sequences and outputs them for comparison with sequences received.

In the case of frequency modulation of signals, the digital frequency discriminator can hold a stable zero and maintain linearity of the demodulation characteristic.

3. Frequency search and automatic frequency tuning in receivers generally involve difficulties in the achievement of high frequency stability of the tunable oscillators. The combination of high frequency stability and frequency tuning ability can be achieved in frequency synthesizers which, however, are too difficult to control except by computer. Frequency synthesis can be performed either using a rigid program, or based on the results of navigational calculations, or in accordance with the changing frequency of the signal from the receiver.

Thus, computer-based construction of frequency and phase synthesizers allows automatic radio reception without optimal methods and improvement of the results of measurement of the parameters of motion of a spacecraft.

4. For a number of problems (radiometry, direction finding, measurement of the energy potential of a radio line, etc.), it is necessary that the on-board receiver measure precisely the level of a signal received, the change in which is much greater than the dynamic range of the receiver. One good solution is

automatic changing of the "scale" (gain) of the receiver, within which the receiver remains linear.

The computer can measure (count) the level received and change the gain of the receiver according to the results of measurement.

5. In principle, a computer can use the indications of sensors (of electron concentration, weighted frequency, radio signal magnitude, etc.) to determine changes in radio communication conditions. Obviously, this can be used to control the selection of communication frequencies.

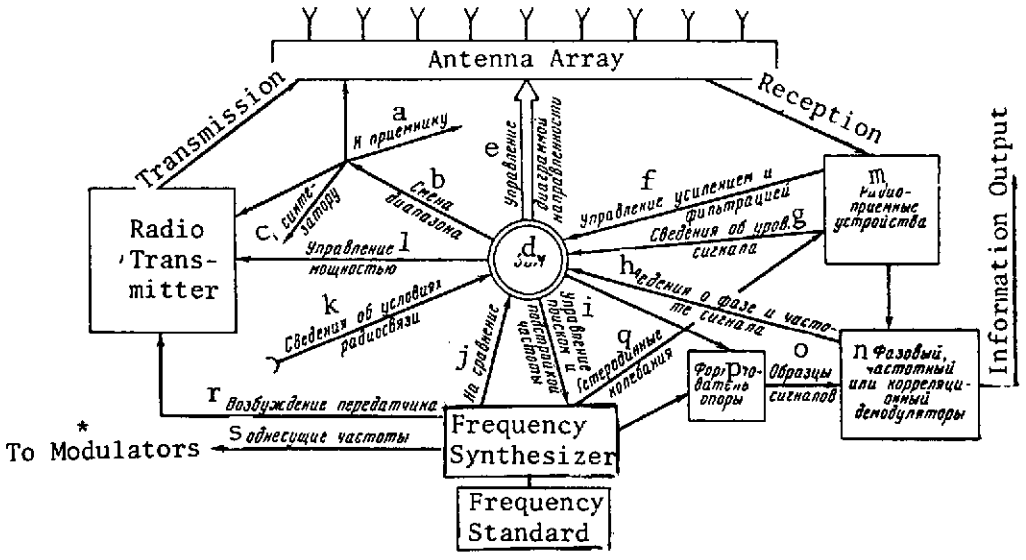


Figure 2. Diagram of Possible Interconnections with Transceiver Apparatus. Key: a, To Receiver; b, Range Change; c, To Synthesizer; d, Computer; e, Radiation Pattern Control; f, Amplifier and Filter Control; g, Information on Signal Level; h, Information on Phase and Frequency of Signals; i, Search and Frequency Tuning Control; j, Comparison; k, Information of Radio Communications Conditions; l, Power Control; m, Receiver; n, Phase, Frequency or Correlation Demodulators; o, Signal Specimens; p, Reference Shapers; q, Beat Frequency Oscillations; r, Transmitter Excitation; w, Subcarrier Frequencies

These capabilities for interaction between transceiver and computers are illustrated by the diagram (Figure 2). /63

The oscillations formed by the frequency synthesizer have a frequency stability determined by the on-board standard. A change in the frequency of the beat-frequency oscillators in the receivers can be called for by the computers according to a program contained in it or can be based on calculated navigational data. In another, more frequent case, the frequency of the beat-frequency oscillators tracks the frequency of the signal using a digital AFC or APC system; the tuning loop is closed through the computer and the frequency synthesizer. The reference shaper for the demodulators is shown separately on Figure 2 for clarity, although in practice the signal samples for demodulation and decoding may be formed by the computer itself. The remaining connections are clear from the diagram.

4. Problems of Practical Realization

The combination of a computer and telemetry apparatus has great potential effectiveness, but requires the joint efforts of specialists in telemetry design, computer engineering, programming and spacecraft design. The nature of the physical measurements, sensors and assumed nature of change in the phenomena studied must also be known.

In our opinion, the primary difficulty lies in the necessity of combined development of equipment. Probably the problem of selection of economical but sufficiently interference-resistant address synchronizers for telemetry signals remains as before.

The combination of a digital computer with a receiver also presents technical difficulties. Long-range spacecraft receivers generally operate with weak signals (both in comparison with the noise present and in absolute terms); it is therefore difficult to create analog-digital converters. Furthermore, a highspeed computer is necessary, particularly for demodulation of wideband signals.

Determination of the range of "duties" of the computer and achievement of self-testing of the computer in order to increase its operational reliability represent independent problems.

Thus, we can state that the on-board computer will first select the most informative measurements, control the operation of an antenna array and synthesize frequencies, changing the operating mode and characteristics of apparatus on the basis of analysis of surrounding conditions and the status of the devices themselves.

REFERENCES

1. Andrews, Davis and Schwartz, "Adaptive Data Compression," Trudy Instituta po elektrotekhnike i radioelektronike [Works of Institute for Electrical Engineering and Radio Electronics] (Russian Translation), Vol. 55, No. 3, 1967.
2. Allen, Prospects for Development of Phased Antenna Arrays, Zarubezhnaya radioelektronika, 1966, No. 7.
3. Draper, Rigly, et al., Navigatsiya, navedeniye i stabilizatsiya v kosmose [Navigation, Guidance and Stabilization in Space], Mashinostroyeniye Press, 1970.

N75 10997

THE EFFECTIVENESS OF CORRECTING CODES IN RECEPTION IN THE WHOLE IN ADDITIVE NORMAL WHITE NOISE

Yu. M. Shtar'kov

Introduction

The effectiveness of correcting codes can be estimated using various criteria. In [1], a number of possible criteria are listed and the energy effectiveness of correcting codes is studied for the case of symbol-by-symbol reception. In this case, effectiveness is taken to mean the ratio $J(p) = E_0(p)/E_1(p)$, where p is the permissible probability of distortion of an individual message, $E_0(p)$ and $E_1(p)$ are the energies which must be expended for transmission of one message to assure probability p with nonredundant transmission and with the correcting code in question respectively. /64

With symbol-by-symbol reception, the receiver contains two decision circuits. The first estimates the value of individual symbols, the second (decoding circuit) establishes the correspondence between the sequence of symbols received and certain code combinations.

The use of two decision circuits simplifies the receiver by reducing interference stability. Therefore, there is significant interest in analysis of the possibilities for a reception method using one decision circuit. The combination of signals corresponding to any code combination in this case is looked upon as one complex signal. Reception of such a complex signal, reflecting an entire code combination, has come to be called reception in the whole.

In the following, we produce expressions for the energetic effectiveness of binary correcting codes for reception in the whole. All conclusions are correct for the case when the signals used to transmit the 0 and 1 symbols differ only in sign (as occurs, for example, with phase telegraphy). However, the result can be extended without significant error to certain other cases, for example the case of coherent and incoherent frequency telegraphy.

1. Asymptotic Energetic Effectiveness where $p \to 0$

Let us represent the code combinations of the correcting (n, k) code used as

$$A_i = (a_{i1}, a_{i2}, \ldots, a_{in}), \quad i = 1, 2, \ldots, 2^k = M, \tag{1}$$

where the symbols a_{il} take on values of 0 or 1. Furthermore, let us list the main properties of the signal and noise to be used subsequently.

For simplicity, we will consider that reception of the first symbol of the code combination is completed at moment in time t = T, of the second signal -- t = 2T, ..., lth -- at moment t = lT, where T is the time required to transmit one symbol. Furthermore, we will assume that the signals s(t) and -s(t) are used to transmit the symbols 0 and 1 respectively. Then signal $A_i(t)$, corresponding to an arbitrary code combination A_i, can be represented as

$$A_i(t) = \sum_{l=1}^{n} (2a_{il} - 1) s(t - lT), \quad i = 1, 2, \ldots, M. \tag{2}$$

We will consider that the elementary signals used at various moments in time are orthogonal, that the energy of each signal is equal to Q, i.e., /65

$$\int_{-\infty}^{\infty} s(t - l_1T) s(t - l_2T)\, dt = \begin{cases} Q, & \text{if} \quad l_1 = l_2, \\ 0, & \text{if} \quad l_1 \neq l_2. \end{cases} \tag{3}$$

Let us assume that useful signal $A_i(t)$ is received against a background of additive normal white noise n(t), so that the summary process y(t) is equal to

$$y(t) = A_i(t) + n(t). \tag{4}$$

In the following we will study the normal random quantities related to n(t)

$$n_l = \int_{-\infty}^{\infty} s(t - lT) n(t)\, dt, \quad l = 1, 2, \ldots, n, \tag{5}$$

for which, as we can easily see, the following relationship is correct:

$$\overline{n_l n_p} = \begin{cases} N_0 Q, & \text{if} \quad l = p, \\ 0, & \text{if} \quad l \neq p \end{cases} \tag{6}$$

(the overline represents averaging of the set, N_0 represents the spectral intensity of the noise, i.e., $n(t)n(\tau) = N_0\delta(t - \tau)$ and where $\delta(x)$ is the delta function). In other words, all values n_l are noncorrelated, the dispersion of each of them is N_0Q, and the mean value, as follows from expression (5), is equal to 0: $\overline{n}_l = 0$.

The task of statistical reception is, by analysis of the process $y(t)$, to determine transmitted signal $A_i(t)$. The optimal procedure consists in the formation of $M = 2^k$ quantities

$$L_j = \int_{-\infty}^{\infty} y(t) A_j(t)\, dt, \quad j = 1, 2, \ldots, M \tag{7}$$

and selection of the maximum quantity [2]. If $L_{j*} = \max_j L_j$, the decision is made that code combination A_{j*} has been transmitted. Thus, in receiving code combination A_i, a false decision will be made only if $L_i \neq L_{j*}$, or, in other words, when the maximum is not L_i.

Obviously, the probability P_e of a false decision satisfies the inequality

$$\max_{j \neq i} P(L_i < L_j) \leqslant P_e \leqslant 1 - \prod_{j \neq i} [1 - P(L_i < L_j)] \approx \sum_{j \neq i} P(L_i < L_j), \tag{8}$$

where $P(L_i < L_j)$ is the probability that L_i will be less than L_j (left portion of inequality corresponds to case of full correlation of all $L_j (j \neq i)$, right portion -- to absence of correlation between L_j). Thus, to produce estimates of the probability of error P_e, we must estimate the probability $P(L_i < L_j)$. According to expressions (4) and (7)

$$L_j = \int_{-\infty}^{\infty} \left[\sum_{l=1}^{n} (2a_{il} - 1)\, s(t - lT)\right]\left[\sum_{p=1}^{n} (2a_{jp} - 1)\, s(t - pT)\right] dt +$$
$$+ \int_{-\infty}^{\infty} \left[\sum_{p=1}^{n} (2a_{jp} - 1)\, s(t - pT)\right] n(t)\, dt. \qquad (9)$$

Representing the first term in the right portion of (9) as R_{ij} and using expression (3), we produce /66

$$R_{ij} = \sum_{l=1}^{n} (2a_{il} - 1)(2a_{jl} - 1)\, Q = (n - 2d_{ij})\, Q, \qquad (10)$$

where d_{ij} is the Hamming distance between code combinations A_i and A_j. Now using formulas (5) and (10), let us convert expression (9) to

$$L_j = (n - 2d_{ij})\, Q + \sum_{l=1}^{n} (2a_{jl} - 1)\, n_l.$$

Then, since $d_{ii} = 0$ and $R_{ii} = nQ$,

$$L_i - L_j = 2d_{ij}Q + 2\sum_{l=1}^{n} (a_{il} - a_{jl})\, n_l.$$

According to the definition of the Hamming distance, $a_{il} \neq a_{jl}$ for d_{ij} values of 1, and therefore,

$$L_i - L_j = 2d_{ij}Q + 2\sum_{s=1}^{d_{ij}} n_s', \qquad (11)$$

where $n_s' = (a_{il} - a_{jl})n_l$ where $a_{il} \neq a_{jl}$ is equal either to n_l or to $-n_l$. Obviously, $\overline{n_s'} = 0$, $\overline{n_s'^2} = N_0Q$, $\overline{n_{s1}'n_{s2}'} = 0$ where $s_1 \neq s_2$.

It follows from (11) that the normal random quantity $\Delta L_{ij} = L_i - L_j$ has a mean value of $2d_{ij}Q$, and a dispersion equal to $4d_{ij}N_0Q$. Therefore

$$p(\Delta L_{ij}) = \frac{1}{\sqrt{2\pi 4 d_{ij} N_0 Q}} \exp\left\{-\frac{(\Delta L_{ij} - 2d_{ij}Q)^2}{8 d_{ij} N_0 Q}\right\},$$

$$P(L_i < L_j) = P(\Delta L_{ij} < 0) = 1 - F\left(\frac{2d_{ij}Q}{\sqrt{4 d_{ij} N_0 Q}}\right) = 1 - F\left(\sqrt{\frac{d_{ij}Q}{N_0}}\right), \tag{12}$$

where

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{t^2}{2}\right) dt.$$

As follows from (12), $P(L_i < L_j)$ depends only on the signal/noise ratio Q/N_0 and d_{ij}, where the maximum $P(L_i < L_j)$ corresponds to the minimum of d_{ij}, which is usually called the minimum code distance of the code and represented by d. Considering this, we can rewrite (8) as

$$1 - E\left(\sqrt{\frac{dQ}{N_0}}\right) \leqslant P_e \leqslant \sum_{z=d}^{n} \alpha_z \left[1 - F\left(\sqrt{\frac{zQ}{N_0}}\right)\right], \tag{13}$$

where α_z is the number of code combinations A_j for which $d_{ij} = z$.

With large values of x

$$1 - E(x) \approx \frac{1}{\sqrt{2\pi}\, x} \exp\left\{-\frac{x^2}{2}\right\}.$$

Therefore where $Q/N_0 \gg 1$, we find that $1 - F\left(\sqrt{\frac{z_1 Q}{N_0}}\right) \gg 1 - F\left(\sqrt{\frac{z_2 Q}{N_0}}\right)$, /67

if $z_1 < z_2$ and (13) can be converted to

$$1-F\left(\sqrt{\frac{dQ}{N_0}}\right)\leqslant P_e\leqslant\alpha_d\left[1-F\left(\sqrt{\frac{dQ}{N_0}}\right)\right]$$

or

$$P_e=\alpha_d'\left[1-F\left(\sqrt{\frac{dQ}{N_0}}\right),\quad 1\leqslant\alpha_d'\leqslant\alpha_d.\right. \tag{14}$$

Upon transmission of k information symbols without redundancy $d = 1$, $\alpha_d = \alpha_1 = k$, and energy Q_0 is expanded on the transmission of one symbol. Substituting these values into (14), we produce for nonredundant transmission

$$P_e=\alpha_1'\left[1-F\left(\sqrt{\frac{Q_0}{N_0}}\right)\right].$$

Since it is assumed that in both cases identical transmission probability is provided,

$$\alpha_1'\left[1-F\left(\sqrt{\frac{Q_0}{N_0}}\right)\right]=\alpha_d'\left[1-F\left(\sqrt{\frac{dQ}{N_0}}\right)\right]. \tag{15}$$

With large values of Q/N_0, we can ignore the difference in coefficients α_1' and α_d', and relationship (15) will obtain only if

$$\frac{Q_0}{N_0}=\frac{dQ}{N_0}. \tag{16}$$

If E_0 is the energy expended in the transmission of k information symbols without redundancy, then $Q_0 = E_0/k$. Similarly, $Q = E_1/n$, and from (16) we produce

$$\frac{E_0}{kN_0}=\frac{dE_1}{nN_0}$$

or

$$\frac{E_0}{E_1}=\frac{kd}{n}.$$

Since E_0 and E_1 correspond to the same value of probability of error P_e, where $P_e \ll 1$, this last relationship can be rewritten as

$$J^* = \lim_{p \to 0} J(p) = \lim_{p \to 0} \frac{E_0(p)}{E_1(p)} = \frac{kd}{n}. \tag{17}$$

It was determined in [1] that with symbol-by-symbol reception of a correcting code $J^* = k(t_0 + 1)/n = k(d + 1)/2n$, while when an interval of erasure was introduced, J^* increased by 1.376 times.

Thus, the asymptotic energetic effectiveness of a correcting code for reception in additive normal white noise is double that of symbol-by-symbol reception. Furthermore, as follows from [1], with a fixed correcting code length n, the maximum asymptotic effectiveness is achieved where $k \approx n/2$. In other words, the use of sequences of maximum length, corresponding to orthogonal signals, is not optimal. The best result is produced when strongly correlated signals are used (d significantly less than n/2). The reduction in quality of reception due to the lower "distance between signals" in this case has less influence than the increase in the number of information symbols k and correspondingly in transmission rate k/n.

2. Reception in the Whole with Finite Signal/Noise Ratios /68

The energetic effectiveness of a correcting code with reception in the whole, defined in the preceding section, corresponds to the case when $p \to 0$ or, in other words, $Q/N_0 \to \infty$. In spite of the fact that the estimate produced in this case is expressed quite simply through the basic parameters of the code, in practice it is always necessary to deal with not just finite, but really quite small ratios Q/N_0. Therefore, it is necessary to at least roughly estimate the behavior of J(p) with finite values of p.

As was noted in [1], with symbol-by-symbol reception of a correcting code and subsequent decoding, the energetic effectiveness J(p) differs significantly from J^* with finite values of p. Thus, where n = 127 and $p = 10^{-10}$, J(p) is less than half J^*. Furthermore, as the code length increases, this relative difference between J(p) and J^* increases with the same values of p. This has been confirmed by calculations.

We can note two formal reasons for this. First of all, ignoring coefficient $C_n^{t_0+1}$, which is possible as $p \to 0$, is impossible with finite p, since it leads to significant errors. Secondly, factor $(Q/N_0)^{-t_0/2}$ is not considered, which is also permissible only where $p \to 0$. Numerical calculation, considering all factors, indicates that these reasons cause a decrease in J(p) as p increases.

In producing relationship (17), only two assumptions were made. The first was that all terms except the first are discarded in the right portion of inequality (13). However, even with sufficiently small Q/N_0, the error introduced is not great (a similar result obtains with symbol-by-symbol reception). The second simplification was that the difference between α_1' and α_d' was not considered in expression (15). This simplification is equivalent to ignoring coefficient $C_n^{t_0+1}$ in symbol-by-symbol reception. However, it is significant that no other assumptions are made.

These discussions allow us to hope that with finite p, the difference between J(p) and J* will be less for reception in the whole than for symbol-by-symbol reception. In many cases, this is almost obvious. For example, in reception of a correcting (32,6)-code in the whole with d = 16 (corresponding to the use of orthogonal signals), $J(10^{-3}) \approx 0.95J^*$. At the same time, with symbol-by-symbol reception, $J(10^{-3}) \approx 0.56J^*$.

However, this means that if where $p \to 0$ the transition from symbol-by-symbol reception to reception in the whole corresponds to a double energy gain, with finite values of p for many codes the energetic effectiveness of reception in the whole increases by more than double, and in many cases quite significantly. This important factor is an additional stimulus for the use of reception in the whole.

Of course, the question arises of the technical realization of reception in the whole. Naturally, with sufficiently large k, the use of 2^k channels of optimal reception is simply impossible, while with smaller k we cannot expect a great energy gain. Therefore, we should orient ourselves toward simplification procedures, equivalent to some extent to reception in the whole. One interesting approximation is related to the use of systems of checks as applied in majority decoding. This system of checks is not difficult to use not only when there is an erasure zone (zero zone), but also a large number of zones of reception [3]. Significant results on the construction of easily realizable

algorithms rather close to reception in the whole were produced by L. F. Borodin [4]. Therefore, we can hope that a realization of reception sufficiently close to reception in the whole will not represent overwhelming difficulties for the near future.

REFERENCES

/69

1. Babkin, V. F., Yu. M. Shtar'kov, The Energetic Effectiveness of Correcting Codes in Certain Communication Channels, Peredacha tsifrovoy informatsii po kanalam s pamyat'yu [Transmission of Digital Information over Channels with Memory], Nauka Press, 1970, pp 52-60.
2. Fink, L. M., Teoriya peredachi diskretnykh soobshcheniy [Theory of Transmission of Discrete Messages], Moscow, Sov. Radio Press, 1964.
3. Kazanov, A. A., One Method of Increasing the Interference Stability of Reception of Binary Signals with Redundancy, Elektrosvyaz', 1968, No. 3, pp. 69-75.
4. Borodin, L. F., Vvedeniye v teoriyu pomekhoustoychivogo kodirovaniya [Introduction to the Theory of Interference-Stable Coding], Sov. Radio Press, 1968.

N75 10998

SELECTION OF AN OPTIMAL PLAN FOR CONSTRUCTION OF A SPACECRAFT ANTENNA SYSTEM

B. A. Prigoda, M. B. Faynshteyn,
V. A. Gur'yev and A. S. Markin

This article presents an attempt to summarize the experience of the design of antennas for various types of spacecraft: fully oriented, partially oriented and with arbitrary orientation. Several examples of the construction of antenna systems are presented and their characteristics are noted.

On the flight trajectory of an interplanetary spacecraft, its on-board radio systems must provide high-quality exchange of information over the communications line between the spacecraft and the Earth.

The antenna-feeder system (AFS) of a spacecraft should be designed so that its basic characteristics, defining the potential of the communications line, are optimal.

In this aspect, the basic characteristics include the effective power radiated by the spacecraft radio system $P_{ef} = P_b \cdot G$ as an energy indicator, for example for a spacecraft-Earth link, providing the required density of energy flux and the required (P_s/P_n) ratio at the input of the receiver of the communication channel, the number of antennas (n_a) as a design figure being in turn a function of P_{ef}, the method of orientation of the spacecraft during communications sessions, the change in sighting angle to the Earth and the fixed reliability of the AFS system. Here G is the gain, P_s is the signal power, P_n is the noise power.

The optimality of these parameters must be evaluated on the basis of the minimum weight, design, power and other "expenditures" for the creation of the AFS system, providing conditions for its normal operation in relationship to expenditures for the creation of the entire spacecraft.

The mathematical expression of this ratio is a complex functional dependence. It is constructed from the relationship of "weights" of components participating in the "expenditures", made up for each specific task according to their individual significance. The "expenditures" for the creation of reserve (emergency) AFS systems, operating with reduced information load in comparison to the main system, may be considered or ignored.

Interplanetary spacecraft can be oriented in space by various methods. The methods depend on the purpose of the spacecraft and its power reserves, both as concerns power supply for on-board systems, and for the exchange of information over various radio channels. In order to provide constant, reliable communications, it is obviously best to have constant orientation of the object in space, with the antenna pointed at the Earth. This requires constant expenditure of working fluid (compressed gas), used by the control system to compensate for various destabilizing forces. /70

Therefore, in certain trajectory sectors, it is sometimes impossible to operate in the mode of partial spatial orientation of the object. This is possible if there is no need for high information capacity of the communications line or when the volume of information transmitted can be transmitted in a limited period of time.

An example of this type of mode is the mode of operation with monaxial spatial orientation of the object or with the spacecraft rotating about its main moment of inertia.

When the spacecraft carries solar batteries, it is oriented with the plane of the solar panels perpendicular to the direction toward the sun and the spacecraft may rotate about a line near this direction.

If the object spends a long time on its trajectory, obviously the position of the axes of the object in space relative to the direction toward the Earth will change according to a given rule and within definite limits. To compensate for this and provide constant and stable communications between the object and the Earth, the directionality characteristics of the on-board antennas must be selected so as to retain the gain of the communications link between spacecraft and Earth above the minimum permissible value within the limits of change of sighting angles from the spacecraft to Earth:

$$G \geq G_{min} \qquad (2)$$

The orbit of an interplanetary spacecraft may be in the form of an ellipse, parabola or hyperbola. Flight over the sector between the Earth and the planet on a semielliptical trajectory tangent to the orbits of the Earth and the planet is considered most economical in the sense of power expenditure.

In various sectors, the trajectories may be sections of ellipses, hyperbolas and parabolas. Since the planets move in definite orbits, in order to calculate the date when a spacecraft

must be launched toward a given planet for a given trajectory, we must consider the mutual influence of the Earth and the planet. If the trajectory is semielliptical, considering the orbit of the planet circular, we can use the following relationship to determine the difference in heliocentric longitudes $l_1 - l_2$ at the moment of launch [1]:

$$\pm(l_1 - l_2) = 180° - n(t_1 - t_0), \qquad (2)$$

where l_1 is the longitude of the outer planet, l_2 is the longitude of the inner planet; n is the mean daily motion of the planet; t_0 and t_1 are the times of flight and descent of the spacecraft.

When flying from an inner planet to an outer planet (for example from the Earth to Mars), the sign used in this formula is reversed. This condition is fulfilled for a given trajectory once in each synodic period of the planet.

Depending on the launch date, the flight trajectory and certain additional conditions, the mutual placement of the spacecraft, planet, Earth and sun will change in certain ways. This change is usually characterized by the angle formed from the center of mass of the spacecraft by the directions to the sun and the Earth, the so-called sun-object-Earth angle or the SOE angle -- (α).

Analysis of the mechanics of flight of an interplanetary spacecraft over a selected trajectory has shown that the SOE angle changes according to a definite rule. The values of these angles will differ for flight to a planet located either inward or outward in relationship to the Earth's orbit. Figure 1 shows that for a flight to Venus, angle α changes over broader limits than for a flight to Mars.

Let us accept for any arbitrary trajectory the curve of changes of SOE angles as a function of flight time as shown on Figure 2. It follows from this figure that the sighting angle to /71 the Earth will shift within angles α_{min} and α_{max}, depending on time. Due to various destabilizing factors, the spatial orientation of the internal axes of the object may differ somewhat from the curve presented. Therefore, the actual curve of the change in sighting angles to the Earth relative to the selected axis on the object (for example, the axis of the antenna system) will form a broad band, pulsating with time (Figure 3).

Considering the great range of change of distances and the requirement to assure a certain information-handling capacity for a spacecraft-Earth communications link, the basic factor used to

define optimality of an antenna system must be considered provision of time-variable (following a certain rule) gain factor G_{min}.

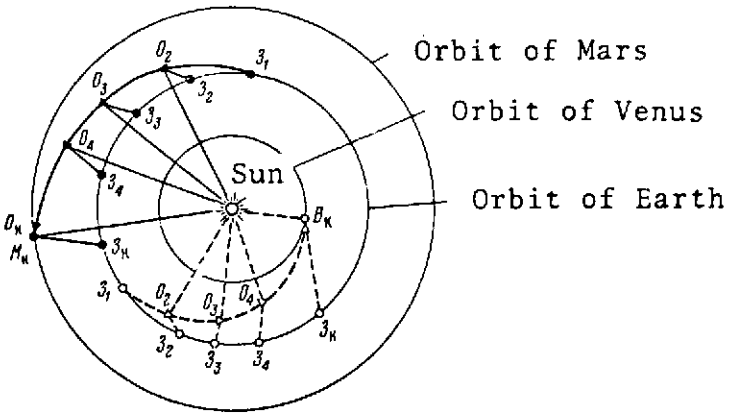


Figure 1. Change in SOE Angles on Flight Trajectory to Venus and to Mars

Considering the constant change in SOE angles, it must be considered that optimal system will provide a variable radiation pattern, assuring the required gain in the direction of the Earth.

If we assume that the interplanetary spacecraft is constructed with monaxial orientation in space, obviously, the most convenient form of antenna radiation pattern will be a funnel shape, as shown on Figure 4.

The radiation pattern is selected in order to provide the required sector of change of sighting angles (consdiering all variations) to the Earth, within which the condition $G \geq G_{min}$ must be fulfilled. The initial condition for the initial orientation of the maximum of the radiation pattern is provision of the required G in the direction of the Earth for any moment in time. Ordinarily, the maximum value of G must be provided over the final sector of the spacecraft flight trajectory, or in communications sessions, during which reception and transmission of the maximum volume of information occur.

Thus, if stabilization of the object is performed precisely along the electrical axis of the antenna, corresponding to the direction to the sun, at any moment in time the Earth communications station will be in the zone of reliable reception (or transmission) of information.

In those cases when the sector of change of SOE angles exceeds the maximum permissible aperture of the beam, where condition (1) is observed, several antennas must be installed on board with different funnel-shaped radiation patterns, intersecting at a level such that $G \geq G_{min}$ (Figure 5).

When the sighting angle to the ground station changes from sector I to sectors III, IV, II, etc., the high frequency section of the AFS must be switched to the proper on-board antenna number 2, 3, etc.

Radiation patterns such as those illustrated on Figures 4 and 5 are most conveniently produced using antennas in the form of multiple log-periodic tapered helices, capable of maintaining constand impedance, polarization characteristics and radiation pattern /7 over a broad range of frequencies. By using a set of such helical antennas with various lead angles with otherwise equivalent conditions, the sighting angles to the Earth can be made to overlap at a fixed G level in any range of change of these angles.

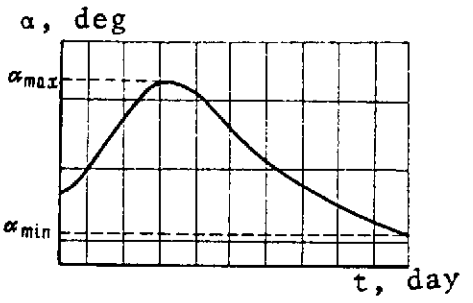


Figure 2. Curve of Change of SOE Angle as a Function of Flight Time

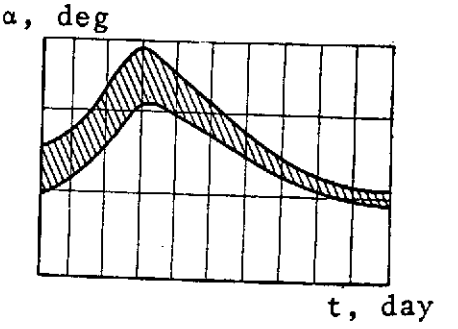


Figure 3. Nature of Change of Sighting Angle of Spacecraft Antenna to Earth Considering Influence of Destabilizing Factors

The primary inconvenience, limiting the possibility of using these antennas, is the fact that the width of the radiation pattern at a fixed level changes little with changing helix parameters and is on the order of 25-45° (at level $P_{0.5}$). The gain factor at the maximum of the pattern is not over 3 to 8. Studies of models of helical antennas of conical type have shown the following relationship is correct:

$$G_{max} = k/\theta_1 \cdot \theta_2,$$

where coefficient k may vary from $3.8 \cdot 10^4$ to $5 \cdot 10^4$, depending on the design of the antenna, and θ_1 and θ_2 correspond to the widths of the pattern in two perpendicular planes.

In order to allow operation in productive communication modes, for example for transmission of a television image to Earth from the surface of a planet or from an orbit around a

planet, highly directional antennas (DA) must be used, with gain factors on the order of some tens, hundreds or thousands.

Obviously, if an interplanetary spacecraft is spinning about its axis, operation with DA is difficult, since it requires a complex antenna aiming system to compensate for the rotation of the object and all destabilizing factors related to orientation inaccuracy.

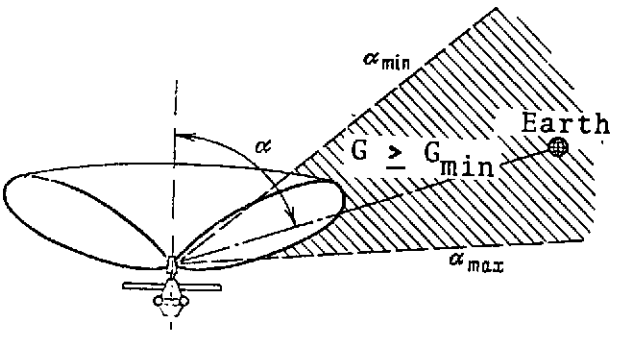


Figure 4. Funnel-Shaped Radiation Pattern

Therefore, high-capacity communications requires a transition from a monaxial orientation mode to the mode of orientation in two or three axes. Obviously, this requires additional expenditures of power. Therefore, over any sector of the trajectory where high rate and volume of transmitted information are not necessary, communications should be conducted with low-directionality antennas and simpler stabilization. /73

For example, communications between the Venera-4, 5, 6 and 7 space probes and Earth stations were conducted in the decimeter waveband using two semidirectional antennas (conical log-periodic helices) and one directional antenna (single-reflector parabolic with 2.3 m paraboloid diameter).

Special work was performed to orient the probe relative to the direction toward the Earth using the sun and Canopus before each communications session.

The level of the signal received from the Venera spacecraft through the directional antenna was several tens of times higher than the level of the signal received from the semidirectional antenna.

To assure operational reliability over the entire flight sector, some of the electronic equipment on board the spacecraft was duplicated.

A typical plan for the antenna-feeder unit to be carried on a spacecraft might be made up as shown on Figure 6. This allows reliable operation in case of failure of any of the transmitters

or receivers, as well as testing of the condition of the antenna device itself and operation during the flight from the Earth to the planet through the orbital-section antenna with the transmitters of the descent apparatus (DA).

In general, the question of the optimal design of a spacecraft antenna system is debatable, and may be answered in various ways for various cases. In particular, the problem can be formulated as follows: what plan of construction of the object and logic of its operation during flight is most energetically suitable and reliable:

1) a flight plan with monaxial permanent solar orientation (PSO) with partial transition to full spatial orientation using a system of switched semidirectional antennas, connected through a complex high frequency circuit;

2) a flight plan with full spatial orientation with the axis of the directional antenna aimed toward the Earth throughout the flight of the spacecraft;

3) a flight plan involving a non-oriented spacecraft using an antenna system providing for omnidirectional transmission (with a fixed gain level).

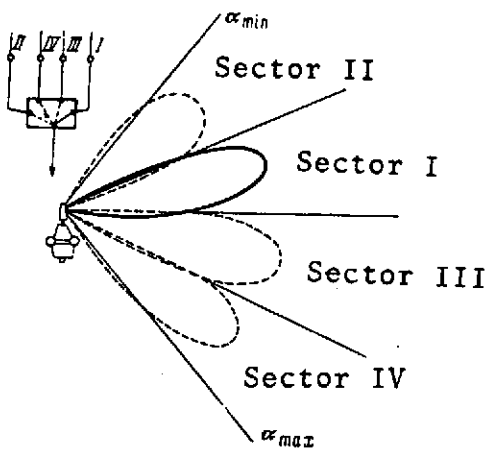


Figure 5. Provision of Required Gain Level Using Several Antennas with Funnel-Shaped Radiation Patterns

In the first case, the power consumption increases due to the increased time of transmission through the semidirectional antenna, the necessity of increasing the power level of on-board transmitters, increasing the consumption of power to stabilize the object, etc. In the second case, the power consumption increases due to the necessity of maintaining constant precise orientation. /74

If constant spatial orientation of the object is maintained, a gain can be achieved by increasing the directionality of the antenna, which is obvious from the following equation

$$P_s/P_n = P_b G_b A_{ef.\ Earth}/4\pi r^2 kT\Delta f,$$

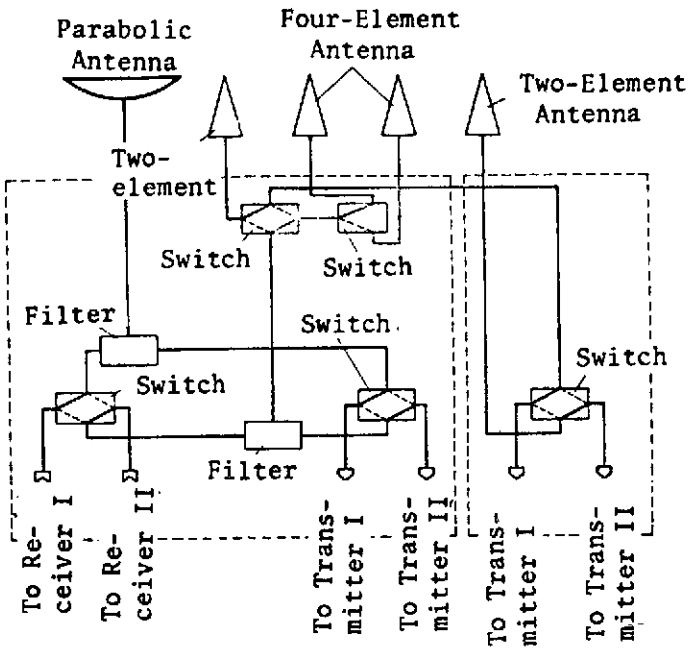


Figure 6. Diagram of Antenna-Feeder System of Spacecraft Including Duplication of On-Board Receiving and Transmitting Equipment

where P_b and G_b are the power and gain of the equipment on the spacecraft, $A_{ef.\ Earth}$, T and Δf are the parameters of the ground station, located at distance r from the spacecraft.

However, it is not desirable to increase the directionality of antennas without limit: first of all, due to the required increases in size and weight and, secondly, due to the limited capabilities of orientation systems. In practice, it is not desirable to use directional antennas with radiation patterns less than 1 degree wide.

When the weight, dimensional and power characteristics of on-board devices, including the AFS, are compared, we find that there is an optimal relationship between them, allowing the most acceptable versions to be selected.

For nonoriented spacecraft (Earth satellites, lunar apparatus) the information capacity is much less than for those studied above. The low gain is compensated by the fact that the distance to the Earth communcations station is much less in these cases. When antenna systems are constructed for these objects, specific difficulties are encountered, related to the provision of omnidirectional radiation patterns. The installation of several antennas operating simultaneously results in interference of their partial fields, creating frequent, deep "holes" in the radiation patterns.

At frequencies where the dimensions of the object are comparable to the wavelength ($l \approx \lambda$), the secondary high frequency fields created by currents flowing over the surface of the object

may strongly distort the fields of the antennas themselves. Various types of decouplers are used to decrease this effect.

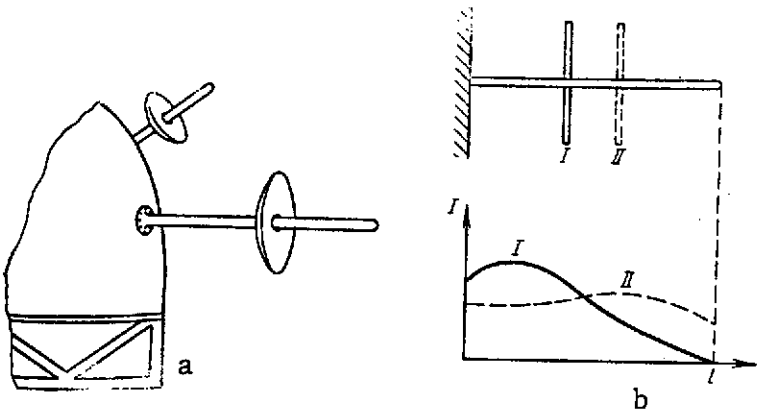


Figure 7. Reactive Discs. a, Installation of Discs on Dipoles; b, Distribution of Current Through Dipole as a Function of Position of Disc (I, II) /75

In many cases, the influence of the body, as a result of its effective excitation, causes the antenna itself to be excited weakly, so that additional devices must be introduced to facilitate more effective and even excitation of the surface of the antenna.

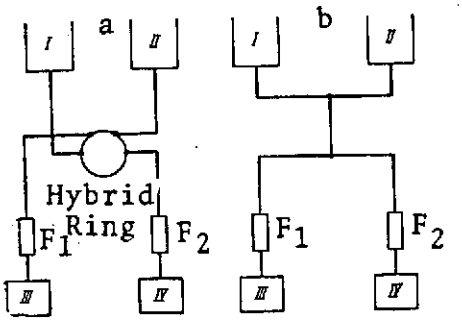
One such device consists of the reactive plates illustrated on Figure 7. As they are moved along the length of the rod, the distribution of amplitude (Figure 7b) and phase of the current along the length of the rod, its Z_{in} and other characteristics change. When there are several rods, these discs can be used to achieve even excitation and mutual phase matching, as well as decoupling from the influence of the body.

For any version of spacecraft, one constant resource available for optimization of the AFS is to increase the efficiency of the high frequency section, which can be achieved by optimal placement of units, reduction of the number of high frequency elements in the AFS system, reduction of high frequency line lengths, as well as selection of optimal designs of antennas with minimum losses in the material and losses due to parasitic radiation outside the useful radiation angles.

The efficiency of AFS sections can be increased also by using hybrid matching circuits (Figure 8a). The advantage of the use of /76
these devices in comparison to ordinary direct connection of the antennas (Figure 8b) is that decoupling is achieved between the

group of antennas III and IV, whereas in the other circuit (Figure 8b), energy may reradiate from section III to IV and back, also distorting the directionality characteristics, decreasing the efficiency of the entire section, etc.

Fig. 8



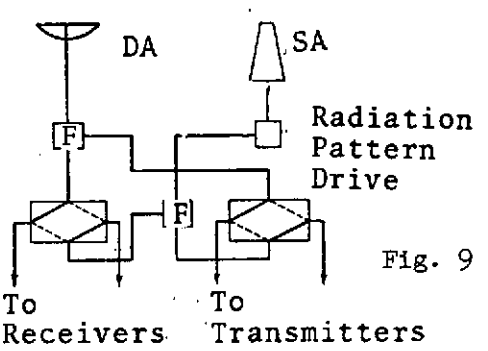


Fig. 9

Figure 8. Connection of Antenna System (I, II) to Units of On-Board Apparatus (III, IV). a, Through Hybrid Ring; b, Direct Connection

Figure 9. Diagram Using Antenna with Adjustable Radiation Pattern. DA, Highly Directional Antenna; SA, Semidirectional Antenna with Adjustable Radiation Pattern; F, Filters

The AFS diagram shown on Figure 6 can be successfully used both for an object with constant orientation of the directional antenna toward the Earth (in this case, the semidirectional antenna is used as an emergency antenna) and in the case of partial orientation, when the semidirectional antenna performs the primary function. The presence of several semidirectional antennas, providing for reliable information exchange by successive connection to the high frequency line over various sectors of the trajectory, suffers from the shortcoming that AFS is generally much more complex and less reliable.

It is expedient to use antennas with adjustable radiation patterns on spacecraft. The development of these antennas allows the number of antennas to be reduced, the feeder system to be simplified, the AFS effeciency to be increased, and the total weight and volume occupied by the spacecraft AFS to be reduced. Thus, the use of antennas with adjustable radiation patterns allows improvement in AFS and increases in effectiveness and reliability.

An example of one such system is shown in Figure 9.

Further optimization of the system can be achieved by increasing the reliability of its individual elements and decreasing the

universality of the individual parts. For example, if we assume that a semidirectional antenna is sufficient for reception of input information by the spacecraft, we can liberate the directional antenna from the function of receiving antenna during communications by means of the directional antenna. This means that the directional antenna is used only for transmission and the system is converted to the form shown on Figure 10, which is simpler both from the standpoint of tuning of the section and from the standpoint of arrangement of AFS elements on board the spacecraft.

Sometimes, circuits similar to those shown on Figure 11 are used. At first glance, this system has the shortcoming that losses will result from radiation from a non-optimal antenna (from the standpoint of orientation toward the Earth) along with operation with the antenna oriented at that moment toward the Earth. However, the system does not have a high frequency switch, which is a quite unreliable element. Systems of this type can expediently be used on spacecraft which will fly over long trajectories and have sufficient power reserve. A more complex system can be produced, consisting of a set of such elementary units, as shown in Figure 11. Each unit contains an antenna with radiation /77
patterns serving a predetermined solid angle of space.

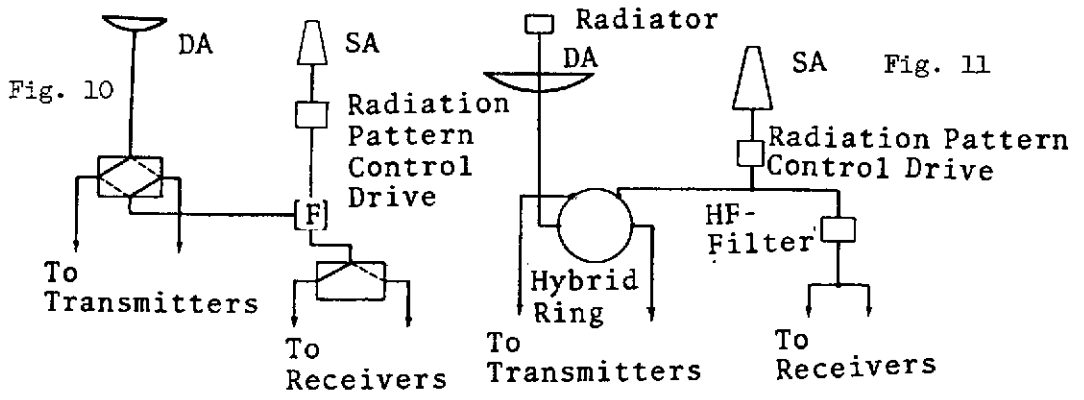


Figure 10. Simplified Version of System with Antenna with Adjustable Radiation Pattern

Figure 11. Antenna-Feeder System without Switch

Due to the difficulty of producing antennas with radiation patterns adjustable over broad limits, the classical plan for construction of an AFS at the present time is a system with high frequency switches. Figure 12 shows an example of such a system for a nonoriented object. As an example, we have taken a portion

of the AFS system of the Luna-15 spacecraft, used during the Earth-moon flight.

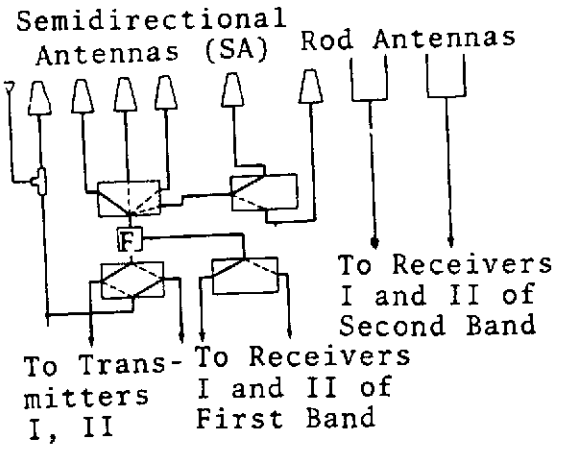


Figure 12. Diagram of Antenna-Feeder System for Nonoriented or Partially (Slightly) Oriented Object

This analysis makes it obvious that the problem of optimal design of AFS systems encounters a number of questions which must be answered during planning of spacecraft:

1) Based on the required value of P_{ef}, we must determine the optimal G for solution of a given specific problem in the light of the relationships with other spacecraft characteristics mentioned above;

2) We must analyze the possibility of separating the functions of reception and transmission of information, studying the power characteristics of the ground communications stations;

3) Analyze the possibility of using two or more frequency bands in spacecraft AFS to separate the functions of reception and transmission of information;

4) In analyzing versions of AFS system design, the optimal version of the AFS providing gain and effective potential P_{ef} which can be varied as time passes (according to a fixed rule).

REFERENCES

1. Astronomicheskiy kalendar' na 1963 god (postoyannaya chast') [Astronomical Calendar for 1963 (Permanent Portion)], Moscow, Nauka Press, 1962, p. 113.
2. Astronomicheskiy kalendar' na 1969 god [Astronomical Calendar for 1969], Moscow, Nauka Press, 1968, p. 219.
3. Space Aeronautics, 1967, Vol. 31, July, p. 94.

N75 10999

STUDY OF THE REDUNDANCY OF TELEMETRY INFORMATION OF THE LUNA-9 AND LUNA-13 AUTOMATIC LUNAR PROBES

/78

A. V. Kantor and T. A. Tolmadzheva

One of the basic problems for the radio telemetry systems of contemporary and prospective spacecraft is reduction of the redundancy of messages on board the spacecraft.

Knowledge of the statistical characteristics of natural telemetry information is very important for the development of principles and methods of reduction of redundancy of telemetry messages, as well as planning of on-board information compression devices (OICD). The main purpose of this work is to determine these characteristics for the telemetry information of the Luna-9 and Luna-13 automatic lunar probes, to allow us to estimate the redundancy of the information.

Along with this, we have set ourselves the task of producing such important characteristics for the planning of OICD as the characteristics of the flow of important interrogations.

Finally, we shall attempt to produce the values of the information compression factors for information transmitted through individual channels and transmitted using the switching device as a whole.

The following quantity will be used as the integral measure of redundancy in this work:

$$U = \frac{f - f_c}{f} \cdot 100\%, \tag{1}$$

where f is the number of interrogations per cycle of operation of the existing switching device; f_c is the number of significant (i.e., nonredundant) interrogations which must be transmitted on the average per operating cycle of the existing switching device when the information is compressed with a fixed value of approximation error. The value of f is equal to the number of channels of the switching device. The value of f_c can be determined as follows:

$$f_c = M\{N_j\}, \tag{2}$$

where N_j is a random quantity, representing the number of significant interrogations in the jth cycle of operation of the switching device (density of the flow of significant interrogations); $M\{N_j\}$ is the mathematical expectation of the value of N_j.

In this work, we will use histograms as well as numerical characteristics, the estimate of mathematical expectation (2) and estimate of dispersion of density of the stream of significant interrogations:

$$D = D\{N_j\} \tag{3}$$

as the main characteristic of the stream of significant interrogations. The value of the information compression factor for information transmitted through individual channels was calculated using the formula

$$K_i^{(n)} = \frac{j_n - j_{n-1}}{N_i^{(n)}}, \tag{4}$$

where $K_i^{(n)}$ is the information compression factor for information transmitted through the ith channel in the nth zone (see Table 1 below) of the file of information; k is the number of the channel in the switching device; n is the number of the zone in the information file; j_n is the number of the operating cycle of the switching device last in the nth zone; j_{n-1} is the number of the operating cycle of the switching device which is last in the (n - 1)th zone; $N_i^{(n)}$ is the number of significant interrogations in the ith channel in the nth zone. /79

The value of the compression factor of information transmitted using the switching device as a whole was calculated using the formula

$$K^{(n)} = \frac{j(j_n - j_{n-1})}{N_n}, \tag{5}$$

where $K^{(n)}$ is the information compression factor for information transmitted using the switching device as a whole, in the nth zone of the file of information; N_n is the number of significant interrogations in the nth zone.

Let us present some information on the telemetry information from the lunar probe, then study the method of investigation and finally present the results produced in the investigations.

Collection of telemetry information on the Luna-9 and Luna-13 during the flight sector was performed using two switching devices: commutator 1 (K1) and commutator 2 (K2). The number of channels in each of the commutators, as we have indicated, is equal to f. The length of a commutation cycle for commutator K1 is 12.8 sec, for commutator K2 -- 6.4 sec. The information is transmitted and recorded at the receiver end in 16 fixed levels, so-called gradations. The minimum value of a signal from the sensor corresponds to the first gradation, the maximum value -- to the 16th gradation. The characteristics of telemetry information for the two probes were studied for two flight sectors for each commutator. Data on files of information processed are presented in Table 1.

TABLE 1

FILES OF TELEMETRY INFORMATION

File Number	Luna-	Flight Sector*	Commutator								
			Number	Number of Cycles							
				In File	In Zone						
					1	2	3	4	5	6	7
1	9	K	K1	311	36	45	118	57	55		
2	9	K	K2	23							
3	9	T	K1	593	33	92	117	89	136	126	
4	9	T	K2	25							
5	13	K	K1	357	14	70	59	117	60	37	
6	13	K	K2	34							
7	13	T	K1	569	55	52	71	50	115	115	111
8	13	T	K2	34							

*L9, ALP "Luna-9," L13, ALP "Luna-13".
**K, Trajectory correctionsession; T, Deceleration session before landing.

As we can see from the table, each file of information is divided into zones. The zone boundaries were selected as follows: each zone encompases a given section of space probe operation, characterized by relative constancy of the dynamics of change of telemetry messages within a zone.

This selection of zones allowed us to calculate the compression factors $K_i^{(n)}$ (4) and $K^{(n)}$ (6), averaged within each zone. For more convenient description of the telemetry messages /80
in the operating cycles of the commutator (in each zone) we calculated the values of N_j -- the number of significant interrogations in the jth cycle of operation of the commutator.

It should be noted that in each of the switching devices, K_1 and K_2, a number of channels are utilized for service information; signals on these channels retain their values unchanged. Furthermore, there are groups of parallel channels, each of which receives a signal from one sensor. In order to simplify the program of the computer, the service channels were looked upon as information channels, and each of the parallel channels was looked upon as a channel through which signals were sent from an independent source of messages[1].

In order to allow the same value[2] of permissible approximation error to be fixed (see below), the values of parameters transmitted through the various channels of the switching devices were introduced not in units of the measured parameter, but rather in gradation numbers.

The information was processed on the M20 computer. Information was input by punch cards. For each file of telemetry information (see Table 1), f sets of punch cards were made. The values of the measured parameter (in gradation numbers) were punched into each set for successive cycles of operation of the commutator. Since the volume of input information is quite large, in order to reduce the number of punch cards required in each file, the data has to be placed compactly on the field of the punch card. This was possible, since many parameters retain their values constant for many cycles of operation of the commutator. Therefore, each card carries the value of a parameter in the given cycle of operation of the commutator and a code number indicating the number of cycles during which this value is retained constant. The computer program includes a sorting subroutine (not analyzed in this article).

Information processing on the M20 computer was performed according to two programs (in two stages): the primary processing program (stage 1) and the secondary processing program (stage 2).

The primary processing program calculated:

1) the values of compression factor for information transmitted through each channel; the mean values of compression factor were determined in each of the zones $K_i^{(n)}$ (4);

---

[1]The presence of parallel channels in the commutators was considered in the work "Digital Computer Modeling of the Process of Reduction of Redundancy of Multichanneled Telemetry Information by the Method of Adaptive Discretization with Associated Sorting," published in this collection, page 90.

[2]This simplifies digital computer calculation.

2) the values of the compression factor of information transmitted using the commutator as a whole; the mean values of compression factors were determined in each of the zones $K^{(n)}$ (5);

3) values of the density of the flow of significant interrogations N_j.

Calculations according to the primary processing program were performed for each of five information compression algorithms:

Algorithm 1 -- zero-order extrapolation;
Algorithms 2 and 3 -- first order extrapolation;
Algorithms 4 and 5 -- first order interpolation.

The secondary processing program calculated:

1) histograms for the density of the flow of significant interrogations;

2) values of estimates of mathematical expectation of the density of flows of significant interrogations;

3) values of estimates of dispersion of density of the flow of significant interrogations.

The initial material for calculation used in the secondary processing program was the results of primary processing, the values of N_j -- the density of the flow of significant interrogations. However, all values of N_{jn} were replaced by values of N_j in preceding or subsequent cycles. This was done to eliminate distortion of historgrams and estimates of the quantities $M\{N_j\}$ and $D\{N_j\}$. /81

In all processing, interrogation was considered significant if the following condition was fulfilled:

$$\varepsilon \geq \varepsilon_{sup} = 1 \text{ gradation}, \qquad (6)$$

where ε is the value of approximation error, i.e., the error which would arise if the interrogation were not transmitted; $\varepsilon_{sup} = 1$ - the permissible approximation error.

Figure 1 shows a flow chart for the program for primary processing for extrapolation of the zero order of 1[1]. This /82

[1]The simplest algorithm for a program for primary processing; for other algorithms, programs are similar to those presented in Figure 1, with the exception of the sections of the program related to determination of the significance of an interrogation (for

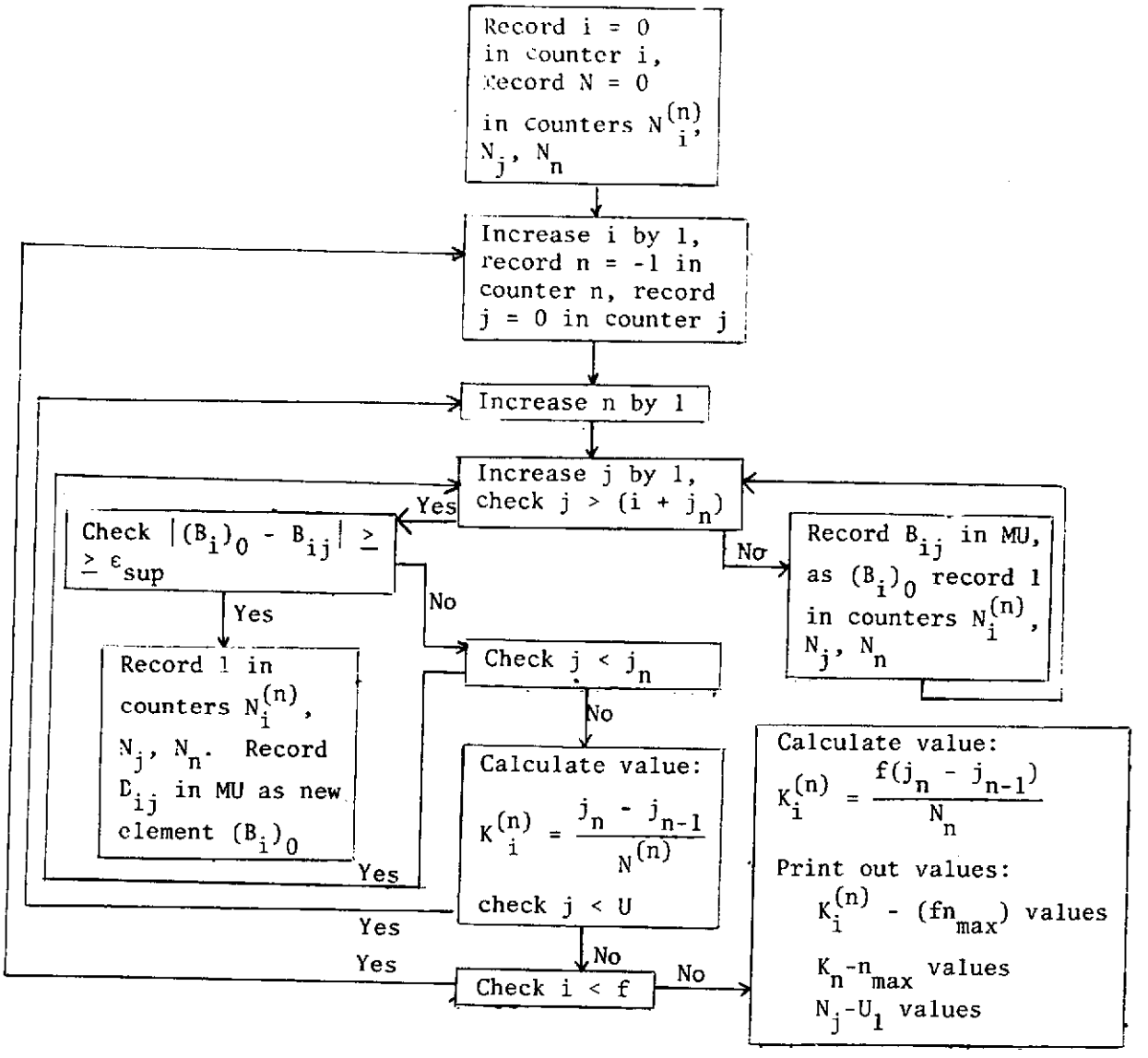


Figure 1. Flow Chart for Program for Primary Processing by Algorithm 1

figure utilizes the following symbols (see also symbols used in formulas (2), (4), (5) and (6)): B_{ij} -- the value of interrogation of a signal transmitted through the ith channel in the jth cycle of

---

the zero order extrapolation, these sections include determination of $|(B_i)_0 - B_{ij}| \geq \varepsilon_{sup}$ and writing of the value of the reference interrogation $(B_i)_0$. We will not study these programs in this work.

operation of the commutator; MU, the memory unit of the device; $(B_i)_0$, the value of the reference interrogation in the ith channel (the reference interrogation is the last interrogation transmitted to the radio channel); U, the number of cycles in a file; $N_i^{(n)}$, the number of significant interrogations in the ith channel of the nth zone. Obviously,

$$N_n = \sum_{i=1}^{f} N_i^{(n)}. \tag{7}$$

The program operates as follows. The information is processed channel by channel, beginning with the first channel. The first interrogation of each zone is fixed as the reference interrogation (the number of cycles corresponding to the first cycles of each zone is defined by the condition $j = I + j_n$, assuming that $j_0 = 0$). Since the reference interrogation is significant in itself, as the reference interrogation in the first cycle of each zone is recorded, a one is simultaneously recorded in counters $N_i^{(n)}$, N_j and N_n. We note in passing that (in the general case) the number of counters $N_i^{(n)}$ is equal to $f \times n_{max}$ (n_{max} being the maximum number of zones $j_{nmax} = U$), the number of counters N_n is equal to n_{max} and counters N_j is equal U. The second and all subsequent cycles are checked for significance. If an interrogation is significant, first of all, a one is recorded in counters $N_i^{(n)}$, N_j and N_n and, secondly, the reference interrogation is changed. This operation (checking of significance) is continued until cycle j, equal to j_n(end of nth zone) is processed. Then, the value of $K_i^{(n)}$ is calculated and a shift is made to the next zone. Processing by zones is completed when j = U. This completes the operation of preliminary processing of information for a given channel, and a shift is made to preliminary processing of information of the next channel. Thus, preliminary processing is performed in sequence for all f channels. Final processing of information is then performed for all f channels: the values of $K^{(n)}$ are calculated and the values of $K_i^{(n)}$ ($f \times n_{max}$ values), K_n (n_{max} values) and N_j (U values) are printed.

Figure 2 shows a flow chart of a program for calculation of a histogram for the density of the stream of significant interrogations. This figure uses the following symbols: j, the number of cycles of commutator K1 (K2), $1 \leq j \leq U$, j is an integer; S, the number of the interval of subdivision of the values of N_j, $1 \leq S \leq 32$, S is an integer; N_j, the value of the frequency

(number) of significant interrogations in the jth cycle of commutator K1 (K2),

$$0 \leq N_j \leq f \text{ interrogations/sec}$$

for any j;

a_S, the right boundary of the Sth interval of subdivision of values of N_j, a_{S-1}, the left boundary of the Sth interval of subdivision of values of N_j, $a_S = a_{S-1} + 4$ for $32 > S > 1$; $a_0 = 0$; $a_1 = 4$ interrogations/cycle; $a_{32} = (f + 1)$ interrogations/cycle; m_S, the number of cycles of commutator K1 (K2) in which the value of N_j satisfies the condition $a_{S-1} \leq N_j < a_S$.

The program operates as follows. The information is processed by operating cycles of commutator K1 (K2), beginning with the first cycle. For each jth cycle, number S of the subdivision interval, corresponding to the value of N_j of this cycle, is sought. After the value of S is found, a one is written in the corresponding counter m_S. Thus, at moment j = U, all values of m_S will be determined, and are then printed out. /83

The estimates of mathematical expectation and dispersion of significant interrogation density are made using the formulas

$$M\{N_j\} = \frac{\sum_{j=1}^{U} N_j}{U}, \tag{8}$$

$$D\{N_j\} = \frac{\sum_{j=1}^{U} [N_j - M\{N_j\}]^2}{U-1}, \tag{9}$$

where N_j is the number of significant interrogations in the jth cycle of commutator K1 (K2); the value of N_j in any cycle may take on any integer value between 0 and f interrogations/cycle; j is the number of the operating cycle of commutator K1 (K2); U is the number of cycles of commutator K1 (K2) in the communications session (i.e., in one file).

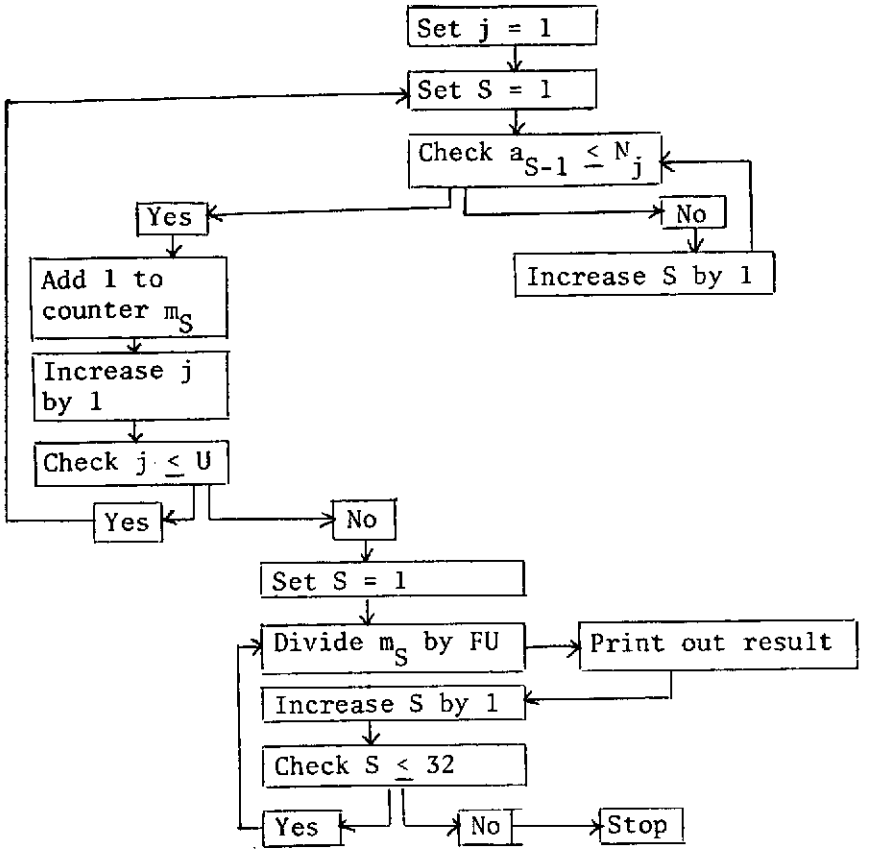


Figure 2. Flow Chart of Program for Calculation of Histogram for Significant Interrogation Density

The primary results of the investigation are presented in Tables 2, 3, 4, 5 and Figures 3a and 3b.

Table 2 presents values of $K_i^{(n)}$ for a number of sensors and algorithms 2, 3, 4 (it is not possible to present the values of $K_i^{(n)}$ for all sensors and algorithms, due to the limited space available), calculated according to formula (4) by the program of Figure 1. /88

Table 3 presents values of $K^{(n)}$ calculated using formula (5) by the program of Figure 1.

Figure 3a and 3b presents histograms for the density of significant interrogations calculated for zone 4 (see Table 1).

Table 4 presents values of estimates of $M\{N_j\}$ and $D\{N_j\}$, calculated using the formulas, also for zone 4. The nature of the histograms and similar values of $M\{N_j\}$ and $D\{N_j\}$ indicate that for the stable sectors, the stream of significant interrogations is similar to a simple stream with the parameter

$$\lambda = M\{N_j\}. \tag{10}$$

TABLE 2*

Commutator I

Algorithm	Zone	2						3						4					
		Channel						Channel						Channel					
		1	10	11	12	13	23	1	10	11	12	13	23	1	10	11	12	13	23
	1	18	18	18	18	18	6	18	18	18	18	18	6	18	18	18	18	18	6
K	2	22,5	22,5	22,5	22,5	22,5	7,6	22,5	22,5	22,5	22,5	22,5	7,5	22,5	22,5	22,5	22,5	22,5	7,5
	3	59	59	59	59	59	19,6	59	59	59	59	59	19,6	59	59	59	59	59	19,6
	4	28,5	3,5	28,5	14,2	28,5	28,5	28,5	3,5	28,5	8,1	28,5	28,5	28,5	28,5	28,5	19,0	28,5	28,5
	5	27,5	27,5	27,5	27,5	27,5	27,5	27,5	27,5	27,5	27,5	27,5	27,5	27,5	27,5	27,5	27,5	27,5	27,5
L9	1	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5
	2	46	46	46	46	46	5,1	46	46	46	46	46	5,4	46	46	46	46	46	5,4
	3	58,5	58,5	58,5	58,5	58,5	19,5	58,5	58,5	58,5	58,5	58,5	19,5	58,5	58,5	58,5	58,5	58,5	19,5
T	4	44,5	44,5	44,5	44,5	44,5	14,8	44,5	44,5	44,5	44,5	44,5	14,8	44,5	44,5	44,5	44,5	44,5	14,8
	5	68	68	34	34	68	22,6	68	68	34	34	68	22,7	68	68	68	68	68	22,7
	6	63	21	31,5	31,5	31,5	15,7	63	25,2	31,5	31,5	31,5	15,7	63	31,5	31,5	31,5	31,5	15,7
	1	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
	2	35	35	35	35	11,6	17,5	35	35	35	35	14	17,5	35	35	35	35	11,6	17,5
K	3	29,5	29,5	29,5	29,5	9,8	7,37	29,5	29,5	29,5	29,5	11,9	11,8	29,5	9,8	29,5	29,5	11,8	14,75
	4	59,5	19,8	59,5	59,5	29,7	11,9	59	23,8	59,5	59,5	29,7	13,2	59,5	29,5	59,5	14,8	59,5	13,2
	5	29	29	29	29	29	29	29	29	29	29	29	29	29	6,4	14,5	29	7,2	29
	6	18,5	6,1	2,3	18,5	3,7	18,5	18,5	7,4	4,6	18,5	7,4	18,5	18,5	4,1	3,3	7,4	9,2	18,5
L13	1	27	4,5	27	4,5	27	6,75	27	6	27	5,4	27	7,71	27	7,7	6	3,4	27	7,71
	2	26	17,7	26	26	26	5,2	26	8,67	26	26	26	5,7	26	12,3	5,2	17,4	26	6,5
	3	35,5	12,5	17,6	17,8	35,5	17,7	35,5	23,7	17,8	23,7	35,5	17,7	35,5	35,5	4,74	17,7	35,5	17,7
T	4	25	11,5	25	25	25	8,3	25	12,5	25	25	25	8,3	25	5	5	25	25	8,3
	5	57,5	5,75	12,8	19,2	57,5	19,2	57,5	16,4	12,8	19,2	57,5	19,2	57,5	57,5	14,4	23	57,5	11,5
	6	57,5	28	28,7	57,5	57,5	11,5	57,5	57,5	38,4	57,2	57,5	11,5	57,5	57,5	19,2	57,5	57,5	12,8
	7	56	12	18,7	28	28	5,09	56	28	18,7	28	28	6,2	56	18,7	10,2	28	28	6,2

*Translator's Note: Commas represent decimals.

TABLE 2 (Continued)

Commutator I

Algorithm	Zone	2						3						4					
		Channel						Channel						Channel					
		43	44	45	48	49	50	43	44	45	48	49	50	43	44	45	48	49	50
K	1	3,6	4,5	18	18	4,5	18	3,6	6,0	18	18	4,5	18	18	18	18	18	5,1	18
	2	1,8	4,5	4,5	2,3	2,0	2,2	2,2	5,0	5,6	2,6	2,2	2,6	15	15	6,5	3,0	2,2	3,0
	3	1,5	1,5	1,5	1,34	1,4	1,8	1,7	1,6	1,5	1,4	1,5	1,6	59	59	1,7	1,7	1,5	1,6
	4	1,1	1,14	1,1	1,23	1,3	1,0	1,1	1,1	1,2	1,3	1,5	1,0	11,4	19	1,2	1,5	1,7	1,2
	5	2,2	1,4	2,0	1,3	1,9	1,5	2,8	1,5	2,3	1,5	2,5	1,7	18,3	27,5	2,5	1,8	2,6	2,1
L9 T	1	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5
	2	2,1	8,4	2,1	3,1	3,7	2,7	2,6	9,2	2,7	3,3	4,4	3,0	46	30,7	2,8	3,4	4,4	3,4
	3	1,1	1,1	1,0	1,0	1,1	1,1	1,2	1,2	1,1	1,1	1,2	1,1	58,5	58,5	1,2	1,2	1,2	1,3
	4	1,1	1,1	1,1	1,3	1,5	1,2	1,2	1,2	1,2	1,6	1,7	1,4	29,7	44,5	1,3	1,8	1,9	1,6
	5	1,2	1,1	1,0	1,1	1,2	1,1	1,3	1,2	1,1	1,2	1,3	1,2	68	68	1,2	1,5	1,5	1,5
	6	2,1	1,2	1,2	1,8	1,2	2,2	2,3	1,4	1,4	2,1	2,3	2,3	63	42	1,7	2,9	3,2	3,2
K	1	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
	2	2,5	2,92	35	2,9	1,7	2,5	3	35	35	3,5	2,1	3,0	35	35	35	3,7	2,4	3,3
	3	2,46	2,68	14,75	3,0	1,4	1,9	2,95	2,68	20	3,7	1,5	2,1	29,5	29,5	29,5	3,7	1,9	2,8
	4	1,35	1,17	1,2	1,3	1,2	1,3	1,43	1,21	1,2	1,5	1,5	1,4	39,7	14,8	1,43	1,7	1,7	1,8
	5	1,1	1,1	1,0	1,5	1,2	1,5	1,1	1,1	1,0	1,7	1,3	1,6	29	29	1,1	1,8	1,7	1,6
	6	18,5	3,08	2,0	2,9	6,2	3,0	18,5	3,36	2,5	1,6	6,2	3,4	3,36	7,4	2,3	1,8	2,3	2,3
L13 T	1	3	27	3	13,5	2,7	9	3,18	27	3,37	13,5	2,84	9	3,6	27	4,5	13,5	2,84	13,4
	2	5,2	26	10,4	3,25	2	3,71	5,2	26	10,4	4	2,74	3,71	6,5	26	17,3	4	2,88	5,2
	3	1,2	1,31	1,16	1,16	1,31	1,22	1,36	1,45	1,25	1,27	1,39	1,22	1,51	1,69	1,27	1,31	1,57	1,39
	4	1,2	1,19	1,14	1,09	1,04	1,13	1,25	1,39	1,32	1,16	1,09	1,13	1,31	1,6	1,68	1,25	1,19	1,25
	5	1,82	1,05	1,09	1,29	1,22	1,3	1,94	1,17	1,14	1,44	1,37	1,3	1,64	1,38	1,22	1,57	1,57	1,53
	6	1,11	1,02	1,03	1,13	1,09	1,28	1,26	1,05	1,05	1,29	1,23	1,28	1,47	1,09	1,08	1,53	1,51	1,47
	7	1,21	1,12	1,33	1,13	1,24	1,3	1,38	1,15	1,38	1,43	1,3	1,3	1,72	1,26	1,51	1,62	1,62	1,55

TABLE 2 (Continued)

Commutator I

Algorithm		2						3						4					
		Channel						Channel						Channel					
Zone		51	67	68	69	70	76	51	67	68	69	70	76	51	67	68	69	70	76
K	1	4,5	5,1	9	18	18	9	4,5	5,1	9	18	18	9	4,3	5,1	18	18	18	18
	2	2,0	11,2	22,5	22,5	22,5	22,5	2,6	11,2	22,5	22,5	22,5	22,5	2,6	11,2	22,5	22,5	22,5	22,5
	3	1,2	9,9	59	59	59	59	1,3	9,9	59	59	59	59	1,5	9,9	59	52	59	59
	4	1,2	28,5	28,5	28,5	28,5	28,5	1,5	28,5	28,5	28,5	28,5	28,5	1,7	28,5	28,5	28,5	28,5	28,5
	5	8,9	27,5	27,5	27,5	27,5	27,5	2,5	27,5	27,5	27,5	27,5	27,5	2,6	27,5	27,5	27,5	27,5	27,5
L9	1	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5	16,5
	2	2,9	15,3	46	46	46	46	3,3	15,3	46	46	46	46	3,7	15,3	46	46	46	46
	3	1,1	58,5	58,5	29,3	58,5	58,5	1,1	58,5	58,5	29,3	58,5	58,5	1,2	58,5	58,5	29,3	58,5	58,5
T	4	1,3	44,5	44,5	22,3	44,5	44,5	1,6	44,5	44,5	22,3	44,5	44,5	1,6	44,5	44,5	22,3	44,5	44,5
	5	1,1	68	34	34	34	34	1,2	68	34	34	34	34	1,3	68	68	68	68	68
	6	2,1	25,2	15,8	25,2	25,2	21	2,3	25,2	15,8	25,2	25,2	21	3,2	25,2	15,8	25,2	25,2	21
K	1	7	7	7	3,5	7	7	7	7	7	4,7	7	7	7	7	7	4,7	7	7
	2	1,6	35	35	17,5	17,5	2,33	2,2	35	35	17,5	17,5	2,5	2,3	35	35	17,5	3,3	2,8
	3	1,4	29,5	29,5	7,4	29,5	1,37	1,8	29,5	29,5	8,4	29,5	1,5	2,2	29,5	29,5	9,8	7,4	1,63
	4	1,2	59,5	59,5	19,8	19,8	1	1,3	59,5	59,5	19,8	23,8	1,1	1,5	59,5	59,5	19,8	23,8	1,27
	5	1,3	29	29	29	14,5	1	1,4	29	29	29	14,5	1,18	1,6	29	29	29	11,6	1,28
	6	4,7	18,5	18,5	18,5	1,2	1,4	2,9	18,5	18,5	18,5	1,5	1,48	2,1	18,4	18,5	18,5	1	2
L13	1	2,57	27	13,5	10,7	9	27	2,57	27	13,5	10,7	10,7	27	2,84	18	13,5	10,7	10,7	18
	2	2,89	4,3	26	8,6	13	13	3,06	4,3	26	8,7	13	13	3,25	6,5	26	12	17,3	26
	3	1,27	11,8	17,7	35,5	35,5	35,5	1,34	14,2	17,7	35,5	35,5	35,5	1,54	14,2	17,7	35,5	25,5	35,5
	4	1,38	25	8,3	25	6,25	25	1,33	25	16,7	25	7,14	25	1,22	25	25	25	8,3	25
T	5	1,16	27,5	14,4	57,5	5,75	57,5	1,22	57,5	16,4	57,5	5,75	57,5	1,45	28,7	16,4	57,5	5,75	57,5
	6	1,13	57,5	57,5	57,5	57,5	57,5	1,25	57,5	57,5	57,5	5,75	57,5	1,51	57,5	57,5	57,5	23	57,5
	7	1,12	28	28	28	28	18,6	1,23	28	28	28	28	18,6	1,47	28	28	28	16	18,6

TABLE 2 (Continued)

Commutator II

Algorithm		2						3						4					
Zone		Channel						Channel						Channel					
		1	10	11	12	19	23	1	10	11	12	19	23	1	10	11	12	19	23
L9	*K*	11,5	5,65	5,75	11,5	5,75	3,83	11,5	5,75	5,75	11,5	5,75	4,6	11,5	5,75	5,75	11,5	4,6	4,6
	T	12,5	6,25	6,25	4,2	6,25	2,5	12,5	6,25	6,25	4,2	6,25	2,78	12,5	6,25	6,25	4,2	6,25	5
L13	*K*	17	17	8,5	17	5,7	5,7	17	17	8,5	17	5,7	5,7	17	17	8,5	17	5,7	3,8
	T	17	17	17	8,5	17	1,21	17	17	17	8,5	17	1,48	17	17	17	8,5	1,7	4,86

Algorithm		2						3						4					
Zone		Channel						Channel						Channel					
		32	35	36	37	38	41	32	35	36	37	38	41	32	35	36	37	38	41
L9	*K*	2,3	5,75	2,3	5,75	5,75	11,5	2,88	5,75	2,88	5,75	5,75	11,5	2,88	5,75	3,28	5,75	5,75	11,5
	T	6,25	4,17	1,79	3,13	1,25	4,17	6,25	4,17	1,79	3,13	1,4	4,17	6,25	4,17	2,08	4,17	1,79	4,17
L13	*K*	8,5	5,7	1,54	8,5	8,5	4,25	8,5	5,7	1,62	8,5	8,5	11,3	8,5	5,7	1,62	8,5	8,5	17
	T	17	8,5	1,17	8,5	1,21	8,5	17	8,5	1,21	8,5	1,26	8,5	17	8,5	1,36	8,5	1,9	8,6

Algorithm		2						3						4					
Zone		Channel						Channel						Channel					
		55	56	67	86	99	118	55	56	67	86	99	118	55	56	676	86	99	118
L9	*K*	2,88	2,88	5,75	5,75	3,83	5,75	2,88	2,88	5,75	7,67	3,83	5,75	3,83	4,6	5,75	7,67	3,83	5,75
	T	3,13	3,13	4,17	12,5	4,17	4,5	3,57	3,57	4,17	12,5	4,17	4,17	4,17	4,17	4,17	12,5	4,17	4,17
L13	*K*	8,5	8,5	17	17	17	4,25	8,5	8,5	17	17	17	6,8	8,5	8,5	17	17	17	17
	T	5,7	5,7	8,5	17	8,5	5,7	6,8	6,8	8,5	17	8,5	5,7	8,5	8,5	8,5	17	8,5	8,5

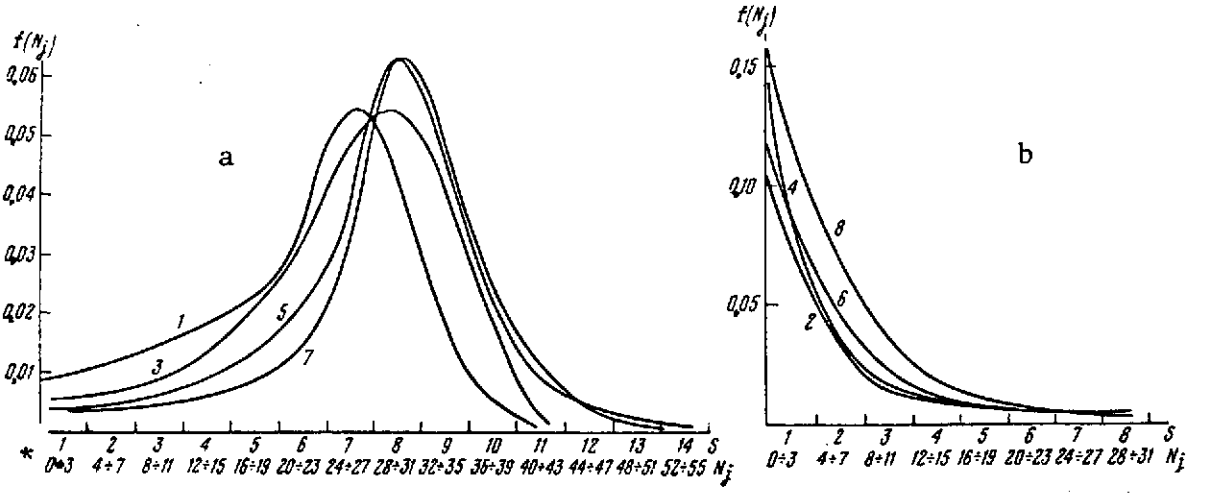


Figure 3. Histogram for Density of Significant Interrogation. a: 1, L9 K KI; 3, L9 T KI: 5, L13 K KI; 7, L13 T KI; b: 2, L9 K KII; 4, L9 T KII; 6, L13 K KII; 8, L13 T KII
* ÷ represents a dash.

Table 5 presents values of redundancy U, calculated using formulas (1), (2) and the data of Table 4. The values of U and $M\{N_j\}$ indicate the significant redundancy of the telemetry information returned by the Luna-9 and Luna-13 spacecraft, and the expediency of using OICD for these spacecraft.

The data on values of $D\{N_j\}$ presented in Table 4 allow us to determine the capacity of the buffer memory necessary as a part of the OICD to smooth the unevenness of the stream of significant interrogations if the OICD operates by the method of adaptive discretization.

In conclusion, the authors express their gratitude to Engineer V. G. Timonin, who was very helpful in performing the studies, as well as Engineers I. I. Golovchenko and L. N. Shatinskaya, who took part in writing the program for processing the information on the M20 computer.

TABLE 3*

Algorithm				2	3	4
Space-craft	Flight Sector	Number	Zone			
L9	K	K1	1	7,6	8,3	8,70
			2	6,87	7,05	7,48
			3	5,25	5,67	6,19
			4	4,06	4,48	4,87
			5	5,03	5,81	6,13
		K2	1	4,76	5,03	5,18
	T	K1	1	16,2	16,21	16,29
			2	6,7	7,07	7,58
			3	5,26	5,73	6,22
			4	3,67	4,21	4,76
			5	4,35	4,95	5,04
			6	3,8	4,76	4,98
		K2	1	6	6,04	6,08
L13	K	K1	1	6,76	6,78	6,83
			2	6,62	6,83	6,98
			3	6,15	6,19	6,24
			4	3,94	4,15	4,56
			5	4,93	5,27	5,34
			6	3,38	6,79	5,10
		K2	1	8,55	8,81	9,25
	T	K1	1	7,8	9,08	8,33
			2	7,14	7,81	8,33
			3	4,35	4,86	5,26
			4	3,85	4,17	4,35
			5	3,45	3,73	4
			6	4,35	4,76	5
			7	3,34	3,57	3,85
		K2	1	13,04	13,21	13,61

*Translator's Note: Commas represent decimals.

TABLE 4*

Spacecraft	L9				L13			
Flight Sector	*K*		*T*		*K*		*T*	
Number	K1	K2	K1	K2	K1	K2	K1	K2
$M\{N_j\}$	26,2	5,3	28,3	6,7	29,1	8	30,5	7,2
$D\{N_j\}$	25	6,3	26,1	7,6	27,3	11	28,1	8,7

*Translator's Note: Commas represent decimals.

TABLE 5*

Algorithm			2	3	4
Spacecraft	Flight Sector	Number			
L9	*K*	K1	79,5	85,5	86,9
		K2	82	84	87,1
	T	K1	81,2	82	83,6
		K2	87,5	86,9	80,6
L13	*K*	K1	79	81,2	82,8
		K2	87,5	88,7	90,6
	T	K1	79	80,5	82
		K2	93,7	92,2	93,7

*Translator's Note: Commas represent decimals.

N75 11000

DIGITAL COMPUTER MODELING OF THE PROCESS OF REDUCTION OF REDUNDANCY IN MULTICHANNEL TELEMETRY INFORMATION BY THE METHOD OF ADAPTIVE DISCRETIZATION WITH ASSOCIATIVE SORTING

T. A. Tolmadzheva, A. V. Kantor,
and L. V. Rozhkovskiy

Analytic study of information measuring systems involves significant mathematical difficulties. This statement applies fully, naturally, to systems for reduction of the redundancy of multichannel telemetry information and, in particular, the class of such systems made using the method of adaptive discretiza-ation. An unavoidable stage in the study and planning of systems for reduction of information reduncancy and, in particular, systems with adaptive discretization, is modeling of the process of reduction by universal digital computer. It should be noted that at the present time, due to a number of objective factors, a tendency has arisen toward performing these studies in the following stages: approximate analytic study; digital computer modeling with the actual information; statistical computer modeling (Monte Carlo method); more complete analytic investigation.

Of course, in some cases the sequence of individual stages of the investigation may change, and some stages may be absent. For example, if statistical characteristics of the actual telemetry information are available, approximate analytic modeling should be followed directly by statistical modeling. In complex measuring information systems, due to the significant mathematical difficulties involved, the final stage is statistical modeling. /91

The present work is dedicated to digital computer modeling of the process of adaptive discretization with associative sorting (ADAS) of actual multichannel telemetry information[1].

The principal methods of ADAS, classification of the structure of systems and statement of the problem of approximate analytic investigation were presented in [1]. Information on

---

[1]Problems of statistical modeling of the ADAS process will form the subject of a special work by the authors of the present article.

the actual telemetry information used in computer modeling was presented in [2]. Therefore, we will now go directly over to a study of the problems of modeling and its methodology, after which we will present the results obtained.

The main task in modeling is production of initial data for determination of dependences describing the operation of the on-board[2] information compression device (OICD) with the ADAS; at the same time, data should be produced as necessary for (later) restoration of the initial information from the compressed information and calculation of the error in the restored information. Furthermore, initial data must be produced for analytic studies and for statistical modeling of the ADAS process.

Also, initial data must be produced for comparison of OICD with three types of adaptive discretization: 1) with buffer smoothing (ADBS) of the irregularities in the stream of significant interrogations; 2) with logical switching (ADLS) for the production of an even stream of significant interrogations, and also to reduce the volume of service information; 3) with associative sorting (ADAS).

The problems of production of dependences characterizing the operation of OICD with ADAS from the initial information, the initial information from the compressed information, the characteristics for analytic study and statistical modeling and comparison of OICD with the three types of adaptive discretization are studied only partially, since they form the subject for an extensive, independent study.

The dependences characterizing the operation of OICD with ADAS for the telemetry information studied are as follows:

1. The dependence

$$P^{(\Pi)}\{\varepsilon^{(\Pi)} < \varepsilon_{per}^{(\Pi)}\} = \phi^{(\Pi)}(\Pi/\tau) \qquad (1)$$

[2]As was demonstrated in [1], this method is designed primarily for use in multipurpose on-board compression devices (of course, the ADAS method can also be used in other types of information measurement systems).

For $E_{AMU} = \infty$, where $\varepsilon^{(\Pi)}$ is the error in approximation in the stream of compressed information with number[3] Π; $\varepsilon^{(\Pi)}_{(per)}$ is the permissible error in approximation in the stream of compressed information with number Π; $P^{(\Pi)}\{\varepsilon^{(\Pi)} < \varepsilon^{(\Pi)}_{per}\}$ is the probability that the error of approximation in the stream of compressed information with number Π will not exceed the permissible error; Π is the number of significant interrogations at the Πth output of the OICD in one subcycle of its operation; τ is the duration of a subcycle of operation of the OICD; $\Pi/\tau = f^{(\Pi)}_{n}$ is the repetition frequency of significant interrogations in the transmission line of the Πth stream of compressed information; E_{AMU} is the capacity of the associative memory unit (AMU). /92

Based on (1) with fixed values of Π, $\varepsilon^{(\Pi)}_{per}$ and $P^{(\Pi)}\{\varepsilon^{(\Pi)} < \varepsilon^{(\Pi)}_{per}\}$, we can produce the required value of $\tau = \tau_0$, the duration of an OICD operation subcycle, for which the maximum error of approximation will not exceed a fixed value with a fixed probability.

2) The dependences

$$P^{(l)}\{\varepsilon^{(l)}_{(per)} < \varepsilon^{(l)}_{(per)}\} = \phi^{(l)}(\Pi/\tau) \qquad (2)$$

for $l = 1, 2, \ldots, (\Pi - 1)$ and for $E_{AMU} = \infty$, where $\varepsilon^{(l)}$ is the error in approximation in the stream of compressed information with number l; $\varepsilon^{(l)}_{per}$ is the permissible error of approximation in the stream of compressed information with number l; $P^{(l)}\{\varepsilon^{(l)} < \varepsilon^{(l)}_{per}\}$ is the probability that the error of approximation in the stream of compressed information with number l will not exceed the permissible error.

---

[3]It is demonstrated in [1] that there are several streams of compressed information in the output of OICD with ADAS. The first stream is produced in the line with the least throughput capacity; the last stream is produced in the line with the greatest throughput capacity. The last stream is formed of all Π significant interrogations, where Π is the number of significant interrogations taken from the OICD in one subcycle of its operation. In the present work, we will assume for clarity that the number of streams of compressed information is Π, i.e., that streams with neighboring numbers are formed in each subcycle of operation of significant interrogations, the number of which differs by one.

3) Dependences:

$$\lambda = U(t) \text{ and } Q = \psi(t) \tag{3}$$

where λ is the density of the stream of significant interrogations at the input to the AMU; Q is the line length, i.e., the number of significant interrogations in the AMU in each subcycle of operation of the OICD at the moment of beginning of transmission of the first of the Π interrogations.

4) The dependence

$$P_s^{(\Pi)}\{\varepsilon^{(\Pi)} < \varepsilon_{per}^{(\Pi)}\} = \phi_s^{(\Pi)}(E_{AMU}) \tag{4}$$

for the fixed values of the quantity $f_n^{(\Pi)}$ where $P_s^{(\Pi)}\{\varepsilon^{(\Pi)} < \varepsilon_{per}^{(\Pi)}\}$ is the probability that the error of approximation in the stream of compressed information with number Π will not exceed the permissible error.

5) The dependences

$$P_s^{(l)}\{\varepsilon^{(l)} < \varepsilon_{per}^{(l)}\} = \phi_s^{(l)}(E_{AMU})$$
$$l = 1, 2, \ldots, (\Pi - 1), \tag{5}$$

where $P_s^{(l)}\{\varepsilon^{(l)} < \varepsilon_{per}^{(l)}\}$ is the probability that the error of approximation in the stream of compressed information with number l will not exceed the permissible error.

In order to restore the initial information from the compressed information and calculate the error of the restored information, we use the corresponding restoration algorithms and formulas for calculation of errors; in this work, we will not discuss this problem.

The following dependences and characteristics are necessary for analytic study and statistical modeling based on fixed telemetric information, in addition to (1)-(5):

1) A description of the random process, the stream of significant interrogations at the input to the AMU

$$V(n, K_\varepsilon, \varepsilon) = \xi(\Pi/\tau), \tag{6}$$

where n is the OICD operational cycle number; K_ε is the number of significant interrogations with approximation error ε; ε is the approximation error; K_ε and ε are random quantities;

2) A description of the random process, the stream of ordered significant interrogations[4] in the AMU /93

$$W(n, K_{\varepsilon y}, \varepsilon) = \eta(\Pi/\tau), \tag{7}$$

where n is the OICD operating cycle number, i.e., the moment of appearance of a group of ordered significant interrogations; $K_{\varepsilon y}$ is the number of ordered significant interrogations with approximation error ε; ε is the approximation error; $K_{\varepsilon y}$ and ε are random quantities.

Formula (7) is necessary both for $E_{AMU} = \infty$ and for finite values of E_{AMU}.

When the ADAS, ADBS and ADLS methods are compared with the same telemetric information, the comparison criteria might be the characteristics listed below; we should note specially that in this work we are speaking only of criteria which can be defined by computer modeling of the ADAS process, not the complete set of criteria:

1. The main criterion is the frequency of transmission of compressed information, necessary to see that the maximum error of approximation does not exceed its permissible value with a predetermined probability.

In order to compare ADAS, ADBS and ADLS, we must determine the quantities

---

[4]Ordered significant interrogations refer to the significant interrogations which have passed through sorting in the AMU, i.e., placed on a list in the order of decreasing approximation error (see [1]).

$$(f_n)_{AS} \text{ for } P\{\varepsilon < \varepsilon_{per} = \varepsilon_0\} = P_0; \quad E_{AMU} = \infty,$$

$$(f_n)_{BS} \text{ for } P\{\varepsilon < \varepsilon_{per} = \varepsilon_0\} = P_0; \quad E_{BMU} = \infty,$$

$$(f_n)_{LS} \text{ for } P\{\varepsilon < \varepsilon_{per} = \varepsilon_0\} = P_0; \qquad (8)$$

where $(f_n)_{AS}$, $(f_n)_{BS}$ and $(f_n)_{LS}$ are the frequencies of transmission of compressed information for ADAS, ADBS and ADLS respectively; ε_0 is the maximum permissible value of approximation error; P_0 is the fixed value of probability $P\{\varepsilon < \varepsilon_{per} = \varepsilon_0\}$; E_{BMU} is the capacity of the buffer memory unit.

2. One important criterion is the buffer memory unit capacity. As we know, in OICD with ADLS, there is no BMU; in OICD with ADAS, the BMU is functionally present: the AMU acts as a BMU; in OICD with ADBS, a separate BMU is present. In order to compare ADAS and ADBS, we must determine the functions

$$(f_n)_{AS} = \psi_{AS}(E_{AMU}) \text{ for } P\{\varepsilon < \varepsilon_{per} = \varepsilon_0\} = P_0,$$

$$(f_n)_{BS} = \psi_{BS}(E_{AMU}) \text{ for } P\{\varepsilon < \varepsilon_{per} = \varepsilon_0\} = P_0. \quad (9)$$

An important particular case for $(f_n)_{AS}$ is the case where $E_{AMU} = \Pi$. The corresponding operating mode of the OICD is called the mode with full erasure (FE), since there is no functional buffer memory in AMU, i.e., $E_{BMU} = 0$ (all information extending beyond limits of the Π locations in AMU is "erased"; it also does not participate in subsequent information sorting cycles.

Let us go over to analysis of the initial data which must be produced as a result of computer modeling to produce dependences (1)-(9) and for restoration of the initial information from the compressed information with calculation of the errors in the restored information.

1. Dependence (1). With a value of permissible error of approximation in the stream of compressed information with number Π of 1 gradation[5], probability (1) can be calculated as follows: /94

[5]The modeling mode was selected so that a significant interrogation is an interrogation of a signal, the approximation error of which reaches or slightly exceeds one gradation (see [2]); transmission of a significant interrogation provides an approximation error of less than one gradation.

$$P^{(\Pi)}\{\varepsilon < 1\} = N_n^{(\Pi)}/N, \tag{10}$$

where $N_n^{(\Pi)}$ is the number of significant interrogations transmitted in the communications session through the Πth line[6]; N is the total number of significant interrogations arriving at the input of the AMU during the communications session. Consequently, the initial data for dependence (1) are values of the numbers of significant interrogations newly arriving and actually transmitted in each OID operation subcycle and the corresponding values of $f_n^{(\Pi)}$.

Note: In modeling the value of $F_n^{(\Pi)}$ is fixed by the number Π_n -- the number of the interrogation transmitted in one cycle of operation of the switching device[6] K1 (or K2).

2. Dependences (2). When the value of permissible approximation error in the stream of compressed information numbered Π is equal to one gradation, probability (2) can be calculated as follows:

$$P^{(1)}\{\varepsilon^{(1)} < 1\} = N_n^{(1)}/N, \tag{11}$$

where $N_n^{(1)}$ is the number of significant interrogations transmitted through the 1th channel during the communications session.

3. Dependences (3). These dependences can be produced directly from modeling. During modeling, in each subcycle of operation of the OICD, we must fix the number of significant interrogations newly arriving and actually transmitted.

4. Dependence (4) is similar to dependence (1); the difference is in the values of E_{AMU}. Therefore, the initial data for dependence (4) are also the values of $N_n^{(n)}$ and N and the corresponding values of Π and τ; however, when each of the modeling modes is assigned, the corresponding value of E_{AMU} is established.

5. Dependences (5) are similar to dependences (2); the differences are in the values of E_{AMU}. Therefore, the initial data for dependences (5) are also the values of $N_n^{(1)}$ and N and the corresponding values of Π and τ, but when each of the modes is fixed, the corresponding value of E_{AMU} is established.

---

[6]See below.

6. Dependence (6). To produce dependence (6), we require the values of K_1, K_2, ..., K_q in each of the subcycles n of operation of the OICD for the infinite and finite values of E_{AMU}.

The initial data for the values of K_1, K_2 ..., K_q and correspondingly for dependence (6) are the values of errors of approximation ε, arriving from the output of the information collection and approximation error calculation device (in modeling -- at the input of the zone of significant interrogations): ε_1, ε_2, ..., ε_q.

7. Dependence (7). To produce dependence (7), we must know K_{1y}, K_{2y}, ..., K_{qy} in each of the n subcycles of operation of the OICD for the infinite and finite values of E_{AMU}. To produce the values of K_{1y}, K_{2y}, ..., K_{qy}, we must, in turn, know the approximation errors ε_y of the ordered significant interrogations in each subcycle n. Thus, the initial data for dependence (7) are the values of ε_{max}; ε_{max-1}, ..., $\varepsilon_{max} - (E_{AMU} - 1)$. For preliminary, rough estimates, we can limit ourselves to the values of ε_{max}; ε_{max-1}, ..., $\varepsilon_{max-\Pi}$.

8. Dependences (8). The initial data for determination of $(f_n)_{AS}$ are the same as for dependence (1) but for $P^{(\Pi)}\{\varepsilon < 1\} = P_0$; $E_{AMU} = \infty$. In order to determine $(f_n)_{BS}$, we must know the values of N_n -- the number of significant interrogations in one subcycle of OICD operation, -- where $\Pi \geq P_{OSI_{max}}$; $E_{AMU} = \infty$. Obviously, /95

$$(f_n)_{BS} = M\{N_n\}/\tau, \tag{12}$$

where $M\{N_n\}$ is the mathematical expectation of random quantity N_n; τ is the duration of one OICD subcycle.

The initial data for determination of $(f_n)_{LS}$ are the same as for (1), but with $P^{(\Pi)}\{\varepsilon < 1\} = P_0$ and with Π equal to 1.

9. Depencence (9). The initial data for $(f_n)_{AS}$ and $(f_n)_{BS}$ are the same as for dependence (8), but the corresponding values of E_{AMU} are established when the mode is fixed.

10. To restore the initial information from the compressed information while calculating the errors in the restored information, we must know the values of the "transmitted" significant interrogations B, with an indication of the channel number and the OICD operating subcycle number.

Thus, to produce dependences (1)-(9) and allow restoration of the initial information from the compressed information and calculation of the errors in the restored information, during modeling we must print out the following data:

1) ε_{max}, ε_{max-1}, ..., $\varepsilon_{max-\Pi}$, $\varepsilon_{max-(\Pi+1)}$ ... $\varepsilon_{max} - (E_{AMU-1})$;
2) $P_{OSI\ max}$;
3) ε;
4) values of significant interrogations B.

All of these quantities should be printed with an indication of the channel number and OICD operating subcycle number (or quantities which unambiguously define this number).

Let us go over to analysis of the ADAS algorithm. As was noted in [1], the ADAS algorithm refers to a rule, used to convert information from the beginning of interrogation of signals from message sources (MS) to the completion of transmission of significant interrogations in each OICD operating cycle, analyzed integrally (without division into subcycles). A flow chart of an algorithm for an ADAS is presented in Figure 1. On this figure, n is the OICD operating cycle number. However, it is not /96 possible to use the ADAS algorithm directly for development of a flow chart for a program modeling the process of reduction of redundancy -- we must consider the specifics of the information processed, consisting in the following:

1) In place of continuous signals from MS, we have discrete readings produced by interrogation during the operation of commutators K1 and K2;
2) The readings are quantized into 16 levels (gradations);
3) The repetition frequency of interrogations is either equal to the length of an operating cycle of the commutator T_{K1} (or T_{K2}) or less than this time. Equality occurs when the signal from the MS is fed to only one of the channels of the commutator K1 (K2), inequality -- if the MS is connected to several channels; this type of connection is called paralleling of commutator channels.

The presence of discrete readings prevents unambiguous determination of the value of a signal from a MS in the interval between neighboring readings. However, with ADAS, these values are needed. The problem is that transmission using commutator K1 (K2) "allows" all f signals[7] from the MS to change from cycle to cycle. OICD with ADAS can transmit only Π channels per OICD

[7]Actually, there are not f MS, but fewer, since some of the channels are used to transmit service information.

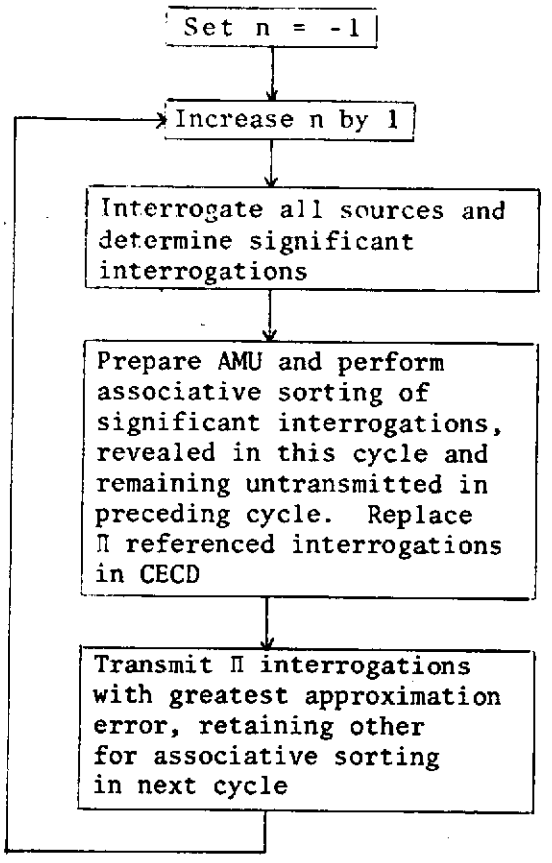


Figure 1. Flow Chart of ADAS Algorithm

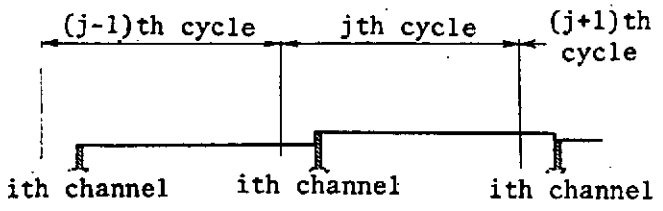


Figure 2. Interpolation of Values from MS in Interval Between Neighboring Interrogations (Showing ith Channel)

subcycle, where $\Pi < f$. When modeling OICD operation, each ensemble of f interrogations (one cycle of K1 or K2) must therefore be interrogated several times, i.e., we must imitate an n-tiple reduction in the length of the OICD cycle in comparison to the length of the cycle of commutator K1. However, this is sensible if the values of interrogations with each OICD subcycle correspond to the actual values at the moment of interrogation in each OICD subcycle, which we do not directly have available. A simple and reliable method of solution of this problem has been suggested. We can use zero-order extrapolation to interpolate the values in the interval between neighboring interrogations: we will consider that the signal in the MS changes suddenly, precisely at the moment of each interrogation, when the change relative to the preceding interrogation is fixed (Figure 2). This means that within the jth cycle of operation of commutator K1 (K2), the level of the signal of this MS retains a value equal to the value of the interrogation in the (j - 1)th cycle up to the moment of the next interrogation, then beginning with the moment of this interrogation to

the end of the cycle, it retains the value of the interrogation in the jth cycle. With this algorithm of interpolation, the f channels can be divided into n groups of m channels each ($n = f/m$). The first OICD operation subcycle with interrogation of all MS is replaced by interrogation of MS 1 through m (since the levels can change during the first subcycle in channels 1 through m, retaining the values of the previous cycle of operation of commutator K1 in the other channels) the second subcycle involves interrogation of MS (m + 1) through 2m (the level of signals in the MS from 1 through m is equal to the values in the first subcycle of the OICD, from (2m + 1) through f, to the level in the preceding cycle of operation of commutator K1), etc.

Quantization of the readings into 16 levels makes it necessary to take values of at least one gradation as $\varepsilon_{per} = \varepsilon_0$.

The presence of parallel channels requires that the values of reference interrogations be replaced for all parallel channels in a given group when a significant interrogation appears (and is /97 transmitted) in any of the channels of the group.

Figure 3 shows a flow chart for a program based on the ADAS algorithm, the list of required initial data, considering the information actually processed. This figure uses the following symbols and abbreviations:

OMU -- operative memory unit of computer;

i -- number of channel of commutator K1 (K2);

j -- number of operating cycle of commutator K1 (K2);

n -- number of subcycle of OICD operation in cycle of operation of commutator K1 (K2);

m -- number of channels of commutator K1 (K2) interrogated in one subcycle of OICD operation;

A_j -- address of memory location containing current value of j;

A_i -- address of memory location containing value of i;

A_m -- address of location containing current value of m;

$\in$ -- "belongs to";

LPC -- lists of parallel channels; number of lists equal to number of groups of parallel channels (LPC-1, LPC-2, LPC-3 respectively);

$A(B)_{oo}^{(LPC)}$ -- list of addresses of common reference interrogations, for example $A(B)_{oo}^{(1)}$ is the list of addresses of common reference interrogations of LPC-1;

ε_{per} -- maximum permissible approximation error;

P_{OSI} -- current (resulting) value of number of interrogations in OSI zone;

$P_{OSI\ max}$ -- maximum value of P_{OSI} in a given OICD operating subcycle; $P_{OSI} = P_{OSI\ max}$ before beginning of transmission of Π OSI;

$P_{OSI\ max\ max}$ -- maximum value of $P_{OSI\ max}$ in a given file of information, i.e., in a given communications session;

SI -- significant interrogation;

Zone of SI -- zone in OMU in which SI are recorded;

OSI -- ordered significant interrogations;

Zone of OSI -- zone in OMU in which OSI are recorded;

Π -- number of OSI which should be output ("transmitted") in each OICD operating subcycle;

AS -- address of OMU location in which each subsequent SI is written;

AAS -- AS address counter (when initial conditions are written, address of location AS_1 is written in AAS; the words "add one to AAS" mean to bring the AAS from value AS_k to value AS_{k+1};

FP -- current value of OSI transmitted;

AFP -- FP counter;

AU -- address of location for reading of next OSI;

AAU -- AU counter;

TMF -- telemetric functions (file of interrogations of MS).

We should note that the flow chart of the program is written for the version in which zero order extrapolation is used to determine the significance or redundancy of an interrogation.

Before analyzing the operation of the program, let us study the individual operations. The operation of the OICD (for each file of information, i.e., for the information from the commutator during one communications session) consists of one search cycle and the working cycles. Each working cycle consists of the following main operations:

1) determination of significant interrogations and their recording into the SI zone; during this operation, the "reference" interrogation for the ith channel is the value $(B_i)_0$ of the last "transmitted" (see below) interrogation from this channel;

2) sorting of significant interrogations and their recording in the order of decreasing approximation error ε in the OSI zone;

3) "transmission" (i.e., printout) of Π OSI (if $\Pi < P_{OSI\ max}$) or $P_{OSI\ max}$ (if $\Pi > P_{OSI\ max}$);

4) replacement of reference interrogation in "transmitted" channels[8]; /99

5) erasure of "excess" interrogations in SI zone;

6) return to the initial state before subsequent subcycle of OICD operation.

---

[8]Figure 3 does not show the logic which changes reference interrogations (see [1]).

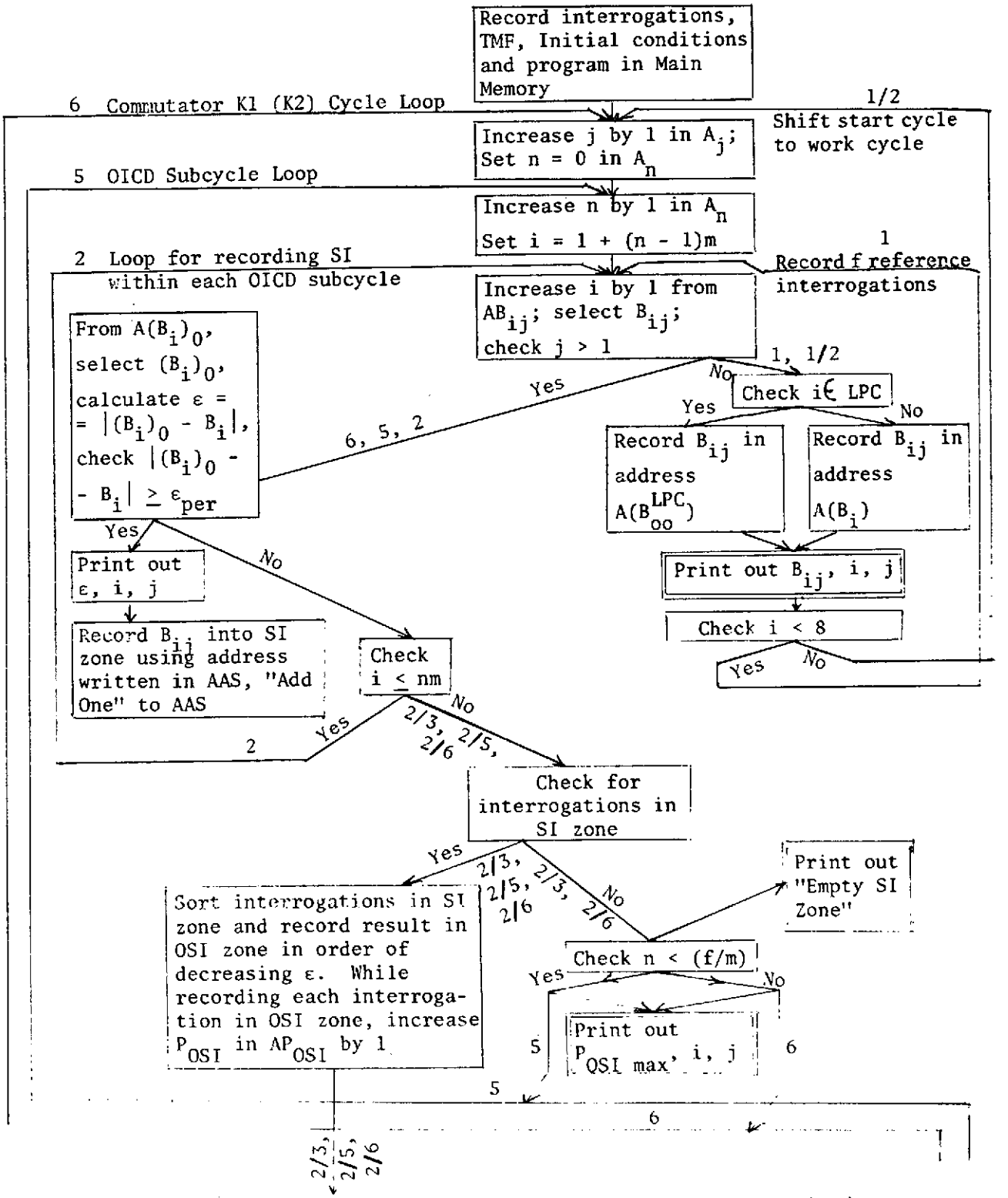
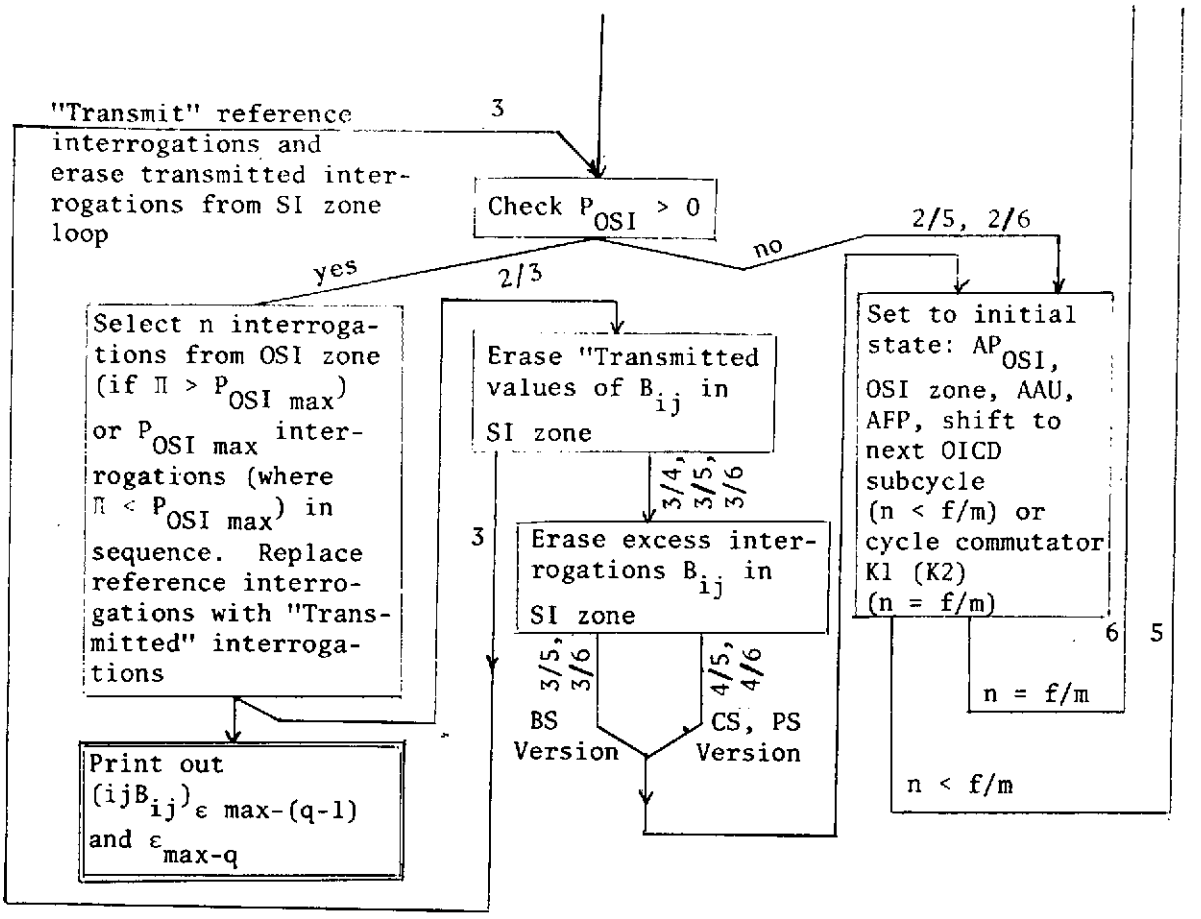


Figure 3. Expanded Flow Chart of Program for Modeling ADAS Process. (Figure continued on following page.)



Let us discuss briefly the operations of erasure of "transmitted" and "excess" interrogations, since the remaining operations require no special explanation (see [1]). These operations are necessary in order that only the SI which were not transmitted will participate in subsequent subcycles of OICD operation ("erasure of transmitted SI"), as well as those which are redundant in the type of mode being used ("erasure of excess SI"). There are three versions of OICD operation: BS, CS and PS.

In the BS version, only the transmitted interrogations are erased from the SI zone, all others participating in subsequent sorting cycles, i.e., there are no "excess" interrogations. In the CS version, interrogations written into the OSI zone beyond the limits[9] of E_{AMU} are considered "excess." In the PS version,

[9]Strictly speaking (see [1]), beyond the limits of $E_{AMU} - N_{SI}$, where N_{SI} is the number of SI prepared for input to AMU as a result of the simultaneous subcycle of operation of the CECD.

all untransmitted interrogations are considered "excess" (in this mode $E_{AMU} = E_{\Pi}$, where $E_{\Pi} = \Pi$).

The start cycle amounts to recording of all interrogations of the first cycle of operation of commutator K1 (K2) into OMU as the reference interrogations.

The program (see Figure 3) operates as follows.

1. Write the f reference interrogations into OMU; $B_{1,1}$, $B_{1,2}$, ..., B_{1f}; this is the start cycle of the OICD. Loop 1.

2. Transition to first working suncycle of OICD. Shift 1/2.

3. m interrogation subcycles, determination of significance and writing of SI into SI zone. Loop 2.

4. If the operations of loop 2 reveal one or more SI, then:
4.1. Sorting of SI and recording of OSI in OSI zone. Shift 2/3.
4.2. "Transmission" subcycles, replacement of reference interrogation and erasure of transmitted interrogation in SI zone. Loop 3. Operation in Loop 3 is interrupted when either FP becomes equal to Π, or P_{OSI} becomes equal to 0.
4.3. If work on loop 3 is interrupted by FP = Π (case when $\Pi < P_{OSI\ max}$), the course of future operation depends on which OICD version is used.
4.3.1. If it is version BS, then:
4.3.1.1. Shift zones OSI, AAU and AFP to initial state. Shifts 3/5 (if occurs during OICD operating subcycle n = f/m, shift 3/6).
4.3.1.2. Shift to next OICD subcycle (if occurs during OICD subcycle n = f/n, shift to next cycle of commutator K1 (K2).
4.3.2. If versions CS or PS, before returning to initial state and shifting to next subcycle of OICD (or K1 (K2)), "excess" values of B_{ij} in SI zone are erased:
4.3.2.1. Erasure in SI zone, loop 4;
4.3.2.2. Return to initial state, shift 4/5 (or 4/6);
4.3.2.3. Operations as described in 4.3.1.1 and 4.3.1.2.
4.4. If operations in loop 3 are interrupted where $P_{OSI} = 0$ (case when $P > P_{OSI\ max}$), subsequent operation is continued according to paragraphs 4.3.1.1 and 4.3.1.2. Shift 2/5 (2/6).

5. If operations in loop 2 produce no SI, and if there are no SI from preceding subcycles of OICD operation (of course, not during the first working subcycle, but during the second or subsequent subcycles), a shift is made to the next subcycle of OICD operation; shift 2/5 (or to the next cycle of commutator K1 (K2), /100
if it occurs during the OICD subcycle with number n = f/m, shift 2/6).

Operations according to paragraphs 2-5 form one working subcycle of the OICD. The shift from one working subcycle to the next is over loop 5. n operating cycles in loop 5 correspond to one cycle of operation of commutator K1 (K2). The shift from one cycle of K1 (K2) to the next is over loop 6. U cycles over loop 6 correspond to a file of information, i.e., the information produced in one communications session.

The results of modeling are presented in Figures 4, 5, 6, 7, 8 and in Tables 1 and 2.

Figure 4 shows dependences (1) and (2) for $\Pi = 3$.

Figure 5 shows dependence (3) for λ.

Figure 6 shows dependences (3) for Q where $\Pi n = 20$ and 16.

Figure 7 shows dependences (4) and (5).

TABLE 1

Output Stream (Values of K_ε)$\cdot \Pi n = 24$. $E_{AMU} = \infty$

	Cycle	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Error	1	1	1	4		4	2	2	2	2	2	2	8	13	7	8	7	8	7	11	9
	2										5	5	2		2		1	1	1	2	1
	3									3	7	7		2	2	2		3			
	4												2				2		2		3
	5												1	1		1			1	1	
	6													1	1	1					1
	7		1								3	3						1		1	
	8										1	1									
	9												1								
	10										4	4									
	11										3	3				1					
	12										1	1									
	13											2									
	14																				

TABLE 1 (Continued)

Cycle		22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Error	1	6	12	9	7	16	10	7	11	10	8	8	5	9	6	10	11
	2	1		2	1		1			1				1	1	1	
	3		1	2								1		2	2		
	4	2	1			1			1		2		3	1			
	5	1		1	1			1						1	2	1	2
	6					2	1	2	1	2			2	1	1	1	
	7				1		2	1	1			1	1	2	1		2
	8											1				2	
	9				1								1			1	2
	10										3				1		
	11				2							2					
	12																
	13																
	14																

NOTE. Empty spaces in Table correspond to 0.

TABLE 2 /101

Output Streams (Values of ε). $\Pi n = 24$. $E_{AMU} = \infty$

Cycle		2							
Subcycle		1	2	3	4	5	6	7	8
Stream	3								1
	2								1
	1								1

Cycle		3							
Subcycle		1	2	3	4	5	6	7	8
Stream	3					7			1
	2					7			1
	1					7			1

Cycle		4							
Subcycle		1	2	3	4	5	6	7	8
Stream	3								
	2								
	1								

TABLE 2 (Continued)

Cycle		5							
Subcycle		1	2	3	4	5	6	7	8
Stream	3								
	2								
	1								

Cycle		6							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1,1	1						1
	2	1,1	1						1
	1	1	1						1

Cycle		7							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1							1
	2	1							1
	1	1							1

/102

Cycle		8							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1							1
	2	1							1
	1	1							1

Cycle		9							
Subcycle		1	2	3	4	5	6	7	8
Stream	3		1						1
	2		1						1
	1		1						1

Cycle		10							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1	1				3	3	
	2	1	1				3	3	
	1	1	1				3	3	

TABLE 2 (Continued)

Cycle		11							
Subcycle		1	2	3	4	5	6	7	8
Stream	3		1		11, 3, 3	7, 7, 3	12, 11, 11	10, 10, 10	10, 8, 7
	2		1		11, 3	7, 7	12, 11	10, 10	10, 8
	1		1		11	7	12	10	10

Cycle		12							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	3, 3, 3	11, 5, 5	13, 13, 12	10, 8, 7	5, 5, 5	4, 3, 3	6, 4, 3	3, 2, 2
	2	3, 3	11, 5	13, 13	10, 8	5, 5	4, 3	6, 4	3, 2
	1	3	11	13	10	5	4	6	3

Cycle		13							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	2, 2, 2	2, 2, 2	9, 5, 1	2, 2, 2	2, 2, 1	4, 1, 1	4, 1, 1	1, 1
	2	2, 2	2, 2	9, 5	2, 2	2, 2	4, 1	4, 1	1, 1
	1	2	2	9	2	2	4	4	1

/103

Cycle		14							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1, 1, 1	1, 1, 1	5, 1, 1	1, 1, 1	6, 1, 1	3, 1, 1	3, 1, 1	1, 1, 1
	2	1, 1	1, 1	5, 1	1, 1	6, 1	3, 1	3, 1	1, 1
	1	1	1	5	1	6	3	3	1

Cycle		15							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1, 1, 1	1, 1, 1	6, 1, 1	2, 2		3, 1, 1	3, 1, 1	
	2	1, 1	1, 1	6, 1	2		3, 1	3, 1	
	1	1	1	6	2		3	3	

TABLE 2 (Continued)

Cycle		16							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1	5, 1, 1	6	1, 1		11, 3, 1	3, 1, 1	
	2	1	5, 1	6	1, 1		11, 3	3, 1	
	1	1	5	6	1		11	3	

Cycle		17							
Subcycle		1	2	3	4	5	6	7	8
Stream	3		1	2	1		4, 1, 1	4, 1, 1	1
	2		1	2	1		4, 1	4, 1	1
	1		1	2	1		4	4	1

Cycle		18							
Subcycle		1	2	3	4	5	6	7	8
Stream	3			1, 1	2, 1	7	3, 1, 1	3, 1, 1	1
	2			1, 1	2, 1	7	3, 1	3, 1	1
	1			1	2	7	3	1	1

Cycle		19							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1	2	5, 1	1, 1		4, 1	4, 1	
	2	1	2	5, 1	1, 1		4, 1	4, 1	
	1	1	2	5	1		4	4	

/104

Cycle		20							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1, 1	1	5, 1	1, 1	7	3, 1, 1	3, 1	1, 1
	2	1, 1	1	5, 1	1, 1	7	3, 1	3, 1	1, 1
	1	1	1	5	1	7	3	3	1

TABLE 2 (Continued)

Cycle		21							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1	1, 1	4, 1	6, 2		4, 1, 1	4, 1, 1	1
	2	1	1, 1	4, 1	6, 2		4, 1	4, 1	1
	1	1	1	4	6		4	4	1

Cycle		22							
Subcycle		1	2	3	4	5	6	7	8
Stream	3		2, 1	1, 1, 1	5, 1, 1		4	4	
	2		2, 1	1, 1	5, 1		4	4	
	1		2	1	5		4	4	

Cycle		23							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1, 1	1	4, 1, 1	1, 1		1, 1	3, 1	1
	2	1, 1	1	4, 1	1, 1		1, 1	3, 1	1
	1	1	1	4	1		1	1	1

Cycle		24							
субцикл		1	2	3	4	5	6	7	8
Stream	3	1, 1	1, 1	1, 1, 1	2, 2, 1		5, 3, 1	3	
	2	1, 1	1, 1	1, 2	2, 2		5, 3	3	
	1	1	1	2	2		5	3	

Cycle		25							
Subcycle		1	2	3	4	5	6	7	8
Stream	3		7	9, 5, 1	2, 1		11, 1, 1	11, 1	1, 1
	2		7	9, 5	2, 1		11, 1	11, 1	1, 1
	1		7	9	2		11	11	1

TABLE 2 (Continued) /105

Cycle		26							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	4, 1	4, 1, 1	1, 1, 1	1, 1, 1		6, 1, 1	6, 1, 1	1, 1
	2	1, 1	4, 1	1, 1	1, 1		6, 1	6, 1	1, 1
	1	1	4	1	1		6	6	1

Cycle		27							
Subcycle		1	2	3	4	5	6	7	8
Stream	3		1, 1, 1	1	2, 1		6, 1, 1	7, 1, 1	1
	2		1, 1	1	2, 1		6, 1	7, 1	1
	1		1	1	2		6	7	1

Cycle		28							
Subcycle		1	2	3	4	5	6	7	8
Stream	3		1, 1	5, 1	1, 1, 1	7	6	6	1
	2		1, 1	5, 1	1, 1	7	6	6	1
	1		1	5	1	7	6	6	1

Cycle		29							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1, 1	1	4, 1	1, 1		6, 1, 1	7, 1	1, 1
	2	1, 1	1	4, 1	1, 1		6, 1	7, 1	1, 1
	1	1	1	4	1		6	7	1

Cycle		30							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1, 1, 1	1, 1, 1	1	2, 1		6, 1, 1	6	
	2	1, 1	1, 1	1	2, 1		6, 1	6	
	1	1	1	1	2		6	6	

TABLE 2 (Continued)

Cycle	31							
Subcycle	1	2	3	4	5	6	7	8
Stream 3		4, 1	10, 4, 1	1, 1, 1		10, 1, 1	10	1
Stream 2		4, 1	10, 4	1, 1		10, 1	10	1
Stream 1		4	10	1		10	10	1

/106

Cycle	32							
Subcycle	1	2	3	4	5	6	7	8
Stream 3		1	3, 1	1, 1, 1	8	11, 7, 1	11, 1	1
Stream 2		1	3, 1	1, 1	8	11, 7	11, 1	1
Stream 1		1	3	1	8	11	11	1

Cycle	33							
Subcycle	1	2	3	4	5	6	7	8
Stream 3	1	1	7, 6, 1	9, 6, 4	1	4, 1	4	
Stream 2	1	1	7, 6	9, 6	1	4, 1	4	
Stream 1	1	1	7	9	1	4	4	

Cycle	34							
Subcycle	1	2	3	4	5	6	7	8
Stream 3	1	3, 1	6, 4, 2	7, 3, 1	1	5, 1, 1	7, 1, 1	1
Stream 2	1	3, 1	6, 4	7, 3	1	5, 1	7, 1	1
Stream 1	1	3	6	7	1	5	7	1

Cycle	35							
Subcycle	1	2	3	4	5	6	7	8
Stream 3		1, 1	5, 5, 2	6, 3, 1		7, 3, 7	10, 1	1
Stream 2		1, 1	5, 5	6, 3		7, 3	10, 1	1
Stream 1		1	5	6		7	10	1

TABLE 2 (Continued)

Cycle		36							
Subcycle		1	2	3	4	5	6	7	8
Stream	3	1	1, 1				8, 7, 1	8, 1, 1	
	2	1	1, 1				8, 7	8, 1	
	1	1	1				8	8	

Cycle		37							
Subcycle		1	2	3	4	5	6	7	8
Stream	1	1	9, 7, 1	5, 5, 1	9, 1, 5		1, 1, 1	7, 1, 1	1
	2	1	9, 7	5, 5	9, 1		1, 1	7, 1	1
	3	1	9	5	9		1	7	1

NOTE. Empty spaces in Table correspond to 0.

/107

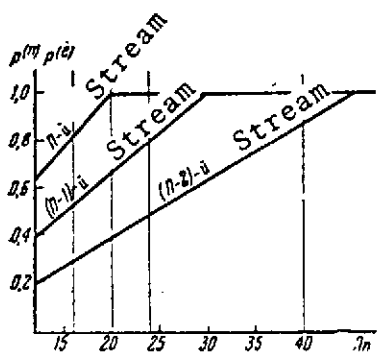


Figure 4. Probability of Transmission of Significant Interrogations as a Function of Transmission Rate ($\Pi = 3$, $E_{AMU} = \infty$)

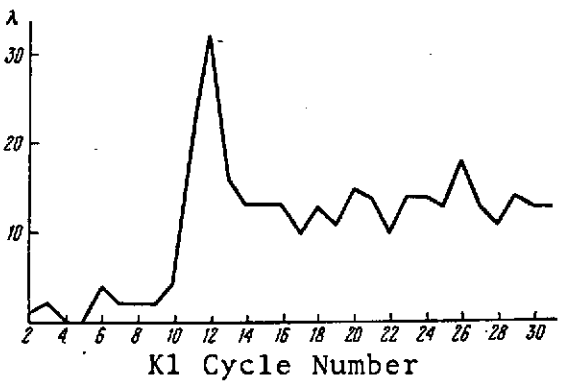


Figure 5. Change in Density of Stream of Significant Interrogations at input to AMU as a Function of Time

To illustrate the results for the initial data for dependences (6) and (7), Table 1 presents values of K_ε, Table 2 -- values of ε_{max}, ε_{max-1}, ..., $\varepsilon_{max-\Pi}$ for $\Pi n = 24$, $\Pi = 3$, $E_{AMU} = \infty$; the data in Table 2 are presented separately for each of the streams. Figure 8 shows data for estimation of $(f_n)_{AS}$ and $(f_n)_{LS}$ for $E_{AMU} = \infty$ (dependence (8)) and for $E_{AMU} \neq \infty$ -- for estimation of $(f_n)_{AS} = \psi_{AS}(E_{AMU})$ (dependence (9)).

The results of modeling lead to the following conclusions.

1. The method of adaptive discretization with associative sorting (ADAS) can produce several streams of compressed information at the outputs of a single conversion system, for transmission through channels with various throughput capacities (see Figure 4 and Figure 7).

2. The method of modeling can produce initial data for determination of the dependences characterizing the operation of an on-board compression device with ADAS, for restoration of the initial information from the compressed information and for analytic investigation and statistical modeling of the ADAS process (see Figures 4-8 and Tables 1 and 2).

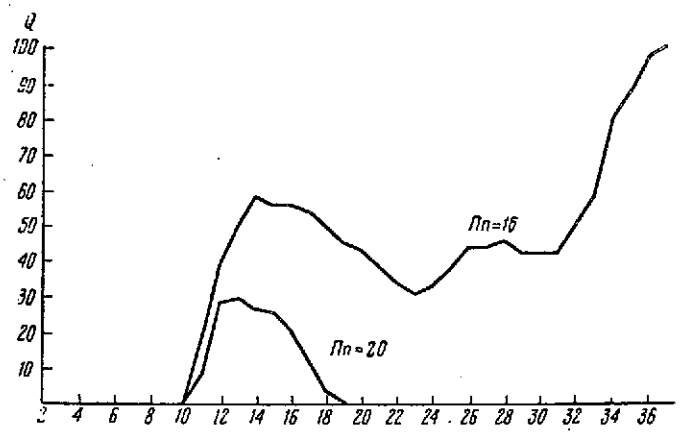


Figure 6. Change of Line Length with Time (Пп = 20 and 16)

3. The actual telemetric information used in modeling has a number of shortcomings (for investigation of the ADAS method):

1) the information is quantized into 16 amplitude gradations; when the permissible error in approximation is one gradation (minimum possible value), it is impossible to realize the potential advantage of the associative sorting system in comparison to a buffer smoothing /108 system, consisting in the lower required value of memory capacity (due to sorting by values of approximation error); therefore, in the results for relationships (8) and (9) for ADAS and ADBS systems, identical data were produced (see Figure 8);

2) the information is quantized with respect to time (which, incidently, is true of all natural multichannel telemetric information with time division of channels); it includes parallel channels; this complicated the modeling program and limited the capabilities for arbitrary selection of modeling modes; for future studies, we must recommend more precise information; this information should be converted in advance to analog form.

4. The system with associative sorting, like the system with buffer smoothing, provides significantly greater compression of information (ignoring service information) than the system with logical switching (see Figure 8).

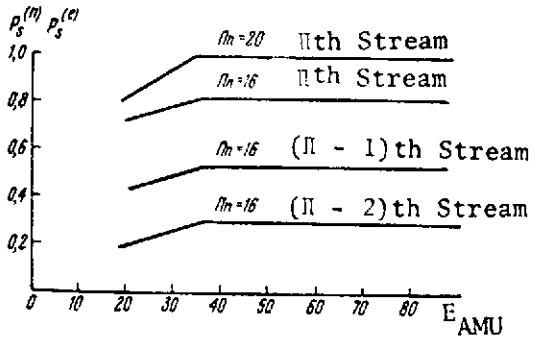


Figure 7. Probability of Transmission of Significant Interrogations as a Function of Value of E_{AMU} (Пn = 20 and 16, Пn = 3)

In conclusion, the authors express their gratitude to their working comrades, who were of great help in the modeling work, most of all to I. G. Kostyuk and R. A. Sagitov, whose help and valuable advice greatly facilitated successful completion of the work.

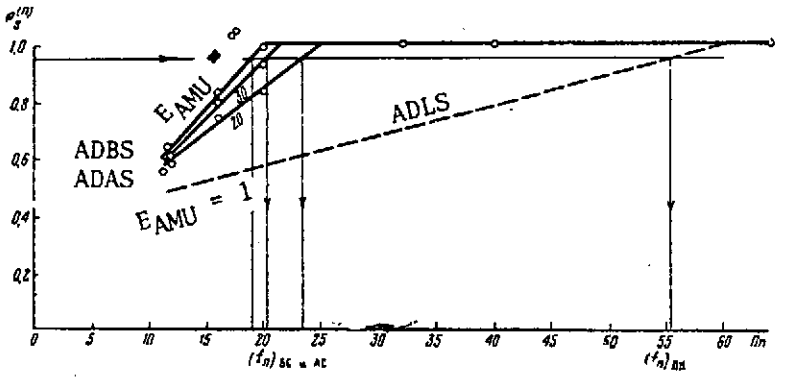


Figure 8. Required Values of Transmission Rate

REFERENCES

1. Kantor, A. V., V. G. Timonin, Yu. S. Azarova, "Reduction in Redundancy of Multichannel Telemetric Information by Adaptive Discretization with Associative Sorting," This collection, p. 114.
2. Kantor, A. V., T. A. Tolmadzheva, "Study of the Redundancy of the Telemetric Information of the Luna-9 and Luna-13 Automatic Lunar Probes," This collection, p. 80.

N75 11001

CIRCULARLY POLARIZED ANTENNAS WITH CONTROLLED RADIATION PATTERNS

B. A. Prigoda, V. A. Gur'yev, N. S. Ul'yanin
and N. N. Kalugin

One of the basic requirements placed on spacecraft apparatus is reliable communcations. /109

If a spacecraft is not oriented in space, omnidirectional antennas must be installed on the spacecraft to assure reliable communications; as we know, these antennas have a low gain factor. Therefore, when an omnidirectional antenna is used, it is impossible to transmit large quantities of information. In order to transmit large quantities of information, the transmitter power must be increased, causing a significant increase in apparatus weight.

As we know, at the present time many spacecraft are partially oriented. Usually, spacecraft rotate about one axis, and this axis is generally oriented toward the sun with some accuracy. Furthermore, during the flight of the spacecraft, the angle between the direction from the spacecraft to the sun and from the spacecraft to the Earth changes.

Partially oriented spacecraft carry antennas with broad radiation patterns. These antennas allow reception and transmission of information over short distances with changes in the sun-spacecraft-Earth angle over broad limits. Furthermore, another antenna is generally carried, the radiation pattern of which has a funnel shape with its maximum power radiation at a predetermined angle to the axis of the antenna (Figure 1b). Antennas of this type are used for long-range communications, when the sun-spacecraft-Earth angle changes over slight limits and almost matches the angle of the direction of the maximum radiation of the antenna. It is desirable that antennas installed on board spacecraft have a polarization which is near circular.

This requirement is rather well satisfied by log-helix conical antennas. A log-helix conical antenna is a self-supplementing structure. The design of log-helix antennas is illustrated in Figure 2. Metal log-helices are placed on the surface of a cone with half angle θ_0 at its tip. Angle θ_0 is 7-15°. The helices are described by the equations /110

$$\rho_2 = \rho_0 \cdot e^{\frac{\sin\theta_0}{\tan\alpha}\varphi},$$
$$\rho_1 = k \cdot \rho_e,$$

where $k = e^{\frac{\sin\theta_0}{\tan\alpha}\delta}$.

The excitation systems of double and quadruple helical antennas are presented on Figure 3.

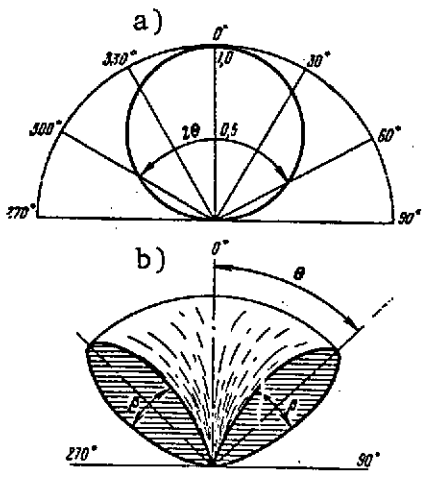


Figure 1. Radiation Patterns of Double Log-Helix Antenna (a) and Quadruple Log-Helix Antenna (b)

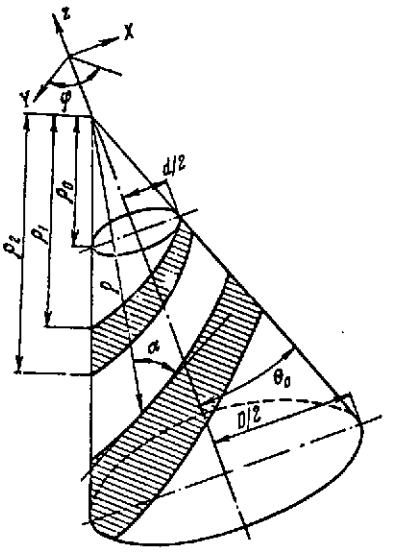


Figure 2. Design of Log-Helix Antenna for Double Antenna $\delta = 90°$ and for Quadruple Antenna $\delta = 45°$

The dimensions of antennas are selected from the following relationships:

for double-helix antenna $d \leq 0.25\ \lambda_1$; $D \geq 0.33\ \lambda_2$, where λ_2 is the wavelength at the lower frequency; λ_1 is the wavelength at the higher frequency;

for quadruple-helix antenna $D \geq 0.66\ \lambda_2$.

The double log-spiral antenna has a radiation pattern in the axial direction (Figure 1a). The width of the radiation pattern of such an antenna depends basically on the lead angle of the helix α. As this angle increases, the width of the

radiation pattern decreases. The approximate dependence of the width of the radiation pattern of a double-helix antenna on the helix lead angle is presented on Figure 4.

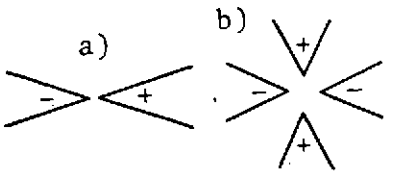


Figure 3. Diagram of Excitation of Double Log-Helix Antenna (a) and Quadruple Log-Helix Antenna (b)

A quadruple log-helix antenna has a funnel-shaped radiation pattern. The direction of maximum radiation in relationship to the axis of the antenna depends on the lead angle of the helix. This dependence is shown in Figure 5.

To provide reliable communications with increasing distance between the spacecraft and the Earth, it is desirable to increase the gain of the antenna. This can be achieved by several methods. For example, several antennas might be installed on board the spacecraft with various gain factors with successive switching of on-board transmitters and receivers to one of the antennas. This increases the weight of the spacecraft, complicates the antenna-feeder system (AFS) and reduces the reliability of the AFS.

Another way to increase the gain of an antenna is to change the width of the radiation pattern for a double-helix antenna or change the position of the maximum of radiation for a quadruple-helix antenna. In some cases, this method can reduce the number of on-board antennas required to one.

Since the width of the radiation pattern of a double log-helix antenna of conical shape with a fixed angle at the tip depends basically on the lead angle of the helix, the gain of the antenna can be increased by changing the lead angle of the helix.

Experiments were performed involving a change in the lead angle of the helix by means of a special mechanism by movement of the helix along the axis of the antenna with simultaneous twisting of the helix. The helices were held tightly against the surface of the cone. The mechanism allows the helix lead angle to be changed between approximately 45 and 46°. The radiation patterns for various helix lead angles for a double-helix antenna are shown in Figure 6, for a quadruple-helix antenna -- in Figure 7. The mechanism moving the helix allows any lead angle to be set with an accuracy of ±3°. The impedance of the antenna changes approximately between 120 and 180 ohms for double-helix antennas and 60-90 ohms for quadruple-helix antennas as the lead angle is changed. The ellipticity factor remains between 0.6 and 0.7 throughout.

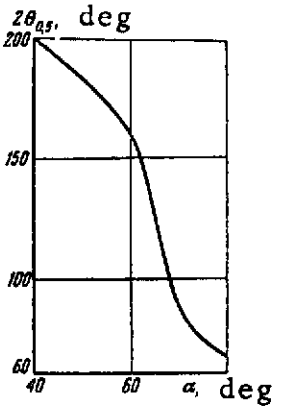


Figure 4. Width of Radiation Pattern of Double Log-Helix Antenna as a Function of Helix Lead Angle

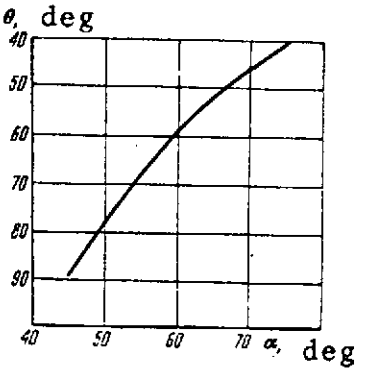


Figure 5. Deflection of Direction of Maximum Radiation as a Function of Lead Angle of Helix /111

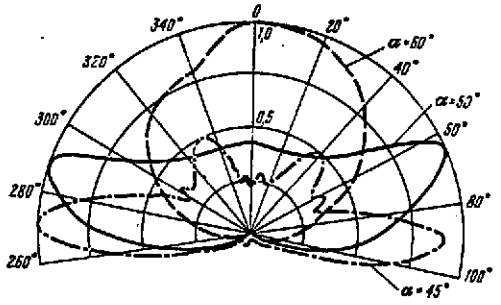


Figure 6. Radiation Pattern of Double Log-Helix Antenna with Various Helix Lead Angles

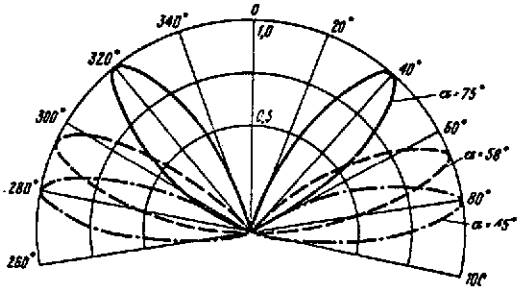


Figure 7. Radiation Pattern of Quadruple Log-Helix Antenna with Various Helix Lead Angles

The design of a multiple-helix antenna allowing the helix lead angle to be changed, due to the properties which we have mentioned and the relatively simple design, can be practically used when the radiation pattern and directivity in the direction of the reception point must be changed during a flight. /112

REFERENCE

1. Sverkhshirokopolosnyye antenny [Superwide Band Antennas], Moscow, Mir Press, 1964.

N75 11002

REDUCTION IN REDUNDANCY OF MULTICHANNEL TELEMETRIC INFORMATION BY THE METHOD OF ADAPTIVE DISCRETIZATION WITH ASSOCIATIVE SORTING

A. V. Kantor, V. G. Timonin
and Yu. S. Azarova

During the process of investigation of space and the planets with spacecraft, one of the primary sources of information is the radio telemetry system (RTS) used to transmit data from the spacecraft to the Earth. As the research tasks become more complicated, the volume of messages to be transmitted using this system during operation of the spacecraft increases. The process of increasing message volume significantly leads the process of increasing RTS information capacity.

Therefore, one of the main problems involved in the design of prospective spacecraft is the problem of reduction of the volume of telemetry messages transmitted from the spacecraft, i.e., the volume of so-called "compression" of information.

Many works on elimination of redundancy from telemetry messages [1] have pointed out that for messages characterized by the signal shape, the method of adaptive discretization is the most promising. The essence of the method of adaptive discretization consists briefly in the following. As we know, in multichannel RTS with time-division multiplexing, the radio channel is multiplexed by discretization (taking discrete readings) of continuous signals from message sources (MS).

The process of discretization is performed cyclically in this case. The period of cyclical interrogation is selected according to the value of the maximum frequency in the spectrum of the signal from the MS and remains unchanged, regardless of the instantaneous value of the frequency of change of the signal. The transition to adaptive discretization of multichannel information consists in halting cyclical interrogation of the signal from each MS and a transition to interrogation of signals from MS which change[1] to the greatest extent. The process of

[1]The term "change" is used here in its broadest sense, i.e., either a change in signal magnitude or a change in signal derivatives or a change in area limited by the signal, etc., or a combination of any of these changes; the approximation error permissible for each MS is also considered in evaluating the degree of change of a signal.

adaptive discretization can be performed by two methods, which we will call first and second order discretization. In first order adaptive discretization, interrogation of signals from MS is performed just as in cyclical discretization; however, only the so-called significant interrogations are transmitted from the output of the device, those necessary for restoration of signals from the MS in the equipment for restoration of the initial information from the compressed information (at the receiving end) with the required accuracy. In order to determine whether a given interrogation is significant or redundant (i.e., insignificant), a first order adaptive discretizer performs the following operations. All interrogations produced by cyclical discretization in the collection device are input to an approximation error calculation device (ECD). The ECD calculates the error of approximation and compares it to the permissible error (for a given MS). If the approximation error is less than the permissible error, the interrogation is redundant; otherwise it is significant. The stream of significant interrogations is uneven. Circuits in which a buffer memory unit (BMU) is used to smooth the stream of information have become common. In second order adaptive discretization, the approximation error of the signals in all MS are calculated in each operating cycle. The number of interrogations taken from the output depends on the type of second order adaptive discretization. In one adaptive discretization device which has been used, one interrogation is transmitted in each cycle, namely the one with the maximum approximation error in the cycle; the cycle length is selected so that with a fixed probability (near 1) this maximum error does not exceed the permissible error. In this type of circuit, the BMU is not used, since the output stream is even. In the rest of this article, we will refer to a first order adaptive discretization system with buffer memory as an ADBS system (adaptive discretization with buffer smoothing), to a second order adaptive discretization system with transmission of one item per cycle as an ADLS system (adaptive discretization with logical switching). /113

The commonly used adaptive discretization systems (ADBS and ADLS) have a number of shortcomings making these systems difficult to produce in on-board information compression devices (OICD) in spacecraft. Some of these shortcomings are fully eliminated, others are significantly reduced in systems involving adaptive discretization with associative sorting (ADAS), studied in this work. ADAS systems make it possible in principle to perform adaptive discretization of both first and second order.

However, in order to simplify synchronization (control) of the interrogation cycles and calculation of approximation error with the transmission cycle, it is expedient to use second order adaptive discretization in ADAS systems; in this case, in order

to reduce the required associative memory capacity, the selection of interrogations with greatest error is performed not of all interrogations of a given cycle (as in ADLS system) but rather of the interrogations, the approximation error of which exceeds a certain threshold value (see below).

The primary element of ADAS systems, the use of which gives these systems a number of specific advantages, is the associative memory unit (AMU) [2]. Development and improvement of ADAS systems is related to the use of an associative computer device (ACD), frequently called (particularly in the foreign literature) the associative processor. Some information on AMU and ACD is presented in [3], which includes a bibliography. Therefore, we will limit ourselves here to a simple notation of the properties of AMU and ACD, the use of which in ADAS systems creates a number of the significant advantages of ADAS in comparison to other systems of adaptive discretization:

1) the possibility of sorting (ordering) of information according to a so-called associative characteristic, performed simultaneously for the entire file of information (not by the sequential sorting method);

2) the possibility of selection of information according to the associative characteristic (not according to address);

3) the possibility of performing large numbers of logic operations; /114

4) the possibility (in ACD) of performing arithmetic operations;

5) the possibility (in ACD) of performing arithmetic and logic operations in parallel (simultaneously) on large numbers (several hundreds or even thousands of parallel channels) of samples using a single algorithm; the algorithm can be changed according to a program or by external instruction;

6) since AMU (ACD) include memory, naturally, they allow information to be recorded.

Property (1) is one of the most important properties for ADAS, allowing multipurpose OICD to be created. One form of multipurpose OICD is a multichannel OICD -- a compression device in which several streams of compressed information are formed at various outputs of a single line, with various values of density of the stream of interrogations and, correspondingly, various accuracies of description of the initial telemetry messages. The associative characteristic is the approximation error. The necessity of creating multichannel OICD results from a number of factors. Let us mention a few of these factors:

transmission of even compressed telemetry information over the radio line between the spacecraft and the Earth does not solve the problem of transmission of information from the spacecraft during the most dynamic flight sectors. These sectors

include maneuvering sectors (trajectory correction, deceleration, docking, etc.) and the operation of the spacecraft on the surface of the planet (particularly, movement of a self-propelled spacecraft over the surface of the planet). The necessity arises, along with transmission, of recording compressed information; the information recorded is transmitted through the communications link during static portions of the spacecraft schedule;

in addition to transmission and recording of information from the on-board telemetry apparatus, the system can be used as a part of the on-board independent spacecraft control system;

transmission of information from the spacecraft to the Earth can be performed over two or more radio links simultaneously, with various link throughput capacities;

information from on-board telemetry apparatus can be used to indicate the status of the system on board piloted spacecraft.

In order to create multichannel OICD using the common adaptive discretization systems (ADBS and ADLS), the number of parallel links (full or partial) must be equal to the number of streams of compressed information. This design leads to impossibly cumbersome circuitry. In multichannel OICD using ADAS, the formation of a number of streams of compressed information occurs in a single link. At the output of this common link, in each cycle a file (list) of significant interrogations is formed, ordered (sorted) in the order of decreasing approximation error: the interrogation with the greatest error stands in first "place" in the line, the interrogation with the least approximation error stands in last place. Of this list, only the smallest portion is included in the first stream (with greatest error), beginning with the first line, the rest of the list falling into the later lines; an intermediate number of lines of the list falls into the streams with intermediate numbers (always beginning with the first line).

The combination of properties (1), (2) and (6) allows ADAS to use AMU as BMU. Elimination of the BMU as a circuit element results in a significant savings in comparison with ADBS systems. ADAS systems do have functional buffer memories. This allows the required repetition frequency of useful information at the output of the OICD to be reduced (i.e., the compression factor increased) in comparison to ADLS systems[2]. The problem is that in ADLS systems, the frequency of repetition of interrogations in the stream of compressed information is determined by the /115
maximum rate of change of information in the file transmitted, while in ADAS systems (as in ADBS systems) it is determined by

[2]It should be noted that service information in ADLS occupies a smaller volume.

the mean rate of change. Furthermore, functional BMU in an ADAS system has one significant advantage over the hardware BMU used in ADBS systems: when the AMU is overfilled (as it performs the function of BMU), the least significant interrogations are lost. In AMU with sorting at the input (see below), this property results from the very principle of sorting at the input of the AMU; in AMU with ordered selection (see below) this property may be realized, for example, as follows: when the state near overflow is reached, an ordered selection program is activated, so that interrogations with the least approximation error are erased as the AMU is overfilled.

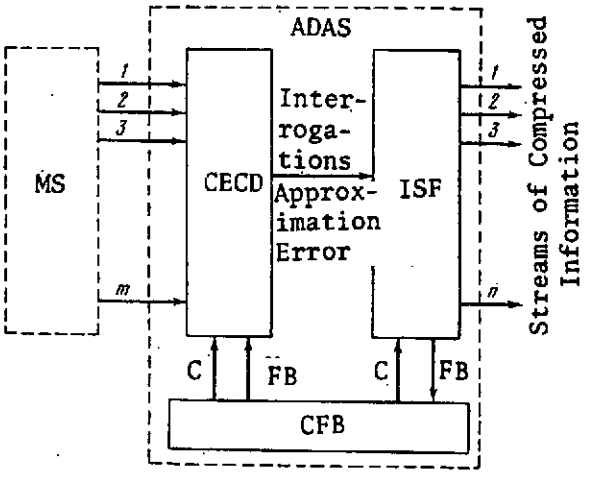


Figure 1. ADAS Block Diagram

Properties (2) and (3) allow simple realization of a number of logic operations involved in collection and conversion of telemetry messages in ADAS, for example such operations as collection by priorities, comparison of the values of telemetry messages from various MS, etc.

Property (4) allows a single ACD to be used both for calculation of approximation error and for information sorting. Combination of these functions into one device can produce a reduction in size, weight and power consumption, very important for OICD.

Property (5) allows ADAS to operate several orders of magnitude more rapidly with a slight increase in hardware. To achieve an equivalent increase in speed in ADBS and ADLS, the number of parallel links (full or partial) would have to equal the number of telemetry messages. An ADAS system with parallel processing is described in [3].

Before going over to classical ADAS systems and their analysis, let us briefly study the operation of OICD with ADAS using the ADAS block diagram of Figure 1. On this figure: CECD is the device for collection of information and error calculation; ISF is the device for formation of streams of information; CFB is the control and feedback unit. The primary element of the ISF is the AMU. As was noted above, the combination of CECD and ISF can be produced using an ACD.

Signals from the MS are sent to the CECD. This device performs primary discretization of signals from the MS (i.e., "collection" of information from MS), calculation of approximation error and determination of significant interrogations, i.e., interrogations of signals, the approximation error of which exceeds the fixed threshold value (see below). The CECD outputs /116 both the values of the significant interrogations and the values of errors. The ISF performs two main functions: 1) sorting of interrogations according to approximation error, and 2) smoothing of the unevenness of the stream of significant interrogation, outputting even streams of compressed information.

As we have noted, the basic element of the ISF is the AMU. The AMU consists of two memory fields: the field of associative characteristics and the field of the corresponding (main) information. When AMU are used in ADAS for sorting of information, the values of approximation error are recorded in the field of associative characteristics, while the values of significant interrogations are recorded in the field of the corresponding information. Thus, each line (location) in memory contains both the value of an interrogation and the corresponding value of approximation error. As a result of the sorting of information during associative search, each row is assigned a number in order for transmission (in a system with sorting at the input -- see below -- the significant interrogation with the maximum error is physically located in the first line, with the next lower error -- in the second line, etc.). When the first interrogation pulse arrives from the transmission section, information is transmitted into that section (for example, into the radio link), the value of the interrogation corresponding to the maximum approximation error; when the second transmit signal arrives, the interrogation with the next highest approximation error is output, etc. In all, Π interrogations are transmitted in each cycle. The untransmitted interrogations are sorted in the next cycle along with the newly arriving interrogations. If the number of significant interrogations is less than Π, insignificant interrogations are input. It must be noted that the existing circuits used in AD compression devices are particular cases of ADAS. For example, in ADBS systems, the CECD determines and outputs to the ISF the significant interrogations (approximation error not output to ISF), and the ISF is a buffer memory (BMU). In ADLS systems, the CECD performs the same functions, as the CECD in ADAS; ISF selects the interrogation of the signal from the MS with greatest approximation error (in the given operating cycle).

In ADAS, in each operating cycle the first stream is formed of the first Π_1 interrogations with errors from ε_{max} to $\varepsilon_{max} - (\Pi_1 - 1)$, where ε_{max} is the maximum approximation error

in the given OICD operating cycle; the interrogation with this error is output first; $\varepsilon_{max} - 1$ is the next lower (after ε_{max}) error; the interrogation with this error is transmitted second; $\varepsilon_{max} - (\Pi_1 - 1)$ is the next lower error (after $\varepsilon_{max} - (\Pi_1 - 2)$); the interrogation with this error is transmitted Π_1th. Obviously, $\varepsilon_{max} > \varepsilon_{max} - 1 > \varepsilon_{max} - 2 > \ldots > \varepsilon_{max} - (\Pi_1 - 1)$. The second stream is formed of the first Π_2 interrogations with errors of ε_{max} to $\varepsilon_{max} - (\Pi_2 - 1)$, where $\Pi_2 > \Pi_1$.

The third stream is formed of the first Π_3 interrogations, where $\Pi_3 > \Pi_2 > \Pi_1$, etc.; the last stream consists of all Π interrogations.

Obviously, in each cycle the first stream contains the greatest approximation error; as the stream number increases, the approximation error decreases. Accordingly, the threshold value mentioned above is determined by the required accuracy of approximation in the stream with the greatest number.

Clearly, the streams differ from each other as to density of the stream of significant interrogations. The least density is in the first stream, the greatest -- in the last stream. The density of a stream is determined by the corresponding value of Π_i. Matters are much more complex as concerns the dependence between approximation errors in the initial telemetry messages in each stream of compressed information.

Our study of ADAS can be expediently begun by producing the characteristics of operation of OICD for the last stream. This expediency is determined by two factors:

1) the threshold value used in the OICD, determined by the permissible approximation error for the last stream of compressed information; /117

2) the mathematical apparatus necessary to study all streams, related to the requirement for analysis of uneven streams; for the study of the last stream, it is sufficient to analyze only a homogeneous stream.

Actually, in studying the operation of OICD for the formation of the last stream, the duration of the cycle is selected from the following condition: the probability that on the average over the time of transmission of the compressed information (for example, during the time of one communications session with the spacecraft) the number of significant interrogations in the cycle will not exceed Π, near unity; this means that with a probability near unity, all Π interrogations in all cycles which exceed the permissible error (for the last stream) will be transmitted to the output during the time of transmission of the

compressed information; as a result of this, the "residual" error (i.e., the approximation error remaining after transmission of Π significant interrogations of all cycles during the transmission time) is equal (with a probability near unity) to the permissible error and is independent of the values of error Π in the significant interrogations. With the selected (on the basis of these requirements) duration of an OICD operating cycle, the value of approximation error in each of the remaining streams of compressed information depends not only on the corresponding values of Π_i and the distribution of moments of appearance of significant interrogations, but also on the distribution of approximation errors characterizing the significant interrogations. The study of OICD for a heterogeneous (two-dimensional) stream of significant interrogations at the input of the AMU goes beyond the scope of this article (this analysis will form the subject of a special work). Therefore, in our subsequent presentation we will study ADAS which produce one stream, keeping in mind that the stream is the last stream (i.e., with the highest number) of the series of compressed information.

One of the primary tasks of investigation of ADAS in the aspect of the stream with the greatest number is production of the dependence

$$E_{AMU} = F(\lambda,\ f\Pi = (\Pi/T),\ P\{\varepsilon_m \leq \varepsilon_{thr}\}), \qquad (1)$$

where E_{AMU} is the capacity of the AMU; $f\Pi$ is the frequency of transmission of ordered significant interrogations with the highest number (sorted in order of decreasing approximation error); T is the length of an OICD operating cycle; Π is the number of ordered significant interrogations selected per OICD operating cycle for transmission; λ is a characteristic (see below) of the stream of significant interrogations at the input to the ISF; $P\{\varepsilon_m \leq \varepsilon_{thr}\}$ is the probability that the maximum approximation error ε_m will not exceed the permissible threshold value ε_{thr}.

Analytic studies of ADAS in general form (even for one stream of compressed information) involve significant mathematical difficulties and are rather cumbersome. Therefore, it is expedient to perform:

1) approximate analytic study;
2) modeling with actual telemetry information using the approximate analytic investigation;
3) statistical modeling (statistical testing using a Monte Carlo method);
4) more complete analytic study.

In this work, we will simply study the statement of the problem of approximate analytic investigation. The problem of modeling with actual telemetry information is the subject of [4].

To perform approximate analytic investigation and modeling with actual telemetry information, we must study the possible versions of design of ADAS systems and, on this basis, select a generalized ADAS algorithm. /118

The classification of ADAS systems in the general case (regardless of the present work) should be performed using the block diagram of Figure 1, according to the following main criteria:

1) time division of two processes: the process of information collection and calculation of discretization error and the process of associative sorting of information;

2) time division of the processes of collection and calculation of signal discretization error from various MS;

3) the type of sorting of information.

Criterion (1) is used to distinguish ADAS systems: a) with time division and b) with time matching of the process of collection of information and calculation of error of discretization and the process of associative sorting of information.

Criterion (2) is used to differentiate ADAS systems: a) with successive and b) with parallel (simultaneous) calculation of discretization error of signals from different MS.

Criterion (3) is used to differentiate: a) ADAS systems with associative sorting at the input to the AMU; b) ADAS systems with sequential associative selection from AMU; these systems are then in turn divided into systems with successive writing of information in AMU and systems with parallel writing. The combination of versions of systems according to all three criteria allows us to differentiate twelve ADAS system versions (Figure 2).

As was noted above, in the aspect of this work, our greatest interest in the classification of ADAS systems is in the possible differences in ADAS algorithms. From this standpoint, the differences between systems with time division of calculation and sorting processes and systems with simultaneous performance of these processes are of no interest; the same is true concerning differences between systems with successive and parallel calculation of discretization error. /11 These differences are of great interest for determination of technical characteristics of OICD and the study of problems related to the planning and manufacture of these devices. As concerns the ADAS algorithm, we are

interested in differences between systems with sorting of the input (systems 1, 4, 7, 10) and systems with ordered selection from memory (systems 2, 3, 5, 6, 8, 9 and 11, 12). Let us study these systems in somewhat greater detail, using the example of system (1) -- sorting at the input and system (9) -- ordered selection[3].

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

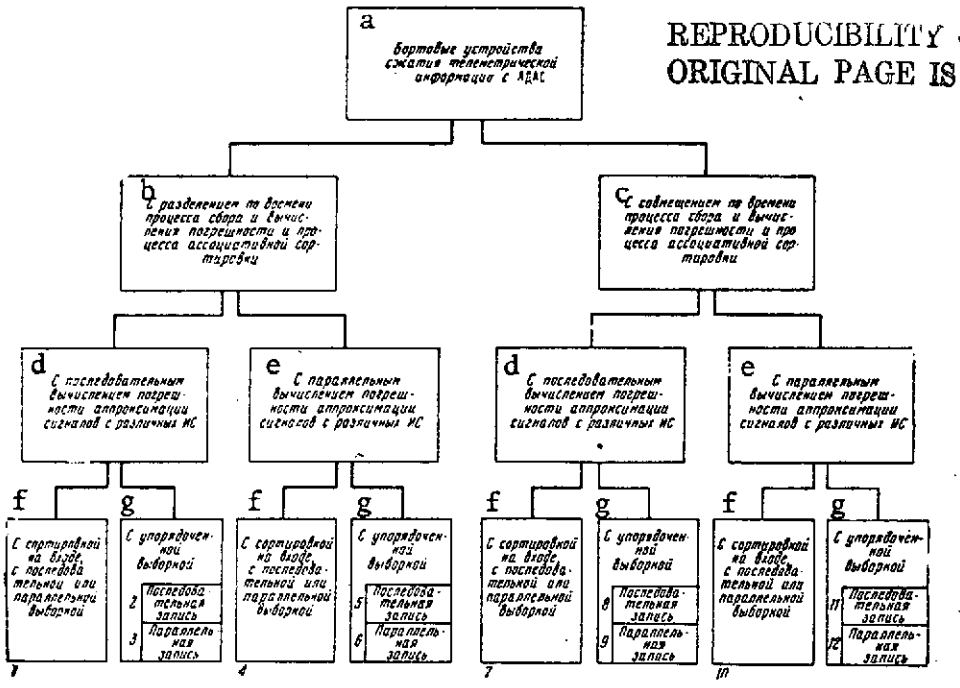


Figure 2. Classification of On-Board Compression Devices with ADAS. Key: a, On-Board Compression Devices for Telemetry Information with ADAS; b, With Time Division of the Process of Collection and Calculation of Error and the Process of Associative Sorting; c, With Simultaneous Collection and Calculation of Error and Associative Sorting; d, With Successive Calculation of Approximation Error for Signals from Different MS; e, With Parallel Calculation of Approximation Error of Signals from Different MS; f, With Sorting at the Input with Successive or Parallel Output Selection; g, With Ordered Selection; 2, 5, 8, 11, Sequential Writing; 3, 6, 9, 12, Parallel Writing

[3]These systems were selected for generality of discussion, as the easiest to understand.

A block diagram of an OICD according to plan (1) and the corresponding cyclogram are presented in Figure 3; a block diagram of an OICD according to plan (9) and the corresponding cyclogram are presented in Figure 4. In both cases, the OICD operates cyclically. The cycle length is equal to the transmission time of 2P significant interrogations. Each cycle consists of two subcycles of equal length.

The first is the subcycle of collection, calculation of error (determination of significance) and sorting. The second is the subcycle of transmission of the Π significant interrogations with the greatest approximation error. During sorting, the signals from all MS are successively interrogated and the approximation error is calculated after each interrogation. If the interrogation is significant, it is sorted: according to the value of approximation error, the "place" of the interrogation in the "list" of preceding significant interrogations is determined; the list is made up in the order of decreasing approximation error. The operation of sorting of information at the input of the AMU is described in the literature (see [5], for example); therefore, we will not discuss it.

It may be found that by the moment of sorting of a significant interrogation in an AMU with sorting at the input, all AMU memory locations are filled. In this case, during the sorting process one significant interrogation, the one with the least approximation error, is erased. The main portion of the subcycle is completed with processing of information from the mth channel (m is the number of channels in the OICD), replacement of reference interrogations in the CECD (with the Π interrogations with greatest approximation error prepared for transmission) and preparation of the AMU, consisting in that where $N < \Pi$ (N is the number of significant interrogations in the AMU after processing of the information from the mth channel), $\Pi - N$ redundant interrogations are input to the AMU.

Where $N < \Pi$, there are not enough significant interrogations for transmission. If no steps are taken, this will disrupt the evenness of the stream of compressed information and the radio communications link will be under utilized. In order to avoid this, the remaining $\Pi - N$ redundant interrogations must be input to the register; these interrogations might be redundant interrogations from earlier defined MS. The number of significant interrogations in the subcycle, as well as the sorting time of each significant interrogation, are random quantities; the length of the main portion of the subcycle is, therefore, also a random quantity. The duration of the subcycle of transmission of significant interrogations is a regular quantity. The subcycle of transmission of a given cycle begins upon completion

of the subcycle of transmission of the preceding cycle. According to what we have described, after the main ("working") portion of the subcycle of collection and sorting, there is a section of information storage ("waiting") until the beginning of the subcycle of transmission or (which amounts to the same thing) the beginning of the next[4] cycle of OICD operation. Immediately before beginning the transmission subcycle (after transmission of the Πth interrogation in the preceding subcycle), the Π significant interrogations with greatest approximation error are moved from AMU to the register, while the remaining $N - \Pi$ remain in AMU. Then, the next subcycle of collection and sorting is begun, these $N - \Pi$ interrogations participating in sorting; this is the buffer action of the AMU.

One important factor, already noted above, is the fact that /122
when the AMU is overfilled, the least significant information is lost, i.e., the information with the least approximation error. This is equivalent to the mode of an ADBS when the BMU is overfilled. However, the ADBS system requires feedback and devices for selection of various values of permissible error to achieve this mode. In the ADAS system with AMU, however, when sorting is performed at the input, no such equipment is required. As we can see from Figure 3, in each OICD cycle, immediately after the collection and sorting cycle, the transmission cycle begins. In this subcycle, the significant interrogations are output from the register each time a cycle pulse arrives from the clock.

For a more detailed description of the operation of the OICD, Figure 5 presents a flow chart for the operating algorithm of an OICD with an AMU with sorting at the input (for one of the two overlapping branches of the cyclogram, see Figure 3). Figure 5 uses the following symbols: i is the channel number (MS number); n is the OICD operating cycle number; N_c is the number of free locations in AMU before the next significant interrogation is recorded in memory; B_i is the value of the significant interrogation of the signal of the ith MS; ε_i is the approximation error of the signal of the ith MS; i_c is the ordinal number of the interrogation transmitted to the radio channel.

With ordered selection (see Figure 4), signals from all MS are interrogated in sequence and the approximation error is calculated after each interrogation (to determine significance). If an interrogation is found to be significant, it is written into the register; both the value of the interrogation and the value of the approximation error are recorded.

[4]As we can see from Figures 3 and 4, neighboring OICD operating cycles overlap.

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

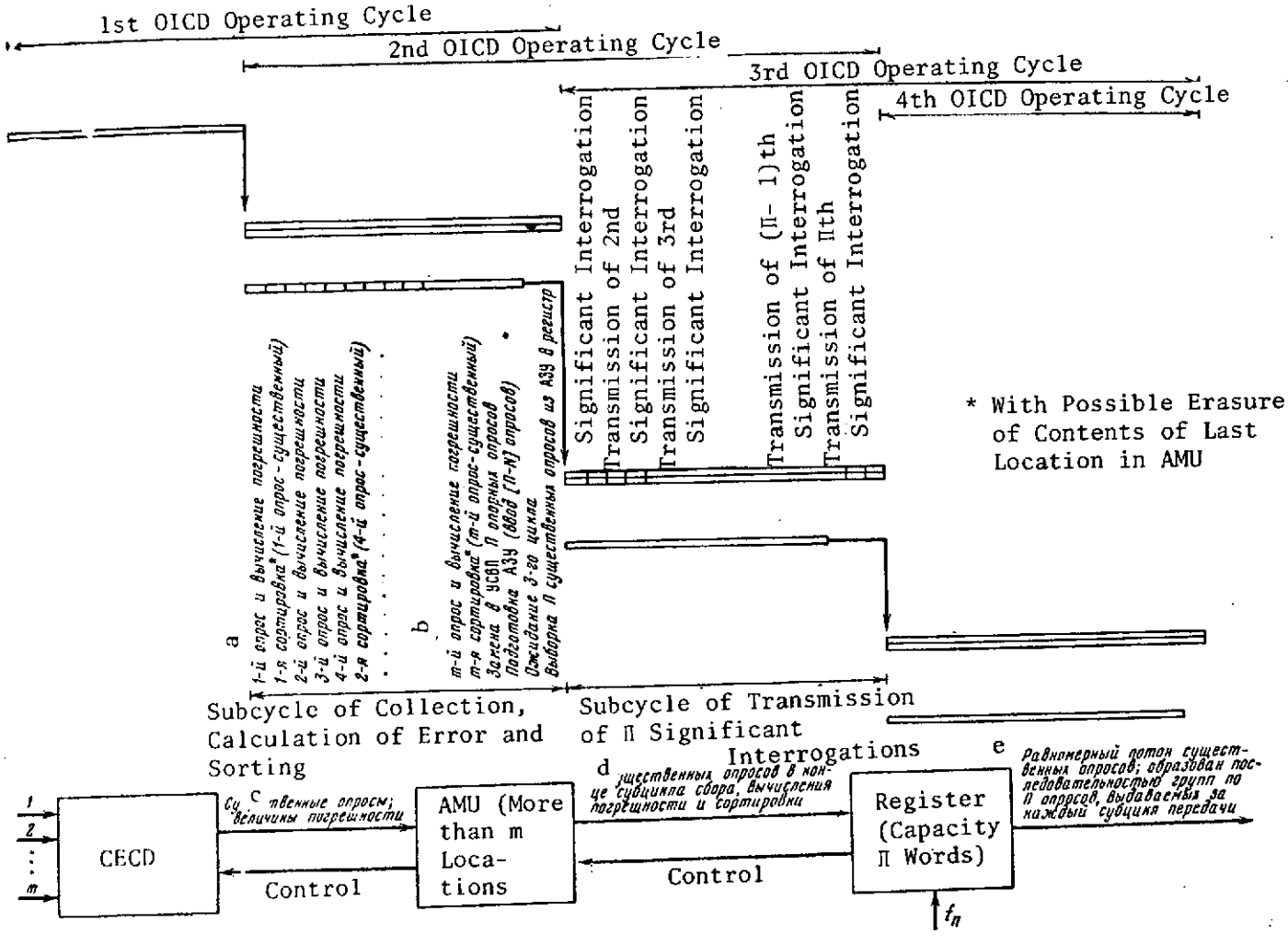


Figure 3. Block Diagram and Cyclogram of OICD Operation with Sorting at Input of System (Key on next page.)

Key (Figure 3): a, First Interrogation and Calculation of Error/ First Sorting* (First Interrogation is Significant)/ Second Interrogation and Calculation of Error/ Third Interrogation and Calculation of Error/ Fourth Interrogation and Calculation of Error/ Second Sorting* (Fourth Interrogation is Significant); b, mth Interrogation and Calculation of Error/ mth Sorting* (mth Interrogation is Significant)/ Replacement of P Reference Interrogations in CECD/ Preparation of AMU (Input of [Π - n] Interrogations)/ Wait for Third Cycle/ Selection of Π Significant Interrogations from AMU and Input to Register; c, Significant Interrogation; Values of Error; d, Π Significant Interrogations at End of Subcycle of Collection, Calculation of Error and Sorting; e, Even Flow of Significant Interrogations; Sequence of Groups of Π Interrogations Formed, Output in Each Transmission Subcycle

Key (Figure 4): a, First Interrogation, Calculation of Error, Writing into Register/ Second Interrogation, Calculation of Error (Redundant)/ Third Interrogation, Calculation of Error (Redundant)/ Fourth Interrogation, Calculation of Error, Writing into Register; b, mth Interrogation, Calculation of Error, Writing into Register/ Preparation of AMU (Erasure or Input of Interrogations)/ Rewriting from Input Register to AMU/ First Ordered Selection*; Replacement of First Reference Interrogation in CECD/ Second Ordered Selection*; Replacement of Second Reference Interrogation in CECD; c, Πth Ordered Selection*; Replacement of Πth Reference Interrogation in CECD/ Wait for Third Cycle/ Rewriting of Π Interrogations from AMU into Output Register; d, Significant Interrogations; Values of Error; e, Π Significant Interrogations; f, Even Stream of Significant Interrogations; Sequence of Groups of Π Interrogations Output in Each Subcycle of Transmission

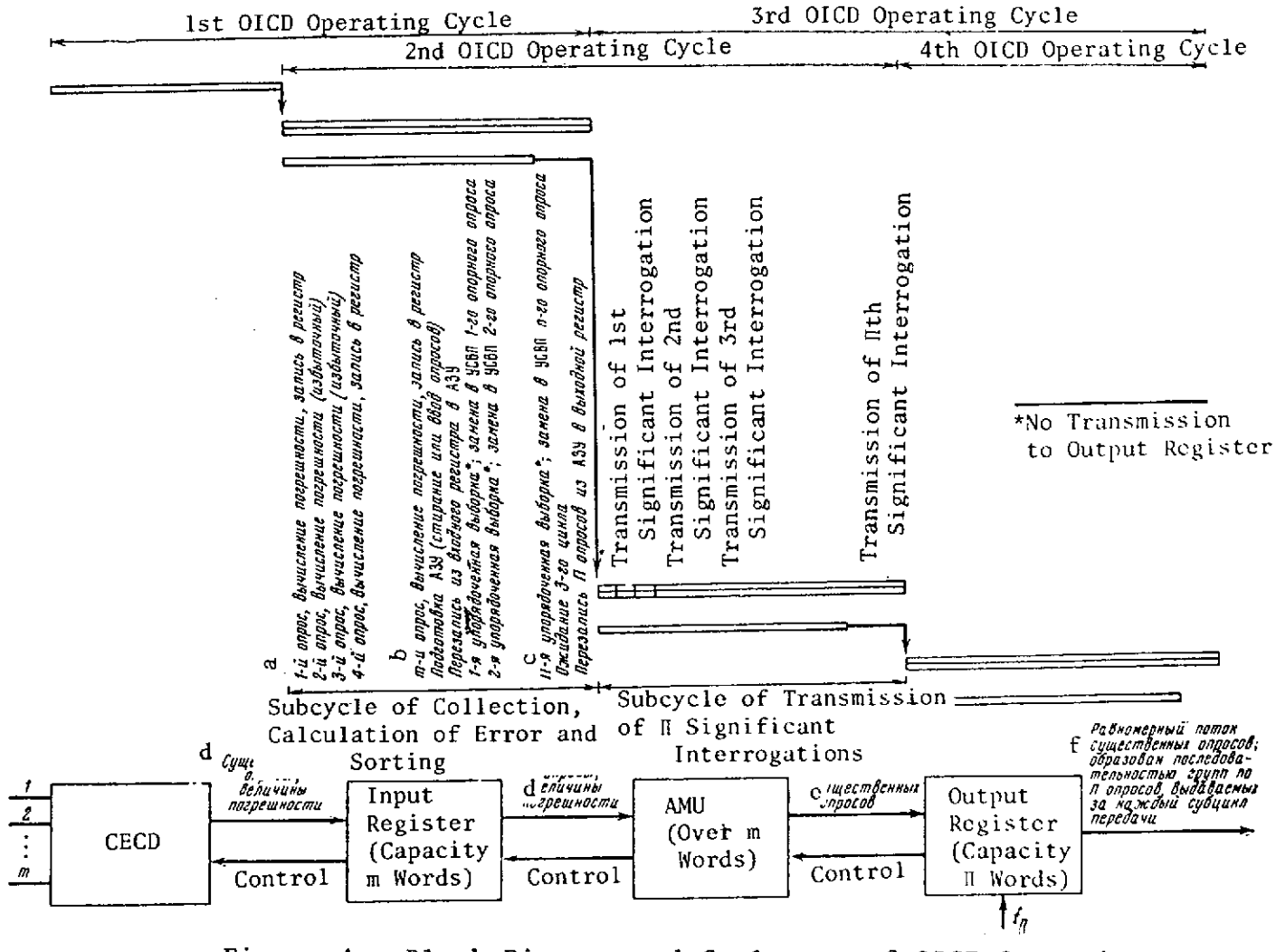


Figure 4. Block Diagram and Cyclogram of OICE Operation with Ordered Selection (Key on preceding page.)

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

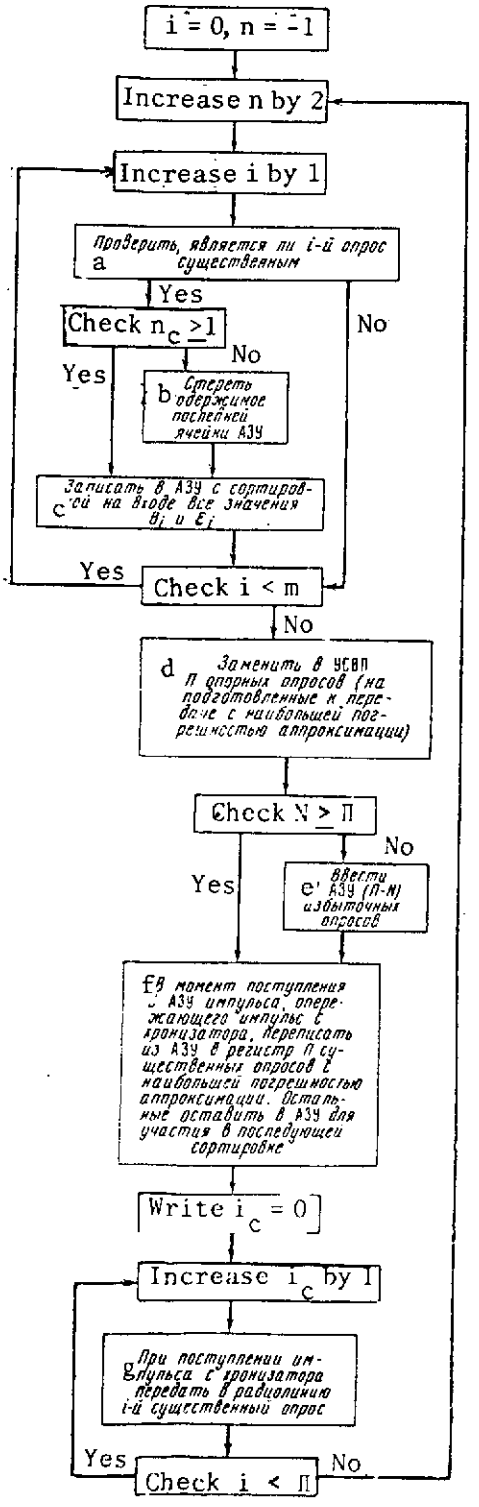


Figure 5. Flow Chart of Algorithm of Operation of OICD with AMU with Sorting at Input. Key: a, Check Whether ith Interrogation is Significant; b, Erase Contents of Last AMU Location; c, Write All Values of B_i and ε_i in AMU with Sorting at Input; d, Replace П Reference Interrogations in CECD (with Values Prepared for Transmission, with Greatest Approximation Error); e, Input (П - N) Redundant Interrogations to AMU; f, At Moment when AMU Receives Clock Pulse, Write П Significant Interrogations with Greatest Approximation Error from AMU into Register. Remaining Interrogations Remain in AMU to Participate in Next Sorting Operation; g, When Clock Pulse Arrives, Transmit ith Significant Interrogation into Radio Channel

After the significance of the interrogation of the mth channel is determined and (if necessary) /123 the interrogation is written in the register, the AMU is prepared to rewrite the information from the input register. Before rewriting, the AMU and input register may be in one of three states, characterized by the following relationships:

1) $(N_p + N) < П$,

2) $П \leq (N_p + N) \leq E_{AMU}$,

3) $(N_p + N) > E_{AMU}$,

where N_p is the number of significant interrogations written in the input register after determination of significance of the interrogation of the mth channel and writing (if necessary) this interrogation; N is the number of existing interrogations remaining in the AMU after completion of the preceding subcycle of collection and sorting.

In state (1), the total number of significant interrogations in the register and AMU is less than Π. In this case, similar to the plan with sorting at the input, an additional $[\Pi - (N_p + N)]$ redundant interrogations must be input to the register. In state (2) the total number of significant interrogations in the register and in the AMU exceeds the number Π, but not the capacity of the AMU, i.e., the number of significant interrogations is sufficient for transmission of Π interrogations and at the same time, all interrogations present in the register can be written in the AMU without overfilling. In state (3), a portion of the interrogations written in the input register may be lost; in order to avoid this, in state (3) before rewriting from the input register to the AMU, significant interrogations with the least approximation error are erased (by ordered selection with erasure of information). The number of interrogations to be erased is $(N_p + N - E_{AMU})$ interrogations. Upon completion of preparation of the AMU, the information is rewritten from the input register into the AMU and Π operations of ordered selection (without transfer to the output register) and replacement of Π reference interrogations in the CECD by the Π interrogations with the greatest approximation error prepared for transmission are performed. Replacement of the reference interrogations in each of the channels is performed only if the interrogation has a greater cycle number n than that of the last interrogation of this channel. This is done by writing cycle number n in which the reference interrogation was replaced into the register of each channel. Replacement of a reference interrogation is performed only where

$$n_{\text{interrogation replaced}} > n_{\text{register}}.$$

The operations involved in replacement of reference interrogations are not illustrated on the flow chart of the algorithm. The operation of ordered selection from the AMU has been described in the literature (for example, see [5]); therefore, we will not discuss this operation.

Input to the AMU in the process of rewriting from the input register is performed into three memory locations, i.e., AMU locations are not assigned to given parameters. As we know,

AMU access can be performed by addressing contents (not location); therefore, this system is possible; the expediency of the system results from two factors: the better[5] use of AMU capacity as functional BMU; and the simpler organization of AMU.

As in the plan with sorting at the input, the duration of the main portion of a collection and sorting subcycle is a random quantity, and therefore is followed by a sector of information storage ("waiting") until the beginning of the transmission subcycle. Immediately before the beginning of the transmission subcycle (after transmission of the Πth interrogation from the preceding transmission subcycle), the Π interrogations selected are rewritten from AMU into the output register.

The significant interrogations remaining untransmitted participate in ordered selection in the next subcycle of collection and sorting together with significant interrogations newly arriving to the AMU; this is the buffer action of the AMU. /124

Following the subcycle of collection and sorting, the subcycle of transmission of the Π significant interrogations from the output register into the radio link begins; transmission is performed under the control of cycle pulses from the radio link synchronizer.

For more detailed explanation of the operation of the OICD, Figure 6 presents a flow chart of the algorithm of operation of an OICD with AMU with ordered selection (for one of the two overlapping branches of the cyclogram -- see Figure 4). The symbols used on Figure 6 are the same as on Figure 5.

Earlier, we twice spoke of an ADAS algorithm without explaining what this means. The ADAS algorithm refers to the rule used to convert information from the beginning of interrogation of signals from the MS to the completion of the transmission of significant interrogations in each OICD operating cycle, looked upon as a unit whole (without subdivision into subcycles).

This concept of the ADAS algorithm allows us to construct a generalized (for all versions of OICD shown on Figure 2) mathematical model of the OICD for analytic study and modeling (see [4]) of the operation of the OICD with actual telemetry information.

---

[5]If the locations were rigidly assigned to parameters after each subcycle of ordered sampling and transmission, the remaining untransmitted interrogations would have to be rewritten into some area, the locations in which were not assigned to parameters.

Actually, we showed above how the circuits with time division of the processes of calculation and sorting and circuits with time combination of these processes do not differ from each other in this respect, or from systems with successive and parallel calculation of approximation error.

The OICD versions with sorting at the input and with ordered selection which we have studied also do not differ from each other. Actually, the same processes of conversion and transmission of information occur during the cycle as a whole in both systems:

1) collection and calculation of approximation error of m signals from MS with determination of the significant interrogations;

2) associative sorting of interrogations of a given cycle and those remaining untransmitted from the last cycle;

3) transmission of the first Π interrogations (from the list), with storage of the remaining interrogations for sorting in the next subcycle.

A flow chart of the ADAS algorithm is presented in Figure 7. In this figure, n is the OICD operating cycle number.

Statement of the Problem of Approximate Analytic Study of ADAS

The apparatus of queueing theory is convenient for the study of OICD. The state of the system refers to the number of interrogations in AMU. The known probabilities of states are used to determine dependence (1) and the other characteristics. The process of operation is not Markovian, which prevents us from using the Chapman-Kolmogorov equation [6]. It is convenient to use the method of Kendall [7], according to which the states of the system form an embedded Markov chain at moments preceding servicing, i.e., the beginning of ordered selection. The Chapman-Kolmogorov equation is correct for the points in the embedded chain.

Let us use $Q_{K(j)}$ to represent the probability that there are K interrogations in the ISF before the beginning of ordered selection in the jth OICD operating cycle.

We must note that in the actual situation, the approximation error of K signals cannot simultaneously (i.e., in a single OICD operating cycle) reach the precise value ε_m. If the error in the jth cycle reaches the precise value of ε_m for any of the K signals, the error of the remaining (K - 1) signals may differ from this value. Strictly speaking, the event "K input signal interrogations present in ISF, approximation error /12 of which has reached ε_m" means that "there are K input signal

interrogations in the ISF, the approximation error of one of which has reached the value of ε_m, while that of the remaining (K - 1) is near ε_m." The degree of "simultaneity" depends on the statistical characteristics of the telemetry messages and can be considered only by digital computer modeling (see [4]). To simplify approximate analysis, we will consider that the error has simultaneously reached ε_m for all K interrogations.

Dependence (1) which we require can be produced from the expression for $Q_K(j)$. Therefore, our task is to find the value of $Q_K(j)$. To do this, we study the next two intervals: the (j - 1)th and jth intervals. Suppose in the (j - 1)th cycle of operation of the OICD, before beginning ordered selection, there are r significant interrogations in the ISF. According to the symbols used above, the probability of this state is $Q_r(j - 1)$. As a result of ordered selection, performed in the (j - 1)th cycle of OICD operation, there are (r - П) significant interrogations in the line in the jth cycle. In order for there to be K significant interrogations in the ISF in the jth cycle of OICD operation before beginning ordered selection, it is necessary that (K + П - r) significant interrogations be input into the ISF in a jth subcycle of CECD operation.

Let us represent by $P_{K + П - r}(j)$ the probability that precisely (K + П - r) significant interrogations arrive at the ISF in the jth cycle of OICD operation.

In order for the system to be in state K in the jth cycle, two events must occur simultaneously: 1) presence of state r in the (j - 1)th cycle; 2) input of (K + П - r) interrogations in the jth cycle. We must consider that difference (R - П) must be positive [it is impossible to take more interrogations from the AMU in the (j - 1)th cycle than were present]. This means that the number of newly arriving interrogations must be:

$$K \quad \text{where } r \leq П,$$

$$K + П - r \text{ where } r > П.$$

Also obviously $r_{max} = П + K$, since where $r_{max} > (П + K)$, in order to achieve state K in the jth cycle, it would be necessary to remove significant interrogations during the CECD subcycle, which is senseless. Correspondingly, state K can be produced with the following combinations of values of r and K + П - r:

r		0	1	2	3		П	П + 1		+ 2	...	П + K
K +	- r	K	K	K	K	...	K	K - 1		K - 2	...	0

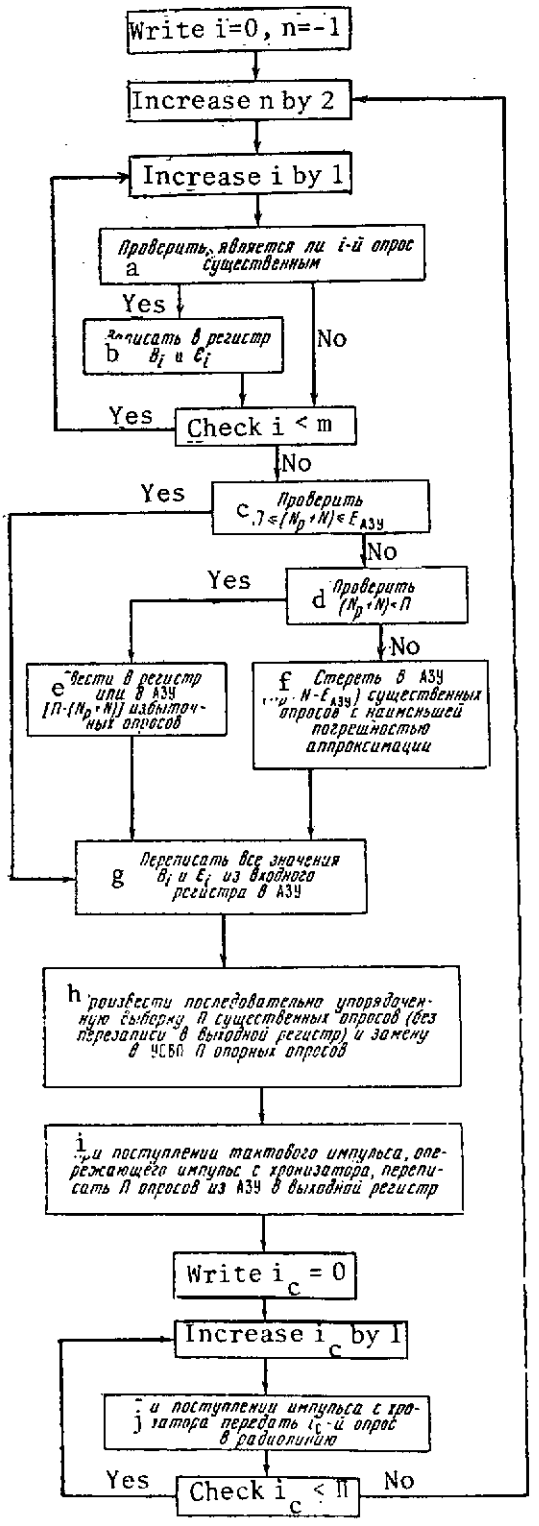


Figure 6. Flow Chart of Algorithm of Operation of OICD with AMU with Ordered Selection. Key: a, Check Whether ith Interrogation is Significant; b, Write B_i and ε_i into Register; c, Check $П \leq (N_p + N) \leq$ $\leq E_{AMU}$; d, Check $(N_p + N) < П$; e, Input $[П - (N_p + N)]$ Redundant Interrogations into Register or AMU; f, Erase $(N_p + N - E_{AMU})$ Significant Interrogations with Least Approximation Error from AMU; g, Rewrite all Values of B_i and ε_i from Input Register into AMU; h, Perform Successive Ordered Selection of П Significant Interrogations (Without Rewriting into Output Register) and Replace П Reference Interrogations in CECD; i, When Cycle Pulse Preceding Pulse from Synchronizer Arrives, Rewrite П Interrogations from AMU into Output Register; j, When Pulse Arrives from Synchronizer, Transmit i_cth Interrogation into Radio Line

ORIGINAL PAGE IS POOR

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

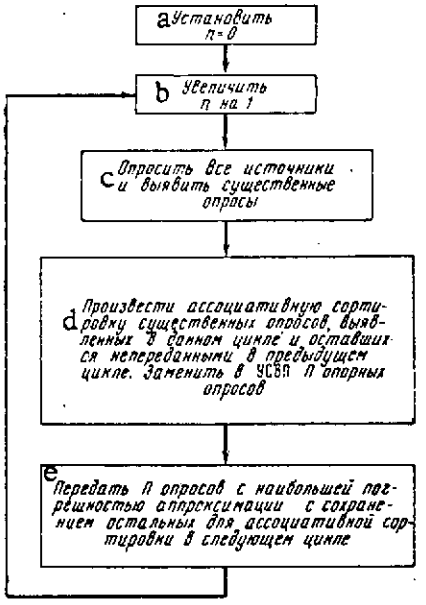


Figure 7. Flow Chart of ADAS Algorithm. Key: a, Set n = 0; b, Increase n by 1; c, Interrogate All Sources and Determine Significant Interrogations; d, Perform Associative Sorting of Significant Interrogations, Determined in This Cycle, and Remaining Interrogations Untransmitted in Preceding Cycle. Replace П Reference Interrogations in CECD; e, Transmit П Interrogations with Greatest Approximation Error, Saving Others for Associative Sorting in Next Cycle

Table 1 presents the values of K + П - r necessary to achieve state K in the jth cycle:

TABLE 1

POSSIBLE VALUES OF П + K - r

r	0	1	2	3 . . . П	П+1	П+2	. . . .	П+K
K	r − П							
	0	0	0	0 . . . 0	1	2	. . .	K
0	0	0	0	0 . . . 0	—	—	. . . .	—
1	1	1	1	1 1	0	—	. . . .	—
2	2	2	2	2 . . . 2	1	0	. . . .	—
. . .	. . . .	. . . .	. . . .	. . . . .	. . . .	. . . .	. . . .	. . . .
K	K	K	K	K . . . K	K−1	K−2	. . .	0

The Chapman-Kolmogorov equations: /127

$$
\begin{aligned}
Q_0(j) &= Q_0(j-1)P_0(j) + Q_1(j-1)P_0(j) + Q_2(j-1)P_0(j) + \ldots \\
&\ldots + Q_\Pi(j-1)P_0(j) = P_0(j)[Q_0(j-1) + Q_1(j-1) + \ldots \\
&\ldots + Q_\Pi(j-1)], \\
Q_1(j) &= Q_0(j-1)P_1(j) + Q_1(j-1)P_1(j) + Q_2(j-1)P_1(j) + \ldots \\
&\ldots + Q_\Pi(j-1)P_1(j) + Q_{\Pi+1}(j-1)P_0(j) = P_1(j)[Q_0(j-1) + \\
&+ Q_1(j-1) + \ldots + Q_\Pi(j-1)] + P_0(j)Q_{\Pi+1}(j-1), \\
Q_2(j) &= Q_0(j-1)P_2(j) + Q_1(j-1)P_2(j) + \ldots + Q_\Pi(j-1)P_2(j) + \\
&+ Q_{\Pi+1}(j-1)P_1(j) + Q_{\Pi+2}(j-1)P_0(j) = P_2(j)[Q_0(j-1) + \\
&+ Q_1(j-1) + \ldots + Q_\Pi(j-1)] + P_1(j)Q_{\Pi+1}(j-1) + \\
&+ P_0(j)Q_{\Pi+2}(j-1), \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
Q_K(j) &= P_K(j)\sum_{K=0}^{\Pi} Q_K(j-1) + \sum_{i=1}^{K} P_{K-i}(j)Q_{\Pi+i}(j-1).
\end{aligned}
\tag{3}
$$

Furthermore, probability Q_K must satisfy the condition of normalization

$$
\sum_{K=0}^{K_{max}} Q_K(j) = 1, \tag{4}
$$

where K_{max} is the maximum possible number of states.

Let us assume that the stream of significant interrogations is simple, i.e.,

$$
P_K(j) = P_K = \frac{(\lambda\tau)^K}{K!}e^{-\lambda\tau} = \frac{a^K}{K!}e^{-a}. \tag{5}
$$

Then for the stable state, equation system (3) can be represented as follows:

$$
\begin{aligned}
Q_0 &= P_0 \sum_{K=0}^{\Pi} Q_K, \\
Q_1 &= P_1 \sum_{K=0}^{\Pi} Q_K + P_0 Q_{\Pi+1}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
Q_K &= P_K \sum_{K=0}^{\Pi} Q_K + \sum_{i=1}^{K} P_{K-i} Q_{\Pi+i}.
\end{aligned} \tag{6}
$$

Let us introduce the generating function [6]

$$
Q(x) = \sum x^K Q_K. \tag{7}
$$

We can then produce the following expression

$$
Q(x) = \frac{Q_0 + xQ_1 + \ldots + x^{\Pi-1} Q_{\Pi-1} - x^{\Pi} \sum_{K=0}^{\Pi-1} Q_K}{1 - x^{\Pi} e^{-a(x-1)}}. \tag{8}
$$

The following additional condition is also set: the numerator must vanish as the value of each of the П roots of the denominator located within a unit circle is substituted into the formula.

Let us introduce the symbols: /128

$$
\alpha = a/\Pi, \; y = \alpha x. \tag{9}
$$

Then the П roots of the denominator represent the roots of the equation

$$
y e^{-y} = \alpha \cdot e^{-\alpha} \cdot \exp(\sqrt{-1} \cdot 2\pi s/\Pi); \quad s = 0, 1 \ldots (\Pi - 1). \tag{10}
$$

Representing

$$y = r\exp(\sqrt{-1}\,\theta) \tag{11}$$

and dividing the real and imaginary parts, we produce the following equations, equivalent to equation (11):

$$\begin{aligned} r\exp(-r\cos\theta) &= \alpha e^{-\alpha}, \\ \theta - r\sin\theta &= \frac{2\pi s}{\Pi}. \end{aligned} \tag{12}$$

These equations can be solved by an iteration method. In order to use the roots of equation (10), the numerator of expression (8) must be:

$$Q_0 + xQ_1 + \ldots + x^{\Pi-1}Q_{\Pi-1} - x^{\Pi}\sum_{K=0}^{\Pi-1} Q_K = -N\prod_{0}^{\Pi-1}(x - x_j), \tag{13}$$

where x_j are the roots of the denominator; $x_0 = 1$. Constant N is determined from the condition $Q(1) = 1$ by the formula

$$N = \sum_{K=0}^{\Pi-1} Q_K = \frac{\Pi - \alpha}{(1 - x_1)\ldots(1 - x_{\Pi-1})}. \tag{14}$$

If

$$\prod_{0}^{\Pi-1}(x - x_j) = x^{\Pi} - a_1 x^{\Pi-1} + \ldots + (-1)^s a_s x^{\Pi-s} + \ldots + (-1)^{\Pi} a_{\Pi}, \tag{15}$$

then

$$Q_s = (-1)^{\Pi-s+1} a_{\Pi-s} \sum_{K=0}^{\Pi-1} Q_K; \quad s \leqslant \Pi - 1. \tag{16}$$

Knowledge of expressions (8), (16) and quantities Q_K in equation system (6) allows us to produce the desired dependence (1).

REFERENCES

1. "Works of IEEE" (Russian Translation), Thematic Publication Sokrashcheniye izbytochnosti [Reduction of Redundancy], No. 3, 1967.
2. Ganderson, D. S., "Use of Associative Memory Units On Board Spacecraft," WESCON/66 Part 5.
3. Kantor, A. V., S. M. Perevertkin, T. S. Shcherbakova, "The Problem of the Possibility of Using Associative Computers for Parallel Adaptive Discretization of Multichannel Telemetry Information," see this collection page 131.
4. Tolmadzheva, T. A., A. V. Kantor, L. V. Rozhkovskiy, "Digital Computer Modeling of the Process of Reduction of Redundancy in Multichannel Telemetry Information by the Method of Adaptive Discretization with Associative Sorting," see this collection page 92.
5. Krayzmer, L. P., D. A. Borodayev, L. I. Gutenmakher, B. P. Kuz'min, I. S. Smelyanskiy, Assotsiativnyye zapominayushchiye ustroystva [Associative Memory Units] Energiya Press, 1967.
6. Riordan, J., Veroyatnostnyye sistemy obsluzhivaniya [Probabilistic Servicing Systems], Svyaz' Press, 1966.
7. Saati, T., Elementy teorii massovogo obsluzhivaniya i yeye prilozheniya [Elements of Queueing Theory and its Applications], Moscow, Sov. radio Press, 1965.

N75 11003

THE QUESTION OF THE POSSIBILITY OF USING ASSOCIATIVE COMPUTER DEVICES (ACD) FOR PARALLEL ADAPTIVE DISCRETIZATION OF MULTICHANNEL TELEMETRY INFORMATON

A. V. Kantor, S. M. Perevertkin
and T. S. Shcherbakova

/129

The inclusion of computers in on-board radio telemetry apparatus in prospective automatic interplanetary probes (AIP) has been dictated by the necessity of processing telemetry information on board the AIP.

The necessity of processing of telemetry information on board the spacecraft results from the necessity of compressing information, programmed selection by use of telemetry information in the control loop and other factors. The area of information compression is an extremely important area. In this work, primary attention is given to problems of information processing for purposes of compressing.

At the present time, the method of adaptive discretization is considered to be the most promising method for radio telemetry systems with time division multiplexing.

This results from the fact that adaptive discretization can transmit the shape of a signal, which is the primary requirement for the most extensive types of measurements currently used. Therefore, there is some interest in problems related to the application of computers for adaptive discretization. In the method of adaptive discretization, the instantaneous values of continuous telemetry functions produced by cyclical interrogation (discretization) are checked for significance according to a predetermined compression algorithm before transmission through the radio communication link. The selection of a frequency of cyclical interrogation is made a priori, based on the assumed dynamics of change of functions and the permissible error of restoration of continuous telemetry functions from their discrete values. For parameters with high information activity, this frequency may be quite high. Processing of the initial information according to the compression algorithms involves a number of arithmetic and logic operations, which must be performed by the on-board computer.

The computer used is a specialized computer (SC), since its structure is determined by the limited range of tasks involved in the algorithms used.

The task performed by the SC for adaptive discretization is processing of a large volume of information according to a single algorithm.

When a computer program is run for a large number of measured parameters in a single discretization cycle, very high requirements may arise as to SC speed. Obviously, with sequential adaptive discretization, the required calculation speed is determined as follows:

$$v_{seq} = in, \tag{1}$$

where i = kf is the information content of the RTS, measurements per second; k is the number of telemetry channels; f is the required interrogation rate, measurements per second; n is the complexity of the compression algorithm (complexity of an algorithm refers to the number of elementary operations in the program used to realize the algorithm).

We can see from expression (1) that the required computation rate with sequential information processing increases in proportion to the increase in information capacity of the RTS and the complexity of the compression algorithm. We note that the effectiveness of a compression algorithm is approximately proportional to its complexity. /130

Approximate calculations have shown that if the information capacity of the RTS is on the order of hundreds of thousands of measurements per second and the complexity of the compression algorithm is on the order of ten or more operations, the required calculation rate may be over one million operations per second. It is difficult to realize this speed by means of on-board SC.

The principle of parallel adaptive discretization differs from sequential adaptive discretization in that discretization of all parameters is performed simultaneously, along with parallel processing according to the compression algorithm selected. The requirements for the SC as to speed are significantly reduced by parallel discretization, since the required calculation rate is now independent of the number of telemetry channels and can be determined as follows:

$$v_{par} = nf,$$

$$\text{i.e., } v_{par} = v_{seq}/k. \tag{2}$$

However, it is obvious that the SC for parallel and independent processing of k telemetry channels must essentially consist of k SC for sequential discretization or, at least, k analog arithmetic units and control devices.

The relative hardware complexity of such a computer is determined from the expression:

$$\lambda \approx kb, \tag{3}$$

where b is a unit of hardware complexity, as which we use the hardware complexity of the computer for sequential discretization, k is the number of telemetry channels and the degree of parallelism of the SC.

As we can see from expression (3), the hardware complexity of SC increases in proportion to the number of channels. Considering the very rigid dimensional and weight limitations for on-board spacecraft apparatus, it may be difficult to produce such a computer.

Parallel-sequential discretization can be used as a compromise between totally sequential and totally parallel methods of discretization.

This method can be achieved by switching telemetry channels using one primary and several local commutators, connected to the primary commutator in parallel. In this case, incomplete combination of processing of the data of k sensors is achieved, and the processing of information occurs simultaneously only for m sensors, where m is the number of sensors in a group (equal to the number of local commutators), sequentially for groups of k/m, where k/m is the number of groups (equal to the number of channels connected to one local commutator).

In this case, the required speed of computation and the relative hardware complexity can be determined by the expressions

$$\lambda = mb \text{ (where } m < k), \tag{4}$$

$$v_{ps} = (k/m)nf = in/m. \tag{5}$$

The selection of number m with a fixed total number of telemetry channels k should be performed as a function of the specific requirements for the telemetry system.

Generally speaking, only the specific requirements for telemetry systems can determine the selection of a method of adaptive discretization of data. However, the method of parallel adaptive discretization, in spite of the difficulties involved in its realization, is the most promising method and allows rather effective compression algorithms to be used for high information-capacity RTS. /131

We suggest that parallel computers based on associative memory units (AMU) be used to realize this method. We call this type of computer an associative computer device (ACD).

Before going over to description of the structure of an ACD, let us discuss certain problems related to the application of AMU [1].

At the present time, address-type memory units are most commonly used in computers. In these memory units, there is a one-to-one correspondence between the coordinate address of a memory location and the information written in the location, and the required information is sought by searching for the location with the required address.

In AMU, on the other hand, information is generally written in arbitrary order, and the required information is sought by simultaneous inspection of the entire information file, seeking a match with some external information characteristic (the interrogation characteristic). This allows some logical processing of information and, furthermore, significantly simplifies programming (since no addresses are required for access to the memory unit).

Several types of logical information processing are possible in AMU. The first of these is extraction of information from the AMU on the basis of some external information characteristic, i.e., simple search. Using the concepts of set theory [2], the problem is reduced to determination of the subset of words B ($B \subset A$) of the set of words A written in the AMU, the elements of which are equivalent to external word C. Set A is a finite, disordered set of words in a binary alphabet, the numerical values of which are determined by the weight function:

$$A = \begin{cases} a_i : a = \sum_{j=0}^{r-1} x_j 2^j \\ i = 1, 2, 3, \ldots, N, \end{cases} \quad (6)$$

where N is the number of words in the field of storage; x_j is a single-position variable with two possible values 0 and 1; r is the number of digits in a word.

Word C (the interrogation characteristic) is a word in the binary alphabet of length r ($\ell C = r$). The desired subset B is defined by the intersection of sets A and C:

$$B = A \wedge C$$

Subset B may be the empty set $\emptyset$; this means that set A contains no elements equivalent to C; subset B may consist of one element equivalent to C; finally, subset B may contain z elements, equivalent to C. In this latter case, a so-called multivalued response is produced, the possibility of which is a distinguishing feature of AMU in comparison to ordinary address memory units.

The process of search for information in AMU may involve either the information written in the locations (information words) or a portion of the information, called the associative characteristics or inherent addresses of the words, in contrast to the coordinate addresses used in addressed devices. In this case, the portion of the word not participating in search is called the basic or accompanying information.

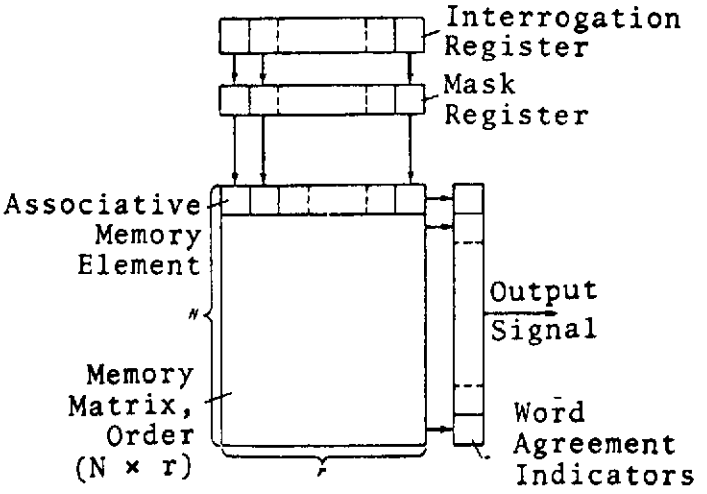


Figure 1. Similified Block Diagram of AMU

Since there is always a one-to-one correspondence between associative characteristics and the basic information, search for the corresponding associative characteristic in memory corresponds to search for the required basic information. /132

In many cases, search for information does not involve all of the bits in the interrogation register, but rather only a portion of these digits, the remaining digits being "masked" using a special mask register. In this case, with simple information search, we are locating a

subset of words, the elements of which are equivalent to the search characteristic with regard to the code of the mask register.

The possibility of masking some of the digits of the interrogation register greatly expands the logic capabilities of AMU with various types of complex search.

In the case of the binary alphabet, masking of some of the digits of the interrogation register can be performed by setting the corresponding bit positions of the mask register at "0" and leaving the other bit positions at "1" state.

Table 1 presents possible values for the ith digits of the interrogation register and mask register.

TABLE 1

ith Digit of Interrogation	ith Digit of Mask Register	Characteristics
1	0	No search
0	0	No search
0	1	Search for "0"
1	1	Search for "1"

A simplified block diagram of an AMU is presented in Figure 1.

Most frequently, information is sought in the AMU either in parallel by words and in sequence by digits or in parallel by words and digits.

The condition of agreement of a word from the storage field with the code of the interrogation characteristic, in correspondence with the code of the mask register, is equality of the match function (equivalence function) f for the word in question to "1":

$$f = \bigwedge_{i=0}^{r-1} f_i = \bigwedge_{i=0}^{r-1} [a_i \wedge c_i] \vee [\bar{a}_i \wedge \bar{c}_i] \vee b_i. \tag{7}$$

Here f_i is the equivalence function for one bit section; /133
a_i is the value of the ith digit of the word in the accumulator field; c_i is the value of the ith digit in the interrogation register; b_i is the value of the ith digit in the mask register.

Table 2 presents values of function f_i for various values of a_i, c_i and b_i.

TABLE 2

a_i	c_i	b_i	f_i	a_i	c_i	b_i	f_i
0	0	0	1	1	0	0	1
0	0	1	1	1	0	1	0
0	1	0	1	1	1	0	1
0	1	1	0	1	1	1	1

If f is equal to "1" for several words in the accumulator field, a multivalued answer is produced. A multivalue indicator should be included in the AMU to allow multivalued answers.

Complex information search in associative memory includes:

1. Search for the word with the maximum (minimum) value in the initial set A. This task consists in the determination of

$$\max_{a_i \in A}\{a_i\} \quad \text{and} \quad \left(\min_{a_i \in A}\{a_i\}\right).$$

2. Search for words less than the interrogation characteristic amounts to isolating from set {A} the subset {B}, all elements of which are less than the fixed external characteristic c, i.e., where

$$\min_{a_i \in A}\{a_i\} \leqslant c \leqslant \max\{a_i\}$$

we must separate subset $B \subset A$ such that

$$\max_{b_j \in B}\{b_j\} < c \quad \text{and} \quad \min(A \setminus B) \geqslant c.$$

Separation of the subset of words, all elements of which are greater than the fixed characteristic is performed similarly.

3. Search for words in the interval of values of interrogation characteristic a and b (b < a). The task consists in

isolating from A the subset A_1, all elements of which are less than characteristic a, and subset A_2, all elements of which are greater than characteristic b. The desired subset is determined by the intersection of the subsets thus produced $A_1 \wedge A_2$. However, the greatest interest is in complex search allowing sorting of information according to a predetermined sorting criterion in the AMU.

There are two types of sorting of information using AMU: sorting of information at the output of memory, or ordered selection; and sorting of information at the input, or ordered writing into associative memory. Let us first study the problems related to ordered selection of information.

The problem of selecting information in the order of decreasing values of the associative characteristics amounts to the following.

In the initial set A, we determine the maximum element

$$a_{1\max} = \max_{a_i \in A} \{a_i\},$$

and then, after excluding the element produced $A_{1\max}$, we determine set $A_1 = \{A \setminus A_{1\max}\}$; the next element of the sample is /134

$$a_{2\max} = \max_{a_j \in A_1} \{a_j\}$$

etc.

The task of selecting information in the order of increasing value of the associative characteristic is performed similarly with the sole difference that in each state the minimum element of the remaining set is determined.

There are a number of algorithmic and hardware methods for ordered selection of information. The algorithmic methods or methods with variable interrogation characteristics consist in successive interrogations of the information written in memory for agreement with an interrogation characteristic which changes in some manner. The code of the interrogation characteristic in each subsequent interrogation cycle is assigned as a function of the result of the preceding interrogation.

Let us discuss the following algorithmic methods of ordered selection, which are of the greatest interest.

The algorithm method of Frey and Goldberg [1] is the simplest method of ordered selection from AMU, although it is not highly effective.

Let us study this method as applicable to the case of selection of information in the order of increasing associative characteristics in the binary alphabet. The first stage of ordered selection consists in search for the least associative characteristic of the entire fixed set. The first stage of the interrogation register is 000...0, of the mask register -- 000...0, i.e., there is no search. Search begins with search for associative characteristics with values of the high-order bit positions of 0, by replacing the 0 in the highest-order bit position of the mask register with a 1 and searching for a 0 in the high order bit positions of the associative characteristics. Depending on the result of the first interrogation, two versions of further search are possible.

1. If there are characteristics with 0 in the high order position among the associative characteristics, i.e., at least one coincidence indicator is shifted to the excited state, the next bit position of the interrogation register is unmasked, i.e., the code in the mask register becomes 110...0, and a 0 is sought in the next bit position of the associative characteristics.

2. If.there are no words with high order bit position zeros, no coincidence indicator is shifted to the excited state, the 0 in the high order position of the interrogation register is replaced by a 1, i.e., the code in the interrogation register becomes 100...0, and the next digit of the interrogation register is unmasked for further search as in the preceding case.

The criterion of the end of search for the word with the minimum associative characteristic is excited state of a match indicator with no zeros in the mask register (all digits of the interrogation register unmasked).

In order to find the next associative characteristic (next higher), the code in the interrogation register remaining after search for the least characteristic is modified by adding a 1 considering all possible carries, with the digits of the interrogation register shifted from the 1 to the 0 state masked by the code in the mask register. Further search continues similarly.

The overall effectiveness of the method, defined as the ratio of the number of successful interrogations to the total

number of interrogations for the binary system of notation is approximately 0.178, i.e., rather low [1].

One advantage of this method is the relative simplicity of its realization. The AMU must only have match indicators for the words. /135

The main shortcoming of the Frey-Goldberg method is the dependence of the total number of interrogations on the number of digits in the interrogation characteristic. We note that the use of more effective algorithmic methods of ordered selection always requires some increase in hardware complexity of the AMU.

Of all known algorithmic methods of ordered selection, the most effective is the Lewin method [3]. The total number of interrogations required by this method is independent of the number of digits in the associative characteristics being ordered and is unambiguously determined by the number of words participating in the ordering process:

$$U(N) = 2N - 1. \tag{9}$$

The Lewin method, in contrast to the Frey-Goldberg method described above, requires that the AMU have match indicators for each word plus digit indicators, showing in which of three states the digits of the associative characteristics participating in the search are found:

1) the digits contain only 0 -- state 0,
2) the digits contain only 1 -- state 1,
3) the digits contain both 0 and 1 -- state X.

Each interrogation, in the Lewin method, is performed in two stages:

1. First the entire set of words to be ordered is interrogated normally, according to the code of the interrogation register considering the code in the mask register. The results of the interrogation are recorded by the word coincidence indicators.

2. Further, for the words revealed in the first stage, interrogations are fed over the digit wires to determine which of the three states hold for each digit (0, 1, X).

The results of interrogation are recorded by the digit indicators, so that the digit indicators with state X mark the digits in the isolated words by which they differ (do not

coincide). Therefore, the next step of search is unmasking of the next higher digit in the interrogation register corresponding to an X state.

The interrogation reveals a new group of N_{i+1} words, where $N_{i+1} < N_i$, N_i is the number of words participating in preceding search, and the states of the digits within this group of words are once more determined. With each interrogation, the number of words participating in the search is decreased. The word with the desired associative characteristic is determined by the word match indicator when there is no X state for the digit indicators in the second stage of search.

We can see from this that analysis of the states of the digits in the group of words selected prevents the danger that two successive interrogations will select the same group of words (which is quite possible for the method of Frey and Goldberg described above).

In conclusion, we note that the introduction of additional hardware to the AMU for the Lewin method in many cases can be justified by the significant reduction in the total number of ordering sequences required.

Tables 3 and 4 present examples of an ordered sample of four words in the binary alphabet: 1011, 0111, 0101, 1100 using the Frey and Goldberg method and the Lewin method respectively.

We can see from the tables that when the Frey and Goldberg method was used, 19 interrogations were required, while the Lewin method required only 7.

In Table 3, $\eta = \bigvee_{i=1}^{N} f$, where f is the equivalence function for each individual word, defined by formula (7).

When information is sorted in AMU with sorting at the input, the information arriving in arbitrary sequence is recorded depending on its numerical value. For example, information with the maximum associative characteristic is recorded in the first memory location, information with the associative characteristic less than the first characteristic but greater than the remaining characteristics is recorded in the second location, etc.

From the standpoint of set theory, the problem is to place the ordered set of memory location numbers

$$M_1 = \begin{Bmatrix} a_i : a = i \\ i = 1, 2 \ldots N \end{Bmatrix}$$

in one-to-one correspondence with the elements of the disordered set of words in the binary alphabet $M = \{b_j\}$ $a_i \leftrightarrow b_j$ so that $b_1 \leq b_2 \leq b_3 < \ldots < b_N$, where, for example, b_1 is an element in set M_2, placed in a one-to-one correspondence with an element of set $a_1 \in M_1$.

Writing in an AMU with sorting at the input allows simple ordered selection of information, requiring only that all locations in the AMU be interrogated in the order of increasing (or decreasing) addresses.

/136

TABLE 3

Number	Interrogation Register	Mask Register	η	Word Selected	Number	Interrogation Register	Mask Register	η	Word Selected
1	0000	0000	0		11	1000	1000	1	
2	0000	1000	1		12	1000	1100	1	
3	0000	1100	0		13	1000	1110	0	
4	0100	1100	1		14	1010	1110	1	
5	0100	1110	1		15	1010	1111	0	
6	0100	1111	0		16	1011	1111	1	1011
7	0101	1111	1	0101	17	1100	1100	1	
8	0110	1110	1		18	1100	1110	1	
9	0110	1111	0		19	1100	1111	1	1100
10	0111	1111	1	0111					

TABLE 4

Number	Interrogation Register	Mask Register	η	Digit Indicators	Word Selected
1	0000	0000	1	XXXX	
2	0000	1000	1	01X1	
3	0000	1010	1	0101	0101
4	0010	1010	1	0111	0111
5	1000	1000	1	1XXX	
6	1000	1100	1	1011	1011
7	1100	1100	1	1100	1100

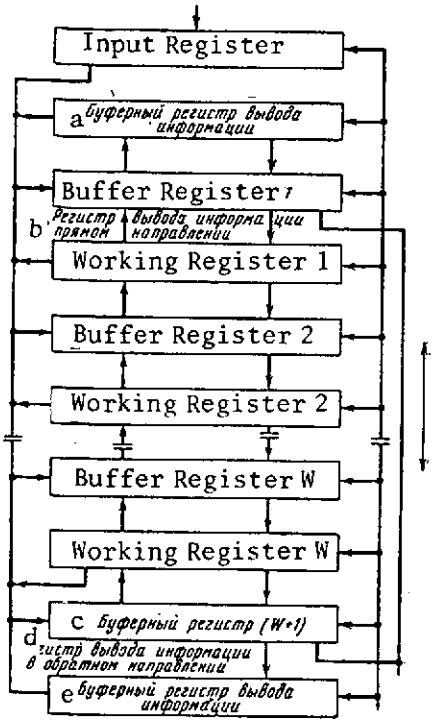


Figure 2. Flow Chart of Sorting of Information in AMU with Sorting at the Input. Key: a, Information Output Buffer Register; b, Forward Direction Information Output Register; c, Buffer Register (W + 1); d, Reverse Direction Information Output Register; e, Information Output Buffer Register

A system which can be used to sort information in an AMU with sorting at the input is shown in Figure 2 [4]. The input number arrives at the input register. Comparison circuits (not shown on the figure) not only compare the contents of each working register to the contents of the input register, but also record the results of the comparison ("greater," "less," "equal"). Thus, the position in AMU in which the next number should be written is between the working registers marked by the "greater" ("equal") and "less" ("equal") symbols.

Recording of information in AMU as a function of time is illustrated in Table 5. During the first half cycle of each period, the next number arrives at the input register, and the comparison circuits determine its numerical value in relationship to the numbers already entered. This determines the position for recording of the next number in AMU. During the second half cycle, recording itself occurs.

In addition to the algorithmic methods described above for ordered selection, there are hardware or circuit methods. The basic difference between these methods and algorithmic methods is that the code of the interrogation characteristic in each stage of search is independent of the results of preceding interrogations.

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

TABLE 5

Location Number	Time Interval							
	τ_1'	τ_1''	τ_2'	τ_2''	τ_3'	τ_3''	τ_4'	τ_4''
4	00	00	00	00	00	00	00	11
3	00	00	00	00	00	11	11	10
2	00	00	00	11	11	10	10	01
1	00	01	01	01	01	01	01	00
Input	01		11		10		00	

Among the various hardware methods for ordered selection, the one which is simplest to realize and rather effective under certain conditions ($N \approx 2^r$) is the method of cyclical review [5]. According to the method of cyclical review, the interrogation characteristic must run through 2^r values (r is the number of digits in the interrogation characteristic) from 000...0 to 11...1, and all associative characteristics recorded in AMU will be read in the order of increasing numerical values. When the associative characteristics are ordered in the sequence of decreasing numerical values, the code for the interrogation characteristic must vary from 111...11 to 00...0.

The hardware realization of the method is quite simple. /138
All 2^r values of the characteristic can be produced by an r-digit binary counter. Masking of the digits in the interrogation register is not required.

Generally speaking, hardware methods of ordering are simpler to realize than algorithmic methods; therefore, algorithmic methods should be used only when hardware methods cannot achieve acceptable effectiveness.

We have studied the logic capabilities of AMU for search operations.

These operations do not change the contents of the information recorded in the AMU.

However, the capabilities of AMU for information processing can be significantly expanded by adding certain logic circuits, with which, in addition to search operations, arithmetic operations can be performed as well. This converts an AMU to an associative computer device ACD. We should note that the additional logic may be evenly distributed through the entire structure of the AMU, so-called AMU with distributive logic, or may be made separately from the AMU memory matrix, AMU with external logic.

AMU with distributed logic, producing a homogeneous structure, are probably more promising, but are at present very difficult to manufacture.

In the following, we describe one possible realization of an ACD based on AMU with external logic, as applicable to the method of parallel adaptive discretization with associative sorting.

A block diagram of the ACD is presented in Figure 3.

In composing the block diagram of the ACD, the following considerations were taken into account:

1) the input of codes of sources to AMU memory must be parallel by digits and sequential by words;

2) arithmetic and logic processing should be performed sequentially by digits and parallel by words;

3) adaptive discretization should be performed according to a first order extrapolation algorithm.

The actual AMU memory device (5) is a matrix ($k \times r$) associative memory element, where k is the number of telemetry channels, r is the number of digits in an information word. In turn, each information word consists of several fields and contains full information concerning a specific message source. Some operating cycles of the device utilize information written in various AMU memory fields.

For a first order extrapolation algorithm, the information word consists of eight fields: the field of the first initial interrogation, the second initial interrogation, the current interrogation, the source address code, flag digits, the extrapolated value, the excess over permissible approximation error and the permissible approximation error.

The logic and arithmetic operations are fully performed by devices external to the memory matrix (section 6). Control of the operation of the ACD is by programmed control circuits controlling the storage and conversion of information (7).

The operation of the ACD consists of a number of separate cycles of information processing. They include startup or preparatory cycles, necessary at the beginning of operation of the ACD, and the working cycles involved in information processing. A working cycle, in turn, is divided into two subcycles: the main and supplementary subcycles.

The number of startup cycles is equal to the number of interrogations used in calculation of the extrapolated values; in the case of first order extrapolation, the number of such

interrogations is equal to 2. Interrogations used in calculation of the extrapolated value are called initial interrogations. Operation of the ACD in the startup mode consists of the following. /140

Signals with sources of analog information (AI) and sources of code information (KI) are sent to the signal inputs of switches 2. The control inputs of switches 2 receive blocking signals from the program-switching device (PSD). The input of switches 2 simultaneously receives the signal either from one of the analog information sources, or from one of the code information sources.

Conversion of the analog signal to a code is performed by a pulse generator, switch, pulse counter, exponential voltage generator and comparison circuit 3. The exponential voltage of the generator is sent to one of the inputs of comparison circuit 3, the other input receives a signal from one of the AI sources. At the moment when the values of the analog signal and exponential voltage agree, a signal is output to write the code from the counter in one of the fields of the corresponding AMU location (5.1.1 or 5.1.2, 5.1.3, etc.). The number of the location and number of the field into which the information is written depend on the source and nature of the information being recorded at the moment.

For example, field 1 of location 5.1 (5.1.1) always records the first initial interrogation, corresponding to the first source of information, field 2 of location 5.1 -- (5.1.2) records the second initial interrogation of the first source, etc.

To write information from AI sources in the field of primary information, the digit wires of the corresponding AMU field must be excited. Selection of the required field is performed by the AI source signal separation system. The separation system consists of three multidigit switches (logical AND circuits). The signal inputs of all three switches receive a code from the counter, the control inputs of the switches receive signals from the PSD ("field 1, field 2, field 3" signals respectively).

The signal from a KI source passes through switch 2 to the KI source signal separation circuit. The code distribution circuit for KI sources operates similarly to the system for AI sources.

The full cycle of ACD operation involved in processing of information in the stable mode consists of the following partial cycles:

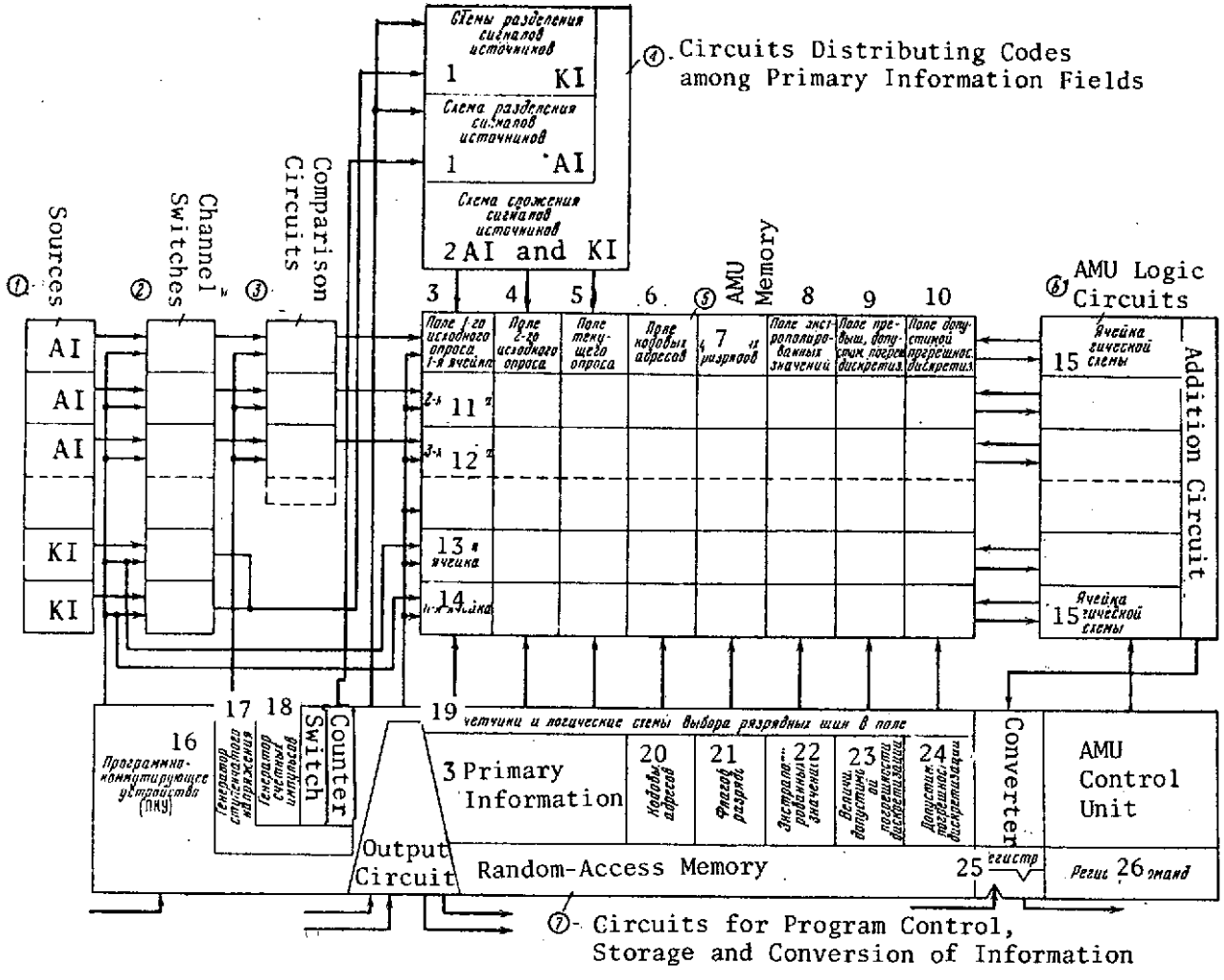


Figure 3. Block Diagram of Associative Computer Device
(Key on next page.)

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

Key (Figure 3): 1, Source Signal Separation Circuits; 2, Source Signal Addition Circuits; 3, Field of First Initial Interrogation First Location; 4, Field of Second Initial Interrogation; 5, Field of Current Interrogation; 6, Code Address Field; 7, Flag Digit Field; 8, Field of Extraploated Values; 9, Field of Excess over Permissible Discretization Error; 10, Field of Permissible Discretization Error; 11, Second Location; 12, Third Location; 13, Location n - 1; 14, Location N; 15, Logic Circuit Location; 16, Program-Switching Device (PSD); 17, Exponential Voltage Generator; 18, Counting Pulse Generator; 19, Counters and Logic Circuits Selecting Digit Wires in Field; 20, Code Addresses; 21, Flag Digits; 22, Extrapolated Values; 23, Quantity Greater than Permissible Discretization Error; 24, Permissible Discretization Error; 25, Register; 26, Instruction Register

1) parallel associative calculation of extrapolated values of readings of all telemetered functions in the nearest cycle of current interrogation;

2) successive current interrogation of information sources and input of results of interrogation to AMU 5 in current interrogation field;

3) parallel associative calculation of values by which all telemetered parameters exceed the permissible approximation error and recording of these quantities into AMU 5 in the field for excess over permissible approximation error;

4) associative search -- associative ordered selection of the results of current interrogation in the order of decreasing excess over permissible error (the interrogation with the maximum approximation error is selected first); search for interrogations of all priority sources is conducted simultaneously.

As was noted above, all computations in the ACD are performed in parallel for words and in sequence for digits. The adder for each word is a single-digit adder which forms the digit sum s_i and records carries c_i.

The adding circuit follows the rule

$$s_i = a_i + b_i + c_i - 2c_{i+1}, \tag{10}$$

where s_i are the values of the digit sum in the ith digit; /141
$a_i b_i$ are the values of the ith digits of the components; c_i is a carry unit in the ith digit; c_{i+1} is the carry unit in the (i + 1)th digit.

Table 6 illustrates the operation of a single digit binary adder for sequential counting.

TABLE 6

a_i	b_i	c_i	s_i	c_{i+1}	a_i	b_i	c_i	s_i	c_{i+1}
0	0	0	0	0	1	0	0	1	0
0	0	1	1	0	1	0	1	0	1
0	1	0	1	0	1	1	0	0	1
0	1	1	0	1	1	1	1	1	1

In subtraction operations in the case of successive counting, it is convenient to represent one of the components in complement code. This component is inverted to form the reverse code, and the "1" of the complement code can be fed to the carry circuit of the adder at the moment of addition of the lowest-order digits.

AMU 5, logic circuits 6 and control circuit 7 participate in subtraction cycles. Successive excitation of the digit wires in the fields of AMU 5 is performed by the counters and logic circuits of the digit wires. The instructions controlling the operation of the counters and logic circuits and the control section are written in the memory device and output in sequence to the instruction register.

As the excess approximation error is output, ones are recorded in the first flag digits of the field of flag digits in AMU 5, corresponding to sources of information with excessive approximation error. This allows ordered selection of interrogations from only those information sources with excessive approximation error during the associative search cycle.

Ones can also be written in the flag digit fields for the locations in AMU 5, interrogations of information sources of which should be selected independently of the approximation error. This allows priority to be given to certain channels, either automatically or by the program.

Associative search can be performed, for example, using the cyclical review method described above. The associative characteristic of the words used in this case is the value of the excess over the permissible approximation error, or when information from priority sources is selected -- the 1 found in the corresponding flag positions. The accompanying information consists of the values of instantaneous readings and addresses of information sources.

One of the basic problems related to the cycle of associative search is the problem of communication with the external devices. In multichannel RTS, the spacecraft with external devices may be the transmitter of a radio set, the RTS memory or the on-board computer.

During the associative search cycle, it is possible to output various numbers of significant interrogations to various external devices. For example, all of the significant interrogations need not be output to the transmitter, but rather only a certain portion with the highest values of excess approximation error may be sent to the transmitter; all significant interrogations may be sent to the on-board memory device; the on-board computer may receive one significant interrogation, /142 that with the maximum error. This flexibility has great practical significance.

Control and conversion in the associative search cycle are achieved by the AMU control unit, the flag digit field, logic circuits 6, the converter, register and output circuits. Since interrogations are not output in the order of the information source numbers, each significant interrogation should be equipped with an address corresponding to its source. Significant interrogations are output together with the address codes digit by digit and sent to the external devices through the output circuits.

During the supplementary cycle of ACD operation, AMU 5 and logic circuits 6 are prepared for the next full working cycle, beginning with the partial cycle of calculation of the extrapolated values. During the supplementary cycle, three successive operations are performed in AMU 5. The information in the field of the first initial interrogation is erased, the information from the field of the second initial interrogation is recorded into the field of the first, and the value of the current interrogation is recorded in the field of the second initial interrogation. Information not required for the next working cycle is erased from AMU 5. At the end of the supplementary cycle, the fields of AMU 5 contain the fields of the first and second initial interrogations, the field of permissible errors, the field of address codes of information sources, the field of

flag digits (the digits of this field marking priority sources are occupied); the remaining fields in AMU 5 should be free for the beginning of the next full working cycle of the ACD.

Along with the starting and working cycles, in some situations a cycle of change of initial data and programs may be used. This cycle is controlled both by external devices (such as the time programming device of the interplanetary space probe, the command radio link of the probe), and from results of analysis of the instantaneous information in the ACD information collection and compression device independently or using external analyzers. In the last case, the ACD, in addition to adaptive discretization, also performs adaptive switching of information sources. During this cycle, the collection and compression of data are interrupted. The change in initial data (for example, values of permissible errors, numbers of priority channels, etc.) is made by recording in the random-access memory with subsequent rewriting in the AMU 5.

The program can be changed by two methods:

1) upon appearance of a certain combination of conditions, determined, for example, by associative search with the participation of external devices, an instruction is developed to make a transition to another program recorded in random-access memory;

2) the new program is written in memory and an external instruction is given to shift to operation of the ACD using the new program; the operating mode of the PSD may also be changed by an external program.

Thus, an on-board computer similar to the ACD described above should be used for collection and compression of telemetric information on board a space probe, allowing:

1) performance of identical arithmetic and logic operations for a rather large volume of information;

2) use of effective algorithms of compression for high information capacity RTS due to parallel processing of information from all telemetry channels;

3) rapid associative search in files of frequently changing information, allowing flexible control of the collection and compression of information before it is transmitted into the radio link.

REFERENCES

/143

1. Krayzmer, L. P., D. A. Borodayev, A. I. Gutenmakher, B. P. Kuz'min, I. L. Smelyanskiy, Assotsiativnyye zapominayushchiye ustroystva [Associative Memory Units], Energiya Press, 1967.
2. Kolmogorov, A. N., S. V. Fomin, Elementy teorii funktsiy i funktsional'nogo analiza [Elements of the Theory of Functions and Functional Analysis], Nauka Press, 1968.
3. Lewin, M. H., Retrieval of Ordered Lists from a Content Addressed Memory, RCA Rev., 1962, Vol. 23.
4. Seeber, R. R., Associative Self-sorting Memory Eastern Joint Computer Conference, 1962.
5. Bazilevskiy, Yu. Ya., "Problems of the Theory of Temporal Logic Functions," Voprosy teorii matematicheskikh mashin [Problems of the Theory of Mathematical Machines], No. 1, Rizmatgiz Press, 1958.

N75 11004

HIGH FREQUENCY SWITCHING DEVICES

B. A. Prigoda, Ye. A. Vlasov, A. S. Markin
and E. I. Yakovlev

Switching devices are widely used in the high frequency lines of ground station and on-board electronic devices.

Depending on the operating conditions and purpose of the switching, commutators may be constructed using various circuits of electronic or mechanical elements.

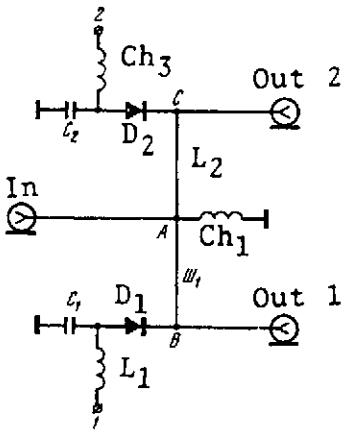


Figure 1. Commutator Using a Diode

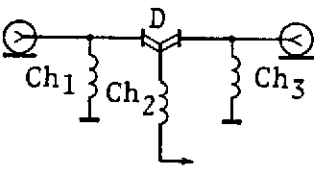


Figure 2. Commutator with NIPIN Structure

At the present time, the most common types of remote controlled antenna switches are:

1) diode switches using low-power p-n diodes;
2) diode switches using PIN or NIPIN structures;
3) electromechanical switches, using electromagnetic relays or electromagnetic relay mechanisms as drive mechanisms;
4) switches using other electromechanical systems such as electric motors as drive mechanisms.

Diode switches are generally made as follows: chokes Ch_1 - Ch_3 are used to decouple the diode control circuits at low frequency from the main high frequency line. The controlled diodes D_1, D_2 are of type P - N or PIN [1]. The elements L and C are quarter-wave loops and loading capacitors respectively.

When diode D_1 is closed, loop L operates in the idle mode. Since its length is one-fourth of the working wavelength, at point A on the circuit of Figure 1 there is a short circuit, /14
i.e., the input-output 1 line is closed at the operating frequency. When diode D_1 is open, loop L_1 is closed at its free end through capacitor C_1, and due to transformation, the idle mode obtains at point A and this line is open at the working frequency.

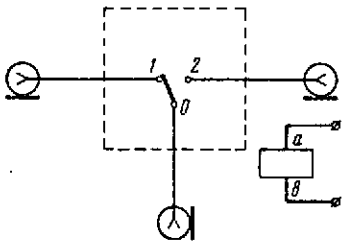


Figure 3. Diagram of Switch Using Relay

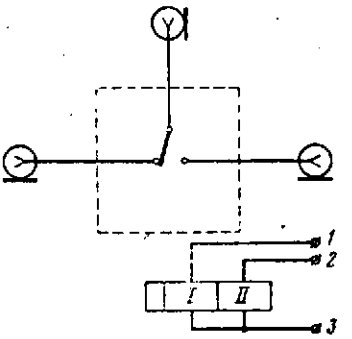


Figure 4. Diagram of Switch Using Polarized Relay

When an NIPIN structure is used, the circuit is as shown on Figure 2. Whereas in the preceding circuit the diode was connected to the free end of the resonant element (loop), in this circuit the structure is directly connected into the line; an increase in bandwidth is achieved by eliminating the resonant elements.

The circuit is open only when a dc voltage is fed to the input of choke Ch_2. To increase decoupling at high frequency in the closed state, the input of Ch_2 is supplied with a reverse polarity dc voltage.

One common shortcoming of these two circuits is the necessity of constant supply of power to the semiconductor device.

Switches constructed as shown on Figure 1 and 2 are generally used to switch the dipoles in antenna arrays with electronic beam scanning.

At comparatively low frequencies (up to 200 MHz), satisfactory electronic characteristics can be obtained by using ordinary low frequency electromagnetic relays types RES-9, RES-10, etc., enclosed in a special shielded body. At higher frequencies, the contacts themselves must be placed in a special shielded body to produce the required wave impedence

of the line. These switches are currently being series produced (for example, types REV-15, RPV-5/7).

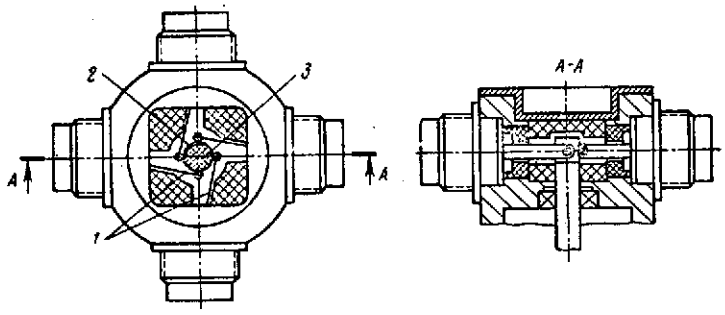


Figure 5. Switch with Commutation Shaft

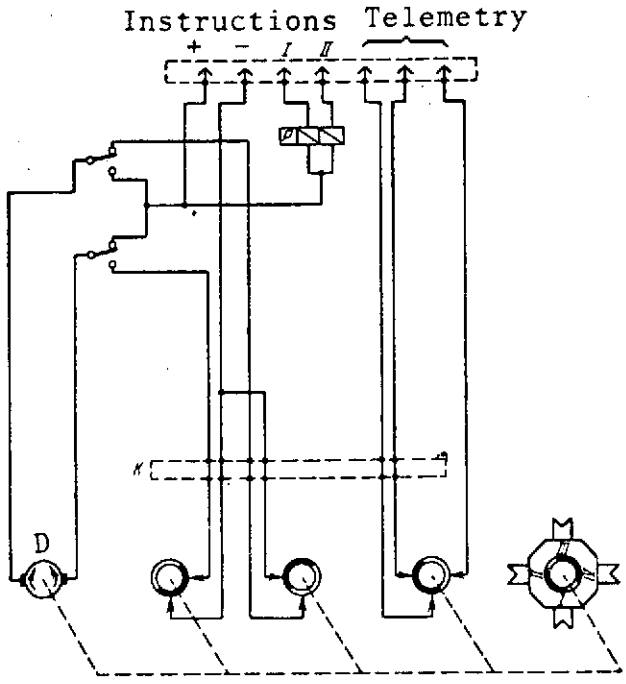


Figure 6. Electrical Diagram of Drive Mechanism

A diagram of a switch of this type is shown in Figure 3. One shortcoming of this type of high frequency switch is the fact that in one of the positions of the contacts, the winding of the relay must be constantly under power. These switches can be expediently used if the times the relay will spend in the position where power must be supplied to the winding are brief. /145

From the standpoint of increasing reliability and saving power, it is more expedient to use switches driven by an electromagnetic polarized relay. One very important advantage of this type of switch is the fact that the current must be transmitted through the winding of the relay only during a very short interval of

time, required for switching, after which the relay remains in its position. Of the relays of this type currently series produced, we might mention the RPV-5/4 (see Figure 4, [2]).

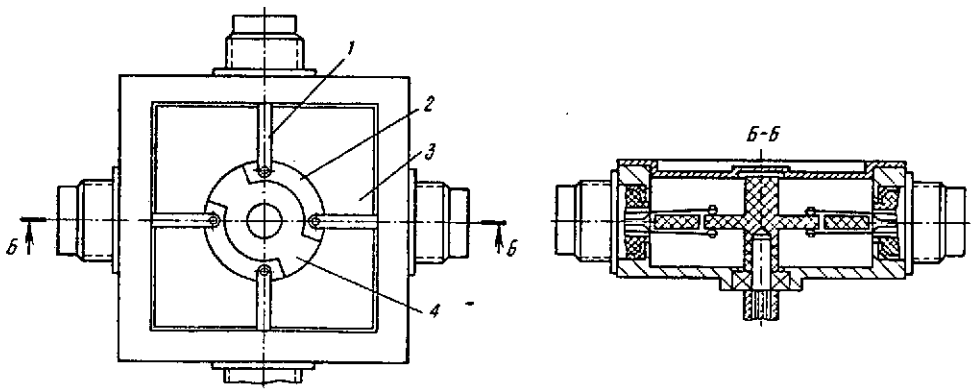


Figure 7. Switch Using Band Line

All of these "relay" switches have significant capacitances /146
between the contacts, limiting their application at high frequencies due to the decrease in decoupling between contacts.

Switches in which the contacts are placed at an angle of 90° to each other (see Figure 5) and contacts are made through a contact strip fastened to the switching shaft operate satisfactorily (at up to 1000 MHz).

Figure 6 shows the electrical circuit of the drive mechanism of such a switch. A reversible electric motor is used as the drive unit, using a polarized electromagnetic relay as the instruction receiver to reverse the motor. At higher frequencies, a switch with a similar drive using a high frequency switching head consisting of a symmetrical band line with a supporting insulator can be used (see Figure 7). In this case, increased decoupling is achieved due to the fact that the capacitance of the coupling through the end surfaces is an order of magnitude less than in the preceding case.

Furthermore, this type of high frequency head achieves an increase in reliability, due to the fact that there are two pairs of parallel one-two-one contacts. In case of failure of one of these contact pairs, the second pair remains operational.

REFERENCE

1. Sestroretskiy, B. V., L. S. Liberman, "Theory of Semiconductor Diode SHF Switches," Poluprovodnikovyye pribory i ikh primeneniye [Semiconductor Devices and their Applications], Edited by Ya. A. Fedotov, No. 12, Sov. radio Press, 1964.

N75 11005

SHF FILTERS IN SPACECRAFT ANTENNA-FEEDER SYSTEMS

Ye. A. Vlasov and B. A. Prigoda

The SHF-band filters used in modern antenna-feeder systems (AFS), depending on the specific application and specific requirements, may be constructed both of elements with lumped parameters, and of elements with distributed parameters [1].

Coaxial or band lines are the elements with distributed parameters most frequently used. Filters made with these lines may be constructed either entirely of sections of such lines, or in combination with lumped reactances. The lines are used in the form of dipoles or quadrupoles.

Obviously, filters constructed of line sections alone will have an infinite number of bands of transparency and opacity for frequencies which are multiples of the low frequency for which the condition of shorting at the input of the filter unit is met.

Frequently, filters act as decoupling devices at two or more frequencies. Let us study an example of design of such a decoupling device for two frequencies (f_1 and f_2), the example most frequently encountered.

Figure 1 shows a diagram of a two-frequency decoupling filter with two band-blocking elements consisting of line sections.

Figure 2 is a diagram of a decoupling filter with two band-pass elements consisting of line sections. /147

Figure 3 shows a hybrid filter circuit, using one band-blocking and one band-pass element.

In all these circuits, the identical line sections are marked with the identical symbols.

Let us study the operation of a band blocking element using the filters of figures 1 and 3 as examples.

In the side which passes the signal frequency f_1, line section 1a, shorted at the end, is selected at a length such

that the shorting condition is met for frequency f_2 at its connection point a. Correspondingly, section 2a is selected from the condition of providing a short circuit at connection point c for frequency f_1.

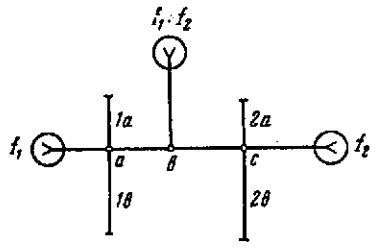


Figure 1. Decoupling Filter with Direct Connection of Loop and Compensating Reactance

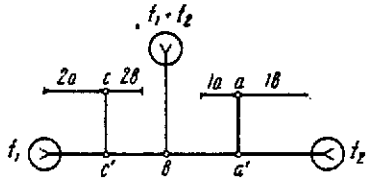


Figure 2. Decoupling Filter with Connection of Loop and Compensating Reactance through Transforming Feeder Section

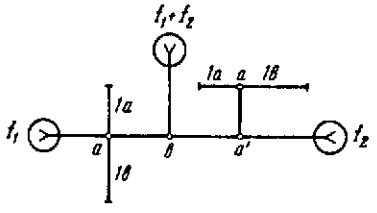


Figure 3. Hybrid Filter Circuit

The lengths of sections 1a and 2a are thus

$$l_{1a} = \frac{\lambda_2}{2}, \quad l_{2a} = \frac{\lambda_1}{2},$$

where λ_1 and λ_2 are the wavelengths in the material of the coaxial line for frequencies f_1 and f_2.

With this selection of circuit elements, mutual decoupling is achieved between channels f_1 and f_2 and the transmission of signals in the common section. To assure the minimum reaction of channels to each other, the lengths of sides a - b and b - c are selected as

$$l_{ab} = \frac{\lambda_2'}{4}, \quad l_{bc} = \frac{\lambda_1}{4}.$$

Thus, the zero impedences $Z_c^{f1} = 0$ at point c and $Z_a^{f2} = 0$ at point a are transformed at junction b to

$$Z_b^{f_1, f_2} = \infty.$$

In order to assure full transmission of energy at frequency f_1 through channel a - b in the direction of the common circuit with otherwise equivalent conditions, we must compensate for the

effect of the reactance of section 1a at this frequency. In this case, this is achieved by introducing additional loop 1b, connected parallel to section 1a at point a. For this parallel connection, we have the resulting impedance at point a for frequency f_1

$$Z_{1a} = \frac{Z_{1a} \cdot Z_{1b}}{Z_{1a} + Z_{1b}} = \infty.$$

For the case of full compensation of reactance, where $Z_{1a} = \infty$, the condition $Z_{1a} + Z_{1b} = 0$ must be observed, i.e., $Z_{1a} = -Z_{1b}$.

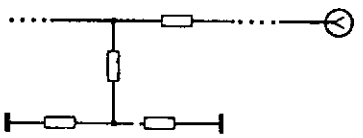


Figure 4. Equivalent Circuit of Filter

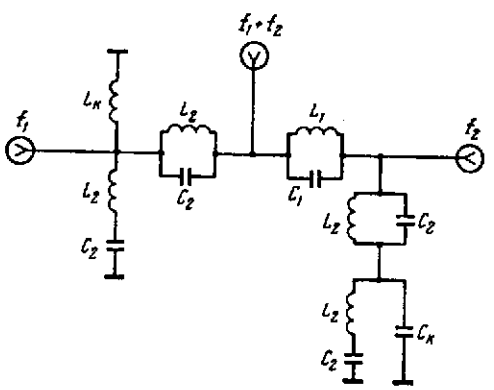


Figure 5. Equivalent Circuit of Hybrid System with Lumped Parameters

The input impedance of a shorted long line sector $Z_{in} = j \tan \beta x$, where $\beta = 2\pi/\lambda = 360°/\lambda$, and x is the length of the line: /148

$$Z_{1a} = j \tan \beta \cdot x_{1a},$$

but

$$\beta_1 = 360°/\lambda_1, \quad \beta_2 = 360°/\lambda_2; \quad \text{since}$$
$$\lambda_1 > \lambda_2, \quad \beta_1 < \beta_2, \; x_{1a} = (\lambda_2/2) < (\lambda_1/2),$$

from which

$$\beta_1 x_{1a} < 180°.$$
$$Z_{1b} = -Z_{1a},$$

$Z_{1b} = -j \tan \beta_1 x_{1a} = j \tan (180° - \beta_1 x_{1a}) = j \tan \beta_1 x_{1b}$.

Let us study the operation of a band pass filter unit (Figure 2.3).

For the side passing the signal at frequency f_2, it is a peculiarity of the circuit that arms a and b change places, i.e., sector 1a, due to the transforming action of sector 1b, provides impedance $Z_{1a} = \infty$ at point a' for frequency f_1 (here, $Z_a = 0$ and decoupling occurs). Section 1b compensates reactivity 1a and frequency f_1 and provides at point a $Z_{2a} = \infty$ (which in turn is transformed at frequency f_2 to some finite impedance at point a'). The selection of section lengths a' - b and b - c' is performed similarly to the preceding case. With full compensation of reactance at point a for frequency f_2, $Z_{2a} = 0$. Suppose the length of sector a - a' is $\lambda'/4$. In this case, the following equation obtains at point a':

$$Z_{2a'} = j\,\mathrm{tg}\,\beta_2 x, \quad x = \frac{\lambda_1}{4}, \quad \beta_2 = \frac{360^\circ}{\lambda_2}, \quad \lambda_1 > \lambda_2, \quad \beta_2 x = \frac{360^\circ \cdot \lambda_1}{4\lambda_2},$$
$$\beta_2 x > 90^\circ, \quad x > \frac{\lambda}{4}.$$

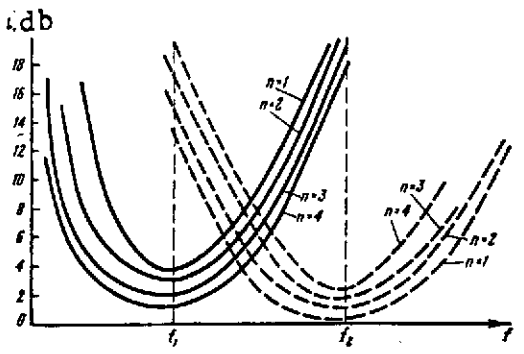


Figure 6. Frequency Characteristics for a Hybrid Version of a Filter

Representing 1 section of the filter f_1 or f_2 individually as an equivalent circuit, as shown on Figure 4, we can easily produce an expression for the impedance of the filter section at the nontransmission frequency. Actually, the condition $Z_b = 0$ indicates that /149

$$Z_б = Z_1 + \left(\frac{Z_2 \cdot Z_3}{Z_2 + Z_3}\right) = 0,$$

i.e.,

$$Z_1 \cdot Z_2 + Z_1 \cdot Z_3 = -Z_2 \cdot Z_3; \quad (Z_1 + Z_2) \cdot Z_3 = -Z_1 \cdot Z_2.$$

Thus

$$Z_3 = -\frac{Z_1 \cdot Z_2}{Z_1 + Z_2}.$$

The practice of operations with these filters at frequencies on the order of 100 mHz has shown that they provide decoupling of at least 25-28 db at very similar frequencies (difference between frequencies about 5%). The direct attenuation introduced for the circuit (Figure 1) is on the order of 2.5-4.5 db, or generally not over 0.5 db for the circuit shown on Figure 2, although in this case the mutual decoupling is only 5-8 db.

The optimal version of the filter is the hybrid version presented on Figure 3. The equivalent circuit with lumped parameters for this version is presented on Figure 5.

In order to increase the steepness of the slopes of the resonant characteristics of filters, we can replace the sectors of length $\lambda/2$ with sectors of length $3\lambda/2$. This increases the input impedance of the sections at the transmission frequency. Compensation is more effective in this case.

Figure 6 shows the frequency characteristics of attenuation for the hybrid version of the filter as a function of n, the number of half wavelengths in the reactive loops. As n increases in the side designed to transmit a given frequency, the pass band expands (due to the increased impedance and decreased Q). At the same time, attenuation at the nontransmission frequency decreases. An experimental study of these circuits has shown that the optimal version is that with $n = 3$.

In the section transmitting the signal at frequency f, as n increases the transmission band becomes narrower, losses at the transmission frequency increase, but attenuation at the blocking frequency also increases.

These filters are successfully used in AFS, where decoupling is required between the transmitter and receivers operating at different frequencies but using a common antenna.

REFERENCES

1. Khalyapin, D. B., Koaksial'nyye i poloskovyye fil'try SVCh [SHF Coaxial Antennas and Filters], Svyaz' Press, 1969.
2. Ayzenberg, G. Z., Korotkovolnovyye antenny [Shortwave Antennas], Chapter 19, Sov. radio Press, 1962.

N75 11006

CHECKING THE PARAMETERS OF AFS IN SPACECRAFT IN FLIGHT

Ye. A. Vlasov, B. A. Prigoda
and A. S. Markin

The use of remote testing devices to check the matching of on-board spacecraft antennas is required by the necessity of receiving timely telemetry information on the quality of operation of the antenna-feeder system (AFS) of the apparatus during long flights, in particular information on the traveling-wave ratio (TWR) and efficiency of the section. /150

In order to allow telemetry testing of these parameters, special devices are included in the AFS: TWR sensors and transmitted power meters.

Naturally, these devices should be designed to be of minimum size and weight while providing the necessary mechanical strength under significant accelerations, and their inclusion in the antenna-feeder system should not cause the characteristics of the system to deteriorate.

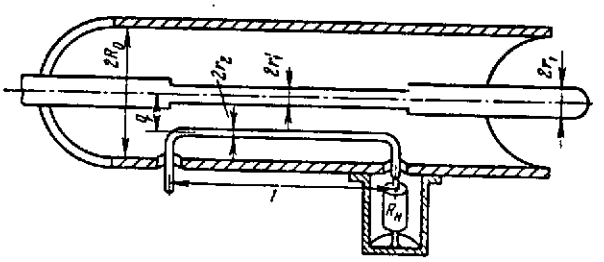


Figure 1. Diagram of Single-Loop Coaxial Directional Coupler

The operating principle of TWR and transmitted power sensors is based on the use of the properties of a directional coupler, calculated approximately below. A single-loop directional coupler is shown schematically on Figure 1.

The following are given in designing directional couplers:

1) on-board transmitter power P;

2) wave impedance of the feeder W_f and secondary line W_2;

3) wavelength λ;

4) power tapped from the primary line to the secondary line D in db or coupling factor between primary and secondary lines for power K_{av}; $D = 10 \log K_{av}$; depending on the power of the on-board transmitter, the coupling factor is 25-45 db;

5) the length of coupling loop l, diameters of the external conductor of the primary line $2R_0$ and central conductor $2R_1$ are selected from structural considerations.

The desired quantities in planning of directed couplers are:

1) coupling factor between primary and secondary lines for voltage K_{cu};

2) coupling wave impedance W_c;

3) distance between coupling loop and central conductor of primary line q;

4) diameter of conductor of coupling loop (secondary line) $2r_2$;

5) diameter of central conductor of primary line over coupling sector $2r'_1$.

The desired quantities are determined as follows: /151

$$K_{cu} = \sqrt{K_{av}}; \tag{1}$$

$$W_c = K_{cu}(W_2/\sin kl), \tag{2}$$

where $k(2\pi/\lambda)\sqrt{\varepsilon}$ is the wave number for a medium with an insulator, ε is the dielectric permittivity of the filler; q is taken from the equation

$$\log\frac{R_0}{q} = \frac{W_c\sqrt{\varepsilon}}{138}; \tag{3}$$

The diameters of the conductors in the coupling loop $2r_2$ and the central conductor $2r'_1$ in the coupling sector are compatible with the solutions of the following equations:

$$W_2 = \frac{138}{\sqrt{\varepsilon}}\left[\frac{\log\frac{R_0}{r'_1}\log\frac{R_0^2-q^2}{r_2R_0} - \left(\log\frac{R_0}{q}\right)^2}{\log\frac{R_0}{r'_1}}\right], \tag{4}$$

$$W'_\Phi = \frac{138}{\sqrt{\varepsilon}}\left[\frac{\log\frac{R_0}{r'_1}\log\frac{R_0^2-q^2}{r_2R_0} - \left(\log\frac{R_0}{q}\right)^2}{\log\frac{R_0^2-q^2}{r_2R_0}}\right]. \tag{5}$$

Solution of equations (4) and (5) gives us:

$$\log\frac{R_0}{r_1'} = \frac{1}{2}\left\{\frac{W'_\Phi\sqrt{\varepsilon}}{138} \pm \left[\left(\frac{W'_\Phi\sqrt{\varepsilon}}{138}\right)^2 + \frac{W'_\Phi}{W_0}\left(2\log\frac{R_0}{q}\right)^2\right]^{1/2}\right\}, \tag{6}$$

$$\log\frac{R_0^2 - q^2}{r_2 R_0} = \frac{W_2}{2W'_\Phi}\left\{\frac{W'_\Phi\sqrt{\varepsilon}}{138} \pm \left[\left(\frac{W'_\Phi\sqrt{\varepsilon}}{138}\right)^2 + \frac{W'_\Phi}{W_2}\left(2\log\frac{R_0}{q}\right)^2\right]^{1/2}\right\}. \tag{7}$$

In formulas (5), (6) and (7), W'_Φ is the wave impedance of the primary line in the coupling sector.

We can easily determine the diameters of the coupling loop and internal conductor of the primary line over the coupling sector from expressions (6) and (7).

The TWR sensor is a two-loop directional coupler with separate high frequency outputs in the secondary line. One of the coupling loops of the TRW sensor reacts to the incident wave in the line, the other -- to the reflected wave. The transmitted power sensors use a single-loop directional coupler, reacting only to the incident wave. A detector-amplifier circuit is used to convert the high frequency signal, proportional to the powers of the incident and reflected waves, to a low frequency signal and amplify it to a voltage convenient for telemetry measurements.

The circuits of TWR and transmitted power sensors are presented on Figure 2 and 3.

The detected high frequency signal from the output of the coupler is amplified by a semiconductor diode voltage multiplier and fed to the telemetry input. The values of U_{inc} and U_{ref}, recorded by the ground telemetry station, can easily be used to determine the traveling wave ratio K in the antenna-feeder system of the spacecraft:

$$K = \frac{1-\Gamma}{1+\Gamma} = \frac{1-\frac{U_{ref}}{U_{inc}}}{1+\frac{U_{ref}}{U_{inc}}} = \frac{1-\sqrt{\frac{P_{ref}}{P_{inc}}}}{1+\sqrt{\frac{P_{ref}}{P_{inc}}}},$$

where Γ is the reflection factor in the AFS; U_{inc} is the voltage of the incident wave; U_{ref} is the voltage of the reflected wave; P_{inc} is the power of the incident wave; P_{ref} is the power of the reflected wave. /152

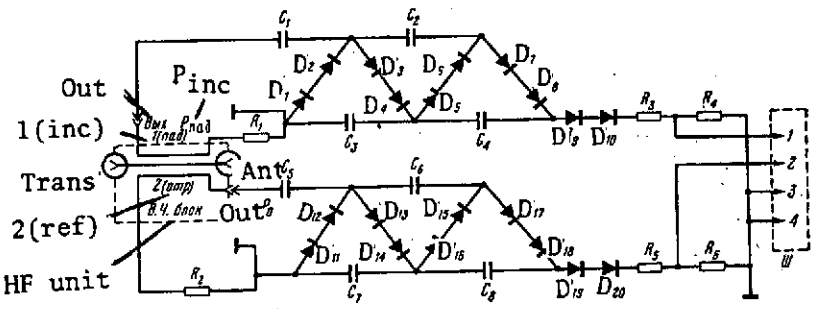


Figure 2. TWR Sensor Circuit

As practice has shown, a diode voltage multiplier, used as an amplifier requiring no additional power supply, produces a quasilinear characteristic U = f(P) through the entire range of powers of on-board transmitters. To increase the accuracy of measurements performed using TWR and transmitted power sensors, calibration graphs $U_{inc} = f(P_{inc})$ and $U_{ref} = f(P_{ref})$ are drawn, as illustrated by Figure 4.

The TWR sensors are generally connected to the AFS in the immediate vicinity of the on-board transmitters, since high values of TWR are particularly important in this area. Transmitted power sensors are usually placed right next to the antennas.

If we know the values of U_{inc} from the TWR and transmitted power sensors recorded by the ground telemetry stations, we can use the calibration graphs of $U_{inc} = f(P_{inc})$ to determine the value of $P_{inc.tr}$ at the output of the transmitter and $P_{inc.ant}$ at the input of the antenna. These values are used to determine the efficiency of the AFS

$$\eta = (P_{inc.ant} \cdot P_{inc.tr}) \cdot 100(\%).$$

The error in determination of the TWR and efficiency using these sensors is not over 10-15%.

The use of two-loop directional couplers with separate high frequency outputs to the secondary line and the use of diode voltage multipliers requiring no power supply and producing the required signal level at the telemetry input in TWR and transmitted power sensors to amplify the detected HF signal have allowed significant reductions in sensor size and weight, as well as simplification of the process of adjustment and regulation.

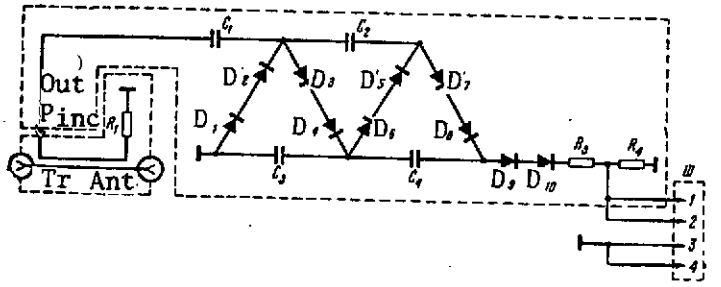


Figure 3. Diagram of Transmitted Power Sensor

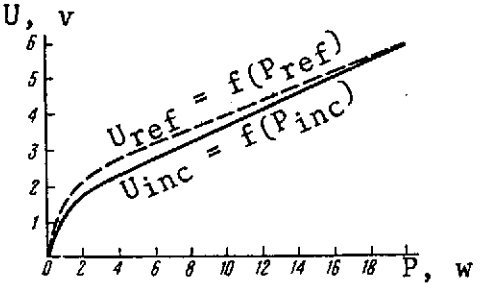


Figure 4. Calibration Graphs for TWR Sensor U = f(P)

The introduction of elements for remote checking of such parameters as TWR deficiency to the AFS has allowed improvement of the quality and reliability of communications with spacecraft and developed telemetric information on the matching of on-board antennas, the efficiency of the feeder system and the power of on-board transmitters during long spacecraft flights. /153

In addition to objective and reliable testing of the AFS parameters of spacecraft in flight, stable radio communications must be constantly maintained between the spacecraft and the Earth, i.e., the potential of the communications link must be above the minimum permissible level. For this purpose, spacecraft carry several antennas operating in turn, depending on which has the maximum gain in the direction of the Earth at each moment. Antennas are generally switched using antenna switches on command from the Earth. However, emergency situations may arise during the flight of a spacecraft, causing antennas of AFS high frequency sections to fail (short circuits, open circuits, etc.), resulting in a sharp drop in the TWR of a section or an increase in the reflected wave level and, therefore, failure of the on-board transmitter, i.e., a complete interruption of radio communications. This defect can be eliminated by using the effect of increased HF energy reflection noted by the TWR sensor as a control signal to switch the on-board transmitter to another AFS.

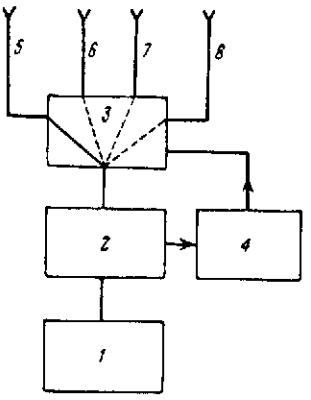


Figure 5. Block Diagram of Automatic Device for Connection of Transmitter to Properly Operating AFS: 1, Transmitter; 2, TWR Sensor; 3, Antenna Switch; 4, Control Mechanism; 5, 6, 7, 8, AFS

Figure 5 shows a functional diagram of a device to perform this switching. The operation of the device is obvious from the block diagram. For example, in case of damage to the AFS with antenna 5, the reflected signal level recorded by TWR sensor 2 will rise to a value exceeding the threshold of operation of control mechanism 4, which will shift antenna switch 3 to position 7, connecting the /154 next AFS, etc. The use of such devices on spacecraft significantly improves the reliability of operation of the on-board radio equipment and of the entire communications system.

The use of the properties of directional couplers may also be helpful for automatic tuning of AFS in spacecraft to achieve the maximum TWR. This purpose also requires feedback to the tuning element control mechanism, for example a reactive loop. The variable voltage of the reflected wave can be used as a control signal.

REFERENCES

1. Altman, J., Ustroystva SVCh [SHF Devices], Moscow, Mir Press, 1968, pp. 184-195.
2. Fradin, A. Z., Ye. V. Ryzhkov, Izmereniye parametrov antenn. [Measurement of Antenna Parameters], Moscow, State Publishing House for Literature on Radio and Communications Problems, 1967.

N75 11007

THE ACCURACY OF AUTOMATIC TRACKING

V. V. Kastrov

In developing automatic tracking systems, it is usually assumed that the accuracy of tracking is a direct function of the rate of increase of the signal in the sensor of the tracking device as the parameter observed increases, i.e., it is considered that tracking accuracy changes similarly to the rate of change of the curve of the measurement conversion. The fact that internal noise increases along with the signals processed by the tracking device (and that tracking accuracy thus drops) in the first stages of processing is frequently ignored. For example, if the measurement conversion $y = \Phi(x)$ is performed in the process of tracking, where x is the instantaneous value of the parameter being observed, y is the signal formed by the sensor in the tracking device, the area where $dy/dx = y' \approx y'_{max}$ is considered most promising for precise tracking. Since any measurement conversion $y = \Phi(x)$ generally has the form of a smooth curve with sectors of insensitivity, maximum sensitivity and saturation (Figure 1), the selection of the working sector for the tracking system is limited to the area adjacent to the point of inflection of this curve.

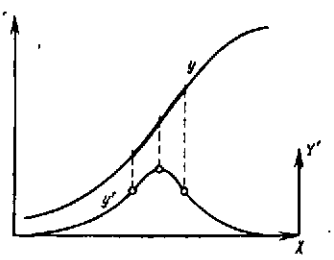


Figure 1. Typical Curve of Measurement Conversion

However, we know from practice that in order to increase the accuracy of tracking, it is frequently necessary to avoid the steepest sector of the measurement conversion curve and decrease the values of the working signals by going over to the lower, nonlinear sector of the curve. The problem of finding the area of the measurement conversion corresponding to the most precise tracking is solved by trial and error and is not always completed.

This problem can be solved unambiguously as soon as the form of the function of the measurement conversion $y = \Phi(x)$ is determined. The main prerequisite for solution is consideration of the dependence of the output signal of the tracking device sensor not only on the measured parameter but on the signal itself: /155

$$y = \Phi(x) \pm \Delta y = \Phi(x) \pm \gamma y, \tag{1}$$

where γ is the relative conversion error.

On Figure 2, the zone of uncertainty of the conversion function, resulting from term $\Delta y = \gamma y$ of equation (1), is shown by the shaded area, expanding with increasing signal value.

As the result of the measurement conversion of process parameter x, the signal takes on a certain value $y \pm \Delta y$, perceived by the user as an erroneous value of parameter $x \pm \Delta x$. We can see from Figure 2 that the error in the parameter is related to the conversion error

$$\Delta x = \frac{\Delta y}{y'} = \gamma \frac{y}{y'}. \tag{2}$$

This equation describes a fact well known from practice: for a specific measurement conversion (for a certain error γ), the absolute value of measurement error is greater, the higher the level of the signal, and less, the steeper the characteristic of the converter. Equation (2) allows us to convert the value of error, calculated or determined experimentally for any point, to any sector of the conversion function. The reciprocal of the error

$$\frac{1}{\Delta x} = \frac{1}{\gamma} \cdot \frac{y'}{y},$$

which is the relative steepness of the conversion function, characterizes the measurement accuracy. The maximum measurement accuracy (and minimum error) can be produced at the point on the parameter where

$$\frac{d(\Delta x)}{dx} = 0 = \gamma \frac{y'^2 - yy''}{y'^2}$$

or

$$y'^2 = yy''. \tag{3}$$

This last equation is the condition of selection of the optimal area for an automatic tracking device.

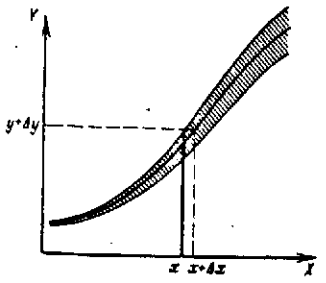


Figure 2. Zone of Uncertainty of Conversion Curve

Analysis of this "criterion of optimality" allows us to draw the following conclusions:

1) the linear sectors of the conversion function are not optimal, since the second derivative should not be equal to 0 with any finite values of y; /156

2) the optimal point is located on the bent section of the curve, since the sign of the second derivatives agrees with the sign of the function;

3) the maximum accuracy can be produced either on the ascending or on the descending branch of the conversion function, since the first derivative may have either sign;

4) conversions passing through the 0 value of the signal have no solution, and each portion of the process (positive and negative) must be studied individually;

5) the quality of the measurement conversion (its relative error) influences the value of measurement error Δx, but does not change the position of the area of the optimal conversion;

6) the presence of a constant "background" (increase in y) not only increases error, but also displaces the area of the optimal conversion.

The optimal area of the measurement conversion for tracking can be found most simply in processes not masked by background noise. When there is no background noise, the initial sector of the measurement conversion always has negligible steepness ($y'_0 = 0$ where $y_0 = 0$). This is explained by the fact that energy is transported in the conversion process, and the efficiency of any transfer device approaches 0 at very low energies -- the energy is dissipated within the device. Therefore, for processes not weighted by background noise the optimality criterion always determines the initial point. The equation $y'^2 = yy''$ where $x = 0$ is fulfilled unconditionally. Furthermore, a continuous monotonic conversion function having $y_0 = y'_0 = 0$ containing optimal points other than the point $x = 0$ cannot be suggested. Thus, the problem of selection of optimal conditions of operation of a tracking device as concerns accuracy is solved unambiguously.

The position of the point of optimal conversion is influenced by the background level, i.e., the portion of the total signal not following the conversion rule $y = \Phi(x)$. The origin of the background noise is insignificant: it may belong to the field of measurement (for example, background light) or may appear in the converter itself due to various leaks. Therefore, in principle, a significant portion of the background signal can always be compensated in the input devices of the converter. The portion of the total background signal which cannot be eliminated for some reason is dangerous.

The conversion function including background noise can be represented as the sum of the variable component and the background signal: $y = \Phi(x) + C$.

The influence of noise on accuracy and the position of the optimal point can be estimated only for specific conversion functions. An example of a rather universal conversion function is a sine wave:

$$y = 1 - \cos x + C$$

where C is the background signal. The position of the optimal point is determined according to (3) from the equation

$$\sin^2 x_{opt} = (1 - \cos x_{opt} + C) \cos x_{opt},$$

since

$$x_{opt} = \arccos (1/1 + C).$$

Figure 3 shows: the form of the function (without background noise, C = 0), the graph of "motion" of the optimal point as a function of background level and the graph of error at the optimal point (relative conversion error γ arbitrarily taken as /157 unity). The error at the optimal point is determined from equation (2) as

$$\Delta x_{opt} = y/y' = y'/y'' = \tan x.$$

The curves of Figure 3 show a very strong dependence of the position of the optimal point on background noise level. For

example, noise amounting to only 1% of the maximum signal ($C = 0.02$, point a_1, Figure 3) displaces the optimal tracking point by $0.2/\pi \cdot 100\% = 6.4\%$ of the entire range of measurement of the parameter. As noise increases further, the optimal point asymptotically approaches $x = \pi/2$, the point of inflection of the function. At the same time, error Δx increases without limit.

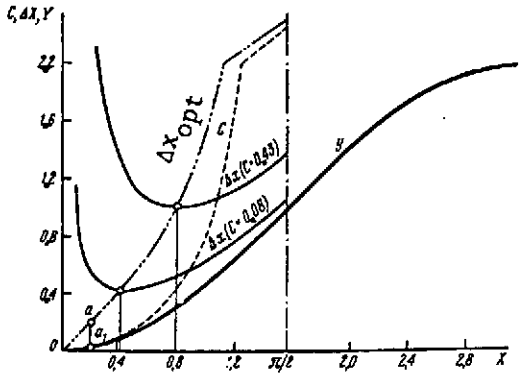


Figure 3. Graphs for Processing of Measurement Conversion Function $y = 1 - \cos x + C$; C is the Noise Level Corresponding to Displacement of the Optimal Tracking Point; Δx_{opt} is the Error at the Point of Optimal Tracking (with Relative Error of Converter $\gamma = 1$)x; Δx ($C = 0.08$) and Δx ($C = 0.43$) are the Errors with Non-Optimal Tracking with Noise Levels $C = 0.08$ and $C = 0.43$ Respectively

During the process of tracking, the measurement converter may operate under conditions significantly different from the design conditions. For tracking devices in particular, the relationship between the actual background noise level and the working area selected is important. When tracking is performed at a point on the parameter which is not optimal for the noise level in question, error increases. This can be seen easily on the example of the same sinusoidal conversion function, for which the error equation can be written as

$$\Delta x = \frac{y}{y'} = \frac{1 + C - \cos\,(x_{opt} \pm \delta)}{\sin\,(x_{opt} \pm \delta)},$$

where x_{opt} is the optimal value of the parameter for the background noise level; δ is the deviation of the parameter from its optimal value. Figure 3 shows two "error curves" for non-optimal tracking, calculated for $C = 0.08$ and $C = 0.43$. The curves illustrate the loss of accuracy arising upon improper selection of the working point of the tracking device quite well. For example, with a noise level $C = 0.08$, selecting the working point near the point of inflection of the conversion function ($x = \pi/2$) instead of point $x = 0.4$ increases the error from $\Delta x = 0.42$ to $\Delta x = 1.08$, i.e., by 2.5 times.

The use of this method of estimating the working parameters of a converter was studied on the example of a solar guide, a device determining the direction to the sun. This type of direction sensor is widely used not only in astronomy but in spacecraft control systems, and the requirements for their accuracy are continually increasing.

This type of converter is based on a telescopic system (Figure 4), a portion of the focal plane of which is obscured by a screen. The working signal is created by the light flux from the edge of the solar disc projecting from behind the screen. The width of this edge (sufficient for operation of the tracking system) is generally taken as the conversion error. /158

Since the analysis which we are performing here is illustrative, we can use the following simplifying assumptions:

1) the image of the sun in the focal plane is a circle of even brightness with sharp edges;

2) the lens of the guide consists of a single plano-convex thin lens;

3) the background noise is created by scattering of sunlight in the lens.

A diagram of the guide is shown on Figure 5, the form of the focal plane -- on Figure 6. In all calculations, the radius of the image of the sun will be taken as 1. The other linear dimensions will be defined in units of solar diameter.

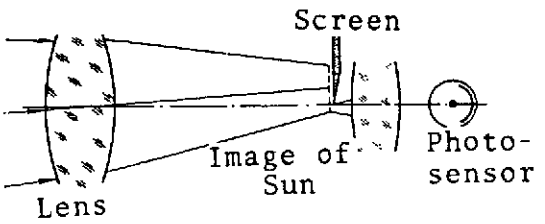


Figure 4. Diagram of Solar Guide

The signal from the guide is proportional to the level of illumination of the image of the sun E_{im} and the area of the portion of the image projecting from behind the screen S:

$$y = G(E_{im}S + C). \quad (4)$$

Here G is a factor considering the sensitivity of the photosensor and the amplification of the signal in the electronic portion of the guide; C is the background light flux. In turn,

$$E_{im} = (d/2)^2 E_{\odot}, \quad (5)$$

where $E_\odot$ is the solar constant; d is the light diameter of the lens. If we base ourselves on a specific angular radius of the sun ε, the diameter of the lens can be selected on the basis of its relative aperture A:

$$d = A/\varepsilon \text{ (with small } \varepsilon). \tag{6}$$

The area of the projecting portion of the image of the sun

$$S = 2\int_x^1 \sqrt{1-x^2}\,dx = \arccos x - x\sqrt{1-x^2}, \tag{7}$$

where x is the distance of the center of the image from the edge of the screen. The formation of the background due to scattering in the lens is reduced to the following: the sunlight is twice reflected from the flat and spherical surfaces of the lens so that the elementary points of light arriving from the sun with cross section δs_1 are distributed over areas δs_2 in the focal plane. The total light flux of the background from the entire transparent portion of the focal plane is thus:

$$C = \frac{\pi}{2} E_\odot R^2 \eta^2 \frac{\delta s_1}{\delta s_2} = \frac{\pi}{2} E_\odot R^2 \frac{(n-1)^6}{(n+1)^4 (3-n)^2}. \tag{8}$$

Here R is the radius of the field of vision in the focal plane; /159

$\eta = \left(\frac{n-1}{n+1}\right)^2$ is the reflection factor on the surface of the lens, made of a material with refractive index n; $\frac{\delta s_1}{\delta s_2} = \left(\frac{n-1}{3-n}\right)^2$ is produced from geometric considerations (see Figure 5), since the radius of curvature of a thin convex lens (r), its focal length (f') and the index of refraction are related by the equation

$$r = f'(n - 1).$$

Thus, the total signal of the guide, considering dependences (4-8), can be written as

$$y = \frac{GE_{\odot}A^2}{4\varepsilon^2}\left[\arccos x - x\sqrt{1-x^2} + \frac{2\pi R^2\varepsilon^2(n-1)^6}{A^2(n+1)^4(3-n)^2}\right] =$$
$$= \frac{GE_{\odot}A^2}{4\varepsilon^2}\left[\arccos x - x\sqrt{1-x^2} + C'\right]. \qquad (9)$$

The position of the optimal point, according to condition (3), is determined by solving the equation

$$C' = \frac{2-x^2_{\text{opt}}}{x_{\text{opt}}}\sqrt{1-x^2_{\text{opt}}} - \arccos x_{\text{opt}} \qquad (10)$$

The error at the optimal point

$$|\Delta x_{\text{opt}}| = \frac{y'}{y''} = \frac{1-x^2_{\text{opt}}}{x_{\text{opt}}}. \qquad (11)$$

The error with non-optimal tracking /160

$$|\Delta x| = \frac{y}{y'} = \frac{1}{2}\left[\frac{\arccos x + C'}{\sqrt{1-x^2}} - x\right]. \qquad (12)$$

In these equations, the quantity

$$C' = \frac{2\pi R^2\varepsilon^2(n-1)^6}{A^2(n+1)^4(3-n)^2}$$

can be called the background factor, since it is the ratio of the background signal to the signal created by half of the solar disc.

Figure 7 shows the initial sectors of the curves $C(x_{opt})$ and $\Delta x(x_{opt})$, corresponding to values of x near unity (tracking of the edge of the image of the sun), and the error curves in the mode of non-optimal tracking of the sun (ε = 0.0045) using a guide with the following parameters: R = 20 (2β = 10°), A = 0.2 and n = 1.6. In the lower portion of Figure 7, the dotted line shows the curve of the actual conversion function (without noise). The angular values of the projecting portion of the solar disc are shown on a separate scale on the X axis.

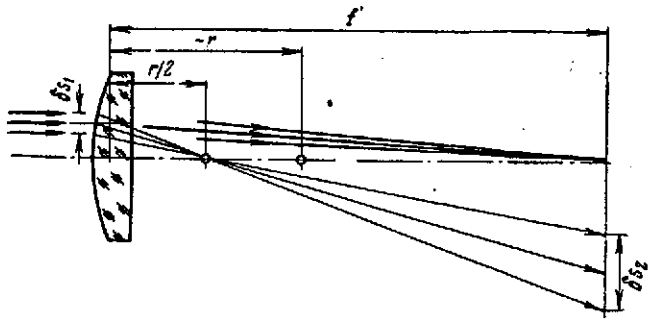


Figure 5. Calculation Plan of Solar Guide

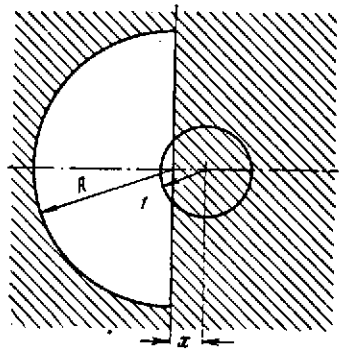


Figure 6. View of Focal Plane of Guide (Circle of Radius 1 is Image of the Sun)

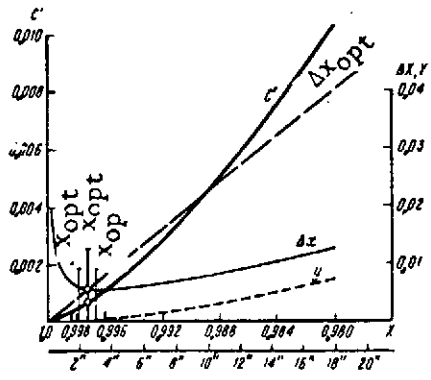


Figure 7. Graphs for Processing of Conversion Function of Solar Guide (for Symbols, see Figure 3 and Text)

Background factor C' may differ significantly from 0.00066, the value calculated in this example. The reasons for variation of C', in addition to changes in parameters of the guide, are complication of the lens (increasing the number of reflecting surfaces), the use of coatings on the surfaces of the lens, scattering of light in the glass and internal walls of the device, the intensity of the background level around the object of observation, etc. Each actual value of background level corresponds to a specific curve Δx, limiting the attainable accuracy.

In this example, the greatest accuracy is produced if the tracking system is adjusted to a signal corresponding to a value of x = 0.9972 (about 2.5 seconds of arc from the edge of the solar disc). However, two working points are necessary for operation of the tracking system: the point of operation and the point of release. The distance between these points must be at least the sum of the errors at the two points. Thus, the total uncertainty of the value

of parameter x (in this case the uncertainty of the distance of the center of the image of the sun from the edge of the screen in the field of vision of the guide) is

$$\gamma \cdot 2(\Delta x_{op} + \Delta x_{rel}).$$

For example, where $\gamma = 0.1$ and $C' = 0.00066$, the working points might be $x_{rel} = 0.9978$, $x_{op} = 0.9966$, symmetrical relative to the optimal point. The total uncertainty of the position of /161 the center of the image of the sun (the tracking error of the system) is approximately $0.1 \cdot 4\ (1.1 \cdot \Delta x_{rel}) = 0.0024$ or 2.2 seconds of arc. Tracking is performed not for the edge of the solar disc, but rather for a point 2.5 angular seconds from the edge.

REFERENCES

1. Novitskiy, P. V., Osnovy informatsionnoy teorii izmeritel'nykh priborov [Principles of the Information Theory of Measuring Instruments], Leningrad, Energiya Press, 1968.
2. Shestov, N. S., Vydeleniye opticheskikh signalov na fone sluchaynykh pomekh [Separation of Optical Signals from a Background of Random Noise], Moscow, Sov. radio Press, 1967.
3. Tudorovskiy, A. I., Teoriya opticheskikh priborov [The Theory of Optical Devices], Acad. Sci. USSR Press, 1948.
4. Yakushenkov, Yu. G., Opticheskiye sistemy fotoelektricheskikh ustroystv [The Optical Systems of Photoelectric Devices], Moscow, Mashinostroyeniye Press, 1966.

N75 11008

COMMUNICATIONS BY LIGHT WITH LONG-RANGE SPACECRAFT[1]

M. Z. Novikov

Communications with spacecraft are usually conducted in the radio band, making them independent of weather conditions and allowing high information transmission rates over practically attainable distances. However, equipment has been developed for light contact (contact in the optical range), primarily consisting of devices for observation and measurement of the angular coordinates of spacecraft. These devices are capable of high measurement accuracies. Furthermore, to the extent that the light communications lines are not connected to the radio lines, failures in radio communcations will not result in total loss of contact with the spacecraft and prevent continued transmission of information from the spacecraft to the Earth. The promise of light communications lines was indicated as early as 1958 by I. S. Shklovskiy in [1], dedicated to the observation of Earth satellites.

The subject of this article is light contact with long-range spacecraft located several hundreds of thousands of kilometers from the Earth, particularly with spacecraft flying between the Earth and the moon. The low brightness of long-range spacecraft makes it necessary to equip them with on-board light signalling devices. In the following, we will study the possibility of creating such devices. We will also touch upon certain problems of the construction of devices for receiving light signals.

1. The Natural Brightness of Spacecraft

Let us estimate the possible stellar magnitudes of long-range spacecraft due to reflected sunlight. In most cases, the spacecraft are cylindrical in shape. The level of illumination E at the edge of the Earth's atmosphere created by a cylinder diffusely reflecting the light of the sun can be expressed by the formula

$$E = 2500 \frac{MB}{R^2} f(\varphi), \tag{1}$$

[1]Work performed in 1969.

where M is the area of an axial cross section of the object; R is the range; ϕ is the angle between the direction from the object to the Earth and the direction from the object to the sun; B is the brightness of the surface at the point on the object perpendicular to the ray from the sun. If we ignore darkening, the value of B can be taken as 3 cd/cm^2. /162

For the case when the axis of the cylinder is perpendicular to the ecliptic plane (illumination E_1), we produce $f(\phi)$ in the form

$$f_1(\varphi) = (\pi - \varphi)\cos\varphi + \sin\varphi. \tag{2}$$

It follows from formulas (1) and (2) that at the ranges of interest to us, the brightness of objects is low. For example, the stellar magnitude of a comparatively large cylinder with $M = 20\ m^2$, 300-400 thousand kilometers from the Earth, is in the range between 14 and 15^m for $\phi = 0$, 21-22^m for $\phi = 170°$.

For a cylinder with its axis in the plane of the ecliptic, we find that illumination E_2 created by its side surface is determined by formula (1) for $f(\phi)$ equal to

$$f_2(\varphi) = \pi \cos\alpha \cos(\varphi - \alpha), \tag{3}$$

where $(\phi - \alpha) < 90°$ or $(\phi - \alpha) > 270°$, α is the angle between the direction from the object to the Earth and the plane of its cross section.

It is also easy to calculate the light reflected from the end surfaces if they are flat. With identical reflection factor, the illumination E_T created by a flat end with area M_T is

$$E_T = 10^4 \frac{BM_T}{R^2} \sin\alpha \sin(\varphi + \alpha).$$

Under the conditions of the example described above, with, let us say, $\alpha = 0$, the stellar magnitudes of the cylinder, as expressions (1) and (3) indicate, vary from 14-15^m where $\phi = 0$ to 17=17.5^m where $\phi = 85°$. In the interval $90° < (\phi - \alpha) < 270°$, there is no brightness at all, since the portion of the side surface of the cylinder turned toward the Earth is not illuminated. Where $\phi = 2\alpha$, illumination E_2 reaches its maximum value,

corresponding to the maximum $f_2(\phi)$, equal to $\pi \cos^2 \phi/2$, and for all values of ϕ the value of E_{2x} is greater than E_1.

There is also interest in the case of a cylinder rotating about its transverse axis, which can take up any position in space. The light flux reflected toward the Earth in this case changes periodically from a minimum, which may approach 0, to an amplitude which determines the possibility of observing the object. With any position of the axis of rotation, there is an instantaneous position of the cylinder which is an average between the position for which the axis is perpendicular to the plane of the ecliptic and the position yielding illumination E_{2max}. It can be shown that upon transition from one of these positions to another, the characteristics of reflection change slowly, and the illumination at any intermediate position will be intermediate between E_1 and E_{2max}.

We can conclude from this that in all possible positions of the axis of rotation, the minimum value of illumination amplitude is E_1, the maximum value is E_{2max}.

2. Operating Conditions of Light Signal Line

We will base our discussions on the condition that the ground observation points are equipped with telescopes with diameters of at least 50 cm. The observation conditions are such that observation time is limited and it is impossible to await favorable atmospheric conditions. We should therefore design for the most probable level of turbulence, which, according to many investigations (for example, see [2]), has a /16
characteristic angular diameter of the flickering disc of about two seconds. The angular diameter γ of a "point" image, which generally depends on the image receiver and lens, can be considered approximately equal to this value in this case.

The data from the preceding section show that the requirements of observation can be satisfied to some extent by visual observation, since a value of 15^m can be visually observed under ideal conditions by a telescope with a diameter of 50 cm. This requires, in particular, that the spot imaging the object in the focal plane of the lens not exceed the resolving capacity of the eye after magnification (about 1 minute), i.e., that the angular diameter of the image spot be not over 0.6 seconds. Then, where $\gamma = 2$ seconds, the optical power of the telescope drops by 2.5^m. Poor results also are produced by photography of long-range spacecraft, since their angular motions prevent exposures from being increased to more than a few tens of seconds (only in the best cases), i.e., objects no weaker than 12^m can be recorded. An additional reduction in the effectiveness of

visual observation in photography may result from insufficient image contrast. Thus, photography is also not suitable for tracking of long-range apparatus and visual observation.

The conditions of observation of long range apparatus can be satisfied by electronic image receivers consisting of high-sensitivity electronic-optical converters (EOC) or television systems which, as we know, can increase the sensitivity of receivers by hundreds of times. For example, a television system designed for astronomy described in the literature can observe stars of magnitude 16.5^m with a 50 cm telescope [3]. This system provides some capabilities for tracking long-range spacecraft using reflected sunlight, but does not cover the entire range of distances and phase angles. Here also, we must expect some drop in effectiveness at low contrast levels.

One common shortcoming of electronic receivers is their low resolving capacity, which may influence the accuracy of measurements of coordinates, particularly in television systems with their line image structure. In order to avoid significant measurement errors, the focal length F of the lens should be rather long. It is easy to see that on a television image, the center of the bright spot of an object will be displaced from its true position by a distance which may be as great as one half of one line width. The line width in the plane of the photocathode for various television systems, like the size of an image element in an EOC, is 20-40 μm. If, for example, this displacement should not exceed 1 second, the focal length must be at least F_{min} = 2-4 m. The focal length for which the diameter of an image element in an electronic receiver d_{el} is equal to the probable diameter of the flickering disc will be represented as F_0 (in this example $F_0 = F_{min}$) as we study the influence of the focal length on the sensitivity of a receiver. The sensitivity, as we know, is determined by two characteristics: the sensitivity to the energy of the light signal received and the degree of harmful influence of the sky background. Each of these characteristics depends on the resulting image element of the receiver, which can be assumed homogeneous. Its diameter is

$$d_{res} = d_{el} + \gamma F.$$

The angle related to an image element, within which a resolvable sector of the sky is observed, is equal to

$$\gamma_{res} = (d_{el}/F) + \gamma. \qquad (4)$$

The illumination of the sky background is proportional to /164

γ_{res}^2. The direct influence of the sky background is a reduction in visual contrast, manifested in the EOC. Furthermore, in both EOC and television receivers, the sky background creates noise, which masks the return from the object. The noise level is proportional to the square root of the illumination, i.e., to γ_{res}.

According to (4), as F increases from F_0 to $2F_0$, the value of γ_{res} decreases by 25%, while a further increase in F causes γ_{res} to drop slowly, leading at the limit ($F \to \infty$) to a decrease in γ_{res} by another 25%. The value of γ_{res}^2 decreases by 44 and 31% respectively. Thus, with the most probable value of γ, an increase in focal length decreases the effect of the background very little. As concerns the sensitivity to the energy of the signal received, an increase in focal length reduces it, and this effect is not limited. The effect of the light radiation received is distributed over the resulting image element and decreases with increasing d_{res}.

Therefore, in the case of "point" images, we can introduce the concept of the effective illumination E_{pc} on the photocathode, equal to the ratio of the light flux received from a point object to the area of the resulting image element. Noise within the area of the image from an object disrupts it, and the sensitivity to the energy is determined by the signal/noise ratio in this area. Since the output effect is proportional to E_{pc}, while the noise is proportional to $-\sqrt{E_{pc}}$, the signal/noise ratio in this area will be proportional to

$$D/\gamma F + d_{el} = A/\gamma_{res}, \qquad (5)$$

where D is the diameter and A is the relative aperture of the telescope. As F changes above F_0, the signal/noise ratio will change in proportion to $1/d_{el}\ (1 + F/F_0)$. Where $F = 2F_0$, the sensitivity to energy becomes 1.5 times less than where $F = F_0$. With a further increase, the decrease continues and, according to (5), in order to maintain constant sensitivity, the diameter must be increased only slightly less than to maintain unchanged relative aperture.

Keeping in mind also the influence of background and the sensitivity to the energy of the signal received, we can conclude that the value of $F = 2F_0$ is near optimal.

It was assumed above that the rate of recording of an image was significantly greater than the rate of its displacement, since the effectiveness of the signal is determined by the illumination which it creates. Electronic recording of images of long-range spacecraft satisfies this condition. In other cases, the exposure is significant which, as we know, must be considered in calculations for photography of Earth satellites. In the latter case, based on expression (4), it is easy to see that the sensitivity is proportional to $D^2/(\gamma F + d_{el})\omega F$, which, where $\gamma F >> d_{el}$, is converted to $A^2/\gamma\omega$, where ω is the angular velocity of the spacecraft.

Let us estimate the brightness of the background created by the night sky under the worst condition, i.e., with a full moon, based on the brightness of the day sky. We can consider it sufficient to provide for observations when the spacecraft is outside the sphere of influence of the moon, i.e., with angular distances of the spacecraft from the moon of at least 10 degrees. According to B. V. Pyaskovskaya-Fesenkova [4], the maximum brightness of the day sky at an angular distance from the sun of over 10 degrees is not over 1.5 cd/cm^2, corresponding to a brightness of 2^m per square second of sky. Conversion to full moon conditions yields approximately 16^m per square second of sky.

Restricting ourselves to existing types of telescopes, we will base our calculations on the comparatively low-power AZT-14 telescope (D = 48 cm, F = 7.715 m). We will also assume that an electronic receiver is used, the resolving capacity of which is used exhaustively ($d_{el} \approx 20$ μm). The solid angle filled by an image element in this system is approximately 5 square angular seconds (γ_{res} = 2.5 seconds), and the brightness of the background, calculated per image element under these conditions, is about 14.5^m. /165

It is important to note here that electronic receivers, reproducing the image with the unavoidable low level of illuminance in this case (on the order of 10^{-5} lx) create a high noise level. According to the studies of A. M. Khalfin, even for a so-called "ideal" light convertor with illuminances of this order of magnitude, the signal/noise ratio approaches 1 [5]. But with this signal/noise ratio, details with diameters of several image elements begin to be distinguishable against the noise background only with a contrast level of about 1 [6].

Let us therefore take the necessary contrast level as 1. Since the angular size of the spacecraft is negligible, in order to produce this contrast level, the brightness of the spacecraft must correspond to 14.5^m.

We will assume the use of a multi-slit photocathode, due to its high sensitivity, low dark current and good state of development. The characteristics of the photocathode presented in [7] are used in our calculations.

The spacecraft should be observed from many points on the Earth's surface, distributed as widely as possible as to longitude, in order to increase the observation time. For on-board light signaling devices, this requirement defines the angle of divergence of radiation β. Keeping in mind the provision of visibility of a light signal from the side of the Earth toward the spacecraft at a distance of 100,000 km or more (at shorter ranges, visibility is generally sufficient without artificial light signals) and assuming an accuracy of aiming of the on-board light source toward the Earth of no worse than $\pm 1.5°$, we can assume $\beta = 10°$. We note that the production of a simple reflector (searchlight type) optical system with a radiation divergence of 10° represents no difficulties. Further discussions will be concentrated on the question of suitable light sources, which may be various devices, some already used in space technology. These include active continuous and pulsed light sources, as well as passive sources such as mirrors and artificial clouds.

In each observation session, in order to allow reliable search, measurement of coordinates and, possibly, transmission of information, the light signal from the spacecraft must be received continually and reliably for a rather long time interval. If we also consider the requirements of small size and weight, we find that all of these requirements are most fully satisfied by constant-burning sources, which we will now analyze.

3. Signalling Using Constant Burning Light Sources

The following effective, powerful constant light sources have been developed: high temperature incandescent bulbs, xenon lamps and mercury lamps. The radiation of the currently popular mercury-vapor lamps (types DRSh and DRL) falls mainly in the shortwave section of the spectrum and is strongly dissipated in the atmosphere. The lamps have long (up to 5 minutes) unstable warm-up times, preventing precise regulation of observation sessions. All of this makes mercury vapor lamps unsuitable for signalling purposes.

In xenon lamps, the spectral distribution of radiation in the area of operation of multi-slit photocathodes is practically constant. The lamps are noninertial, have a low luminescent volume (3-5 mm) and rather high light output (at least 30 lm/W). /166

However, they ignite at a voltage higher than the on-board voltage, and require a separate voltage transformer. Furthermore, as in all cases when high-current gas discharge devices are used, additional equipment is required to eliminate high frequency interference. In all, the prospects for the use of xenon lamps should not be forgotten.

There is doubtless interest in incandescent bulbs. The fact that they can be plugged directly into the on-board power supply and the lack of interference together with their stability and small size all make incandescent bulbs attractive possible light sources. In particular, small lamps with the iodine cycle, type KIM, can be used. These lamps have high effectiveness and better reliability than ordinary lamps, and can operate at a color temperature of 3300°K with a light output of 30 ℓm/W. Let us calculate the power of incandescent and xenon lamps necessary for operation of a light signalling device.

The spectral sensitivity curve v_λ of the eye, it has been discovered, can be considered roughly symmetrical, if we take the ordinate corresponding to the wave length $\lambda_0 = 0.558$ μm as the center of symmetry. This fact allows us to simplify calculations for a filament lamp as concerns determination of light values by the following approximation of the curve for its spectral distribution $\Phi(\lambda)$ in the visual wave range. Setting the coordinate origin λ at point $\lambda = \lambda_0$, we can represent $\Phi(\lambda)$ as the sum of a flat spectrum with intensity $\Phi(\lambda_0) = \Phi_{av}$ and components Φ' near an odd function of λ. Light flux L is equal to

$$L = 683 \int_{-\infty}^{\infty} \Phi(\lambda) v_\lambda \, d\lambda = 683 \int_{-\infty}^{\infty} (\Phi_{cp} + \Phi') v_\lambda \, d\lambda.$$

Taking this integral within symmetrical limits and considering Φ' an odd function, we produce $L_{(lm)} = 73\Phi_{av}$ (W/μm).

Representing Φ_{max} as the maximum value of spectral distribution and $k = \Phi_{av}/\Phi_{max}$, we produce

$$L = 73k\Phi_{max}. \tag{6}$$

In the working range of color temperatures of incandescent bulbs (2800-3300°K), the ratio of visual brightness of tungsten to visual brightness of an absolutely black body is practically

constant at 0.44 [8]. The curves of spectral distribution of radiation of tungsten and of an absolutely black body in the visual area are identical -- component Φ' in both cases is approximately linear; therefore, the ratio of the ordinates of these curves can also be considered constant, equal to 0.44. This factor allows us to use the characteristics of an absolutely black body for calculations for incandescent bulbs and, in particular, for determination of coefficient k. A check using specific examples has shown that the error in calculation with formula (6) for 500-1000 w bulbs is not over 10%. Let us determine the integral sensitivity ε of a multi-slit photocathode exposed to radiation $w(\lambda)$ of various spectral compositions:

$$\varepsilon = g_{max} \int_0^\infty w_\lambda g_\lambda \, d\lambda / 683 \int_0^\infty w_\lambda v_\lambda \, d\lambda, \tag{7}$$

where $w_\lambda = w(\lambda)/w_{max}$; w_{max} is the maximum value of spectral distribution of radiation; g_{max} is the maximum value of spectral sensitivity $g(\lambda)$ of a photocathode: $g_\lambda = g(\lambda)/g_{max}$.

For incandescent bulbs, according to (6), /167

$$683 \int_0^\infty \Phi_\lambda v_\lambda \, d\lambda = 73k,$$

and the value of ε is determined from the formula

$$\varepsilon = \frac{g_{max} \int_0^\infty \Phi_\lambda g_\lambda \, d\lambda}{73k}. \tag{8}$$

The results of calculations using formulas (7) and (8) for various radiations are presented in Table 1.

TABLE 1

Type of Radiation	ε, μa/lm
Radiation of incandescent bulb with color temperature T_c = 2848°K	122.5
Radiation of incandescent bulb with color temperature T_c = 3300°K	126.6
Radiation of xenon bulb	177.3
Solar radiation	101.5

The calculations were performed in the 0.4-0.85 μm wave range. The sector $\lambda < 0.4$ μm was not considered in the calculations due to the strong absorption of this radiation by the atmosphere and glass. Due to this, the value of ε_{st} calculated for T_c = 2848°K (standard source) was found to be somewhat below that presented in [7], ε = 132 μa/lm.

We note that modern highly sensitivit electronic light receivers produce acceptable image quality for extended objects with illumination on the photocathode $\varepsilon_{pc} = (1\text{-}2)\cdot 10^{-5}$ lx, i.e., with a photoelectric current density $\varepsilon_{st}E_{pc} = (0.12\text{-}0.24)\cdot 10^{-6}$ μa/cm^2. Recognition of point images is much more difficult for an observer than recognition of extended images, since the noise is also present in the form of points. This factor hinders subjective comparison of these images. However, effective comparison of point and extended images can be performed using the value of E_{pcd}, which can be calculated as

$$E_{pcd} = E\tau(A^2/\gamma_{res}^2).$$

where τ is the transmission factor of the atmosphere and the objects, which can be taken as 0.4 in the worst case. A brightness of 14.5^m in the solar spectrum corresponds to a level of illumination at the boundary of the atmosphere $E_{sol} = 4.38\cdot 10^{-12}$ lx, i.e., $E_{pcd} = 4.6\cdot 10^5$ lx, and the photoelectric current density $j_{sol} = \varepsilon_{sol}E_{pcd} = 0.47\cdot 10^{-6}$ μa/cm^2, which is greater than the value mentioned above.

The light flux of a source can be found from the following expression, considering the necessity of conserving $j = j_{sol}$:

$$L = E_{sol}(\varepsilon_{sol}\pi\beta^2R^2/\varepsilon_{lamp}4\eta),$$

where η is the light efficiency of the light source, $\eta \approx 0.8$. Let us assume R = 400,000 km, since this is the maximum distance for a spacecraft travelling to the moon.

According to (9), the light flux of an incandescent bulb with T_c = 3300°K and a xenon lamp should be L_{3300} = 16,700 ℓm, L_x = 11,900 lm. With a light output of 30 lm/w, the power of the light source incandescent bulb should be about 600 w. With a power of 600 w, eight sessions of light contact (which can be considered sufficient) of 2.5 min each consume an energy of 0.2 kw·hr, requiring an increase in battery weight of not over 2.5 kg. /168

A flux of 11,900 lm can be provided by a xenon lamp with a power of 0.4 kw. Power consumption will be increased by the separate power supply to approximately 0.6 kw. Of the series-produced xenon bulbs, the closest to this power level is the DKShS-500 bulb. The dimensions of this bulb are diameter 300 mm, light source body 2.5 mm, quite suitable for use in a spacecraft light source.

Finally, concerning the possibility of information transmission. The light source will not limit the transmission rate if non-inertial light flux modulators are used. The use of a xenon lamp yields great possibilities for transmission by modulation of the anode voltage (in the 20 KHz frequency band). In these cases, the information should be received using non-inertial devices -- photomultipliers. Even without modulators, some information can be transmitted by manipulating the light source power supply voltage or by mechanical interruption of the light beam (for example, with a gate) at a rate of less than 1 cycle per second. In this case, the message can be received directly from the image receiver.

4. Other Means of Signalling

Passive signalling equipment using solar radiation has been used to some extent in space technology. One such means consists of artificial clouds, studied by I. S. Shklovskiy [9], who suggested the method of increasing brightness by means of an atomic gas cloud creating resonant radiation, successfully used in 1959. It was shown in [9] that with the same weight of diffused substance, a gas cloud is more effective than an artificial dust cloud both in the energetic respect (intensity of scattered radiation) and, more particularly, in the respect of contrast as the brightness of the background increases. However, extremely high pressures and temperatures develop in a gas cloud generator, so that it must have a massive design, which is not

always possible. A dust cloud, on the other hand, is created using comparatively light pneumatic sprayers.

It is currently possible to produce powders with mean particle diameters of about 1 μm. Particles of this size create practically even scattering throughout the cloud, except for a small sector near the direction of the initial light ray, which is not of interest in this case, and an effective scattering cross section considering the evenly scattered portion of the light equal to the cross sectional area of a particle [10].

This means that the effective cross section Σ of even scattering for the entire cloud is

$$\Sigma = P/pa, \tag{9}$$

where P and p are the weight and specific gravity of the powder dispersed, a is the mean particle diameter.

In the case of corundum powder, currently used to form aerosols, we find that 1 kg of powder with a particle diameter of 1 μm will create a dust cloud with $\Sigma = 330\ m^2$ (specific gravity of corundum 3 g/cm^3). It is realistic to assume the following parameters of a dispersed gas: volume about one liter, pressure up to 10 atm. Then, with a powder weight of 1 kg, the rate of expansion of the cloud will be not over 45 m/sec, and at 200-400,000 km range the cloud will expand beyond the limits of one image element (2.5 seconds) over at least 25-50 sec. Up to this moment, a cloud with $\Sigma = 330\ m^2$ would be equivalent to a diffuse reflecting cylinder with a longitudinal cross sectional area $M = 150\ m^2$ at $\phi = 0$, i.e., at a range of 400,000 km the cloud would have a stellar magnitude of 13^m. Design development performed by Yu. F. Perevertov has shown that the weight of one such charged device with an attachment for spraying from the spacecraft would be about 3 kg. /169

However, it is important to note that creation of a dust cloud encounters the problem of conserving the fractional composition of the powder, which changes due to clumping of the particles into aggregates. Studies of glass and quartz powders [11] have shown that aggregates are formed, hundreds and thousands of times greater in volume than the initial particles, and come to make up the main portion of the powder. Similar results were produced at the Scientific Research Institute for Photochromic Studies imeni Karpov and at our institute using corundum powder. If this problem can be solved, either by creation of stable powders, or by the use of some sort of device to break up the aggregates [11], which would complicate the design

of the sprayer, the dust cloud method, in spite of certain shortcomings such as the brief time of existence of the bright cloud, great expenditures of weight to achieve a few sessions of observation, could be significant for spacecraft which are not oriented in flight, when directional devices are unusable.

There is great interest in the use of another passive signaling device, a mirror, which has been used to produce photographs of spacecraft in orbit around the moon, and is described in [12].

The action of the mirror for signaling can be compared to the action of the aperture in a camera obscura (the idea of this analysis of the phenomenon was suggested by A. S. Dorofeyev). It follows from this analogy that in this case when the distance between the surface of the Earth and the mirror is incomparably greater than b/β_C, where b is the projection of the diameter of the mirror on the plane of the sun and $\beta_C = 0.53^\circ$ is the angular diameter of the sun, the radiation will diverge in a cone with a peak angle $\beta = \beta_C$. Considering this fact, it is easy to determine the illumination created by an ideally flat and totally reflecting mirror at the boundary of the Earth's atmosphere:

$$E = (2.02 \cdot 10^9/R^2) M_m \cos(\phi/2) \text{ (lx)} \qquad (10)$$

where M_m is the area of the mirror.

According to (10), at a distance of 400,000 km, a brightness of 14.5^m at $\phi = 0$ will be created by a mirror with an area of 3.5 cm^2. However, the slight angle of divergence of the radiation requires high accuracy of stabilization of the position of the mirror in space or the use of scanning [12], and also limits the sector on the Earth's surface where observation can be conducted. Mirrors to reflect sunlight with an angle $\beta > \beta_C$ are possible. Where $\beta = 10°$, a mirror with an area of not over 1300 cm^2 would be required. However, the effectiveness of the mirror decreases with increasing ϕ and, furthermore, passive devices do not work in shadow areas.

Spacecraft in near-Earth space have utilized signaling devices with pulsed xenon lamps, creating pulses several milliseconds long, with pulse repetition frequencies of less than 1 Hz. This method is suitable when there is no requirement for orientation (since the image is to be photographed) and at /170 short ranges, when direct photography can be used and wide-angle lenses eliminate the difficulties of search. In the area of long-range flights, this advantage disappears. Furthermore,

measures may be required to synchronize the pulse generator on the spacecraft and the on-board receiver, making the system more complex. Information cannot be transmitted with a pulse lamp, due to the low pulse repetition frequency. Furthermore, when pulse signals are used it is necessary to observe static images which are less interference resistant than continuous images, since they do not allow movement to be observed. Experience has shown that movement is very important for the recognition of objects on the order of $14\text{-}15^{m}$.

In our work with low-frame-rate (static) images of actual subjects, we have noted that the necessary signal/noise ratio is 1.5-2.5 times higher than for continuous images. For point images, the least interference resistant, we can assume a drop in interference stability of 2.5 times as a first approximation, requiring that the exposure be increased by 6.25 times (the spectral composition of radiation of continuous and pulsed light sources is the same). The increase in exposure is achieved partially by the decrease in the area of an image element. With an exposure time on the order of several milliseconds, γ can be estimated as 1 second, i.e., $\gamma_{res.p} = 1.5$ sec and $d^2_{res.p} = d^2_{res}/2.8$. We can assume that the additional exposure increase of 6.25:2.8 = 2.2 times will be sufficient, particularly since the contrast of the image also increases (by 2.8 times). The time of formation of the image, the frame time of a television receiver, is 0.08 sec. In the case of an electronic-optical converter, the time constant of the eye, which is on the same order of magnitude, is also significant. The electrical power consumed by a continuous burning xenon lamp (400 w) during the time of formation of an image is about 32 J. The efficiency of the radiation of pulsed lamps is approximately 20% lower than that of continuous lamps.

Thus, the energy which must be stored in the storage capacitor of a pulsed light source and for which the lamp must be designed is $32 \times 2.2 \times 1.2 = 85$ j. The closest series-produced is the IFK-120. The presence of an individual rectifier and storage capacitor will increase the weight of the light source with a pulsed lamp, but the power consumption may be very slight. Thus, with 100 pulses in a session and a rectifier efficiency of 0.5, eight sessions will consume about 0.05 kw·hr.

We can conclude that the main advantages of electronic light receivers -- high sensitivity, immediacy of measurements, possibility of receiving information -- are largely lost when a pulsed lamp is used. The accuracy of each measurement is also slightly reduced, since turbulent deviations are not averaged.

Let us briefly study the capabilities for signaling from a spacecraft against the background of the disc of the moon. Against the background of the sunlit moon, signaling is actually possible only using quasimonochromatic light, allowing deep filtration at the receiving end. A gas cloud or laser generator can serve as sources of this type of light.

In order to produce a bright spot with a contrast level of 1, the reaction of the photocathode to the power of the signal entering the telescope and to the power of the sunlight reflected from the moon per image element of the receiver device must be identical. It is assumed here that the exposure time for both radiation sources is identical. In the case of a pulsed laser, this requires that the spacecraft and receiver device be synchronized, as was noted above. The condition of the production of unit contrast level in the general case is: /171

$$\frac{\pi D^2}{4\omega R^2}\int_{\lambda_1}^{\lambda_2} g(\lambda)\tau(\lambda)w(\lambda)\,d\lambda = \frac{\pi r_{res}^2 F^2}{4}\,\frac{A^2}{4}\int_{\lambda_1}^{\lambda_2} g(\lambda)\tau(\lambda)\rho(\lambda)f_c(\lambda)\,d\lambda, \tag{11}$$

where λ_1 and λ_2 are the boundaries of the filtration range; ω is the solid angle of divergence of the beam; $\rho(\lambda)$ is the reflection factor of the lunar surface; $\tau(\lambda)$ is the transmission factor of the optics and atmosphere; $f_c(\lambda)$ is the intensity of solar radiation, $w(\lambda)$ is the radiation power of the light source.

If the radiation used for signaling is in the sector of wavelengths less than $\Delta\lambda = \lambda_2 - \lambda_1$, and $\Delta\lambda$ is sufficiently slight, condition (11) is simplified:

$$w = \int_{\lambda_1}^{\lambda_2} w(\lambda)\,d\lambda = \frac{r_{res}^2}{4}R^2\omega\rho f_c(\lambda)\,\Delta\lambda. \tag{12}$$

Using formula (12), we can see that a decrease in $\Delta\lambda$ to a few angstroms can produce unit contrast with comparatively easy requirements for on-board devices: for a sodium cloud -- the weight of the gaseous sodium is on the order of a few hundreds of grams, for a laser generator -- the energy stored in the storage capacitor is on the order of a few hundreds of joules. The difficulties created are: in the case of the gas cloud, the necessity of repeated observation; in the case of the laser, the great range.

Let us estimate the power of a pulsed ruby laser necessary to signal the Earth on the flight trajectory to the moon, at the moon or on the moon. Pumping of a laser by a flash tube involves a reduction of efficiency by a factor of approximately 40 (about 1% for a laser as opposed to 40% for a flash tube). The effectiveness of ruby laser light and the integral radiation of a xenon lamp acting on a multi-slit photocathode is about 0.247 storage capacitor must be 36.5 times greater when a laser accumulating condensor must be 36.5 times greater when a laser is used than when a lamp is used, i.e., should amount to 3000 j. Devices are also required for artificial spreading of the laser radiation beam, plus a storage capacitor weighing at least 40 kg. Thus, the use of a laser is possible only if angle β can be significantly reduced, reducing the area of reception of the signal on the Earth and making signaling impossible at relatively short ranges.

We note that signaling by mirror from the illuminated surface of the moon is also unrealistic. Based on condition (11), considering ρ constant and considering that in the case of the mirror $w(\lambda) = f_c(\lambda)M_m$, we produce the following mirror size necessary to create unit contrast:

$$M_m = (1/4)\gamma_{res}^2 \rho \omega R^2,$$

which indicates that the area of an ideal flat mirror should be 30-60 m^2.

It is significantly easier to perform signaling from the dark portion of the moon. The brightness of the lunar radiation is 13.7 m from each square second, and signaling could be provided by pulsed lamp of 700 J, the storage capacitor of which would weigh about 10 kg or, for example, an incandescent bulb with a power rating of 5 kw. By increasing the complication of the system, introducing time selection, the energy of a flash can be reduced to 100-120 J.

Conclusions /172

Comparing various methods of light signaling, we can conclude that on the flight trajectory to the moon, signaling using constant burning lamps is most favorable to allow observation of a spacecraft and transmission of information from the spacecraft. Calculations indicate that it is generally possible to construct such a system.

Signaling using mirror devices, requiring minimum expenditures of on-board power and therefore allowing continuous transmission, is also potentially very important.

For spacecraft not oriented in space, light signaling for observation purposes can be achieved by the creation of artificial clouds.

The author thanks Doctor of Physical and Mathematical Sciences V. G. Kurt and Doctor of Technical Sciences B. N. Rodionov for their discussion of this article and their useful suggestions.

REFERENCES

1. Shklovskiy, I. S., Optical Methods of Observation of Artificial Earth Satellites, Iskusstvennyye sputniki Zemli, 1958, No. 1, p. 44.
2. Koyper, J., B. Middlehurst, Teleskopy [Telescopes], Moscow, Foreign Literature Press, 1963.
3. Anisimov, V. F., et al., Experience in the Use of Television Equipment for Observation of Stars, Izv. Krymskoy astrofiz. obs., 1963, Vol. 30, p. 3.
4. Pyaskovskaya-Fesenkova, Ye. V., Issledovaniye rasseyaniya sveta v zemnoy atmosfere [Study of Scattering of Light in the Earth's Atmosphere], Acad. Sci. USSR Press, 1957.
5. Khalfin, A. M., Electronic Amplification of Images, Tekhnika kino i televideniya, 1958, No. 9, p. 10.
6. Krasil'nikov, N. N., Pomekhoustoychivost' televizionnykh ustroystv [Interference Stability of Television Devices], Moscow-Leningrad, Gosenergoizdat Press, 1961.
7. Rabotnova, T. N., L. V. Kononchuk, Some Data on Photosensors with Multiple-Slit Cathodes, Svetotekhnika, 1959, No. 9, p. 3.
8. Fabry, C., Obshcheye vvedeniye v fotometriyu [General Introduction to Photometry], Leningrad-Moscow, State Technical and Theoretical Press, 1934.
9. Shklovskiy, I. S., An Artificial Comet as a Method of Optical Observation of Rockets, Iskusstvennyye sputniki Zemli, 1960, No. 4, p. 195.
10. Shifrin, K. S., Rasseyaniya sveta v mutnoy srede [Scattering of Light in a Turbid Medium], Moscow-Leningrad, State Technical and Theoretical Press, 1951.
11. Fuks, N. A. A. N. Selin, Pneumatic Spraying of Powders, Inzhenerno-fizicheskiy zhurnal, Minsk, 1964, No. 1, p. 122.
12. First photographs from Earth of a lunar satellite, Sky and Telescope, 1968, Vol. 35, No. 4, p. 220.

N75 11009

DATA COMPRESSION

V. F. Babkin, A. B. Kryukov
and Yu. M. Shtar'kov

Space studies always require transmission of data produced to the Earth. Each year, mankind penetrates ever deeper into space, increasing the complexity and number of experiments and, as a result, the volume of data produced. This tendency has already resulted in a number of serious problems. The stream of data is so intense that significant difficulties have arisen in transmission, storage and processing. Therefore, one pressing problem related to the development of space studies is reduction of the volume of data transmitted from spacecraft. Many studies of actual telemetry data have demonstrated the possibility of significant reduction in data volume; in many cases, reductions by factors of some tens or even hundreds have been achieved. The remaining data are still sufficient to /173 restore the information sufficiently to satisfy the demands of the consumer.

Obviously, the problems of transmission, storage or processing of large volumes of data do not arise only during space studies. In the construction of complex measurement systems with storage of data produced in the performance of investigations, the task of reduction of the volume of data is also very pressing. However, most works dedicated to its solution involve the development of studies of space to some extent.

The term "data compression" is used rather freely as applicable to any method of conversion of data allowing the energy or bandwidth required to transmit a message with a given message quality from a source of a given type, or the volume of memory required to store the message, to be reduced. However, this rather broad definition is not exhaustive.

Primarily, a reduction in energy, bandwidth or memory volume assumes that there is some initial representation of the data, relative to which compression is performed. But it is then obvious that the effect of compression depends to a great extent on the method of initial representation of the data and cannot be defined unambiguously. The less efficient the initial representation, the greater the effect which can be expected by a given compression method. It is therefore worth our time to look into another approach, in which the effect achieved by compression is evaluated not by the degree of reduction of the initial volume of data, but by the closeness of the

approach to the minimum possible volume (if, of course, the minimum volume can be determined for the message source in question).

The case when the consumer is interested not in the output of the source of the data, but only in certain individual characteristics (for example, when the actual realization of a process being studied is not so important as its dispersion of mean value, etc.) is rather specific. Replacement of the set of parameters with certain of its characteristics is usually referred to as parameter separation. Naturally, separation of parameters in many cases can sharply reduce the volume of data transmitted or stored. Nevertheless, serious doubts arise as to whether this procedure can be looked upon as data compression, since irreversible conversion of the initial data takes place. In other words, since a qualitative transition is made to another set of data, it is hardly sensible to compare the description of the two different sets. However, according to the definition presented above, the separation of parameters is one method of data compression. In many works this is used, since the separation of parameters and compression of data have many points of similarity. For example, such a method of parameter separation as calculation of quantiles can be looked upon as a method of compression, but not of the initial sample, but rather of its histogram.

It should be noted in general that when we discuss statistical processing, we can reduce the volume of data accumulated, by discarding values which do not contain information concerning the parameter being estimated. The correctness of this procedure is explained by the existence of so-called sufficient statistics, i.e., sets of functions of observed values which contain all the information on the parameter contained in the set of these values. Therefore, in order to estimate the parameters, it is sufficient to know only the sufficient statistics, not the entire set of observed values [1].

Various approaches and opinions on problems of data compression have been partially reflected in the terminology used. In addition to the term "data compression" many other terms are used in the same sense: "reduction of redundancy," "data reduction," "frequency band reduction," "adaptive sampling," etc. This review uses the term "data compression," since it /174 seems more general than the other terms, and also sufficiently brief, which is of some significance from the standpoint of compression itself.

The terms "statistical coding," "effective coding," "entropy coding," "coding related to the source" and "separation of parameters," "conversion decreasing the entropy of the

source" are used in parallel for two particular methods of compression. In both cases, the first of these terms can be used.

Recently, the number of works dedicated to data compression and the number of periodicals in which they have been published have increased sharply. Therefore, of course, some works may be omitted from any study. In preparing the present review, we analyzed some 200 works in the major periodicals up to mid-1969. Unfortunately, among these works we cannot note a single review yielding a sufficiently complete concept of the current state of the problem. This fact stimulated the authors to attempt to systematize the known results on data compression. Of course, we will not refer to nearly all of the works just mentioned in this review. This results from the fact that in certain areas (related to statistical coding) monographs are available, and there is no need to refer to the original works. In those cases when there are several works covering essentially the same areas, references are given only to the most important ones.

The review is constructed as follows. In the first section, we describe methods of data compression based on statistical coding and rooted in information theory. To realize these methods, a rather complete knowledge is required of the statistics of the source and of complex coding devices. In the second section, we study the application of methods of interpolation and extrapolation for compression of data describing a certain process. These methods are in many cases less effective than statistical coding, but recently have been given primary attention. This is explained by the fact that they are significantly simpler to realize and can be successfully used if the statistics of the soruce are not completely known. The third, relatively short section, is dedicated to a specific compression method, related to description of the histogram of a sample by means of a set of quantiles. The fourth section studies some criteria for effectiveness of data compression and methods of representation of the service information required for identification of compressed data. The fifth section discusses some models of sources of messages, suggested for theoretical analysis of data compression. We note at the outset that this review does not include various problems concerning the design of compression systems, methods of realization, design of buffer memories, etc.

With this design of the review, the complete set of information characterizing any compression method is distributed through several sections: the essence of the method is presented in one of the first three sections, an evaluation of its effectiveness in the fourth section, the models of sources using which its effectiveness can be determined -- in the fifth section. However, by concentrating all information relating to a given method of data compression in one location, the review would be

converted to a simple list of various methods and the results produced for them. The total number of methods is extremely great, and the results are frequently incomplete, determined to a great extent by the effectiveness criterion and source model used. Therefore, this type of presentation could not produce a general picture. On the other hand, the placement of the material which we have used does make our review general, systematic and orderly.

The authors thank L. V. Dmitriyev, N. Ye. Ryb'yev, Z. K. Fedotov for their help in selection and preparation of the materials for the review.

Statistical Methods of Compression of Discrete Data /175

Statement of the Problem

Let us study a discrete source, the output of which records events in succession. The total number of different events is fixed. One of the most important tasks consists in providing economical description or coding of these events by means of a finite set of different symbols. This task arises both during recording and during transmission of data through communications channels.

In analyzing the various methods of coding, we must recall that the code should satisfy the condition of unambiguous decodability. In other words, it should be possible to restore the initial sequence of events from the sequence of symbols describing these events.

In certain particular cases, the condition of unambiguous decodability is weakened. This type of statement of the problem is justified for a number of sources. Furthermore, in some problems it is sufficient (or necessary) to require that the condition of unambiguous decodability be provided not always, but rather with a probability rather close to unity. This statement results, for example, in the problem of coding for a source with a fixed rate [2]. However, in our analysis we will generally study the case of unambiguous decoding, since the other areas are primarily in the stages of development.

Suppose q different symbols are used to describe events. It is quite simple and very convenient in practice to use an even code, when each event corresponds to a code number (sequence of q-nary symbols) of identical length m_0, where m_0 satisfies the condition $q^{m_0-1} < M \leq q^{m_0}$. Obviously, this type of code is unambiguously decodable.

When an even code is used, it is impossible to expend less than m_0 symbols in the description of each event. However, the problem is frequently stated otherwise. It is necessary to find a coding method for which the mean number of symbols $\bar{m}$ expended in description of one event is minimized. This statement of the problem is not always correct. For example, for a discrete Markovian source, for which there are no limiting probabilities, it simply has no sense. However, in the following we will study only sources for which the concept of mean code word length can be used.

The tendency to decrease the mean length of a code word generally makes it necessary to use an uneven code, for which the condition of identical length of all code words is not fulfilled. This brings up the problem of construction of an uneven code, always providing for unambiguous decoding and minimizing the mean code word length.

For unambiguous decoding of an uneven code, it is necessary to divide the sequence of symbols into code words without knowing the length of each word in advance. A code for which this type of decoding can always be performed is called an unambiguously decodable code. Many codes do not have this property.

Sardinas and Patterson have precisely formulated the problem of unambiguous decoding and found necessary and sufficient conditions which must be satisfied by an unambiguously decodable code [3, 4]. One of the most important classes of unambiguously decodable codes consists of prefix codes [2, 4, 5]. The prefix of a code word refers to the code word itself and all sequences /176 of symbols produced from it by removal of the last characters (for example, the code word 101110 has the prefixes 101110, 10111, 1011, 101, 10, 1). A code is called a prefix code (or a code which has the prefix property) if no code word is a prefix of another code word.

It can be shown [4] that for unambiguous decodability of a code, it is sufficient that the prefix property be fulfilled either for the initial code word or for inverted code words (words written in the reverse order). Prefix codes, for which the inverted words also have the prefix property, are called codes with the strong prefix property (for example, even codes have this property).

One of the most common forms of prefix codes consists of codes with a separating character [5]. By definition, in a code with t separating characters of length T

$$
\begin{aligned}
K_1 &= k_{11},\ k_{12}, \ldots,\ k_{1T},\\
K_2 &= k_{21},\ k_{22}, \ldots,\ k_{2T},\\
&\ldots\ldots\ldots\ldots\\
K_t &= k_{t1},\ k_{t2}, \ldots,\ k_{tT}
\end{aligned}
$$

any code word $A = a_1, a_2, \ldots, a_n$ must satisfy the following conditions:

1) the last T symbols of this code form one of the separating characters $K_1, K_2, \ldots, K_t$;

2) in the sequence

$$k_{i2}, k_{i3}, \ldots, k_{iT}, a_1, a_2, \ldots, a_{m-1} \quad 1 \leqslant i \leqslant t$$

no set of T symbols following in sequence forms a separating character.

For example, for q = 2, t = 1 and K = 10 (T = 2), code words might be 10, 010, 110, 1110, 0110, 0010, ... In the simplest case, the last symbol of each code word is one and the same, and this symbol is not used in any other position in the word. For example, where q = 3, possible code words are 2, 02, 12, 002, 012, 102, 112, 0002, ..., while where q = 2, -- 1, 01, 001, 0001, ...

One important advantage of prefix codes is the simplicity of decoding. Determination of the end of a code word is possible using knowledge only of the symbols (prefixes) which precede it; knowledge of the last symbols is not required. In other words, immediately after arrival (analysis) of the last symbol, any code combination can be separated (decoded).

For nonprefix codes satisfying the Sardinas-Patterson condition, this type of decoding is impossible. We must know not only the preceding, but also the subsequent symbols in the sequence in order to separate a code word. For example, suppose a nonprefix code contains only three words -- 0, 01, 11 (the prefix of the second word is equal to the first word) [5]. Then, in decoding the sequence 000111..., two different versions of separation are possible -- 0, 0, 0, 11, ... and 0, 0, 01, 11 ... In order to determine which of the two versions is correct, we need only calculate the number of ones in a row. If it is even, the first version of division is correct; if it is odd, the second version is correct. Consequently, to determine the proper version of division of code words, even in this simple case, we must determine the number of ones received in a

row, no matter how many they may be. Thus, nonprefix codes involve a decoding delay by the number of symbols following a given code word which must be known in order to separate the code word in question unambiguously.

In addition to this very significant shortcoming, nonprefix codes have no common method of construction. The use of the Sardinas-Patterson method results in very cumbersome transformations. Therefore, only a few individual examples of nonprefix codes are known at the present time [4, 5]. /177

The great attention given to nonprefix codes in this section is explained by the fact that optimal methods of construction of prefix codes have been found for a broad class of information sources [2].

Let us now go over to presentation of the main known results on the construction of uneven prefix codes for various message sources.

Methods of Statistical Coding of Independent Messages

Let us assume that events A_{i1}, A_{i2}, ..., sequentially recorded at the output of a source, are statistically independent and have probabilities p_{i1}, p_{i2}, ... respectively. In this case, the mean number of symbols expended in description of this message exists and is equal to

$$\bar{m} = \sum_{i=1}^{M} p_i \cdot m_i, \tag{1}$$

where m_i is the length of the code word corresponding to A_i. It can be shown [2] that for any prefix code $\bar{m} \geq H_q(A)$, where $H_q(A)$ is the ordinarily defined entropy of the set of events A_1, A_2, ..., A_M, calculated in q-nary units. If $p_1 = p_2 = \ldots = p_M$, a simple even code is most efficient. In this case, $\bar{m} = m_0 = H_q(A)$. If, however, the probabilities are not even, an uneven code is used, and some m_i are less than m_0, while others are greater than m_0 (the case when all m_i are greater than m_0 is of no interest). Roughly speaking, event A_i, for which $p_i > 1/M$, corresponds to a code word of length $m_i < m_0$, while event A_j, for which $p_j < 1/M$, corresponds to $m_j > m_0$. As a result, we produce $\bar{m} < m_0$.

An algorithm for code construction suggested independently by Shannon and Fano [2, 4, 5], is widely known. In spite of its simplicity, its description would require a great deal of space. Considering the detailed discussion of this question which has been conducted in many sources, we note only that this method allows us to produce $\bar{m}$ satisfying the inequality

$$H_q(A) \leqslant \bar{m} \leqslant H_q(A) + 1. \tag{2}$$

The equality in the left portion is fulfilled if and only if all p_i are negative powers of the number q.

In most cases, statistical Shannon-Fano coding is not optimal. A method of optimal prefix coding for a fixed set of independent messages has been suggested by Huffman [2, 4, 5]. It is impossible to strengthen condition (2) for this code, but for any prefix code (including the Shannon-Fano code), the value of $\bar{m}$ is no less than that of the Huffman code. The Huffman algorithm is more complex than the Shannon-Fano algorithm, and various solutions are possible in the construction of a code.

If $H_q(A) \gg 1$, the Shannon-Fano method and the Huffman method yield quite similar results. However, if $H_q(A) \ll 1$, the difference may be significant. Furthermore, even for the Huffman method in this case it may be found that $\bar{m} \approx H_q(A) + 1 \gg H_q(A)$.

In order to produce $\bar{m}$ differing arbitrarily little from $H_q(A)$, the events should be "enlarged." To do this, we analyze not the initial events $A_1, A_2, \ldots, A_M$, but rather groups of successive events, each of which contains $r > 1$ initial events. /178
Naturally, the number of different groups is equal to $M^r \gg M$, and the number of code words increases sharply. However, on the other hand, as we perform statistical Shannon-Fano or Huffman coding for these groups of events, we find that the description of one initial message on the average requires

$$H_q(A) \leqslant \bar{m} \leqslant H_q(A) + \frac{1}{r} \tag{3}$$

q-nary symbols. For any distribution of probabilities of initial events and any value of $H_q(A)$, we can select a value of r such that $\bar{m}$ differs arbitrarily little from $H_q(A)$.

Generally speaking, it is not at all necessary that the number of initial events in each group be identical (r). It is

necessary only that the requirement of completeness be fulfilled: the set of groups used must be so selected that any sequence can be divided into groups belonging to this set.

One widely known method of coding, in which the number of events in different groups is different, has come to be called "series length coding."

This method of coding was studied by Shannon, then by Elayes [6] and many other authors. A rather detailed analysis and summary of the results of this problem can be found in [4, 5].

Let us assume that there are only two independent events -- A_1 and A_2 (which we will represent arbitrarily as 0 and 1) and that the probability p_0 of appearance of event 0 (A_1) is much greater than the probability p_1 of appearance of event 1 (A_2): $p_0 \gg p_1$. Let us select the following sequence of different groups: 1, 01, 001, 0001, ... The number of such groups is infinitely great. If the maximum size of a group is limited by a certain quantity ω, the last group is

$$\underbrace{000\ldots01}_{\omega}.$$

To make this set complete, we must add a group consisting of ω zeros. Otherwise, in case of a sequence of ω or more zeros in a row, it would be impossible to divide the sequence into groups.

The groups of events produced differ little in probability. For the first group we produce p_1, for the second -- $p_0 \cdot p_1$, but since $p_0 \gg p_1$, $p_0 \approx 1$ and $p_0 \cdot p_1 \approx p_1$, etc. Thus, the probabilities decrease monotonically and very slowly. The exception is the group consisting of zeros alone. Actually, this is the group which determines the minimum value of ω; for effective statistical Shannon-Fano or Huffman coding for these groups of events it is desirable that the probability of appearance of the last group (p_0^{ω}) be not over 0.5.

The introduction of non-identical (in length) groups allows us to bring the mean length of a code word $\bar{m}$ close to $H_q(A)$ more rapidly than when the same number of groups of identical size is used. Furthermore, there is generally a rather simple relationship between the dimensions of a group coded and the type of code word, significantly simplifying coding [5].

In practice, a simplified version of coding of series lengths is frequently used. The groups of events studied above are coded by an even code, where m_0 satisfies the condition $\omega = 2^{m_0} - 1$. The group of events 000...0 corresponds to code word 000...0, any other group -- to a word such that its binary notation is equal to the simple number of events in the group (series length). This simplification greatly facilitates the operations of coding and decoding at the expense of some deterioration of the code itself (in comparison to the Huffman or Shannon-Fano code). /179

Obviously, for each value of p_0 (or p_1) there is a certain optimal limiting length of series ω, providing the most effective coding. This was first noted by Blokh [7].

This method of coding of series lengths is applicable in any case when a sequence of zeros and ones is studied such that $p_0 >> p_1$. Let us assume, for example, that there are M events and the probability of appearance of one of them is much greater than all the others. If we first perform coding using an even code in which the most probable event corresponds to the code word 000...1, as a result we produce a sequence of symbols in which $p_1 >> p_1$. Series length coding can then be used for the sequence produced.

It is interesting to note that a source can be artificially organized in which the probability of appearance of zeros and ones differs significantly [8, 9]. For example, in satellites, sets of multidigit binary counters are used to record the number of different particles. When the satellite is in an area of high particle density, the counters are filled rapidly, so that almost all of the high-order digits contain ones. However, when the satellite is in an area of low particle density, most of the high-order digits of the counters contain zeros. In [8], it is suggested that "coding in a digit plane" be used to transmit the contents of these counters, i.e., that reading could be performed not by the ordinary method, but rather "crosswise." To do this, first the highest-order digit of all counters is read out, followed by the next highest order digit, etc. right down to the lowest-order digit. This method of reading allows a sequence of zeros or ones to be organized (depending on the location of the satellite). Series length coding can then be applied to this sequence.

Obviously, coding in a digit plane can be used not only for counters, but in other cases as well, for example for compression of a television signal. The image can be looked upon as a matrix, each element of which is an m-digit binary number (m being determined by the number of brightness gradations used). Each such matrix with m-digit elements can be replaced

by m matrices with one-digit elements, which can be easily compressed, for example, using series length coding. Actually, if two elements have different brightnesses, this difference does not appear in all of the digits of the numbers describing the brightness of the elements. Therefore, when coding is used in the digit plane, series are formed in the high-order digits of the numbers. Obviously, the results will be better, the fewer the number of digit positions in which neighboring numbers differ. To improve the situation, the image might first be coded by a Grey code, which is more effective for subsequent coding in the digit plane than a normal binary code.

The Shannon-Fano and Huffman algorithms for statistical coding are based on full knowledge of the probabilities of various events. If the actual probabilities differ somewhat from the assumed probabilities (used in coding), the code constructed will either correspond or differ little from the optimal code corresponding to actual probabilities. But in practice, very frequently, nothing or almost nothing is known concerning the probabilities of events. Furthermore, a finite sequence of events is always actually studied, and the frequency of appearance of events, of primary importance in statistical coding, may differ significantly from the probabilities. In [10], algorithms are suggested for coding for the case when the statistics of a source are unknown and the observed frequencies of appearance of various events can be used. The realization of these algorithms involves significant difficulties. However, /180 there is another means. It consists of multiple repetition of a simple coding algorithm. This iterational procedure continues as long as the number of symbols describing the finite number of events at the output of the source in question continues to decrease. The first such iterational procedure was suggested by Blokh [11]. Suppose there are only two independent events (as before 0 and 1), $p_0 >> p_1$ and $q = 2$. We will perform coding using an uneven prefix code following the rule

$$\begin{aligned} 00 &\rightarrow 0 \\ 01 &\rightarrow 11 \\ 1 &\rightarrow 10. \end{aligned} \tag{4}$$

If $p_0 >> p_1$, description of t events will involve the expenditure of approximately t/2 binary symbols, since the combination of events 00, coded by a single symbol, will be most frequently encountered. However, it may be that the number of zeros will continue to remain much greater than the number of ones in the resulting sequence of code word symbols. Then, repeated appli-

application of conversion (4) can still greater reduce the number of symbols. This operation can be recurrently repeated as long as it continues to yield the desired effect -- to decrease the total number of symbols in the sequence. At the moment when this number reaches its minimum, the iterational procedure is stopped. For example, suppose the initial sequence of events A_1 and A_2 contains 34 binary symbols

0000000000010010000000110000000000.

Successive repetitions of transformation (4) yield

0000010011000111000000	(22 symbols)
0011010100111010000	(19 symbols)
0101011110101010110	(20 symbols)
1111111010101111111110	(23 symbols)

Analysis of further iterations is senseless, since the number of ones in the last sequence is significantly greater than half. Thus, in this example, the best result is achieved after two iterations, allowing the sequence to be reduced from 34 to 19 symbols.

As is the case when coding series lengths, not all of the groups contain the identical number of events (two groups contain two events each, one group contains one event). Finally, as in coding series lengths, the method is applicable in all cases where $p_0 >> p_1$ or, naturally, $p_1 >> p_0$ (in which case the zeros must be replaced by ones and the ones by zeros in the coding table (4)).

The method described above was the first effective example of an iterational statistical coding procedure, the analytic methods of which were developed somewhat later in [12]. This same reference gives a few more examples of permutations. We should note that studies of the effectiveness of iterational procedures are far from complete.

All of the methods of statistical coding which we have described assume unevenness of appearance of events at the output of the source. However, messages frequently consist of readings of a certain process performed at even time intervals. The use of statistical coding is impossible in this case without a device to equalize the rate of arrival of bits at the output of the coder to the rate of transmission in the communications channel.

We must be quite careful in our selection of the capacity of /181 such a device, since it may be overfilled with a certain probability; in this case, normal operation of the communications channel is interrupted.

Another method of coding for this type of source, not requiring the use of a smoothing device, is described in [2]. In this case, effective coding is possible if we compare the various code words of all messages with the exception of the subset of messages with negligibly small probability of appearance. Here, an inequality such as (3) can be produced for the mean length of a code word.

Coding Methods for Statistically Dependent Events

The task of statistical coding is significantly complicated if the events recorded at the output of the source are statistically dependent. However, on the other hand, obviously the presence of statistical dependence still further increases the possibility of economical description of events.

The simplest solution to this problem consists in the following. We divide the sequence of events into groups of arbitrary length r. We calculate the probabilities of appearance of each group, considering the statistical dependence of the events included in it. After determining these probabilities, we perform ordinary statistical Shannon-Fano or Huffman coding. We must note that the result in this case will depend significantly on the method of subdivision of the sequences.

This approach fully considers all the statistical dependences within the group. However, since various groups are studied in isolation from each other, the statistical dependences at the boundaries of neighboring groups are ignored. In certain cases, it may be that the statistical dependence of events (at the boundaries of groups) not considered is much less than that which is considered (in calculating the probabilities of each group).

This discussion does not pretend to be strict, but only illustrates the idea of the method. It is important, however, to note that the r-dimensional distributions of probabilities (and, consequently, codes which should be used in each step of the subdivision) will depend on the number of division groups. For a particular type of sources, which [2] calls "stationary," we can construct a code such that a relationship such as (3) is satisfied for the mean length of a code word. The difference is that the entropy $H_q(A)$ of the source of

independent messages must be replaced by the corresponding entropy of the source of statistically dependent messages, defined in some suitable manner. Furthermore, the second term in the sum in the right portion of the inequality is equal not to 1/r, but rather to a certain function of r, not less than 1/r. This last difference is related to the fact that this method does not consider the statistical independence of events at the boundaries of neighboring groups.

Thus, the method of "enlargement of events" and subsequent statistical Shannon-Fano or Huffman coding fully solves the problem both for a source of independent and for certain types of sources of dependent messages. However, in the case of independent events it is sufficient to select an r such that the entropy of the groups produced is significantly greater than one. At the same time in the second case the primary difficulty is related to determination of an effective method of subdivision of the sequence of events into groups. This is particularly true of non-ergodic sources, for example periodic Markovian sources. In both cases, however, the number of codes used is great. This shortcoming is quite significant. However, when a source of statistically independent messages is Markovian, a rather economical code can be used [2].

Let us assume that the first message A_1 has been transmitted /182
by some method. If, for simplicity, we study a first order Markovian source, the probabilities of appearance of any message at the next moment in time depends only on A_1. If these probabilities $p(A_1/A_i)$, $p(A_2/A_i)$, ..., $p(A_M/A_i)$, are known, then by using them we can construct a Shannon-Fano or Huffman code to describe the second message.

Suppose the second message is A_j. Once again, probabilities $p(A_1/A_j)$, $p(A_2/A_j)$, ..., $p(A_M/A_j)$ are known, and we can construct a code for description of the third message. This discussion can be continued. However, we know immediately that the number of different sets of probabilities used is equal to M:

$$\begin{array}{l} p(A_1/A_1),\ p(A_2/A_1),\ \ldots,\ p(A_M/A_1) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ p(A_1/A_M),\ p(A_2/A_M),\ \ldots,\ p(A_M/A_M). \end{array}$$

Therefore, M codes are used, each of which contains M words.

With this method of coding, the statistical dependence of the events is fully considered. When a new event appears,

"instantaneous adjustment" of the uneven code to the new situation occurs. At each step, coding is performed by the best possible means.

In spite of the fact that various prefix codes are used in chaotic sequence, the problem of decoding brings up no difficulties. Actually, a prefix (for example an even) code is used to describe the first event A_i. Then, at the beginning of decoding, it is not difficult to separate the first code word and find that the first event was A_i. However, after this, it becomes quite clear which code was used to code the next event. Since the code has the prefix property and all code words are known, the second code word can also be separated. Thus, the second event is known and, consequently, the code used in the third stage is known. All of this continues cycle after cycle.

This procedure can be extended without difficulty to the case of νth order Markovian sources, when the probability of appearance of any event depends on the last ν (not the last one) events which have appeared. All of the discussion presented above is correct, but it is necessary here to have codes for any set of ν preceding events. Thus, in all M^ν codes will be used, each of which contains M code words, i.e., in all $M^{\nu+1}$ code words.

Naturally, in this case "enlargement of events" can be used as well. However, the purpose of this enlargement is but one: to provide a sufficiently small difference between $\bar{m}$ and the entropy in each stage of coding. In this case, the coding procedure analyzed above is performed for groups of r events, since in place of M events we use M^r groups, the total number of code words used is $M^{(\nu+1)r}$.

With this method of coding, the mean length of a code word satisfies a condition similar to (3), with the entropy of the source of independent messages replaced by the entropy of the Markov source.

A certain simplification of the coding procedure is possible for a Markov source, by reduction of the number of codes used [4]. Let us illustrate this possibility using a first order Markov source as an example.

Let us assume that for any A_i, the set of probabilities $p(A_1/A_i)$, $p(A_2/A_i)$, ..., $p(A_M/A_i)$ is formed of the same numbers, in different order for different A_i. With statistical coding by the Shannon-Fano or Huffman method, the specific form of the code words is determined only by the values of probabilities of /183 events, not by their order. As a result, we find that in any

stage of coding, the same set of code words is used. The different sequence of these probabilities leads to the fact that, depending on the event which appears last, the same code word corresponds to different events. Therefore, it is sufficient to recall the set of code words of only one code and the table which indicates the correspondence between code words and events as a function of the preceding event.

Of course, it is not always so that sets of probabilities for different A_i differ from each other only in the sequence of the different numbers. However, in many cases, with slight differences in the set of probabilities, the code words have the same form and this approach is applicable.

As an example, Table 1 presents an analysis for the case of three events -- A_1, A_2 and A_3. Table 1a presents the probabilities of appearance of various events (first column) as a function of the preceding event (first row). Thus, the numbers in the three columns form three sets of probabilities, and the sets differ not only in the placement, but also in the values of the numbers. Table 1b presents the code words from each set. Since the same code words are contained in each column, coding and decoding can be performed using Table 1c, showing which event corresponds to which code word, depending on the preceding event.

TABLE 1

a

	A_1	A_2	A_3
A_1	0.1	0.3	0.6
A_2	0.2	0.5	0.1
A_3	0.7	0.2	0.3

b

	A_1	A_2	A_3
A_1	10	11	0
A_2	11	0	10
A_3	0	10	11

c

	A_1	A_2	A_3
0	A_3	A_2	A_1
11	A_2	A_1	A_3
10	A_1	A_3	A_2

This situation is also not very characteristic. However, there is a method for which the same code is used in each stage, regardless of the degree of difference of the sets of probabilities [2, 5]. Regardless of the specific form of these sets, we will always represent the most probable event with the same code word, the second most probable event with the same code word, etc., right down to the least probable event. Then, the same set of code words is used in each stage, and, depending on the preceding event, the same code word corresponds to different events.

In other words, we code not events, but transitions from one event to another [5], the most probable transition always corresponding to the same code word, the second most probable to another constant code word, etc.

In many cases, this method is called "decorrelation" [4, 5], although this term is quite debatable. All coding methods related to Markov sources are called "coding with prediction" in [2], which seems more precise.

It is now quite obvious that in coding it is necessary to calculate the probabilities of the most probable transition, the second most probable transition, etc. Looking upon each transition as an independent event and knowing the probabilities of all such events, it is not difficult to code the transitions by any one of the methods suggested for independent events.

Thus, the method suggested allows the statistical dependence /184 of events to be considered to some extent; it is therefore better than coding based on one-dimensional distribution of probabilities, applicable to sources with stable, one-dimensional distribution. On the other hand, this method fails to consider the dependence of the probability of the most probable (or any other) transition on the last event A_i. Consequently, for a Markov source, this method is never better than coding using M different codes. However, the possible losses of effectiveness in many cases are compensated by the transition from M codes to one code. As an illustration of this method, let us return to Table 1. Where $A_i = A_1$, the most probable transition is the transition to A_3, where $A_i = A_2$ -- the transition to A_2, where $A_i = A_3$ -- the transition to A_1. Let us represent the most probable transition as B_1. This refers to the transitions $A_1 \to A_3$, $A_2 \to A_2$, $A_3 \to A_1$. Transition B_2 (the second most probable) corresponds to the transitions $A_1 \to A_2$, $A_2 \to A_1$, $A_3 \to A_3$; transition B_3 (least probable) corresponds to the transitions $A_1 \to A_1$, $A_2 \to A_3$, $A_3 \to A_2$. Suppose these transitions correspond to the code words 0, 11 and 10 respectively. Then in coding a specific sequence of events we produce

A_3	A_1	A_1	A_2	A_3	A_1	A_2	A_2	A_1	A_3	A_3	A_3
B_1	B_3	B_2	B_3	B_1	B_2	B_1	B_2	B_1	B_2	B_2	
0	10	11	10	0	11	0	11	0	11	11...	

(We do not show the description of the first event here.)

Summing up, we must note that all of the methods of statistical coding which we have studied have a serious shortcoming, that distortion of even one symbol of the sequence received in many cases results in distortion of a significant portion of the message. Attempts to increase the interference stability of the coding method and the simultaneous lack of complete statistics concerning the source have led to other methods of economical description of sequences of events. The specifics and wide popularity of these methods require special analysis, contained in the next section.

Methods of Prediction and Interpolation

In this section, we study certain methods of reduction of the volume of data necessary for description of the behavior of an observed process, the realization of which is a certain function of time. It is assumed that the process is evenly quantized and consists of a sequence of readings $\{s_i\}$, where the subscript i is the time or reading number. The task of compression in this case consists of piecewise approximation of the initial process by some known function of simpler form than the realization of the process over the approximation sector. Here, instead of transmitting all readings of the process, only the parameters of each sector of the approximating curve are transmitted, reducing the volume of transmitted data if the approximating curve can be described more economically than the realization of the process by means of the readings. Obviously, approximation introduces some error, since the precise values of readings are replaced by approximate values; therefore, after this procedure the process is represented by a sequence of readings $\{y_i\}$, equal to the sum

$$\{y_i\} = \{s_i\} + \{e_i\}, \tag{5}$$

where $\{e_i\}$ is the error of reproduction of readings $\{s_i\}$.

The approximation accuracy criteria may be various, but most frequently the compression algorithms utilize the criterion of the guaranteed maximum error, since it is comparatively /185 simple to produce in hardware. In this case, the precise value of reading s_i is replaced by an approximate value (estimate) $\hat{s}_i$, under the condition that the approximation error $|s_i - \hat{s}_i|$ does not exceed a fixed value γ. Then the reproduction error e_i in expression (5) will have an absolute value no higher than γ.

Methods of compression of this type can be divided into two large classes: prediction and interpolation. The most frequently used approximating function is an mth power polynomial:

$$s(i) = b_0 + b_1 i + b_2 i^2 + \ldots + b_m i^m. \tag{6}$$

In this case, the compression algorithm is called a polynomial algorithm, although this is not quite true, since polynomial (6) represents only a narrow class of polynomials. However, the term "polynomial algorithm" is very frequently used in the literature in this sense; therefore, compression algorithms using polynomial (6) will also be called polynomial algorithms here. Methods of prediction and interpolation are subdivided into zero order, first order, etc., depending on the power of the approximating polynomial.

In the prediction algorithms, for each subsequent reading s_{i+1}, an estimate $\hat{s}_{i+1}$ is formed on the basis of preceding readings y_i, y_{i-1}, ... We note that the estimates are formed using not the readings $\{s_i\}$ known only at the transmission point, but rather the values $\{y_i\}$, since they are known at both the transmitting and receiving ends of the communications system. The number of preceding readings used to estimate a subsequent reading and the rule for formation of the estimate are determined by the approximation method selected. In mth order polynomial algorithms, the estimate of a subsequent reading is sought as

$$\hat{s}_{i+1} = \sum_{j=0}^{m} (-1)^j C_{m+1}^{j+1} y_{i-j}. \tag{7}$$

This expression allows the value of the approximating polynomial to be determined from the known values at the preceding (m + 1) points, using the method of finite differences. After an estimate is calculated, it is compared with the true value of the reading. If the prediction error of the reading is less than a fixed quantity, the reading is discarded and is not transmitted, since it can always be restored with the required accuracy. In subsequent calculations, the precise value of the reading is replaced by its estimate, which is produced at the reception point, as well as at the transmission point to follow the value of error. If the prediction error is greater than the fixed quantity, the precise value of the reading is transmitted and subsequently used. Therefore, readings $\{y_i\}$ are either

the initial readings of the process or their estimates, depending on the prediction error.

In prediction algorithms, an estimate of the next reading is calculated in each step, then remains unchanged, so that as new readings arrive the approximating curve does not change its position, but rather is simply elongated as long as possible. During interpolation, the position of the approximating curve may change as each new reading arrives, so as to provide precise interpolation (within the limits of the fixed error) of all readings belonging to a given approximation sector. In this case, estimates (points on the approximating curve) for all interpolated readings are calculated in each step, and the interpolation accuracy is checked. If the approximating curve cannot be constructed so as to provide the required accuracy of interpolation of all readings when a new reading arrives, the position of the approximating curve in the preceding moment /186 in time is transmitted, i.e., the position when interpolation was still possible. After this, interpolation of the next group of readings begins.

Prediction Methods

The simplest type of prediction polynomial is the zero order prediction algorithm. Assuming in (7) m = 0, we produce

$$\hat{s}_{i+1} = y_i,$$

i.e., the estimate of the next reading is simply equal to the preceding reading. It is easy to note that reading y_i, in turn, is always equal to the last transmitted reading. Therefore, prediction of all subsequent readings requires simply knowledge of the value of the last transmitted reading. A horizontal tolerance band (aperture) is set for the predicted reading, the width of which is determined by the fixed maximum error γ. If the values of readings do not go beyond the aperture, they are considered redundant and are discarded, since the prediction error will not exceed the fixed value. If a reading falls outside the limits of the aperture, it is considered significant and is transmitted. Then this newly transmitted reading is used to form the estimate of the next subsequent reading, a new aperture is established, etc.

The methods of determining the tolerance band may differ. Three versions of zero order prediction are known, differing in the method of determination of the aperture. For example,

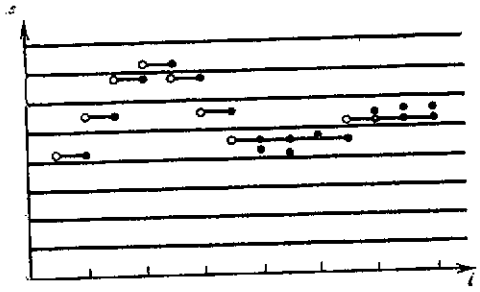


Figure 1. Zero Order Prediction with Fixed Aperture

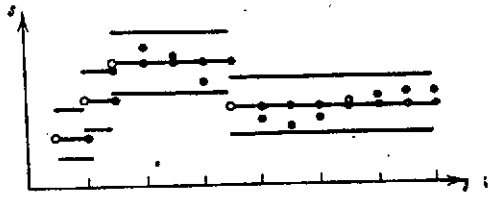


Figure 2. Zero Order Prediction with Floating Aperture

the entire scale may be divided into horizontal bands of width γ, the position of which is strictly fixed and independent of the values of readings. Readings falling within the same tolerance band as the last transmitted reading are considered redundant. This method is called zero order prediction with fixed aperture (Figure 1). One advantage of this method is the exceptionally simple technical realization which can be achieved, since membership of readings in a given band can be established without arithmetic operations. It is sufficient to compare the numbers defining the values of the readings, eliminating one or more of the low-order digits, depending on the value of γ. One shortcoming of the method of fixed aperture is the dependence of the permissible deviation on the value of the reading. For example, if the last transmitted reading was in the region of a boundary between two tolerance bands, even a slight increment might result in transition to the neighboring zone and, consequently, transmission of the subsequent reading, even though the difference between it and the preceding reading is much less than γ.

The method of zero order prediction with floating aperture /187
is more effective from this standpoint, but somewhat more difficult to realize. In this case, a tolerance band of width 2γ is established symmetrically relative to the value of the last transmitted reading, so that the aperture shifts along the scale together with the position of the last transmitted reading. Obviously, only those readings which truly differ from the last transmitted reading by more than γ will be transmitted in this case. With the same maximum error, the tolerance band here is twice as wide as when a fixed aperture is used; therefore, the effectiveness of this method is higher than that of the preceding method. The operation of this algorithm is illustrated by Figure 2.

One modification of the floating aperture method is the method of zero order prediction with shift. Here, the conditions of departure of the last transmitted reading from the aperture are considered. The predicted value in this case is the last transmitted reading plus a shift. The sign and, possibly, the magnitude of the shift are determined by the direction and amount by which the last transmitted reading fell outside of the last tolerance band. A symmetrical floating aperture is established around the predicted value (Figure 3).

Figure 4, borrowed from [13], presents the results of investigation of the effectiveness of these algorithms in the compression of actual telemetry data, produced during the powered sector of flight of a satellite. We must note that the effectiveness of compression can be determined according to various criteria. The numerical values of effectiveness of algorithms studied in this section we define as the ratio of the total number of readings to the number of transmitted readings. The question of determining the effectiveness of a compression algorithm will be analyzed in detail below.

Let us now go over to first order polynomial prediction algorithms. Setting m = 1 in (7), we produce

$$\hat{s}_{i+1} = 2y_i - y_{i-1}.$$

Thus, the predicted value is a point on a straight line passing through the last transmitted reading s_i and the preceding reading y_{i-1} which, possibly, was predicted. A symmetrical sloping aperture of width 2γ is established around this line. If the new reading does not fall outside the aperture, it is considered redundant and is discarded. When a reading does fall outside the aperture, it is transmitted, and a new line is constructed for prediction of subsequent readings, passing through this transmitted reading and the preceding (predicted) reading. The operation of this algorithm is illustrated by Figure 5. Thus, prediction is performed using one precise value of the reading and one predicted value. A modification of this method is suggested in [14], in which the estimate of the first predicted reading is formed using the precise values of both preceding readings. In this case, the reading s_i which falls outside the aperture is not transmitted, but rather the preceding reading s_{i-1}. A straight line is drawn through readings s_{i-1} and s_i, of which reading s_{i-1} was transmitted, to predict the next reading s_{i+1}; the remaining readings are transmitted as in the preceding case. We note that in this

case the precise value of the preceding reading s_{i-1} must be stored, which was not required in the preceding method.

Investigation of the effectiveness of first order prediction algorithms shows that they are generally inferior to zero order prediction algorithms with floating aperture, but may be superior to other types of zero order prediction [13-15]. First order prediction using the precise values of both readings is better than ordinary first order prediction, but the difference is insignificant.

A second order polynomial prediction algorithm is studied in [13]. In this case $s_{i+1} = 3y_i - 3y_{i-1} + y_{i-2}$.

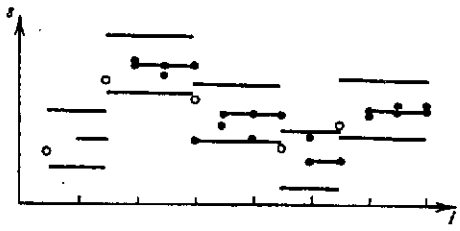


Figure 3. Zero Order Prediction with Shift Equal to Half the Tolerance (Floating Aperture)

Three readings are /188 used to form the estimate here which, of course, complicates the realization of this algorithm. Investigation of the effectiveness of this method has shown that it is significantly inferior to zero and first order prediction algorithms.

Higher order polynomial prediction algorithms are still less effective. This can be explained by the unsuitability of this finite-difference polynomial for extrapolation.

In the prediction methods studied above, the estimate of a subsequent reading is a linear combination of the values of preceding readings with coefficients determined by equation (7). These coefficients are fixed and independent of the autocorrelation function of the compressed signal, although it this function which determines the relationship between values of neighboring readings It is therefore more efficient to seek an estimate of each subsequent reading in the form

$$\hat{s}_{i+1} = \sum_{j=0}^{N} \alpha_j y_{i-j}, \tag{8}$$

where $\{\alpha_j\}$ is the optimal set of coefficients minimizing the mean square prediction error. This method of formation of

estimates is called optimal linear prediction. A theoretical analysis of optimal linear prediction is produced in [16]. It is pointed out here that this method is more effective than polynomial prediction. /189

It is significant to note that determination of the coefficients in this case requires knowledge of the autocorrelation function, which is not always possible. Otherwise, a certain autocorrelation function approximately corresponding to the process must be fixed and a suboptimal prediction algorithm constructed. Another solution of this problem might consist in construction of a self-teaching algorithm. In [17-19], such an adaptive prediction algorithm is suggested and studied for a Gaussian sequence with unknown spectral density. In this case, coefficients $\{\alpha_j\}$ are not selected in advance, but calculated in the process of operation of the algorithm. To do this, the compression system requires a memory capacity of N previous readings, which are used to calculate coefficients $\{\alpha_j\}$ according to the behavior of the compressed signal. Estimates of each subsequent reading are made according to (8).

Interpolation Methods

Zero order interpolation is simplest to do. A zero power polymonial is used to approximate the input signal, i.e., a horizontal straight line, around which a horizontal symmetrical aperture of width 2γ is constructed. In contrast to prediction in which the position of the approximating curve and aperture remains unchanged over the entire interval of approximation and depends only on the initial reading of this interval, with interpolation the interpolating line is adjusted to the values of the readings on each step. As each subsequent reading arrives, the approximating line and aperture are shifted in the vertical direction so that all readings beginning with a certain s_i remain within the aperture. When a reading s_{i+h} appears such that the aperture cannot be set so as to contain all readings including s_{i+h}, the position of the interpolating line at the preceding moment in time $i + h - 1$ is transmitted. We note that the position of the interpolating line is found by calculation and, generally speaking, may not pass through any of the readings $\{s_i\}$. Therefore, in contrast to prediction in which values of readings falling outside the aperture are transmitted, here the readings themselves are generally not transmitted at all.

Interpolation of subsequent readings is performed in the same manner, beginning with s_{i+h}. The operation of this algorithm is shown in Figure 6.

C 4

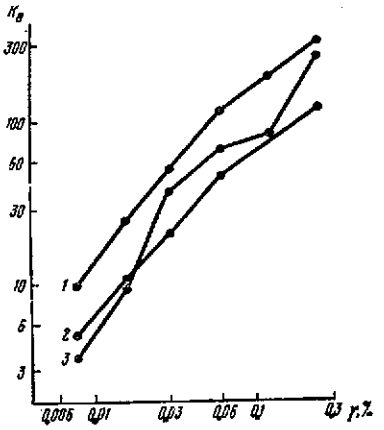


Figure 4. Ratio of Total Number of Readings to Number of Readings Transmitted (Sample Compression Factor) as a Function of Tolerance. 1, Zero Order Prediction with Floating Aperture; 2, Zero Order Prediction with Fixed Aperture; 3, Zero Order Prediction with Shift

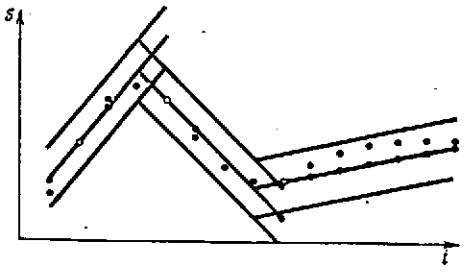


Figure 5. First Order Prediction

Obviously, zero order interpolation is always no less effective than prediction with floating aperture, since a significant reading appears in interpolation only after the signal fills the entire aperture; with prediction, a significant reading may appear even if only half (the upper or lower portion) of the aperture has been "used." Therefore, the effectiveness of zero order interpolation has a lower limit equal to the effectiveness of prediction with a floating aperture of the same size, and an upper limit equal to the effectiveness of this prediction with an aperture twice as large. The advantage of interpolation has been confirmed theoretically [20] and experimentally [21, 22].

Let us study further first order interpolation algorithms, characterized by the fact that the interpolating line may be a straight line of any inclination. A symmetrical inclined aperture of width 2γ is constructed around this line, and the problem consists in determination of a position of the interpolating line for which the /190 greatest possible number of neighboring readings falls within the aperture. Various types of first order interpolation are known. If no limitations are placed on the rule for construction of the interpolating line except the interpolation accuracy, the method is called first order interpolation with four degrees of freedom, since there is freedom in the selection of the initial and final points of the line. As a result of this, the initial signal is a set of straight line sectors

which do not mate in time or value. Therefore, the coordinates of both the initial and final points on each sector must be transmitted. In spite of the high effectiveness of this method, it is apparently only of theoretical interest, since its realization requires an excessively large number of computations.

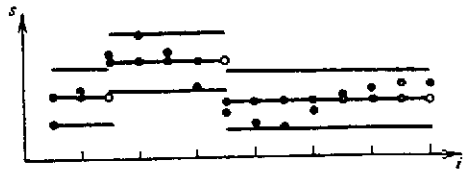


Figure 6. Zero Order Interpolation

The construction of an interpolating line is simplified if we require that the initial point of each subsequent interpolating line mate with the end point of the preceding line. This method has come to be called interpolation with mating sectors. This condition, of course, somewhat limits the capabilities of the algorithm, but at the same time reduces the volume of computation required. Furthermore, it is now sufficient to transmit the coordinates of a single point (the mating point) for each approximation interval. Thus, from the practical standpoint this method may be more effective than the preceding method.

In the methods of interpolation studied above, the points of the beginning and end of each interpolating line are found by computation. A further simplification of the interpolation algorithm might consist in matching of these points with values of readings. Construction of the interpolating line in this case is performed as follows (see Figure 7). The initial reading s_i, the final point of the preceding straight line sector, is joined by a straight line with reading s_{i+2}. An aperture is established around this line, i.e., two more lines are drawn: one through points $s_i + \gamma$ and $s_{i+2} + \gamma$ the other through points $s_i - \gamma$ and $s_{i+2} - \gamma$. If reading s_{i+1} differs from the interpolation value by less than γ, i.e., falls within the aperture, it is considered redundant and a new interpolating line is drawn through s_i and s_{i+3}, relative to which a new aperture is fixed and the accuracy of interpolation of readings s_{i+1} and s_{i+2} is checked. The procedure is continued until the interpolation error of any reading exceeds γ. When any reading falls outside the aperture established around the interpolating line, connecting readings s_i and s_{i+h}, reading s_{i+h-1} is transmitted.

This method of interpolation has become most common, due to the relatively simple realization and high effectiveness it

provides. In transmitting electrocardiograms, for example, the number of transmitted readings can be reduced by a factor of 15 where γ = 3% of the entire scale [22]. The results of study of the effectiveness of various types of first order interpolation are presented, for example, in [15, 21].

Many works have described the so-called fan method of compression [23]. It is essentially one possible method of realization of first order interpolation with mating sectors, where no limitations are placed on the position of the mating points. The interpolation process is performed as follows. Two first order predictors are used, one of which fixes the upper boundary of the tolerance zone, the other -- the lower boundary. The operative principle of this algorithm consists in the following. Two straight lines are drawn from initial point s_i: one through $s_{i+1} + \gamma$, the other through $s_{i+1} - \gamma$, forming a fan aperture. /191

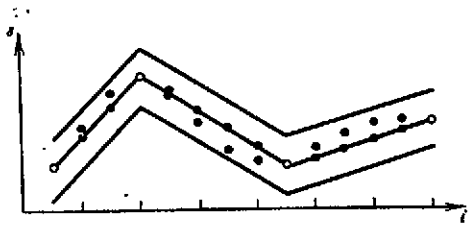


Figure 7. First Order Interpolation. Mating Points of Interpolating Line Sectors Matched with Values of Readings

If one of these lines passes by s_{i+2} at a distance of less than γ, s_{i+2} is considered redundant, even though it may be outside the aperture. The line which provided a precise prediction of the reading, within limits of $\pm\gamma$, remains unchanged for the estimate of the next reading. The line which passed by s_{i+2} at a distance of greater than γ is replaced by a new line passing by s_{i+2} at a distance of γ. To do this, a straight line is drawn from s_i either through $s_{i+2} + \gamma$ or through $s_{i+2} - \gamma$, depending on which of these points was closer to the old line. Thus, the fan aperture may change with each redundant reading, but the new aperture is always within the old. Therefore, as in prediction methods, there is no need to check the accuracy of approximation of intermediate readings. Any line with its origin at s_i drawn within the aperture assures interpolation of all intermediate readings with an accuracy of $\pm\gamma$. Consequently, on each step it is sufficient to check the accuracy of prediction of the last previous reading. The process is continued until some reading s_{i+h} lies at a distance of greater than γ from the straight lines drawn during the previous stage. In this case, the position of the center (or of any other point) of the aperture at moment in time i + h - 1 is transmitted. Interpolation of subsequent readings is then performed with the origin at the transmitted point. The operation of the algorithm

is illustrated graphically in Figure 8. Here, for definition, the initial point is placed on the first reading. Generally, the mating points are found by calculation and may not correspond to the values of any readings, such as the second or third transmitted points. The method of fan interpolation is highly effective and does not require storage of intermediate readings or checking of the error of their interpolation, as is required in all of the interpolation algorithms described above.

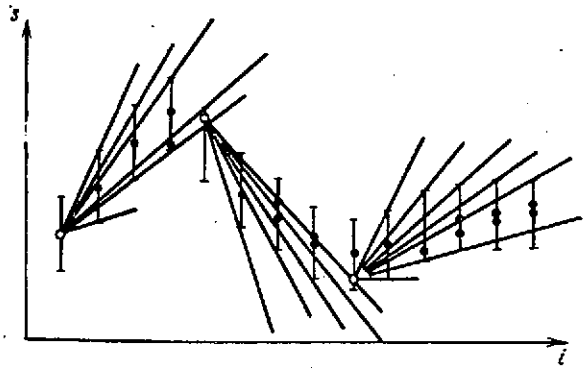


Figure 8. Fan Interpolation

In conclusion, let us study one more interpolation method, which differs in principle from those described above in that all readings are transmitted. /192 Furthermore, the maximum error is not fixed, but the mean square interpolation error is minimized. The operating principle of this algorithm, called "multistage interpolation" by its inventor [24], consists in the following. Each fourth reading, regardless of its value, is considered significant and transmitted by the ordinary method, i.e., as a complete binary number. The intermediate readings are interpolated by a special method, so that one ternary digit of information is required to transmit each one. Let us study a group of readings s_i, s_{i+1}, s_{i+2}, s_{i+3}, s_{i+4}, where readings s_i and s_{i+4} are significant. First, the middle reading s_{i+2} is interpolated (quantized). It is assigned the closest of the three values

$$\frac{s_i + s_{i+4}}{2},$$
$$\frac{s_i + s_{i+4}}{2} + f(s_{i+4} - s_i),$$
$$\frac{s_i + s_{i+4}}{2} - f(s_{i+4} - s_i),$$

where $f(\cdot)$ is a certain function which we will call the interpolating function. Then the interpolated value s^*_{i+2} of this reading can be written as

$$s^*_{i+2} = \frac{s_i + s_{i+4}}{2} + a_{i+2} f(s_{i+4} - s_i),$$

where coefficient a_{i+2} may take on one of three values: 0, +1, -1. After this, reading s_{i+1} is quantized similarly, with the single difference that s_i and s^*_{i+2} are considered to be the significant readings for it, and another interpolation function $\phi(s^*_{i+2} - s_i)$ is used. Reading s_{i+3} is quantized using s^*_{i+2} and s_{i+4} and the interpolating function $\phi(\cdot)$.

Obviously, the interpolation error depends essentially on the type of interpolating functions, which must correspond to the nature of the input signal. In [24], the type of interpolating functions was selected experimentally. Either way, the form of functions f and ϕ is known in advance, so that for each group of four readings s_{i+1}, s_{i+2}, s_{i+3} and s_{i+4}, it is sufficient to transmit the precise value of one significant reading s_{i+4} (reading s_i is included in the previous group) and the three ternary symbols a_{i+1}, a_{i+2} and a_{i+3} for the intermediate readings. If a binary channel is used for transmission, the three ternary symbols are recorded into five binary symbols. Then, transmission of one reading requires an average of (6 + 5)/4 = 2.75 bits, if the significant readings are transmitted by 6-bit words. Compression of various television images using this algorithm was modeled by computer. With a signal/noise ratio of up to 30 db, the image quality was not noticeably reduced in comparison to ordinary digital transmission.

In comparing the compression algorithms studied here, we must note that usually the best results are yielded by first order interpolation. This is particularly true when television signals are compressed [21]. Some works have suggested more complex algorithms. These include, for example, extrapolators, calculating the conditional mode or conditional mean to estimate the next reading [25, 26]; interpolators in which the intermediate readings are approximated not by points on a line but by conditional means [27].

These complex compression algorithms, based on approximation of the initial signal, also include a group of compression methods using digital filters [15]. In these methods, the initial function is expanded into a series, and the coefficients /193 of the expansion are transmitted. In spite of the fact that these methods can produce maximum effectiveness, they are of merely theoretical interest, due to the exceptional complexity of their realization. Generally, the results of studies performed by many authors have shown that even a slight increase

in effectiveness is purchased at the expense of a significant complication of the compression algorithms. This comment also applies to interpolation; on the one hand, it is more effective than extrapolation, but on the other hand, it is more complex for realization. Actually, first and higher order interpolation algorithms require calculations with lower quantization error than simple quantization of the input signal. Furthermore, the values of all preceding readings must generally be stored and the accuracy of their interpolation checked in each step.

Prediction algorithms are slightly inferior to interpolation as to effectiveness, but do not have these shortcomings. Generally, the selection of either method of compression may be based on various factors, and it is impossible to name a single best method. In each specific case, one must be guided by the requirements on the entire system as a whole, determining the requirements for the compression algorithm and only then making a selection as to the specific method to be used.

Compression of the Volume of Statistical Data

In some experiments, the task of statistical investigation of an observed phenomenon is set. For example, these include experiments involving the study of characteristics of a stream of particles such as temperature, mass spectra, etc. The experimentor is not interested in the value of the observed parameter at each moment in time, but rather in its statistical characteristics, for example the mean value, dispersion, etc. It may also be necessary to check the agreement of data produced during the course of the experiment with a supposed mathematical model of the phenomenon. In these cases, rather large quantities of experimental data are accumulated and transmitted to the reception point, where they are processed by statistical methods. In problems of this nature, the parameter observed during the course of an experiment is looked upon as a certain random quantity ξ with distribution function $F(x)$. Since the task of reduction of the volume of transmitted data is closely related here to mathematical statistics, let us briefly discuss certain definitions which will be required in our further presentation. A full and strict presentation of the statements encountered from statistics can be found, for example, in [28].

To produce the necessary statistical data, n independent measurements are made, yielding n observed values of random quantity ξ. Let us represent these values as $x_1, x_2, \ldots, x_n$. This sequence of observed values of random quantity ξ is called a sample of n values of a distribution corresponding to $F(x)$ or a sample of a set with distribution function $F(x)$. The term

"set" here refers to all the values of random quantity ξ. The actual values $x_1, x_2, \ldots, x_n$ are called the sample values. An empirical distribution function $F^*(x)$ can be constructed for the sample. This will be an exponential function with jumps in height which are multiples of $1/n$ at each point x_i. If we represent the number of sample values not exceeding x by α, then

$$F^*(x) = \frac{\alpha}{n}.$$

Thus, $F^*(x)$ is the frequency of production of values $\xi \leq x$ in each sequence of n measurements. With unlimited increase in the volume of a sample, the empirical distribution function $F^*(x)$ approaches the distribution function of set $F(x)$. If n is sufficiently great, it can be considered for practical purposes that $F^*(x)$ yields a good approximation of function $F(x)$. Consequently, the distribution function of sample $F^*(x)$ can be looked upon as a certain statistical analog of the distribution function of the set. /194

A statistical analog of the probability density can also be constructed for the sample. To do this, the sample values are first grouped, i.e., the individual values are dropped, and only the number of sample values falling in the interval corresponding to each given group is retained. (This is generally true of samples of continuous distributions, since in the case of a discretely distributed random quantity, ξ can only take on a finite number of values.) Each interval is taken as the base of a rectangle of height $\varepsilon/(l \cdot n)$, where l is the length of the interval, ε is the number of sample values falling within the interval. Obviously, the area of each such rectangle is equal to the frequency with which the process falls within the corresponding interval. With large n, this area can be considered equal to the probability of falling within the interval. The figure produced as a result of this construction is called a histogram of the sample. Experimental data are frequently presented in the form of histograms.

Knowing the sample values, we can calculate various characteristics of the sample: the moments, asymmetry coefficient, correlation coefficient, etc. As the volume of the sample approaches infinity, the characteristics of the sample converge with the corresponding characteristics of the actual set. This allows us to look upon the characteristics of the sample as an estimate of the corresponding characteristics of the set. Generally speaking, there are many different functions of the sample values which can be used as estimates of some unknown

parameter of the distribution of the set. We must therefore be able to compare the properties of various estimates for a given parameter in order to select the best estimate. Problems of this type are studied in the theory of estimates [28].

To produce reliable results, a large quantity of experimental data is required, but storage and transmission of these data represent serious technical difficulties. The quantity of data can be reduced by representing the data by a small number of descriptive characteristics. The ordinary characteristics of the distribution of sample values can be used for this purpose, for example the mean value, dispersion, etc. Representation of data in the form of histograms is another method of reduction of the volume of statistical data.

Thus, the problem of reduction of redundancy in these cases consists in replacement of the initial set of data (sample values) with a small number of descriptive characteristics allowing us to produce the necessary information concerning the object of study. There also exist optimal functions of sample values in the sense that nothing other than these functions need be known to estimate the parameters of a distribution. These functions are called sufficient statistics [1]. It is important to note that in selecting descriptive characteristics, we must consider not only information content, but also computation complexity. This represents a serious obstacle to the application of such common characteristics as mean values, dispersion, etc.

One possible solution to the problem of selecting descriptive characteristics was suggested in [29, 30]. Here, in place of all observed sample values, only certain quantiles (percentage points) of the sample are transmitted, knowledge of which /195 allows us to calculate asymptotically unbiased estimates for various distribution parameters.

Let us study a continuous distribution with distribution function $F(x)$ and probability density $f(x) = F'(x)$. The order p quantile ζ_p of this distribution refers to the root of the equation

$$F(\zeta_p) = p, \quad (0 < p < 1).$$

If this equation does not have a unique solution, the quantile may refer to any one of its roots. The distribution quantile can also be determined through the probability density

$$\int_{-\infty}^{\zeta_p} f(x)\,dx = p, \quad (0 < p < 1).$$

Thus, the sense of quantile ζ_p consists in that random quantity ξ takes on any value $x \leq \zeta_p$ with probability p. For example, quantile $\zeta_{0.5}$ is well known as the median of the distribution.

Similarly, for a sample from a set having distribution function F(x), we can determine the sample quantile Z_p, corresponding to the quantile of set ζ_p. To do this, we place the sample values in non-increasing order:

$$x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n.$$

Then the sample quantile

$$Z_p = x_{\langle np\rangle+1}, \quad (0 < p < 1),$$

where <np> represents the greatest integer not exceeding np. Thus, quantile Z_p is simply equal to the (<np> + 1)th sample value. Obviously, np·100% of sample values will not exceed $x_{\langle np+1\rangle} = Z_p$. [If np is an integer, there is some uncertainty, and Z_p may take on any value in the interval (x_{np}, x_{np+1}).] Thus, we can see from the definition that arithmetic operations are generally not required to determine the quantile of a sample.

The sample quantiles themselves are random quantities with distribution function approaching the normal distribution function with mean value

$$E(Z_p) = \zeta_p$$

and dispersion

$$D(Z_p) = \frac{p\cdot(1-p)}{n\cdot f^2(\zeta_p)},$$

where $f(\cdot)$ is the probability density as the volume of the sample approaches ∞. Therefore, for a sufficiently large sample volume, it can be considered in practice that quantities Z_p are normally distributed.

If we have a set of sample quantiles, we can calculate estimates for such parameters of the distribution as the mean value μ and mean square deviation σ. The estimates of these parameters are linear combinations of quantiles taken with constant coefficients, calculated in advance. The measure of accuracy of an estimate is its dispersion. Here the accuracy of an estimate is characterized by a relative indicator, the effectiveness of the estimate, defined as the ratio of the minimum possible dispersion determined by the lower boundary of the Rao-Kramer inequality (using all sample values) to the actual value of dispersion of the estimate (using the given set of quantiles). Obviously, the effectiveness of estimates depends on the number of quantiles k and their position; therefore, quantiles should be selected particularly carefully. /196 Furthermore, the quantiles should be selected so as to assure production of all necessary information using a single set of quantiles, since a change in the number and order of quantiles during an experiment requires great complication of apparatus. The problem of selection of quantiles is studied in [29, 30], on the assumption that the distribution of the set is normal. It is shown that the best estimates for μ and σ are produced with symmetrical placement of an even number of quantiles ($p_i + p_{k-i+1} = 1$). As the number of quantiles increases, the effectiveness of estimates increases. Thus, according to [29], where k = 2

$$\mathrm{Eff}\,(\hat{\mu}) = 0{,}8098 \quad \text{and} \quad \mathrm{Eff}\,(\hat{\sigma}) = 0{,}6522,$$

where k = 10

$$\mathrm{Eff}\,(\hat{\mu}) = 0{,}9808 \quad \text{and} \quad \mathrm{Eff}\,(\hat{\sigma}) = 0{,}9496,$$

where k = 20

$$\mathrm{Eff}\,(\hat{\mu}) = 0{,}9943 \quad \text{and} \quad \mathrm{Eff}\,(\hat{\sigma}) = 0{,}9839,$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the estimates of μ and σ respectively. We note that the effectiveness of the estimates does not increase linearly with increasing number of quantiles, but rather has a tendency toward "saturation." Therefore, the use of a large

number of quantiles is ineffective, since the gain in effectiveness of estimates is not great, while the volume of data transmitted increases in proportion to the number of quantiles. According to [29], in practice four quantiles are sufficient for calculation of estimates of μ and σ. In this case, the effectiveness of the estimates with optimal placement of the quantiles is

$$\mathrm{Eff}\,(\hat{\mu}) = 0{,}9201 \quad \text{and} \quad \mathrm{Eff}\,(\hat{\sigma}) = 0{,}8244.$$

It is important to note that the optimal sets are different for μ and σ. Slight deviations from the optimal placement do not result in serious reduction in effectiveness, but the use of quantiles optimal in relationship to σ for estimation of μ reduces the effectiveness of the estimate by 25.7%. If σ is estimated using a set of quantiles optimal for μ, the effectiveness of the estimate of σ is reduced by 19.2%. Thus, the problem arises of compromise selection of a certain suboptimal set of quantiles providing the necessary effectiveness for both estimates. In [29], it is suggested that a set of quantiles be sought for this purpose to minimize the sum of dispersions of the estimates

$$D(\hat{\mu}) + bD(\hat{\sigma}),$$

where b is a weight factor selected according to the relative importance of each estimate. For example, with increasing b the effectiveness of the estimate for σ might increase, the effectiveness of the estimate for μ decrease.

Selection of quantiles is also complicated by the fact that as their number increases, in order to increase the effectiveness of estimates, the extreme quantiles (of order p_1 and p_k) are shifted ever further out onto the "tails" of distribution. This is undesirable for two reasons: first of all, the distribution of the set generally differs from the normal; secondly, the volume of a sample is always limited. Therefore, even with normal distribution of the set, the distribution of the sample quantiles is not normal. These deviations from the normal rule are particularly extreme on the "tails" of the distribution and lead to bias in the estimates, i.e., the mean value of an estimate converges not on the parameter being estimated, but rather on some other value, shifted relative to the value of the parameter. In order to avoid or at least limit the bias, we must limit the minimum value p_1 and the maximum value p_k.

Knowledge of the quantiles allows us not only to calculate estimates for the parameters of the distribution, but also to check simple hypotheses, i.e., to establish agreement of the experimental data to the advance hypothesis concerning the nature of the distribution or its parameters. For example, it may be required to check the hypothesis that a normal distribution with known mean has dispersion σ^2. Usually, the criteria establishing this agreement are based on all sample values; however, criteria may be constructed knowing only a small number of sample quantiles. In this case, naturally, the optimal quantiles for criteria related to μ are those providing the maximum value of $\mathrm{Eff}(\hat{\mu})$, while criteria related to σ require quantiles maximizing $\mathrm{Eff}(\hat{\sigma})$. According to [30], a set of four quantiles is sufficient to check a simple hypothesis concerning the parameter of the normal set. /197

Quantiles can also be used to check more complex hypotheses. For example, the purpose of an experiment is to study the energy of particles. In this case, it is interesting to know whether the distribution has one maximum or two, since bimodality of a distribution indicates the presence of particles of two different types. This problem is also solved using only four quantiles [30].

The degree of reduction of transmitted data using a set of four quantiles can be illustrated by an example borrowed from [30]. This work studied the use of quantiles for transmission of the results of an experiment involving determination of the intensity of a stream of particles. A sample of 1024 values was taken, the number of particles recorded per unit time taking on values of 0 to 127, i.e., seven bits were used to describe each sample value. This means the transmission of all sample values would require $1024 \cdot 7 = 7168$ bits. Transmission of the four quantiles of this sample required only $4 \cdot 7 = 28$ bits, i.e., the volume of data transmitted was reduced by a factor of 256!

A method was developed for selection of quantiles and numerical results were produced in [29, 30] on the assumption that the distribution of the set was normal or near-normal. It is possible to find the optimal sets of quantiles and calculate the effectiveness of the corresponding estimates for other distributions as well, if the distribution rule is known, and only its parameters are unknown. In some cases, the use of quantiles or similar statistics is convenient and natural. For example, with even distribution of a set, the statistics of extreme values of the sample (minimum and maximum values) is sufficient.

In the case when the distribution is unknown, we cannot say anything definite as to how the quantiles should be selected

and what the effectiveness of estimates will be. The necessity of knowing the distribution is a significant shortcoming preventing broad use of this method.

We must note that the method which we are studying here cannot be considered universal for all problems. It is good enough for estimation of the parameters and construction of parametric criteria, but this is only a portion of the problem of statistical processing of experimental data. However, this does not mean that when more complex statistical problems must be solved, all sample values must be transmitted. The literature on mathematical statistics includes analysis of methods of performing certain tasks allowing the volume of transmitted data to be reduced; these methods are not always so complex that they cannot be performed with on-board apparatus. However, the problems related to the application of more powerful statistical methods for reduction of the volume of telemetry data have not yet been solved.

Problems of Determination of the Effectiveness of Data Compression /198

In the preceding sections, we have studied various methods of data compression. One of the most important questions arising in the selection of a method is its effectiveness, i.e., the degree of reduction of the volume of data. However, the problem of estimating the effectiveness of compression of data is quite complex. As was noted in the introduction, two approaches are possible to the evaluation of effectiveness. First of all, the effectiveness of compression can be estimated by the degree of approximation to the minimum possible volume. This approach, apparently, is most desirable, but at the same time is the most complex approach. Furthermore, the limit which should be approached in the reduction of data volume cannot always be found for every source. Secondly, the effectiveness of compression may be determined by comparing the volume of data after compression with the initial volume. This method, obviously, is simpler than the first method, but this type of estimation is quite subjective, since it depends on the efficiency and economy of the initial presentation of the data. Nevertheless, this type of evaluation does allow various methods of compression to be compared as to their effectiveness and is useful in this sense. Of course, the initial representation of data should be identical for all the methods being compared. Most authors evaluate the effectiveness of suggested compression methods by this means; therefore, let us study this question in greater detail.

The traditional system presently used for the transmission of telemetry data is a pulse-code modulation (PCM) system, in which all messages from a given source (readings) are independently coded by words of equal length, which are then transmitted through the communications channel. Therefore, as a rule, the effectiveness of data compression is estimated by comparing it with the volume of data required for transmission in the PCM form. The method of evaluation of the effectiveness of algorithms based on prediction and interpolation follows directly from the compression procedure itself. Of the sequence of all available readings, a few (significant) readings are selected, which are then transmitted. The remaining (redundant) readings are discarded. Obviously, the most natural and simplest estimate of effectiveness of such an algorithm is the ratio of the total number of readings (transmitted in a PCM system) to the number of significant readings selected by the compression system. This ratio is broadly used in the literature and is called the "data compression factor" or "sampling compression factor." In the preceding sections, which discussed problems of the effectiveness of compression methods, it was this simplest estimate which was referred to as the effectiveness. In our subsequent presentation, we will use the term "sampling compression factor" and the symbol K_B.

Unfortunately, the area of application of the sampling compression factor is quite limited, since restoration of the realization of the process requires more than simple knowledge of the readings; it is also necessary to know the moments in time corresponding to these readings. In the PCM system, this requirement is automatically fulfilled, since all readings are transmitted as they appear, forming a regular sequence. In a system with compression, the appearance of significant readings is irregular in nature; therefore, transmission of readings as they appear in this case results in ineffective utilization of the communications channel and creates serious technical difficulties upon reception. In order to smooth the irregularity of appearance of readings, a buffer memory device can be used. The significant readings are stored in the buffer memory as they appear, then taken from memory at a constant rate (the transmission rate). The use of the buffer brings up its own specific difficulties. For example, when information activity occurs in bursts, the rate of arrival of significant readings is high /199 during these bursts and the buffer memory may be overfilled, resulting in loss of some of the information. On the other hand, when information activity is low, between bursts, the buffer memory may be completely emptied, causing a pause in transmission. The buffer memory, however, is not a necessary link in every system. For example, when the problem is not one of transmission, but rather one of storage of large volumes of

data, the buffer memory is unnecessary. In telemetry systems with data compression, the buffer can be avoided if transmission is performed not in real time, but rather after completion of a group of measurements, using a long-term memory device. Therefore, we will not give further attention to buffer memories. Some problems related to the design of buffers for transmission systems are studied in [15, 31, 32].

As the significant readings are written in the buffer or any other memory device, it is also necessary to record the time when the readings occurred, since otherwise it is impossible to establish when a given reading was taken. Furthermore, in multichannel telemetry systems, we must also know the channel numbers to which the stored readings correspond. The volume of this additional information, necessary for restoration, may vary, depending on the order of collection, the compression algorithm and the method of representation of data for transmission.

The collection algorithm may be adaptive, i.e., the sequence of interrogation of channels established by the collection system itself, depending on the information activity of the channels. Then it is necessary to know the time and channel number for each reading transmitted, since the channel interrogation program is not known at the reception point. Modern telemetry systems generally use a fixed program of collection with time separation of channels. Obviously, in this case the number of the position of the channel switch (address) where a significant reading was produced unambiguously determines the time and the number of the telemetry channel for this reading. Therefore, only the address of each reading need by known to restore the information.

We note that the addresses of significant readings may also be compressed, by zero order prediction with shift, where the value of the shift is always +1, i.e., each significant reading is represented by a pair "reading-address" only if the previous reading was redundant. If the previous reading was significant, only the value of the new reading is transmitted, since its address can be established by adding 1 to the address of previous reading. Obviously, when this method of address transmission is used it becomes necessary to distinguish between the code combination for a reading and an address.

Other methods of reduction of the volume of service information transmitted are known, if channels are interrogated according to a rigid program. For example, after each cycle of interrogation, a transmission frame is composed, consisting of the significant readings and a binary sequence, with one bit for each channel interrogated in the cycle. Thus, one bit per

channel is used to indicate the presence of a significant reading in the channel, i.e., the numbers of the bit positions in the sequence which contain ones correspond to the numbers of channels in which significant readings were produced. Decoding requires that the sequence of symbols received be broken down into frames; therefore, in calculating the volume of service information, frame synchronization information must be included. Unfortunately, this method is not always effective in comparison to simple transmission of "reading-address" pairs. Thus, if the number of significant readings in a frame is low in comparison to the number of channels, this method is ineffective.

One improved method of coding groups of addresses using a binary sequence was suggested in [33]. In this method, it is considered that the sequence of addresses of significant readings within each frame is strictly monotonic. The number of all strictly increasing sequences is C_n^i, where n is the number of channels in the commutator, i is the number of significant readings in a given frame. Therefore, with any possible sequence of i significant readings, $\log_2 C_n^i$ bits are sufficient to code the addresses of the entire group of significant readings. A transmission frame in this case consists of the readings themselves and a binary sequence of length $[\log_2 C_n^i]$ (where the brackets represent the least integer no less than $\log_2 C_n^i$). The length of this address sequence changes from frame to frame, depending on the number of significant readings. Therefore, here, in contrast to the preceding case, when reception is formed we must not only perform division into frames but also must somehow separate the address sequence within each frame. However, if the length of the code combination of a reading is greater than $\log_2 n$, separation of the address sequence is possible without additional expenditures for synchronization, since in this case the length of a frame, i.e., the number of bits including the values of readings of the address sequence, is a strictly monotonic function of the number of significant readings i. Therefore, after reading the number of bits in the frame received, we can unambiguously determine the number of significant readings in a given frame and, consequently, the length of the address sequence. This method of address coding is always more effective than the preceding method and is never worse (if we ignore expenditures for frame synchronization) than ordinary transmission of "reading-address" pairs. Actually, the length of the address sequence changes with changing numbers of significant readings, so that the relationship /200

$$\log_2 C_n^i \leqslant i \cdot \log_2 n,$$

is always fulfilled, where the right portion carries the volume of service information required to transmit "reading-address" pairs. In a rough comparison, we can ignore the expenditures for frame synchronization, since when pairs are transmitted, synchronizing bits are also required. Otherwise, we must number all readings not from the beginning of each frame, but rather from the beginning of the measurement session, but in this case the number of bits in the code combination for the address must be increased.

A method of address coding and decoding using sequences of variable length was described in [33]. The method requires that the memory of the system (at the receiving and transmitting ends) contain a table measuring n × n. With a large number of channels, the dimensions of the matrix and the numbers which are its elements are quite large; therefore, large memory volumes are required to realize this coding method.

As we can see from these examples, attempts to reduce the volume of messages transmitted always result in more complex methods of coding and decoding. Nevertheless, in some cases we can increase the complexity of apparatus and processing algorithms if justified by higher system effectiveness. However, the effectiveness of the system is determined not only by the effectiveness of the compression algorithm, but by the method of representation of service information as well. Therefore, the sampling compression factor is not suitable for comparison of data transmission systems, since it does not reflect the degree of reduction of the number of symbols transmitted. The total volume of transmitted data may vary with the same value of sampling compression factor, if various methods of representation of service information are used.

In order to estimate the effectiveness of data transmission systems using compression, we have introduced the "number of bits compression factor," K_{bit}, sometimes called the "frequency band compression factor." It is defined as the ratio of the total number of bits transmitted in the system without compression to the total number of bits transmitted in the system with compression. Here, in addition to the readings themselves, all of the additional (service) information transmitted in both systems is considered. /201

The number of bits compression factor is a good representation of the effectiveness of compression in data transmission systems, where no errors are introduced upon transmission through the communications channel. However, in many practical cases, the level of errors in a communications channel is rather high, so that errors must be considered. Compression increases the "sensitivity" of data transmitted to errors. Actually,

compression reduces the "disorganized" redundancy included in the initial set of data. It is the presence of this redundancy which allows data to be compressed; at the same time, this redundancy allows detection and correction of errors arising during transmission. However, this ability is practically lost by compression.

Consideration of the influence of error on a system using data compression is complicated by the fact that the errors in the readings themselves and in the service information have a different degree of influence upon restoration. For example, let us study a system in which all significant readings are transmitted by reading-address pairs. If an error occurs in the word determining the value of a reading, the number of readings distorted upon restoration is determined by the sampling compression factor. However, if the error occurs in the word determining the address, the number of readings distorted depends not only on the compression factor but on the magnitude of the error. Great errors, i.e., errors disrupting the monotonous increase in address values for significant readings within a frame, can be detected [34].

In a system in which addresses are coded using a binary sequence of constant length, an error in this sequence results in incorrect restoration of the addresses of all subsequent readings of the frame. However, all errors in a sequence resulting in a change in the number of ones in the sequence are easily detected by the fact that the number of ones does not correspond to the number of significant readings. When a variable-length sequence is used, errors are generally not detectable, since all code combinations in the address sequence in this case are permissible.

It follows from these discussions that in comparing data transmission systems with compression we must require that the system with compression provide the same "quality" of restoration of data as the system without compression when the same communications channel is used. By quality here we mean a certain function of cost, for example the probability of error in a restored reading. It is then necessary to reduce the probability of error of significant readings transmitted in the system with compression. We know that the probability of error is lower, the higher the ratio of signal energy to spectral density of noise. Therefore, in order to provide the same quality of restored data, the quantity of energy expended on transmission of a significant reading must be greater than the quantity of energy expended on transmission of a reading in a system without compression. Thus, under these conditions compression can result in an increase in the energy of symbols.

Therefore, for channels with noise, the criterion for comparison of data transmission systems should be the quantity of energy per reading with a fixed quality of restoration of data and a fixed noise level in the communications channel. This energy approach is particularly justified when the power resources of the transmitter are limited. In these cases, the main purpose of data compression is to decrease the power consumed for information transmission. [1]Davison [35] has introduced the /202 "energy compression factor" K_e to estimate the effectiveness of compression as to power consumption. It is defined as the ratio of the power expended in the transmission of one reading in the PCM system to the power required to transmit a restored reading in a compression system with identical noise level and quality of restored data. The energy compression factor is the most reliable criterion of effectiveness, since it characterizes the entire data transmission system as a whole. Unfortunately, calculation of the energy compression factor is a practically difficult task. This results primarily from the calculation of the value of error in the system for discrete data transmission.

In order to illustrate the application of this energy criterion, let us study an example borrowed from [35]. We will use a zero order prediction algorithm and transmit readings in reading-series length pairs, where the series length (number of repetitions of a reading) can take on 16 values. The input signal is also quantized into 16 levels. Transmission is performed through a channel acted upon by white Gaussian noise. The readings are transmitted in blocks separated by synchronization words, with a maximum number of readings in a block equal to N. It is assumed that errors are not accumulated from block to block, i.e., synchronization is not subject to errors. At the same time, however, an error in a word defining the length of a series results in false restoration of all subsequent readings in the block. In order to reduce the number of bits transmitted, the synchronization words must be transmitted as rarely as possible. On the other hand, long intervals between synchronization words result in an increase in restoration errors. The power approach to solution of this problem allows us to find the optimal length of a block and establish the true gain produced by compression.

The expression for the energy compression factor can be written as

---

[1]An energy criterion for evaluation of the effectiveness of compression was suggested earlier in [36]. However, the energy compression factor in the formulation of Davison is more strict and natural.

$$K_e(P_{er}, N) = (\lambda^2_{PCM}/\lambda^2_c)\cdot K_{bit}(N),$$

where λ^2_{PCM} and λ^2_c are the signal energy to spectral noise density ratios for ordinary PCM and the system with compression respectively. The quality of reception is determined by the probability of error P_{er}, averaged over the length of a block.

Figure 9 presents curves of the dependence of energy compression factor on block length for various error probabilities. We can see from the figure that for any value of error probability there is an optimal block length, for which the energy compression factor is maximal. We note that in this example with a sampling compression factor of 4.96 and 2.43 bits, consideration of channel errors results in an energy compression factor of only 1.6 with a probability of error of 10^{-3}. Furthermore, Figure 9 leads us to the very important conclusion that the gain produced in compression may be totally lost if the block length is improperly selected.

This example shows clearly that ignoring the noise in a communications channel may result in a false conclusion concerning the effectiveness of the use of compression and improper system design. In the example which we have studied here, the gain produced from the use of compression (K_{bit} = 2.43) was reduced by approximately 1.5 times due to the necessity of increasing power upon transmission. However, these additional expenditures of energy can be slightly decreased. In order to reduce the probability of an error, we can use correcting coding rather than increasing signal energy. The expediency of this solution can be studied using the same example. Figure 10 shows a graph of the energy compression factor as a function of block length when values of readings and the numbers of repetitions of readings are coded using a (7,4) code. In this case, the energy compression factor with optimal block length is approximately 2.2[2], i.e., approaches the number of bits compression factor. The power gain produced in this example as a result of correcting coding is not accidental. It has been demonstrated [38] that for values of error probability of practical interest, a "good" correcting code can always be constructed, providing a power gain in comparison with a direct increase in signal energy. /203

---

[2]This high value of K_e = 2.2 does cause some doubt, since the energy gain of 1.4 times for the (7,4) code under these conditions is probably elevated [37].

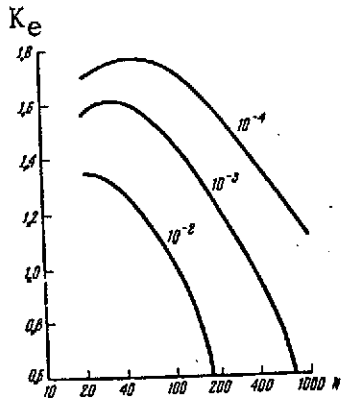


Figure 9. Energy Compression Factor as a Function of Block Length with Various Probabilities of Error per Symbol

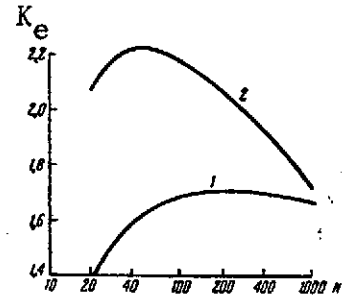


Figure 10. Energy Compression Factor as a Function of Block Length for Coding of Reading Values alone (1) and Coding of Both Reading Values and Numbers of Repetitions (2) with a (7,4) Code. Error Probability per Symbol 10^{-3}

Another approach to evaluation of the effectiveness of the use of a system with compression and the related selection of a data transmission system was suggested in [34]. Its essence is as follows. Three versions of construction of a communications system are analyzed: 1) a system without compression and without coding (meaning correcting coding); 2) a system with compression but without coding; 3) a system with compression and coding. The energy of the symbols in all systems is assumed identical. Each system is characterized by two parameters: the relative "rate" R and the relative "distortion" D. The selection of these parameters, similar to the rate and distortion introduced by Shannon [39], allows a comprehensive and complete characterization of any communications system. The problem consists in selecting a system satisfying fixed values of maximum mean distortion D_{max} and maximum relative rate R_{max}. The value of D_{max} is fixed by the information consumer, the value of R_{max} is determined by the throughput capacity of the communications channel and the cost of the complication of the apparatus and processing algorithms.

The relative rate, corresonding to the rate of generation of messages by the source (in systems 2 and 3, the source includes the compression device) for various systems is: /204

$$R_1 = 1,$$

$$R_2 = 1/K_{bit},$$

$$R_3 = (1/K_{bit})\beta,$$

where $\beta > 1$ considers the increase in the number of bits due to correcting coding. The numerical subscripts indicate the system involved.

Mean distortion D for any system is determined by the equation

$$D = \sum_{M,Z} P(M) \cdot P(Z/M) \cdot D(M, Z),$$

where M is the set of messages at the input of the channel, Z is the set of messages at the output of the channel; P(M) is the distribution of probabilities of source messages; P(Z/M) is the matrix of arbitrary probabilities of transition from M_i to Z_j upon transmission, D(M, Z) is the matrix fixing the measure of distortion upon transition from m_i to z_j.

Probabilities P(M) are identical for systems 2 and 3, but differ from the probabilities of messages of system 1, since the sources of messages differ in nature. The transition matrices P(Z/M) are different for all systems, since when these matrices are constructed the possibility of correction of certain errors is considered. In system 1, due to the natural redundancy of the messages transmitted, certain errors can be corrected; in system 3, correction of errors becomes possible due to the redundancy of the correcting code; in system 2, no error correction is performed. Matrices D(M, Z) for all systems are assumed identical. We can then write the mean distortion for each system:

$$D_1 = \sum_{M,Z} P_1(M) \cdot P_1(Z/M) \cdot D(M, Z),$$

$$D_2 = \sum_{M,Z} P_2(M) \cdot P_2(Z/M) \cdot D(M, Z),$$

$$D_3 = \sum_{M,Z} P_2(M) \cdot P_3(Z/M) \cdot D(M, Z).$$

These expressions, however, do not consider the additional distortions introduced by errors in the service information in systems 2 and 3. These distortions D_2' and D_3' are defined similarly:

$$D_2' = \sum_{M,Z} P_2(M) \cdot P_2'(Z/M) \cdot D(M, Z),$$

$$D_3' = \sum_{M,Z} P_2(M) \cdot P_3'(Z/M) \cdot D(M, Z).$$

Matrices $P_2'(Z/M)$ and $P_3'(Z/M)$ in these expression also consider the possibility of detection and correction of certain errors, but differ from matrices $P_2(Z/M)$ and $P_3(Z/M)$. The possibility of detecting certain errors in the words defining the addresses of readings in system 2 was mentioned above. The resulting distortion D_Σ for systems 2 and 3 is found as the sum of the distortions of readings and addresses; for system 1, the resulting distortion is equal to the distortion of the readings.

We note that the mean distortion in systems with compression /205
is a function of the compression factor. Therefore, the maximum permissible mean distortion determines the maximum permissible compression factor. On the other hand, the use of a system with compression is justified only if it provides a significant compression (decrease in rate). But since the relative rate is inversely proportional to the compression factor, the maximum value of relative rate determines the minimum permissible value of compression factor. Thus, the maximum and minimum values form the area of permissible values of compression factor. Obviously, it may occur that the maximum value is less than the minimum value, i.e., a system with compression is unacceptable. Therefore, attempts to produce the greatest possible compression factor are justified only within the area of its permissible values.

Selection of a system using these parameters is performed as follows. For each source of messages, the consumer fixes the maximum value of mean distortion D_{max} for a given measure of distortion $D(M, Z)$. For systems 2 and 3, the quantity $R_{max} < 1$ is determined from considerations of expediency of the use of compression. Then, the values of R and D_Σ are calculated for each system. As a result of comparison of these parameters, the system is selected which provides distortion less than D_{max}. However, if systems with compression, while fulfilling the condition $D_\Sigma < D_{max}$, yield no significant reduction in rate, i.e., fail to fulfill the condition $R < R_{max}$, system 1 is selected (if, of course, $D_{1\Sigma} < D_{max}$). The gain in rate in this case is not

great, but the realization of the system is simpler than that of a system with compression. If the condition $D_\Sigma < D_{max}$ is not fulfilled for any system, we must either improve P(Z/M), i.e., reduce the probability of error in the channel, using a more powerful correcting code if necessary in system 3, or increase D_{max}.

This approach to the comparison of data transmission systems yields an objective concept of the relationships between systems. The parameters selected here allow us to consider all factors influencing effectiveness. The primary advantage of this approach is that with this statement of the problem a system is selected which satisfies the required conditions in the best manner. However, here also we find the primary shortcoming of this method, limiting the area of its practical application. It is the necessity of knowing the distribution of probabilities of message sources and the maximum value of mean distortion for each source. Unfortunately, we can not always fix the value of D_{max} in advance, and the statistics of sources are generally not known sufficiently. We must then produce a model of the message source and take the distortion in an ordinary PCM system as the maximum value of mean distortion. System 2 in this case obviously must be eliminated from consideration, since $D_{2\Sigma} > D_{1\Sigma}$, and comparison of systems 1 and 3 under these conditions is equivalent to calculation of the energy compression factor.

Models of Message Sources

In the preceding section, we studied criteria of effectiveness for data compression systems. A numerical value of effectiveness can be produced and a compression method selected only for a specific message source, since the same algorithm may have different effectivenesses, depending on the nature of the source. The set of information on the source necessary to make a judgment as to the applicability of a specific compression method and determine its effectiveness is determined by the essence of the compression method itself. Generally speaking, this a priori information may differ significantly both in nature and in completeness; therefore, comparison of compression methods should be based on the presence of all necessary information concerning the source. In practice, the a priori information is generally quite limited; sometimes, for example, almost nothing is known of a message source at all in a new space experiment. Under these conditions, naturally, preference must be given to a method which, although possibly less effective, requires minimum a priori information. /206

The effectiveness of a compression method can be determined by two means. First of all, this can be done analytically for a known mathematical model of a source; secondly, it can be performed experimentally by testing the compression algorithm in practice. Each of these methods has its advantages and disadvantages. Experimental study of the effectiveness of a compression algorithm is easier to perform, but can produce only a partial result for a given realization of the process, yielding no definitive information concerning the results of repeated testing. Knowledge of a mathematical model of the process observed allows theoretical analysis of the data compression system and production of exhaustive characteristics. However, as has been noted, a priori information on a source is usually quite modest, making it impossible to construct an adequate model. Furthermore, even if the necessary information is available, mathematical description may be so difficult that it will be practically impossible to work with the model. Due to these factors, models must be used which provide only a certain approximation of the original process, sufficient for purposes of theoretical study of compression systems.

Many works have studied the effectiveness of certain compression methods in relationship to fixed models [13, 16-20, 40]. In [13], for example, results are presented on modeling of the process of compression of the output signal of a shot noise generator, fed through filters to produce the desired frequency spectrum. This analog signal was then discretized and compressed. Three forms of signals with different high-frequency contents were studied. It was found that for the widest-band signal, the sampling compression factor was only 1.3. Reducing the high frequency content increased the compression factor somewhat; for a spectrum with predominantly low frequencies, the compression factor was 2.6. (The results here relate to a zero order prediction algorithm with a floating aperture with a width of ±5% of the entire scale, which was found to be most effective.)

This same work presents results of compression of actual telemetry data under the same conditions. The compression factor produced was approximately 100! This great divergence indicates that the model used was far from reality.

A great contribution to the theoretical analysis of data compression systems was made by Davison [16-19, 40]. For example, in [16], a theoretical study is performed of certain methods of prediction and approximate theoretical relationships are produced for a stable Gaussian model, which agree well with the results of statistical modeling. The compression factor produced in this case is also on the order of 2-3. A self-teaching prediction algorithm for a stable Gaussian sequence

with unknown characteristics is studied in detail in [17-19]. Approximate expressions are produced for the dispersion of prediction errors, which agree well with experimental data. The next step in theoretical analysis was examination of a Gaussian Markov process, but the compression factor produced for this model was also very low. Investigation of first and second order Markov processes with special distributions resulted in a sampling compression factor on the order of 5 [40]. /207

An excellent work was written by Erman [20], who produced analytic expressions of the mean and mean-square time between significant readings at the output of a compression system for a general Markov process and several methods of compression. Numerical results were produced only for a Wienerian process; it was shown that the results practically coincide for Wienerian and Gaussian Markov processes if the frequency of the initial readings is sufficiently high. Analysis of unstable processes is undoubtedly a great help to this work. Furthermore, consideration of not only the mean time between readings, but also the dispersion of this time, gives us a more complete conception of the effectiveness of the compression method. It is also interesting that the comparative effectiveness of Erman compression is determined not in relationship to an ordinary PCM system, as most authors do, but in relationship to a PCM system with a quantization step equal to the width of the aperture established in compression. The possibility of optimal interpolation during restoration is also considered. The correctness of selection of this system as a standard for comparison is somewhat debatable, but the fact of consideration of the possibility of optimal restoration and reduction of the accuracy of quantization in the standard PCM system is doubtless progressive.

Thus, the models used at the present time for analysis of data compression systems are far, as concerns compression factors, from actual telemetry data. Apparently, this is explained by the fact that most actual sources are essentially unstable and replacement of such a source by a model with averaged characteristics leads to reduction in the capabilities of compression and, as a result, to very great differences in values of compression factor. The introduction of significant instability to the model apparently can lead to rather good agreement between models and actual sources. When there is a serious difference between models and actual sources, it is apparently a good idea to set forth the question of correctness of utilization of such imperfect models for comparison of various compression methods, since this may lead to qualitatively incorrect conclusions. For example, in [13], when actual telemetry data were used, the first order prediction algorithm was the second most effective, the zero order

prediction algorithm with shift was the least effective. When actual data are replaced by a model, the relationship between them changes: all zero order algorithms make up a group significantly more effective than first order algorithms, which are the worst.

The problem of construction of such a model of a message source, allowing comparatively easy theoretical analysis of data compression systems and simultaneously providing good agreement with the results of experimental studies, of course, is quite complex. Up to the present time, unfortunately, no such model is known, but work on construction of new models for problems of analysis of the compression data is under way and, apparently, new and interesting results can be expected in the near future.

In spite of the fact that the problem of compression of data has been discussed in a comparatively large number of works, it is still impossible to answer a number of questions arising in the development of a specific system. In conclusion, let us list the most important of these problems, or, in other words, areas for further study.

One of the main difficulties is related to analysis and checking of existing compression algorithms. This results from the lack of good mathematical models of information sources in general and sources which actually exist in particular. The /20
widely used model of the stable Gaussian source was useful to some extent in the first stage of the study, but is now clearly insufficient. Practically all real sources of information are unstable, the events recorded are dependent, etc. Therefore, one necessary condition for further progress is introduction of nontrivial mathematical models of sources.

No less difficulties arise in evaluating the quality of reproduction of data received. Obviously, in the design of a compression system the algorithms for restoration and processing of information at the reception point must be considered. The effectiveness of a compression system, the expediency of using correcting coding and many other questions arising in planning are related to evaluation of the quality of restoration of data. Presently, in most cases either the mean square error criterion or various semi-intuitive considerations are used. However, even in these cases it is practically impossible to find answers to certain questions: how much should the quality of transmission be increased when data compression is used, should a correcting code by used, how can the frequency of initial readings be selected, etc.

Yet another significant gap is related to the lack of studies on development and analysis of flexible combined algorithms. In the simplest case, for example, we can use two "opposite" algorithms for compression and during each time interval select the one which is more effective. Flexible compression algorithms would allow us to get by with lower a priori values of source statistics and limit the possible loss (and not the gain) which could occur in certain realizations. Furthermore, in parallel with the development of flexible algorithms it is apparently expedient to develop specialized algorithms for the compression of scientific data, since each scientific experiment has its own set of unusual phenomena which are of the greatest value and should be revealed by the compression algorithm.

REFERENCES

1. Leman, E., Proverka statisticheskikh gipotez [Checking of Statistical Hypotheses], Moscow, Nauka Press, 1964.
2. Fano, R., Peredacha informatsii. Statisticheskaya teoriya svyazi [Information Transmission. The Statistical Theory of Communications], Moscow, Mir Press, 1965.
3. Sardinas, A., J. Patterson, "Necessary and Sufficient Condition for Unambiguous Expansion of Coded Messages," Kibernetic heskiy sbornik, No. 3, Foreign Literature Press, 1961.
4. Novik, D. A., Effektivnoye kodirovaniye [Effective Coding], Moscow, Energiya Press, 1965.
5. Meshkovskiy, K. A., N. Ye. Kirillov, Kodirovaniye v tekhnike svyazi [Coding in Communications Equipment], Svyaz' Press, 1966.
6. Elayes, P., "Coding with Prediction," Teoriya informatsii i yeye prilozheniya [Information Theory and its Applications], Moscow, Fizmatgiz Press, 1959, pp. 275-328.
7. Blokh, E. L., "Transmission of an Uneven Binary Sequence with an Even Code," Elektrosvyaz', 1959, No. 1, pp. 76-77.
8. Schwartz, D. V., R. C. Barker, Bit-Plane Encoding: A Technique for Source Encoding, IEEE Trans. on AES, 1966, No. 2, p. 4.
9. Kanefsky, M., Data Compression Using Bit-Plane Encoding. In 1966 Session of Goddard Summer Workshop, 1967, p. 1-21.
10. Fitingof, B. M., "Optimal Coding with Unknown and Changing Message Statistics," Problemy peredachi informatsii, 1960, Vol. 11, No. 2, pp. 3-11.
11. Blokh, E. L., "Sequential Coding," Problemy peredachi informatsii, 1960, No. 5, pp. 55-69.
12. Laemmel, A. E., Application of Lattice Ordered Semigroups to Codes and Finite State Transducers, Proc. Symp. Math. Theory of Automata, New York, 1962, pp. 241-256.
13. Medlin, J. E., Sampled Data Prediction for Telemetry Bandwidth Compression, IEEE Trans. on SET, 1965, March, p. 11.
14. Gardenhaire, L. W., Redundancy Reduction -- the Key to Adaptive Telemetry, Proc. Nat.'l Telemet. Conf., 1964.
15. Andrews, Davis and Schwartz, Adaptivnoye szhatiye dannykh [Adaptive Data Compression], TIIER Press, 1967, No. 55, p. 3.
16. Davisson, L. D., An Approximate Theory of Prediction for Data Compression, IEEE Trans. on I. T., 1967, Vol. 13, No. 2. /209
17. Davisson, L. D., The Adaptive Prediction of Time-Series, Proc. Nat'l Electronics Con., 1966.
18. Davisson, L. D., The Prediction Error of Stationary Gaussian Time Series of Unknown Covariance. IEEE Trans. on I. T., 1965, Vol. 11, No. 4.

19. Davisson, L. D., Theory of Adaptive Data Compression. In "Advance in Communication Systems," New York, Academic Press, 1966, Vol. 2.
20. Erman, Analiz nekotorykh sposobov szhatiya polosy chastot putem ustraneniya izbytochnosti [Analysis of Certain Methods of Compression of the Frequency Band by Elimination of Redundancy], TIIER Press, 1967, Vol. 55, No. 3.
21. Khochman, Katsman and Veber, Szhatiye polosy chastot televizionnykh signalov za schet sokrashcheniya izbytochnosti [Reduction of the Frequency Band of Television Signals by Elimination of Redundancy], TIIER Press, 1967, Vol. 5, No. 3.
22. Specht, D. F., Drapkin, D. E., Biomedical Data Compression, Proc. Nat'l Telemet. Conf., 1965.
23. Glover, E. B., Signal Redundancy Reduction Utilizing Slope Limiting Lines (Radiation Inc.), U. S. Patent No. 3311909, Published 28 March 1967.
24. Fukushima, K., Television Bandwidth Compression by Multimode Interpolation. NHK Laboratories Note, 1966, October, Tokyo, Japan.
25. Davisson, Teoreticheskiy analiz sistem szhatiya dannykh [Theoretical Analysis of Data Compression Systems , TIIER Press, 1968, Vol. 56, No. 2.
26. Balakrishnan, A. V., An Adaptive Nonlinear Data Predictor, Proc. Nat'l Telemet. Conf., 1962.
27. Palermo, C. J., R. V. Palermo, H. Horowitz, The Use of Data Omission for Redundancy Removal, Rec. Internat. Space Electronics and Telemetry Symp., 1965.
28. Kramer, G., Matematicheskiye metody statistiki [Mathematical Methods in Statistics], Foreign Literature Press-1948.
29. Eisenberger, I., E. C. Posner, Systematic Statistics Used for Data Compression of Space Telemetry, J. Am. Stat. Assoc., 1965, March, Vol. 60 pp. 97-113.
30. Anderson, T. O., I. Eisenberger, W. A. Lushaugh, E. C. Posner, Demonstration of a Quantile System for Compression of Data from Deep Space Probes, IEEE Trans. on AES, 1967, AES-3, pp. 57-65.
31. Medlin, J. E., The Prevention of Transmission Buffer Overflow in Telemetry Data Compressors, IEEE Trans. on COM, 1968, Febr., 16.
32. Hollenbaugh, R., A Study of Data Compression Buffering, Conf. on Adaptive Telemetry, 1967.
33. Lynch, T. J., Sequence Time Coding for Data Compression, Proc. IEEE (Letters), 1966, October, 54.
34. Lynch, T. J., Performance Measures for Compressed and Coded Space Telemetry Systems, IEEE Trans. on AES, 1967, Vol. 3, No. 5.
35. Davisson, L. D., The Effect of Channel Errors on Data Compression, WESCON Techn. Papers, 1967, N 3, 6/2.

36. Raga, G. L., A Unified Approach to Digital Television Compression, Nat'l Telemet. Conf., 1965.
37. Babkin, V. F., Yu. M. Shtar'kov, "The Energy Effectiveness of Correcting Codes in Certain Communications Channels," Peredacha tsifrovoy informatsii po kanalam s pamyat'yu [Transmission of Digital Information over Channels with Memory], Moscow, Nauka Press, 1970.
38. Fink, L. M., Teoriya peredachi diskretnykh soobshcheniy [Theory of Transmission of Discrete Messages], Moscow, Sovetskoye radio Press, 1963.
39. Shannon, C., "The Mathematical Theory of Communications," Raboty po teorii informatsii i kibernetike [Works on Information Theory and Cybernetics], Foreign Literature Press, 1963.
40. Davisson, L. D., Data Compression Using Straight Line Interpolation, IEEE Trans. on I. T., 1968, Vol. 14, No. 2.

N75 11010

CALCULATION OF THE EFFECTIVENESS OF A GAMMA TELESCOPE

Yu. A. Rylov

This work contains a calculation of the effectiveness of recording of gamma quanta by a gamma telescope. The device in question [1] is a modification of a gamma telescope described earlier [2] and is designed for recording and determination of the energy of gamma quanta with energy $E \geq 50$ MeV outside the atmosphere.

Calculation was performed by computer by the method of statistical testing. Our calculations differed from other similar calculations [3] in the respect that in place of the effectiveness of recording, we calculated the effectiveness matrix. Furthermore, we considered the isotropic nature of the stream of gamma quanta reaching the gamma telescope. Considera- /210
tion of isotropy revealed a curious effect, in that the low energy gamma quanta are recorded relatively more effectively. This occurs because they are recorded from a wider solid angle than higher energy gamma quanta.

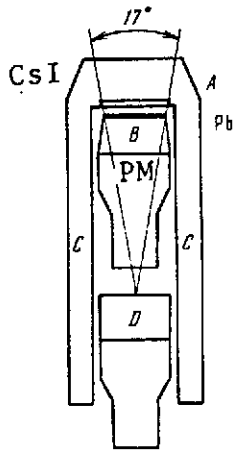


Figure 1. Schematic Diagram of Gamma Telescope

The problem is solved of determining the effectiveness of recording of gamma quanta and determination of their energies by a gamma telescope as shown on Figure 1. The operating principle of the device is as follows: a gamma quantum striking the telescope from above converts an electron-positron pair either in the CsI crystal (1.8 g/cm^2) or in the lead plate (2.26 g/cm^2). The electron and positron, moving downward, pass through scintillation counter B and form an electron-photon avalanche in lead glass D (2.5 t units), the radiator of Cerenkow counter D. The intensity of the Cerenkow radiation created by the charged particles is used to judge the energy of the gamma quantum. The gamma telescope is surrounded

by a cap consisting of a scintillator, which, together with the PM, forms anticoincidence counter C, which is triggered when a charged particle passes through it. The coincidence circuit is as follows: $AB\overline{C}D$ for the first channel and $\overline{A}B\overline{C}D$ for the second channel. The letters represent counters (A is the Cerenkow counter, reporting the appearance of a charged particle in the CsI crystal). The overline above a letter indicates anticoincidence. Switching of the first channel indicates conversion in the CsI crystal, switching of the second channel -- conversion in the lead converter. Each time a gamma quantum is recorded, a signal of a certain amplitude u develops in the counter. Depending on the amplitude of the signal in relationship to certain thresholds γ_1, γ_2, γ_3, γ_4, γ_5, the γ quantum is placed in a certain experimental interval. For example, if amplitudes u of the signal satisfies the inequality

$$\gamma_i \leqslant u < \gamma_{i+1}, \tag{1}$$

the gamma quantum is considered to be recorded in the ith interval $\Delta\gamma_i$. During the process of an experiment, the number of gamma quanta n_i (i = 1, 2, 3, 4, 5) in the various intervals is recorded.

The problem is to use the measured values of n_i to determine the spectrum of gamma quanta, i.e., determine the mean number of gamma quanta N_i in the energy intervals ΔE_i corresponding to the electronic intervals $\Delta\gamma_i$. In principle, the ΔE_i intervals may be selected any magnitude. However, it is practically more convenient to select ΔE_i such that the gamma quanta from energy interval ΔE_i are recorded primarily in electronic interval $\Delta\gamma_i$. The energy thresholds were selected as follows: E_1 = 50 MeV, E_2 = 90 MeV, E_3 = 146 MeV, E_4 = 465 MeV, E_5 = 2000 MeV, E_6 = 16,000 MeV. Intervals ΔE_i are defined by the relationship

$$\Delta E_i = E_{i+1} - E_i. \tag{2}$$

The gamma quanta with energies less than E_i are considered to be in energy interval ΔE_0. The gamma quanta not recorded were considered to be recorded in energy interval $\Delta\gamma_0$. The number of recorded events n_0 corresponding to interval $\Delta\gamma_0$ is considered unknown.

To determine N_i (i = 1, 2, ..., 5) using the measured numbers n_i (i = 1, 2, ..., 5), it is sufficient to know the statistical characteristics of the gamma telescope. These /21

characteristics are represented by matrix χ_{ik}, i = 0,1, ..., 5, k = 1, 2, ..., 5. χ_{ik} is the probability that a gamma quantum striking the CsI crystal with its velocity vector directed within a certain solid angle Ω and energy within the limits of the energy interval ΔE_i, will be recorded in electronic interval $\Delta\gamma_k$. We will call χ_{ik} the effectiveness matrix. Of course, χ_{ik} depends on the design of the gamma telescope and the selection of the electronic and energy thresholds. However, when the effectiveness matrix is known, it is a characteristic of the gamma telescope. Knowledge of χ_{ik} is sufficient to determine the spectrum N_i from the measured values of n_k. The details of the design of the gamma telescope are of no significance. Effectiveness matrix χ_{ik}, the energy and electronic thresholds E_i and γ_i are all that must be known concerning the gamma telescope to use it. We note that χ_{ik}, generally speaking, depends on the distribution of gamma quanta by energy within each energy interval ΔE_i. This dependence can be considered, assuming the spectrum of the gamma quanta to be exponential within each interval ΔE_i and calculating χ_{ik} for two different values of spectral exponent.

The task of determining N_i from n_k is performed as follows. Suppose N_i is the mean number of gamma quanta in energy interval ΔE_i, and n_k (k = 0, 1, ..., 5) is the mean number of events recorded in electronic interval $\Delta\gamma_k$; then, according to the definition of matrix χ_{ik}, we have

$$n_k = \sum_{\beta=0}^{5} N_\beta \chi_{\beta k}. \tag{3}$$

Suppose matrix A_{kl} is the matrix which is the inverse of χ_{ik}, i.e.,

$$\sum_{k=1}^{5} \chi_{ik} A_{kl} = \delta_{il}, \quad i, k, l = 1, 2, \ldots, 5. \tag{4}$$

Then, multiplying relationship (3) by A_{kl}, using expression (4) we produce

$$\sum_{k=1}^{5} N_0 \chi_{0k} A_{kl} + N_l = \sum_{k=1}^{5} n_k A_{kl}, \quad l = 1, 2, \ldots, 5. \tag{5}$$

n_0 is unknown and using the five equations of (5) it is impossible to determine six quantities N_β (β = 0, 1, ..., 5) unambiguously. We must make an additional assumption relative to N_0. We extrapolate the spectrum beyond the lower threshold E_1. The number N_0 is related to N_1 by the formula

$$N_0 = BN_1, \tag{6}$$

where B is a certain quantity which depends on the spectrum of the gamma quanta below threshold E_1 and on the sizes of intervals ΔE_0 and ΔE_1. In particular, with the thresholds selected, E_0 = 25 MeV, E_1 = 50 MeV, E_2 = 90 MeV, the value of B is expressed by the relationship

$$B = \frac{E_0^{-\alpha} - E_1^{-\alpha}}{E_1^{-\alpha} - E_2^{-\alpha}},$$

where α is the exponent of the spectrum. If α changes within limits $1 < \alpha < 2$, B varies from 2.22 to 4.3. Since the spectral exponent is not precisely known, the value of B is not precisely defined and is determined after the spectrum is measured; before that, its approximate value is used. Let us substitute relationship (6) into equation (5). Solving resultant 5 equations/212 produced for N_1, we produce

$$N_1 = \frac{\sum_{s=1}^{5} n_s A_{s1}}{1 + B\sum_{k=1}^{5} \chi_{0k} A_{k1}},$$

$$N_l = \sum_{s=1}^{5} n_s A_{sl} - \frac{B\sum_{s=1}^{5} n_s A_{s1} \sum_{j=1}^{5} \chi_{0j} A_{jl}}{1 + B\sum_{k=1}^{5} \chi_{0k} A_{k1}}, \tag{7}$$

$$l = 2, 3, 4, 5,$$

Formulas (7) solve the problem of determining the energy spectrum of the gamma quanta with respect to the spectrum of the apparatus and contain quantity B, which is known approximately.

They make it possible to calculate N_1 for various values of B, corresponding to various spectral indicators, and then to produce a more precise value of B from the resultant spectrum and, using this precise value, to refine the value of the spectral indicator, etc. In practice, relationship (6) is a method of consideration of threshold effects (see also [4, 5]).

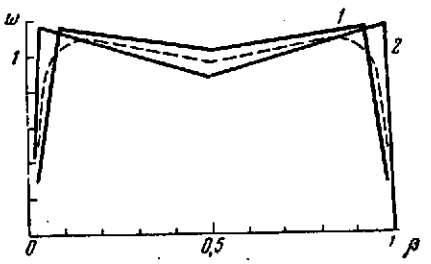


Figure 2. Distribution of Energy Between Electron and Positron. β, Share of Energy Carried by Electron; ω, Probability Density that Electron Will Carry Away Energy β

The problem of calculating the effectiveness of a gamma telescope is reduced to calculation of effectiveness matrix x_{ik}. The calculation has been done by computer by a Monte Carlo method. Essentially, modeling of the operation of a gamma telescope was performed. The following assumptions were made in the calculation.

1. The gamma telescope cannot record gamma quanta with energy of less than 25 MeV. According to this, threshold E_0 was selected as E_0 = 25 MeV and ΔE_0 corresponded to the 25-50 MeV interval.

2. The gamma quanta strike the gamma telescope isotropically.

3. Only gamma quanta striking the gamma telescope at an angle to the axis of the instrument of less than $\theta_0 = 0.3$ ($\approx 17°$) could be recorded. The corresponding solid angle $\Omega = 2\pi(1 - \cos\theta_0) = 0.28$.

4. A gamma quantum can generate an electron-positron pair either in the CsI crystal or in the lead plate. The electron and positron of the pair fly in the direction of the gamma quantum. The distribution of energy between the electron and positron was fixed by the curves on Figure 2. Curve 1 approximated the distribution of energies in a pair for a gamma quantum with an energy E < 1000 MeV, curve 2 -- for E > 1000 MeV. The dotted line illustrates the curve for 100 MeV, taken from [6]. All other forms of action of gamma quanta with matter were ignored.

5. It was assumed that the electron and positron of the pair, after their formation, interact with matter. This interaction is manifested only in ionization losses and repeated

scattering in the CsI, lead and plate (5.3 g/cm^2) of counter D, and scattering in plate D and in the PM located beneath it (2 g/cm^2) was effectively replaced by scattering in the lead 1.1 g/cm^2. Losses to radiation of the electron-positron pair formed were ignored. The basis for this was as follows. When /213 a gamma quantum is converted in lead, the electron-positron pair is formed at an average depth of 1.1 g/cm^2 from the upper portion of the lead converter. The thickness of the remaining portion of the lead, the plastic scintillator and the PM equipment is about 0.4 radiation units. As an electron passes through 0.4 t units of matter, it losses an average of 30% of its energy to radiation. However, the overwhelming portion of the radiation is emitted in the direction of motion of the electron (mean square angle ≈1° for an electron with an energy of 100 MeV) and causes formation of an avalanche in the Cerenkow counter radiator. On the average, the effect produced is as if the avalanche began not at the upper portion of the radiator D, but 0.4 radiation units higher. Consideration of radiation of electrons would lead on the average to an increase in the indications of counter D at high electron energies (+6% with an electron energy of 100 MeV) and a decrease of the indications at low energies (-5% with an electron energy of 25 MeV). This type of error can be ignored, keeping in mind that at high energies they lead only to a slight error in determination of the energy, and that ignoring losses to radiation significantly simplifies calculation.

At low energies of the gamma quantum $E_\gamma \approx 60$ MeV, the losses of energy become significant. Each particle in the electron-positron pair born in the lead converter at a depth of 1.1 g/cm^2 loses about 15 MeV on the average to ionization. With a gamma quantum energy of 60 MeV, in 50% of cases, only one electron reaches the radiator of the counter D. This electron has an energy of 30-45 MeV and is strongly scattered. The mean square scattering angle θ is $\theta \approx 0.4$ (≈23°). Absorption of one particle results in effective recording of gamma quanta striking at an angle on the order of θ to the axis of the gamma telescope. When two particles are scattered, the probability of scattering of both particles in the same direction is low; therefore, one is scattered to the side and energizes anti-coincidence counter C. Single-electron scattering increases the effectiveness of recording of gamma quanta at low energy by increasing the effective solid angle from which gamma quanta can be recorded. Consideration of losses to radiation would increase the effectiveness of recording of 60-MeV gamma quanta by about 7%, since electrons with energies of over 15 MeV will be absorbed completely. On the other hand, consideration of radiation leads to the fact that the intensity of radiation in radiator D will be an average of 6% less, which slightly decreases the effectiveness of recording due to the closeness

to the threshold. These two factors compensate each other to some extent.

The ionization losses were assumed constant, 1.5 MeV/gcm^{-2} in the CsI and lead, 1.8 MeV/gcm^{-2} in plate B and the PM. Multiple scattering of electrons and positrons in the CsI and lead was calculated on the assumption that the scattering layer was thin. The method of Sneider and Scott [7] was used. The electrons with energies less than 3 MeV were considered stopped [8].

6. It was assumed that when an electron struck the lead glass (2.5 t units) of the Cerenkow counter, an electron-positron avalanche was formed, and the amplitude of the signal at the output of counter D was proportional to the number of charged particles in the avalanche. The curve on Figure 3 was used to determine the proportionality factor. This curve was produced by measurement of the intensity of relativistic μ-mesons at sea level using the gamma telescope. The abcissa shows the amplitude of the signal of the counter D in relative units, the ordinate shows the relative number of events causing this signal amplitude. The position of the maximum on the curve was selected as the first electronic threshold $\gamma_1 = 1$. The other thresholds were selected relative to the first threshold as follows: $\gamma_2 = 1.8$, $\gamma_3 = 3.5$, $\gamma_4 = 8.25$, $\gamma_5 = 18.9$, $\gamma_6 = 45$. It was assumed that a μ-meson passes through the lead /214 glass without forming an avalanche, and that the dispersion in the amplitude of the signal in the Cerenkow counter results primarily from fluctuations in the electronic apparatus, approximated by a normal distribution with dispersion

$$D(u) = 0{,}045 \cdot u, \tag{8}$$

where u is the signal amplitude of counter D in relative units, in which the maximum of the curve of Figure 3 corresponds to u = 1. The signal in these units, arising in counter D when the radiator was struck by an electron with energy E, was assumed equal to the ratio of the area beneath the corresponding cascade curve and the area beneath the curve $n_1 = 1$ (see Figure 4). The cascade curves were taken from [9] for the case when lead is the absorber. Figure 4 shows t, the thickness of the absorber in radiation units, on the abcissa, and n_e, the mean number of electrons (and positrons) with energy greater than E = 0 at depth t, on the ordinate. The fluctuation resulting from the operation of the apparatus is added to the signal. Fluctuations in the cascade avalanches themselves were not considered.

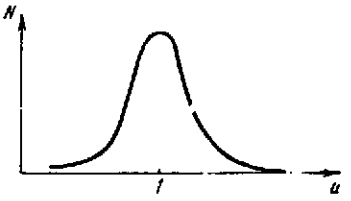


Figure 3. Counting Rate as a Function of Counter Amplitude D

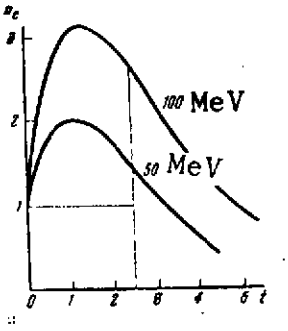


Figure 4. Cascade Curves for Lead

Calculation of the effectiveness matrix was performed as follows. For each energy interval ΔE_i ($i = 0, 1, \ldots, 5$), M_i events of random incidence of gamma quanta in the gamma telescope were recorded. In each of the intervals ΔE_i, the energy spectrum of the gamma quanta was exponential with integral exponent α. Two cases were studied, $\alpha = 1$ and $\alpha = 2$. When the Monte Carlo method was used, the following quantities were fixed at random: energy of gamma quantum, two incident angles of the gamma quantum, its impact point and the depth at which the electron-positron pair was converted. If the pair was converted in the CsI crystal, the energy of one particle of the pair, two scattering angles of each of the two charged particles leaving the CsI, two scattering angles upon leaving the lead were fixed at random (scattering in plate B and the PM were effectively considered by scattering in an additional layer of lead 1.1g/cm^2 thick), and the fluctuation in the electronic apparatus was assigned a random quantity. If the pair was converted in the lead, the following were fixed at random: energy of one particle of the pair, two angles of departure from the lead for each particle and fluctuation in the electronic apparatus. The fluctuation was fixed separately for each particle. Thus, for each gamma quantum it was calculated whether it would be recorded, and if so, in which interval $\Delta\gamma_i$. The result was effectiveness matrix χ_{ik}.

Calculations showed that matrix χ_{ik}, corresponding to various integral exponents of the spectrum $\alpha = 1$ and $\alpha = 2$, differs only within the limits of statistical error of the calculation. Therefore, the effectiveness matrices corresponding to $\alpha = 1$ and $\alpha = 2$ were averaged. The following effectiveness /215
matrices χ_{ik} (in percent) were produced for conversion of a pair in the CsI and in the lead ($\theta_0 = 0.3$):

$$\chi_{ik\,CsI} = \begin{Vmatrix} 0 & 0 & 0 & 0 & 0 \\ 0{,}65 & 0{,}10 & 0 & 0 & 0 \\ 0{,}35 & 1{,}07 & 0{,}05 & 0 & 0 \\ 0 & 0{,}50 & 0{,}95 & 0 & 0 \\ 0{,}05 & 0{,}4 & 0{,}95 & 2{,}3 & 0 \\ 0 & 0 & 0 & 1{,}6 & 2{,}05 \end{Vmatrix}$$

$$i = 0, 1, \ldots, 5; \quad k = 1, 2, \ldots, 5.$$

$$\chi_{ikPb} = \begin{Vmatrix} 0 & 0 & 0 & 0 & 0 \\ 1{,}57 & 0{,}325 & 0 & 0 & 0 \\ 0{,}40 & 2{,}00 & 0{,}1 & 0 & 0 \\ 0{,}3 & 0{,}55 & 2{,}1 & 0 & 0 \\ 0 & 0 & 1{,}55 & 3{,}4 & 0 \\ 0 & 0{,}005 & 0{,}35 & 2{,}25 & 2{,}55 \end{Vmatrix}$$

$$i = 0, 1, \ldots, 5; \quad k = 1, 2, \ldots, 5.$$

In the χ_{ik} matrices, the first subscript numbers the row and represents the number of the energy interval ΔE_i, the second numbers the column and represents the number of electronic interval $\Delta\gamma_k$. The statistics here are as follows: for intervals ΔE_0, ΔE_1, ΔE_2, $M_0 = M_1 = M_2 = 4000$ cases were run, for intervals ΔE_3, ΔE_4, ΔE_5 -- $M_3 = M_4 = M_5 = 2000$ cases were run of gamma quanta at the gamma telescope for each energy interval. As we have noted, with low energies of gamma quanta, the effective solid angle from which the gamma quanta are recorded may be high. In order to check this, a calculation was performed for intervals ΔE_1 and ΔE_2. The solid angle from which gamma quanta were recorded was increased by 5.3 times ($\theta_0 = 0.7$). The summary effectiveness of recording in interval ΔE_1 was increased by 30%, in ΔE_2 it was practically unchanged. The corresponding summary effectiveness matrix χ_{ik} for the two channels (in percent), converted to angle $\theta_0 = 0.3$, is as follows:

$$\bar{\chi}_{ik} = \begin{Vmatrix} 0 & 0 & 0 & 0 & 0 \\ 2{,}85 & 0{,}52 & 0 & 0 & 0 \\ 0{,}6 & 2{,}85 & 0{,}15 & 0 & 0 \\ 0{,}3 & 1{,}05 & 3{,}05 & 0 & 0 \\ 0{,}05 & 0{,}4 & 2{,}50 & 5{,}7 & 0 \\ 0 & 0{,}05 & 0{,}35 & 3{,}85 & 4{,}6 \end{Vmatrix} \begin{pmatrix} 0 \\ 3{,}38 \\ 3{,}6 \\ 4{,}4 \\ 8{,}65 \\ 8{,}85 \end{pmatrix}$$

$$i = 0, 1, \ldots, 5; \quad k = 1, 2, \ldots, 5.$$

The statistics here are as follows: $M_0 = 5500$, $M_1 = 5500$, $M_2 = 5500$, $M_3 = M_4 = M_5 = 2000$. (The last column presents the

summary effectiveness of recording of particles from interval ΔE_i.) The inverse effectiveness matrix A_{ik}, defined by relationship (4), is as follows:

$$A_{ik} = \begin{Vmatrix} 36.2 & -6.25 & 0 & 0 & 0 \\ -7 & 34.2 & -1.68 & 0 & 0 \\ 0 & -11.1 & 33.3 & 0 & 0 \\ 0 & 0 & -14.4 & 17.5 & 0 \\ 0 & 0 & 9.3 & -14.3 & 22.7 \end{Vmatrix}$$
$$i = 1, 2, \ldots, 5; \quad k = 1, 2, \ldots, 5. \tag{9}$$

The first subscript numbers the rows, the second numbers the columns. In calculating matrix A_{ik}, we replaced small elements far from the main diagonal with zeros. /216

The fact that all elements χ_{ok} (k = 1, 2, ..., 5) are equal to zeros simplifies the task. Formulas (7) become

$$N_k = \sum_{s=1}^{5} n_s A_{sk}, \quad k = 1, 2, \ldots, 5. \tag{10}$$

Here n_s is the number of events recorded in intervals $\Delta\gamma_s$ in time T of operation of the device, N_k is the number of gamma quanta with energies in interval ΔE_k flying in at solid angle Ω_0 = 0.28 sterad and striking the CsI crystal in time T.

Formula (10), together with matrix (9), solves the problem of determination of the intensity of gamma quanta from the events recorded by the gamma telescope.

Let us present for comparison the diagonal inverse effectiveness matrix which is usually used. It considers only the fact of recording, not the fact that a gamma quantum from interval ΔE_i might be recorded not only in electronic interval $\Delta\gamma_i$, but in other intervals as well. This matrix is as follows:

$$\eta = \begin{Vmatrix} 29.8 & & & & \\ & 27.8 & & & \\ & & 23.6 & & \\ & & & 11.55 & \\ & & & & 11.3 \end{Vmatrix} \tag{11}$$

Let us estimate the statistical error in determination of N_i. In this estimation, we will assume $N_i \gg 1$ and the statistical error is slight. Under these assumptions, probability density $W(n/N)$ of recording $n = \{n_1, n_2, \ldots, n_5\}$ gamma quanta when $N = \{N_1, N_2, \ldots, N_5\}$ gamma quanta had struck the telescope is given approximately by the normal distribution

$$W(n/N) = A(N) \exp\left\{-\sum_{i=1}^{5} \frac{(n_i - \bar{n}_i)^2}{2\bar{n}_i}\right\}. \tag{12}$$

It is considered here that the mean value of n_i

$$\bar{n}_i = \sum_{k=1}^{5} N_k \chi_{ki}, \tag{13}$$

and the dispersion is

$$\overline{n_i n_k} - \bar{n}_i \bar{n}_k = \delta_{ik} \bar{n}_i. \tag{14}$$

Factor $A(N)$ is defined by the normalization condition

$$\int_0^\infty \ldots \int W(n/N)\, dn_1 \ldots dn_5 = 1. \tag{15}$$

We are interested in the distribution of N with fixed n. This distribution differs from $W(n/N)$ by the normalizing factor $B(n)$.

$$\mathcal{P}(N/n) = B(n)\, W(n/N), \tag{16}$$

which is determined from the normalization condition

$$\int_0^\infty \ldots \int \mathcal{P}(N/n)\, dN_1 \ldots dN_5 = 1. \tag{17}$$

With sufficiently large n_i, distribution (16) will be narrow and we can approximate it with the normal distribution /217

$$\mathcal{P}(N/n) = D(n) \exp\left\{-\sum_{i,k=1}^{5} \frac{1}{2} B_{ik} (N_i - \bar{N}_i)(N_k - \bar{N}_k)\right\}, \tag{18}$$

where the exponential function is the expansion with respect to powers of N_i of the exponential function of (12). Here, we limit ourselves to terms of the expansion of not over second order with respect to N and ignore the dependence of coefficient A(N) on N. This is permissible with sufficiently large $\bar{N}_i$.

$$\bar{N}_i = \sum_{k=1}^{5} n_k A_{ki}, \tag{19}$$

where A_{ki} is matrix (4).

For B_{ik}, we can easily produce the expression

$$B_{ik} = \sum_{l=1}^{5} \frac{\chi_{il}\chi_{kl}}{n_l}. \tag{20}$$

We now produce the following expression for the dispersion of distribution (16)

$$\overline{(N_i - \bar{N}_i)(N_k - \bar{N}_k)} = \sum_{l=1}^{5} A_{li} A_{lk} n_l. \tag{21}$$

The mean square error in determination of N_i is given by the expression

$$\Delta N_i = \sqrt{\overline{(N_i - \bar{N}_i)^2}} = \sqrt{\sum_{l=1}^{5} |A_{li}|^2 n_l}. \tag{22}$$

We can see from Table (9) that the greatest contribution to sum (22) is given by the term with $\ell = i$. The contribution of other terms, at least for i = 1, 2, 3, is insignificant.

Formula (22) gives us the absolute error in determination of N_i. It is assumed here that matrix A_{ik} is known quite precisely, that the error results only from fluctuations in distribution n_i. Let us now calculate the error in determination of N_i, resulting from imprecise determination of matrix A_{ik}. This inaccuracy results from the fact that in calculating the effectiveness matrix χ_{ik}, a finite number of cases of incidence of gamma quanta at the gamma telescope was run. Let us represent the number of cases run by the letter $M = \{M_1, M_2, \ldots, M_5\}$. Suppose when M_i gamma quanta strike the gamma telescope from interval ΔE_i, n_{ik} events were recorded in $\Delta\gamma_k$. The matrix of effectiveness is then determined by the relationship

$$p_{ik} = \frac{n_{ik}}{M_i}, \tag{23}$$

where

$$\bar{p}_{ik} = \chi_{ik}. \tag{24}$$

According to relationships (14) and (23), the dispersion of the effectiveness matrix is

$$\left.\begin{aligned} &\overline{\Delta p_{ik}\Delta p_{lj}} = \delta_{il}\delta_{kj}\frac{p_{ik}}{M_i}, \\ &\Delta p_{ik} = p_{ik} - \chi_{ik}, \quad \overline{\Delta p_{ik}} = 0. \end{aligned}\right\} \tag{25}$$

The error in determination N_i arises due to the fact that in place of $A_{ik}(\chi)$ (matrix inverse to χ_{ik}), we use matrix $A_{ik}(p)$, inverse to p_{ik}. The relationship between them is given by the formula /218

$$A_{ik}(p) = A_{ik} - \sum_{l,s=1}^{5} A_{il}\Delta p_{ls}A_{sk} + \sum_{j,l,s,r=1}^{5} A_{ij}\Delta p_{jl}A_{ls}\Delta p_{sr}A_{rk} + \ldots \tag{26}$$

The right portion of equation (26) is the expansion of $A_{ik}(p)$ with respect to powers $\Delta p_{ik} = p_{ik} - \chi_{ik}$. A_{ik} in the right portion of (26) represents $A_{ik}(\chi)$. In producing the expansion of (26), we used the identity

$$dA_{ik} = -\sum_{l,s=1}^{5} A_{il}\, dp_{ls}A_{sk}. \tag{27}$$

To determine the dispersion of distribution N_i, we use the relationship

$$N_i = \sum_{l=1}^{5} n_l A_{li}(p). \tag{28}$$

Substituting expansion (26) into (28) and using the mean values of relationship (25) for the calculation, we can calculate the dispersion of distribution N_i. Calculation gives us (with an accuracy of Δp_{ik} of not over second power)

$$\Delta N_i = \sqrt{\overline{(N_i - \overline{N}_i)^2}} = \sqrt{\sum_{l,k=1}^{5} N_l^2 |A_{ki}|^2 p_{lk} M_l^{-1}}. \tag{29}$$

The summary mean square error in determination of N_i, resulting both from the statistics and the inaccuracy in determination of matrix A_{ik}, is given by the relationship

$$\Delta N_i = \sqrt{\sum_{l,k=1}^{5} N_l^2 |A_{ki}|^2 \chi_{lk} M_l^{-1} + \sum_{l=1}^{5} |A_{li}|^2 n_l}. \tag{30}$$

Thus, the problem of determination of the true spectrum N_i of the gamma quanta from the measured apparatus spectrum n_k is solved by formula (10) with matrix A_{ik} (9). The mean square error in determination of N_i is given by formula (30).

We would like to take this opportunity to thank L. S. Bratolyubova, L. F. Kalinkin, Ye. A. Pryakhin and I. L. Rozental' for discussing with us the problems related to calculation of the effectiveness of the gamma telescope.

REFERENCES

1. Bratolyubova-Tsulukidze, L. S., N. L. Grigorov, L. F. Kalinkin, A. S. Melioranskiy, Ye. A. Pryakhin, I. A. Savenko, V. Ya. Yufarkin, Kosmicheskiye issledovaniya, 1970, No. 8, p. 136.
2. Grigorov, N. L., L. F. Kalinkin, A. S. Melioranskiy, V. Ye. Nesterov, Ye. A. Pryakhin, I. A. Sysoyev, I. V. Estulin, Kosmicheskiye issledovaniya, 1967, No. 5, p. 124.
3. Dennis, B. R., G. H. Share, Nuclear Instruments and Methods, 1967, Vol. 50, No. 100.
4. Murzin, V. S., Izv. AN SSSR, seriya fiz., 1965, Vol. 29, No. 1644.
5. Rozental', I. L., V. V. Shestakov, Izv. AN SSSR, seriya fiz., 1965, Vol. 29, No. 1751.
6. Rossi, B., Chastitsy bol'shikh energiy [High Energy Particles], Moscow, GITTL Press, 1955.
7. Snyder, H. S., W. T. Scott, Phys. Rev., 1949, Vol. 76, No. 220.
8. Wilson, R. R., Phys. Rev., 1952, Vol. 86, No. 261.
9. Samosudov, B., Candidates Dissertation, Scientific Research Institute for Nuclear Physics, Moscow State University, 1960.

N75 11011

THE PROBLEM OF CONTROLLING THE DIRECTIVITY OF LOW-DIRECTIVITY ANTENNAS ON SPACECRAFT IN FLIGHT

V. I. Mashkov and B. A. Prigoda

Reliable communications with objects in space as they maneuver relative to the ground station is achieved by installing complex antenna systems on the spacecraft including, generally, several antennas, switching, matching and other high frequency elements which solve the problem but greatly complicate the entire on-board electronic system. /219

Lattice structures are currently being used, and for various specific cases these structures permit regulation of the form of the radiation pattern and directivity in the direction of the ground communications station. However, the practical application of arrays on spacecraft is limited by a number of factors, including the complexity of the switching system, large size and weight. There is therefore interest in the creation of antennas allowing adjustment of the radiation pattern and directivity in flight without using complex switching and matching circuits.

As concerns a certain class of conical log-periodic helical antennas, this adjustment of the radiation pattern can be achieved by changing the parameters of the helix, for example, its lead angle, in flight.

This report presents the results of certain experimental tests performed with log-periodic conical helical antennas to determine the possibilities for regulation of their radiation patterns without changing the structure of the current-carrying elements, supply connections, etc.

The regulated element selected consisted of circular metal discs and rings. Each disc (ring) was located at the vertex of a conical (or hemispherical) log-periodic helix, as shown on Figure 1. Selection of the distance from the vertex of the helix to the disc (ring) and the diameter of the regulating element can be used to regulate the radiation pattern without changing the polarization and impedance characteristics of the antenna.

Figure 2 shows the radiation patterns of a hemispherical double-helix antenna with a base diameter of $D/\lambda = 0.6$.

Figure 2 also shows for comparison the radiation pattern of the same antenna in the initial state and with the regulation disc at its vertex. The initial radiation pattern is shown by the solid line.

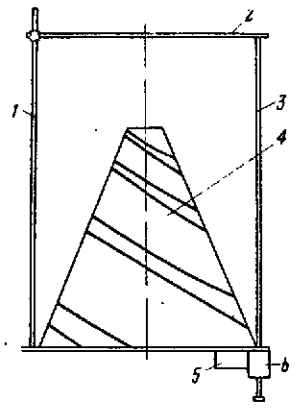


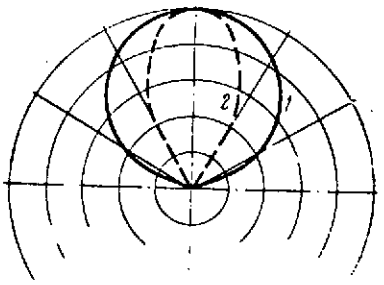
Figure 1. Installation of a Regulating Disc. 1, Guide (insulator); 2, Disc; 3, Pusher (insulator); 4, Antenna; 5, Motor; 6, Reducer

Gradual conversion of a solid metal disc to a ring by cutting increasingly large apertures in its center causes deformation of the radiation pattern, as is shown by Figure 3. The radiation pattern is converted from a single-lobe pattern with its maximum oriented along the axis of the antenna to a funnel-shaped radiation pattern with a gap in the axial direction. /220
The effect of changing radiation pattern is similar for hemispherical and conical double helices.

Obviously, the installation of a device such as a series adjustable diaphragm (such as those used in camera lenses) would allow smooth regulation of the radiation pattern and directivity of the antenna by changing the diameter of the internal aperture.

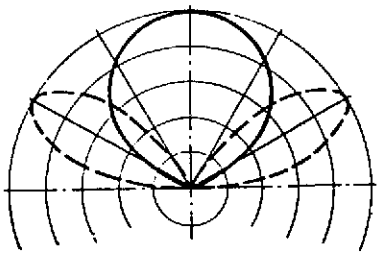
In those cases when the problem of maintaining elliptical polarization is not important, the capabilities of these devices are even broader, since movement of a solid disc along the axis of the antenna can regulate the width of the radiation pattern over broad limits. An example of this is shown on Figure 4. The effect of changing shape of the radiation pattern observed requires further study. However, even the first results indicate that this method of adjustment of the radiation pattern is promising and can be of practical use.

The authors consider it their duty to express their gratitude to N. F. Fomin, N. M. Galkin, M. B. Danilov, A. I. Ilyashenko, I. N. Pozdnyakov, Z. M. Solodova and others, who took part directly in the experimental studies.



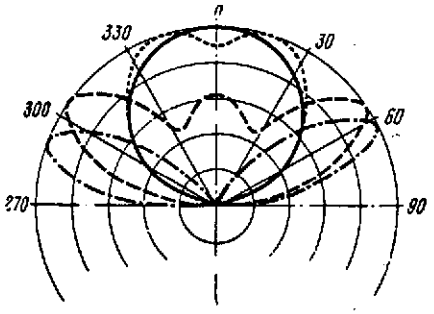
2

Figure 2. Radiation Pattern of Hemispherical Double Helix
1, Pattern without Disc; 2, Pattern with Disc



3

Figure 3. Influence of Rings on Radiation Pattern



4

Figure 4. Deformation of Radiation Pattern as Disc is Moved along Antenna Axis

REFERENCE

1. Benenson, L. S., "Slightly Directional Broad Band Antennas," Sovremennyye problemy antenno-volnovodnoy tekhniki [Modern Problems of Antenna and Waveguide Technology], Moscow, Nauka Press, 1967, pp. 202-214.

N75 11012

SOME PROBLEMS OF THE DESIGN OF HIGHLY DIRECTIONAL SPACECRAFT ANTENNAS

B. A. Prigoda

The primary parameter determining the information capacity of a spacecraft-Earth communications line and characterizing the potential power capabilities of the on-board spacecraft apparatus and, in particular, its antenna systems, is the density of the stream of radio wave energy at the point of reception, which can be expressed by the coupling equation

$$\Pi = P_b G_b \gamma / 4\pi r^2. \qquad (1)$$

Here P_b is the power of the on-board transmitter; r is the distance from the spacecraft to the reception point; G_b is the gain of the on-board transmitting antenna; γ is a factor considering the total attenuation of energy due to absorption and scattering in the medium.

In designing a communications line, the power of the useful signal at the input of the receiver (P_r) equal to the product of the energy flux density times effective area of the antenna (A_{ef}), related to the directional gain of the antenna by the formula

$$A_{ef} = G_r \lambda^2 / 4\pi$$

is of interest.

As a result, the following expression determines the power at the input of the receiver for propagation through free space

$$P_r = P_b G_b G_e \lambda^2 / (4\pi r)^2, \qquad (2)$$

where G_e is the directional gain of the ground station receiving antenna; λ is the wavelength

In the case of isotropic antennas ($G_b = G_e = 1$), the ratio P_r/P_b, characterizing the primary losses in the communications line according to expression (2), will be

$$(P_r/P_b) = \frac{\lambda^2}{(4\pi r)^2} = \frac{C^2}{(4\pi r f)^2}$$

or in decibels

$$\Gamma = 10 \log (P_r/P_b) = 20 \left(\log \frac{C}{4\pi} - \log r - \log f\right) =$$
$$= 33 - 20 (\log r\ [\text{km}] + \log f\ [\text{MHz}]).$$

Figure 1 shows a graph of the dependence of an attenuated signal for nondirectional antennas on the length of the communications line for various frequencies f. The use of directional antennas, according to coupling equation (2), is similar to increasing the spacecraft transmitter power by $G_b \cdot G_e$ times. The minimum value of P_r providing reliable exchange of information is defined as

$$P_{r\ min} = kT_0 \Delta f N (P_s/P_n), \qquad (3)$$

where $k = 1.38 \cdot 10^{-23}$ [w/deg·Hz]; $T_0 = 300°K$; $(kT_0\Delta f)$ is the noise power level at the input of the receiver in band Δf; N is the receiver noise factor.

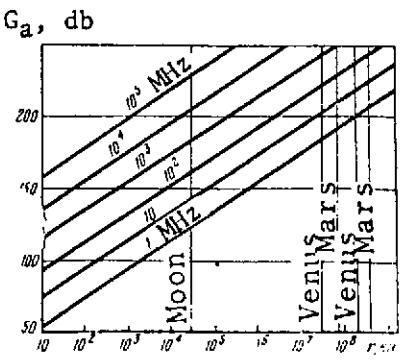


Figure 1. Attenuation of a Signal as a Function of Communications Range (for $G_b = G_e = 1$)

Simultaneous solution of formulas (2) and (3) produces the expression /222

$$r_{max} = \frac{C}{4\pi f} \sqrt{\frac{G_b \cdot G_e \cdot P_r}{kT_0 \Delta f N \left(\frac{P_s}{P_n}\right)}}, \qquad (4)$$

characterizing the dependence of communications range on all parameters of the communications line. Similarly, for the ratio of the useful signal to noise (P_s/P_n) we have

$$\eta = P_s/P_n = P_b G_b A\gamma/4\pi r^2 kT_0 \Delta f,$$

and consequently the minimum power necessary for communications is

$$P_{r\ min} = 4\pi r^2 \eta kT_0 \Delta f/G_b A_{ef}\gamma. \quad (5)$$

It is not expedient to increase r_{max} by constricting the frequency band Δf, since this causes a decrease in the quantity of information transmitted per unit time from the spacecraft, i.e., reduces the information capacity of the communications line.

The influence of the antennas on communication conditions is characterized in (5) by the quantity $G_b A_{ef}$. The directivity of the on-board antenna G_b is determined by the dimensions of the antenna as related to the wavelength. For example, for a parabolic, single-reflector antenna

$$G_b = \alpha \left(\frac{D}{\lambda}\right)^2 \quad (6)$$

where α = 4-5.

The effective area of the ground antenna (A_{ef}) is related to the geometric area of its aperture by the dependence

$$A_{ef} = \xi S, \quad (7)$$

where the surface utilization factor (ξ) may have values of 0.5 to 0.75. This last expression, if the surface of the antenna is of high quality, is almost independent of wavelength.

Thus, in addition to increasing the power of the on-board transmitter, one effective means for increasing the maximum communications range r_{max} is to use on board the spacecraft antennas with high directivity or, in other words, sharply directional antennas.

The use of such antennas on spacecraft involves a number of difficulties, resulting from the peculiarities of the general operation of spacecraft.

First of all, to increase G_b by narrowing the radiation pattern requires precise orientation of the antenna toward the Earth during communications sessions, which places additional requirements on the stabilization system and the adjustment of the antenna in flight. /223

The accuracy of orientation of this axis, consequently, and the accuracy of the stabilization system must be determined by the sector of solid angles within which condition $G_b \geq G_{b\ min}$ is met, where $G_{b\ min}$ is the minimum permissible directional gain of the on-board antenna for a given value of r. Assurance of the necessary accuracy of orientation involves additional expenditures of reaction mass and spacecraft electric power supples. In the final analysis, this requires some increase in the size and weight of the spacecraft. The problem of optimization of the design of the spacecraft antenna system must be solved together with problems of design of the electric power supply, receiving and transmitting system, motor system, automation system and other equipment which make up the spacecraft.

Actually, even simple analysis of the equations presented above indicates the mutual relationship of a number of on-board systems and their parameters. The power of the on-board transmitters depends on the capacity of the on-board power supplies which, in turn, is determined by the maximum permissible size and weight and, finally, the power consumption of the orientation motor devices. The gain (G) of the on-board antenna is directly related to size, which determines the moments of inertia and, consequently, consumption of orientation mass per unit time to maintain the required accuracy of orientation and the number of communications sessions with the Earth, based on the total reserve of orientation mass., etc.

In other words, the quantity of information transmitted in one communications session is expressed by the functional dependence

$$M = F(P_b,\ G,\ K_{of},\ P_{ps},\ B,\ D), \qquad (8)$$

where P_b, G, K_{of}, P_{ps}, B and C are complex, interrelated functions, characterizing the power of the on-board radio systems, gain of the antenna, volume and consumption of orientation mass , capacity of power supplies, moment of inertia and parameters of the orientation system and the characteristics of the ground communications station respectively. An attempt is made in [4] to relate some of these parameters and, with certain initial conditions, an expression is produced for M. However,

it is not possible to produce a universal expression, suitable for all cases, and this problem should be solved in various particular aspects, when a number of functional dependences are assumed fixed and unchanging in various situations.

At the present time, directional antennas with continuous and discrete distribution of sources are used on spacecraft. The first type includes primarily dish antennas, the second type -- arrays, consisting of sets of radiators distributed over the surface of the spacecraft.

The selection of a given type of antenna depends on its characteristics of size, weight, potential and other data.

Parabolic antennas are simplest in design, and do not require complex switching and feeder systems. On the other hand, antenna arrays have advantages from the standpoint of the possibility of switching the beam shape -- expanding and constricting the radiation pattern as a function of operating conditions of the spacecraft. Furthermore, analysis of the electrodynamic excitation systems of these antennas shows that whereas only the surface of the aperture participates in formation of the directivity characteristic of a mirror antenna, /224
the directivity characteristic of an antenna array is created by the entire volume of the antenna. Actually, in the expression

$$F(\theta, \varphi) = F_1(\theta, \varphi) \sum_{n=1}^{N} A_n e^{ikr_n \cos \theta_n}, \tag{9}$$

where $F_1(\theta, \phi)$ is the radiation pattern of one radiator; A_n is the combined amplitude of current in the nth radiator; r_n is the modulus of the radius-vector of the nth radiator relative to the center of the array; θ_n is the angle between $\bar{r}_n$ and the direction toward the observation point, defined by angles θ and ϕ, the second term accounts for the distribution of sources over the surface of the array, while the first term considers the directional properties of each source which, as a rule, are determined by the longitudinal and transverse dimensions of the radiators [5], i.e., the actual volume which they occupy.

In solving the problem of the selection of a given type of antenna from the standpoint of volume and dimensions, we must fix the necessary value of G_{max}, desirable to be achieved from a given volume, surface, etc.

For parabolic antennas, there are optimal relationships between the diameter, focal length f and depth of the parabola

$$G_{max} = \pi^2 k \left(\frac{D}{\lambda}\right)^2,$$
$$\theta_{0.5} = \beta \left(\frac{\lambda}{\alpha}\right)^{0.5},$$

where k is the reflector surface usage factor; $\theta_{0.5}$ is the width of the radiation pattern at the half power level.

For actual reflectors, (d/f) changes within the limits $(d/f)_{opt}$ = 1.6-3.0. Coefficient β changes from 0.5 to 0.73, k from 0.5 to 0.6 [5]. It is shown in [4] that the directional gain in direction θ to the axis of the parabola is expressed by the equation

$$G_\theta = \underbrace{\pi^2 k}_{F_1\left(\frac{d}{f}\right)} \cdot \underbrace{\left(\frac{d}{\lambda}\right)^2 \cdot \exp\left[-\frac{\alpha}{\beta^2}\left(\frac{d}{\lambda}\theta\right)^2\right]}_{F_2\left(\frac{d}{\lambda},\theta\right)}. \tag{10}$$

In studies of the maximum of function F_2, we find that the maximum directional gain (DIR) is provided for

$$\left(\frac{\gamma}{p}\right) = \left(\frac{\gamma}{p}\right)_{opt} = \frac{\beta}{\sqrt{0.69}} \frac{1}{\theta}, \tag{11}$$

i.e.,

$$DIR = DIR_{opt} = \frac{\pi^2 \beta^2 k}{e \cdot \theta^2 \cdot 0.69}. \tag{12}$$

Therefore, for a fixed θ, (11) defines the optimal value of (d/λ) assuming β ≈ 0.6.

$$\left(\frac{d}{\lambda}\right)_{opt} = \frac{0.6}{\sqrt{0.69}} \frac{1}{\theta}, \quad DIR_{opt} = \frac{5.5 \cdot 0.36}{e \cdot \theta^2 \cdot 0.69}.$$

Assuming (d/f) = 2.4 (the mean value usually used as an initial value), we can calculate the volume occupied by a parabolic antenna. The general expression for volume limited by contours ABCDE shown on Figure 2, will be /225

$$V = \frac{\pi d^2 h}{4} + \frac{\pi d^2}{8}(f - h).$$

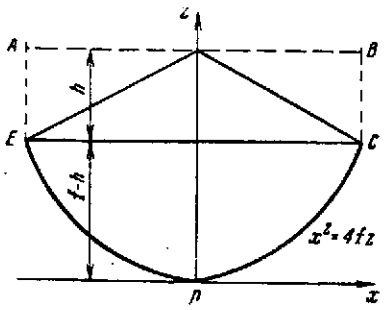


Figure 2. Geometric Relationships for a Parabola

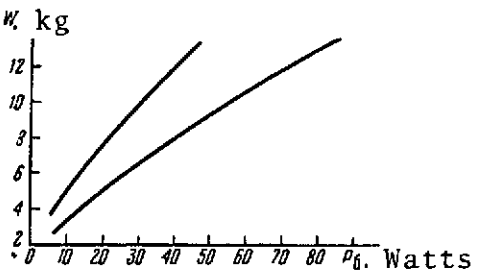


Figure 3. Weight W of an On Board Transmitter as a Function of its Power P_b (cm and dm Range)

It follows from the equation for a parabola that $d^2/4 = 4f(f - h)$. Then where d/f = 2.4, the volume will be

$$V = 0{,}085 \cdot \pi d^3.$$

Thus, we have produced an expression for the optimal volume of a parabolic antenna with a given optimal aperture diameter and value of DIR.

Turning to antenna arrays, we can make the initial assumption that they are composed of N discrete radiators of traveling-wave antenna type, placed evenly over the surface or a portion of the cylindrical surface of the spacecraft.

The initial quantities for the array are selected as a function of that which is important in each case -- minimum aperture area or minimum antenna depth.

The use of arrays on spacecraft for the centimeter waveband is generally not justified due to the cumbersomeness, great weight and losses introduced by the switching and feeder system. In the decimeter waveband, an array may be preferable to a parabolic antenna. As an example, let us take a parabolic antenna with an aperture $d/\lambda \approx 8$ and $\theta_{0.5} = 8.75°$, the directivity of which will be [5]

$$G = \frac{37\,000}{\theta^2_{0.5}} \simeq 483.$$

The total height from the vertex of the mirror to the radiator in this case is approximately $(f/\lambda) = 3.3$.

If we take a radiator height $h = 3.3$ as an initial quantity, we can produce the optimal aperture dimensions for an array of traveling-wave radiators: $d_{opt.ar} = 2.4\lambda$ as opposed to $d_{opt} = 8\lambda$ for the parabola.

In addition to dimensions, great significance in the construction of a spacecraft antenna system must be given to weight characteristics.

The possibilities for regulating output parameters of on-board spacecraft transmitters in the direction of increasing P_b are limited. The efficiencies of on-board transmitters lie between 10 and 30%, so that an increase in P_b causes an increase in the power supply capacity and, consequently, power supply /226
dimensions and weight. There is a direct dependence (quasi-linear) between P_b and its weight W. The slope of the characteristic $W = f(P_b)$ depends on the wave band: upon transition to shorter wave lengths, the slope becomes steeper with otherwise equivalent conditions. Figure 3 shows two average characteristics for the centimeter and decimeter wave bands. The quasi-linear nature of curve $W = f(P_b)$ will break down beginning at a certain point, resulting from the fact that the capabilities of existing electronic devices are limited, and the fact that at high power levels, the weight required for the cooling and power supply systems, etc. begins to increase more rapidly than P_b. Radiated powers of on-board spacecraft transmitters P_b on the order 5-100 w are currently considered optimal, depending on the purpose of the spacecraft.

Within these limits, the dependence $P_b = f(W)$ is linear.

Using the example of a parabolic antenna, we can perform a comparative evaluation of the expediency of increasing antenna weight and transmitter weight in order to increase the potential of a communications line.

In the case of an all-metal parabola, its weight can be estimated from the formula

$$W = W_{rad} + S\delta p, \tag{13}$$

where W_{rad} is the weight of the radiator with its supports; S is the surface area of the reflectors; δ is the thickness of the reflector material; p is the specific gravity.

The value of W_{rad} can be assumed constant, independent of aperture diameter. Then, formula (13) is rewritten as

$$W = \text{const} + S\delta p = \text{const} + \delta \cdot p \cdot f(D),$$

i.e., the weight of the parabola is determined by a function of the form f(D) where parameter D is the diameter of the aperture of the reflector.

All actual parabolic antennas will be somewhat larger than their calculated size, since this formula does not consider the addition of rigidity ribs, elements of the mechanisms used to open individual portions of the antennas, etc.

Actual designs may not necessarily be all metal. Frequently, an antenna is made so that the reflector can fold up completely or partially (for example, its outer portion, as on the Venera 4, 5 and 6 spacecraft). During the powered flight sector, before the nose fairing is jettisonned, the antenna is folded, after which it opens. The antenna may also be made as a metal-coated dielectric paraboloid. Table 1 shows the size and weight characteristics of certain types of parabolic antennas (folding and nonfolding) considering the radiator and remaining elements. These data indicate that actually W is related to D by the dependence W[kg] = D^n [m], where, up to $D \leq 3$m, n is between 2.5 and 3.

Figure 4 shows curves $W = D^2$, $W = D^3$ and the actual curve W = f(D) based on points taken from Table 1.

Analyzing the curves on Figures 3 and 4, we can select the optimal version, providing the greatest gain in communication

line potential. For cases when the parabolic antenna simultaneously services several transmitters for different frequencies, considering duplication, one must always use directional antennas with maximum possible aperture, since the increase in weight due to increased diameter is not as great as the total weight increase which would be required for the N transmitters to increase their power.

In designing and planning directional spacecraft antenna, be they parabolic or antenna arrays, one must consider a number of factors, in particular such factors as losses in the channels, the influence of side lobes, variations of the actual antenna profile from the theoretical profile, etc. As a result, the actual gain G_a will differ from the calculated gain G_c by $G_c - G_a = G_{ir}$, where G_{ir} is the reduction in G caused by various factors. /227

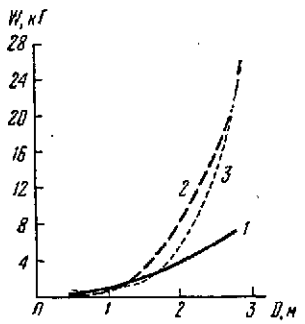


Figure 4. Weight of Parabolic Antenna as a Function of its Diameter, 1, $W = D^2$; 2, $W = D^3$; 3, Actual

The limitations on the maximum possible directional gain are particularly felt when long antenna arrays are created. It was shown in [7] that the true gain of an SHF antenna as a function of phase errors arising due to deviations in the actual profile from the design profile (influence of tolerances) alone is

$$G_a(db) = 10 \log (12.5\, K_S \cdot p \cdot n^2) - 4.3\, (2\pi/x)^2 n^2 m^2.$$

Figure 5 shows graphically the dependence of the true gain (G_a) on antenna length $n = L/\lambda$, expressed in wavelengths.

The curves are given for the case of the maximum relative manufacturing accuracy m = 0.0001 and for two values of usage factor of the surface of the array p = 1.0 and p = 0.7. The optimal length of an antenna considering the relative manufacturing accuracy can be determined by the formula [7] /228

$$n_{opt} = \frac{0.43}{\sqrt{K_S p} \cdot m}.$$

TABLE 1

Design Data	Diameter of Paraboloid D, mm	Weight of Antenna W, kg
Plastic-Foam Based Parabola:	1100	1.354
Plastic Foam δ = 7 mm		
Glass Fabric δ = 1.2 mm		
Foil δ = 0.005 mm		
Plastic-Foam Based Parabola:	1100	1.177
Plastic Foam δ = 7 mm		
Glass Fabric δ = 0.06 mm		
Metal Foil δ = 0.23 mm		
All-Metal Parabola:	1100	1.170
MA8M δ = 0.3 mm		
MA8M δ = 0.4 mm		
Honeycomb Structure:	1100	1.330
Glass Fabric δ = 1.3 mm		
Honeycomb δ = 4 mm		
Foil δ = 0.05 mm		
Folding Rigid All-Metal Structure, Central Portion Ø = 1100 of AMg6M δ = 2.0 mm. Folding portion of metalized fabric	1100	10.220
All Metal Parabola of AMg6BM δ = 1.5 mm	484	0.595
All Metal	1450	1.550
of AMg6M δ = 0.8 mm		
of MAG-1 δ = 0.8 mm	2800	22.500

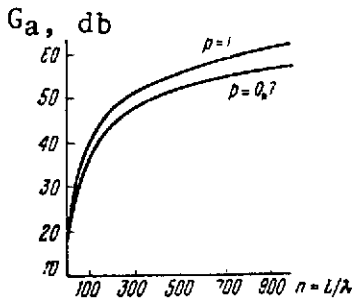


Figure 5. Actual G as a Function of Length of Antenna, Expressed in Wavelengths

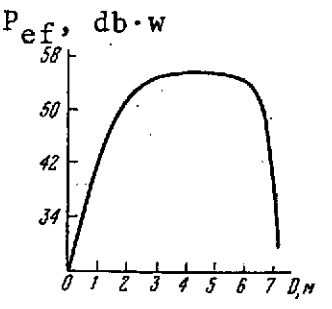


Figure 6. Effective Potential as a Function of Dimensions of a Parabolic Antenna

The reduction in gain of an antenna is strongly influenced by various types of losses. The influence of antenna length L, linear attenuation α and other factors on G is analyzed in [8] and [9].

The limiting length and optimal length L_{opt} of arrays are determined in [10], considering the optimal distribution of field over the aperture. It is expressed by the formula

$$L_{opt} \simeq -\frac{5}{\alpha} \log P_{\ell},$$

where α is the linear attenuation in the feeder; P_{ℓ} are the active losses in the load. No further increase in length of the array will cause an increase in G. When the length of the array is greater than the optimal length $L > L_{opt}$, losses in the feeder α (db/m) and losses in the absorbing load predominate, no matter what the losses in the protective fairing, excitation system, antenna material, etc.

Problems of optimization and selection of the most expedient forms of design of directional antenna systems are among the most significant problems encountered in the stage of design of a spacecraft. This is related both to the energy capabilities provided by the use of directional antennas and to their weight and size characteristics, which significantly influence the arrangement of the spacecraft as a whole.

The effectiveness of directional antennas on spacecraft is usually evaluated on the basis of effective potential (P_{ef}), the product of G of the on-board antenna times the radiated power of the on-board transmitter.

Taking this product as the initial quantity ($G_b \cdot P_b$), we can estimate the expediency of using antennas of various sizes. It is pointed out in [6], in particular, that with a fixed weight of the on-board system, the effective potential increases with increasing antenna size. Within certain limits, therefore, it is suitable to use larger antennas rather than smaller antennas, with the corresponding increase in on-board transmitter power P_b. Beginning at certain values, the curve of P_{ef} as a function of antenna size reaches a saturation point and then begins to drop (Figure 6), i.e., after a certain limit, increasing the size of the antenna yields no increase in effective potential, and in fact causes it to decrease. This also /229
agrees with the curve of Figure 4.

Obviously, the factors we have noted here can be regulated within certain limits, depending on successes achieved in the direction of decreasing the weight of on-board apparatus, increasing its efficiency and other characteristics, directly or indirectly influencing the total potential of the spacecraft.

The author is grateful to V. N. Rychkov for his help in selecting materials for the composition of this article.

REFERENCES

1. Grudinskaya, G. P., Rasprostraneniye radiovoln [Propagation of Radio Waves], Vysshaya shkola Press, 1967.
2. Zarubezhnaya radioelektronika, 1960, No. 12, p. 22.
3. L'Onde electrique, 1965, Vol. 45, No. 458, p. 568.
4. Kosmicheskiye issledovaniya, 1968, No. 5, pp. 765-771.
5. Ayzenberg, G. Z., Antenny UKV [USW Antennas], Svyaz'izdat Press, 1957.
6. Space Aeronautics, 1967, Vol. 31, July, p. 94.
7. Izv. VUZov. Radiotekhnika, 1960, No. 4, pp. 471-476.
8. IRE Transactions, 1954, October, AP-3, p. 147-152.
9. IRE Trans. on Aut. and Propagation, 1957, January AP-5, p. 144-145.

N75 11013

POTENTIAL CAPABILITIES FOR COMPRESSION OF INFORMATION OF CERTAIN DATA PROCESSING SYSTEMS

Yu. K. Khodarev, V. P. Yevdokimov
and V. M. Pokras

1. System Data Compression Factors

One natural requirement involved in the planning of telemetry systems with data processing on board the spacecraft is determination of the potential capabilities of the system as concerns compression quality. Knowledge of the extreme values of possible compression factor of a system allows us to compare various compression systems and select the proper one for practical use.

Let us study a generalized block diagram of a data collection and processing system of a spacecraft, in which a number of sensors or outputs of scientific instruments are cyclically interrogated by a commutator. We will consider that the values in the channels at the output of the commutator are represented by code words consisting of an identical number of characters. This is not a significant limitation, since all subsequent conclusions can be extended to words of variable length as well. The quantity of information of the system will be defined, therefore, as the number of code words. A block diagram of the system is shown on Figure 1.

In each interrogation cycle, the commutator forms a telemetry frame, containing an information portion plus service supplement, occupying several channels and used for purposes of synchronization, recognition and time correlation of frames upon decommutation and decoding on the Earth. Each frame is sent to the central data processing device, which includes a buffer memory. One task performed in the central processor is elimination of redundancy of the information transmitted, i.e., data compression. One significant feature of the information collection and processing system of scientific research spacecraft is that it must be capable of accumulating processed data in the on-board memory device for long periods of time, with subsequent transmission of the information to the Earth during communications sessions.

Let us study data processing systems which, after screening out the redundant values, transmit only significant values /230

according to some criterion, to the telemetry frame channels. The different channels of the system may have different criteria for significance. The criterion for determination of redundancy for a given operation is insignificant. A commutator channel in the telemetry frame which contains a significant sample at any moment in time will be called active for that moment.

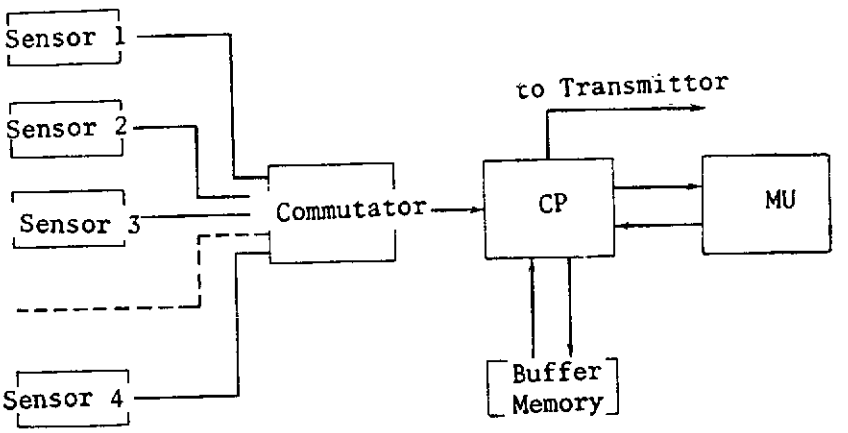


Figure 1. Spacecraft Data Collection and Processing System. CP, Central Data Processor -- Computer; MU, On-Board Memory Unit

In order to make the process of decommutation on the Earth unambiguous, it is necessary to transmit the values of the samples from active channels plus indication as to which channel (address) a significant sample belongs and as to the time of its appearance. Depending on the method of recording the address and time of a sample, a service supplement of variable length must be transmitted. This necessity of transmitting additional information on addresses and times along with the values of the telemetry data, naturally, reduces the effectiveness of compression of data in the system.

Let us place a reasonable limitation on the operation of the data processing system, consisting in that the number of words at the output of the processing system, including the service supplement, must not exceed the number of words at its input. Upon transition through this threshold, compression of the stream of data should be interrupted and the raw information should be simply transmitted to the output of the system. This limitation means that the system will never increase the volume of data transmitted rather than compressing it.

Let us now introduce the concept of the ideal and actual data compression factors in the collection and processing system. We will define the ideal data compression factor as the ratio of the total number of code words in N frames arriving at the input of the system in time T of accumulation of information I_{tw} to the number of code words at the output corresponding to active channels in these N frames I_{aw}. Thus, in determining the ideal compression factor, the supplementary service information is not considered:

$$R_i = I_{tw}/I_{aw} = \frac{Nk}{\sum_{i=1}^{N} n_i} = \frac{N\gamma m}{\sum_{i=1}^{N} n_i}, \tag{1}$$

where k is the total number of channels in a frame (including service channels); m is the number of information channels in a frame; γ is the frame length increase factor resulting from the service channels; n_i is the number of active channels in the ith frame. /231

In formula (1) we can separate the component parts of the compression factor, since there are frames in which none of the channels are active. Let us represent by N_a the number of frames of N which have at least one active channel, and write the quantity

$$I_{aw} = \sum_{i=1}^{N} n_i = \sum_{i=1}^{N_a} n_i.$$

The empirical mean number of active channels in an active frame is

$$\bar{n} = \frac{\sum_{i=1}^{N_a} n_i}{N_a}. \tag{2}$$

Then

$$R_i = \frac{Nk}{N_a \bar{n}} = R_a \bar{r}_i, \tag{3}$$

where $R_a = N/N_a$ is the compression factor resulting from the presence of frames with no active channels; $\bar{r}_i = \bar{k}/\bar{n}$ is the ideal mean compression factor of the frame.

In an actual data collection and processing system on board a spacecraft, in addition to the samples in the active channels, supplementary address-time information is also transmitted. Therefore, after processing N frames in the system, the memory device will contain I_a code words

$$I_a = \sum_{i=1}^{N} k_a(n_i) = \sum_{i=1}^{N_a} k_a(n_i),$$

where $k_a(n_i)$ is the number of words in an active frame, dependent on the number of active channels in the frame and the form of recording of the supplementary address and time information.

The empirical mean number of code words in an active frame is

$$\bar{k}_a = \frac{\sum_{i=1}^{N_a} k_a(n_i)}{N_a}. \tag{4}$$

Then

$$I_a = N_a \bar{k}_a. \tag{5}$$

If the distribution of probabilities of the number of active channels in an active frame p(n) were known, the value of $\bar{k}_a$ could be defined as:

$$\bar{k}_a = \sum_{i=1}^{N_a} k_a(n_i)\, p(n_i).$$

The distribution of probabilities p(n) is usually unknown, since it depends on the statistics of activity of channels in each specific case. Therefore, it is expedient to determine the range of change of the length of a compressed frame, i.e., the frame compression factor, with arbitrary statistics of the number of active channels in a frame for any form of distribution of probabilities p(n). Thus, in the present work we study the behavior of the second component part of the data compression factor of the system -- the frame compression factor. /232

Let us write the actual data compression factor of the system

$$R = \frac{Nk}{N_a \bar{k}_a} = R_a \bar{r}_a, \tag{6}$$

where $\bar{r}_a$ is the actual mean compression factor of an active frame.

The value of $\bar{r}_a$ is influenced by the activity statistics and by various methods of writing the supplementary information. Various methods of writing the address and time information in a frame are studied in the next section.

2. Methods of Writing Address and Time Information in a Compressed Frame

The methods of writing the supplementary information in a frame should be studied on the example of a certain hypothetical telemetry system. Let us assume that each frame in the system consists of 64 channels (k = 64) Of these, 58 channels are used to transmit scientific information (m = 58), while 6 channels make up the service portion of the frame, where three words are a frame pause, two words are the current time, one word is the program code, defining the composition of the information channels. All of our subsequent discussions can be extended without difficulty to telemetry systems with other characteristics.

Let us study certain specific methods of construction of a compressed frame.

System 1. In forming a compressed frame, for each active channel its number is recorded in the commutator (address), plus the full on-board time of onset of activity and the value of a sample, i.e., 1 + 2 + 1 = 4 words per channel. The length of

a compressed frame for n active channels is

$$k_a(n) = 4n.$$

This recording does not contain the program code or frame pause. It is therefore necessary at the beginning of operation of the system and periodically thereafter to transmit to memory either a full uncompressed frame or its service portion. This is also required for separation of channel groups of words (here groups of four words) for decoding on the Earth. The periodic transmission of service data for restoration of synchronization and recognition slightly reduces the degree of compression.

In many practical cases, it may be possible to use only the low-order digits of the full on-board time, which allows a modification of the system

$$k_a(n) = 3n. \tag{7}$$

The necessity then arises of restoring unambiguity of time as the low-order digits of the time counter overflow. Additional transmission of data on the full time reduces compression quality somewhat. In the remainder of this article, we will study only the modified system without considering these factors which slightly reduce data compression quality.

We can see from (7) that where $n > n_{gr} = 21$, the length of a compressed frame becomes greater than the length of an uncompressed frame. It is therefore expedient for those n where $k_a(n) \geq k$ to write the initial uncompressed frame into memory, since it is actually shorter. We produce /233

$$k_a(n) = \begin{cases} 3n & \text{where } 1 \leqslant n \leqslant 21, \\ k & \text{where } 21 < n \leqslant 58. \end{cases} \tag{8}$$

One shortcoming of this system is the small value of n_{gr}, reducing the length of a compressed frame only when less than one third of the total number of channels are active.

Equation (8) is illustrated on Figure 2 (curve 1).

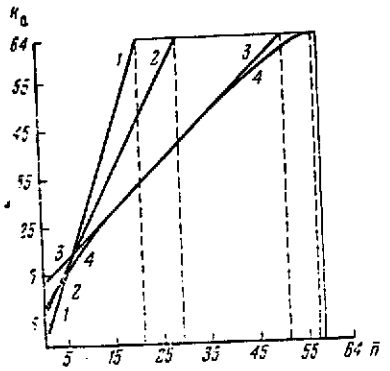


Figure 2. Length of a Compressed Frame. n, Number of Active Channels in a Frame; 1, System 1; 2, System 2; 3, System 3; 4, System 4

System 2. The service portion of the primary frame -- 6 words -- is written completely in each compressed frame. The values of the active channels plus their address (channel number) are then added. In this case, $n_{gr} = 28$. We produce

$$k_a(n) = \begin{cases} 6 + 2n & \text{where } 1 \leqslant n \leqslant 28, \\ k & \text{where } 28 < n \leqslant 58. \end{cases} \tag{9}$$

Formula (9) is illustrated in Figure 2 (curve 2).

System 3. As in system 2, we write the service portion of the primary uncompressed frame -- 6 words. Then, all of the values of active channels are written in each frame -- n words. After this, an address sequence of 58 bits is formed, each of which corresponds to one channel. A one in a given bit location indicates activity of the channel with the same number.

The number of words necessary to write 58 binary digits depends on the number of bits per word used in the system. For example, if we assume that in the telemetry system a word consists of ten bits, nine of which are information bits and the tenth is a parity bit, the address sequence requires [58/9] = 7 words, containing 63 information bits and 7 parity bits. Since the address sequence includes 58 information symbols rather than 63, the last five bits are unused. The length of the service portion of the frame is then 6 + 7 = 13 words. We produce

$$k_a(n) = \begin{cases} 13 + n & \text{where } 1 \leqslant n \leqslant 50, \\ k & \text{where } 50 < n \leqslant 58. \end{cases} \tag{10}$$

Formula (10) is illustrated on Figure 2 (curve 3).

We might comment that all of these systems are linear systems. Other modifications of linear systems are also known. We can use the following general form for the compressed frame in a linear system:

$$k_a(n) = a + bn.$$

However, there are also nonlinear systems, from which we can anticipate more economical recording of the address and time information. Let us analyze one such system. /234

System 4. This system for recording address and time information is a brief version of the last system.

We note that in system 3 with any number of active channels the address sequence always has the same length -- 58 bits.

Suppose there are n active channels in a frame with arbitrary numbers. There are then $C_m^n = C_{58}^n$ varieties of address sequences of 58 bits with n ones. Each of these sequences could be assigned a number between 1 and C_{58}^n and expressed by a binary number with $[\log_2 C_{58}^n]$ bits. The method of numbering (coding) of address sequences of this type was studied in detail in [1]. This method is complex, but is of doubtless theoretical interest.

The greatest number of possible varieties of address sequences, i.e., the highest binary number which can express it, corresponds to n = m/2 = 29 active channels in a frame. The binary number for $[\log_c C_{58}^{29}]$ contains 55 bits, i.e., always $[\log_2 C_{58}^n]$ less than m.

Since when a compressed frame with an address sequence of this type is received on the Earth, the value of the number n in this frame is not known, i.e., it is not known how many of the bit positions contain $[\log_2 C_{58}^n]$, for $n_{max} = 58$ we must add $[\log_2 58] = 6$ bits to the address symbols to transmit the value of the number of active channels in the frame n.

If we do not insist on the limitation that each frame consist of words of constant length, where n = 29 active channels we must transmit in place of 58 a group of 55 + 6 = 61 characters. The loss is slight, equal to

$$\frac{61-58}{58} \cdot 100\% = 5{,}2\%.$$

Since it was assumed earlier that each frame consists of words of constant length, the loss disappears, since [61/9] = = [55/9] = 7 words. For various n we produce

$$k_a(n) = \begin{cases} 6 + n + W(n) & \text{where} \quad 1 \leqslant n \leqslant 55, \\ k & \text{where} \quad 55 < n \leqslant 58, \end{cases} \tag{11}$$

where $W(n) = W\{[\log_2 C_{58}^n] + 6\}$ is the number of words required to transmit the address characters.

As each frame is decoded on the Earth, we must first separate the first 6 characters from the address sequence, allowing us to determine the number of ones in the primary address sequence of 58 bits. Decoding the number $[\log_2 C_{58}^n]$ allows us to determine the number of the sequence, the placement of the ones among the 58 characters, i.e., the addresses of the active channels. Forumla (11) is illustrated in Figure 4 (curve 4). The values of C_m^n and $[\log_2 C_m^n]$ were taken from [2]. The values of $[\log_2 C_{58}^n]$ and $W(n)$ are presented in Table 1.

We should note that systems 3 and 4 must be protected from errors in the communications channel to a much greater extent than the other systems. An erroneously decoded character in the address sequence makes an entire frame unsuitable for use.

TABLE 1 /235

n	1	2	3	4	5	6	7	8	9	10	15	20	29	43	48	52	54	56
$[\log_2 C_{58}^n]$	6	11	15	19	23	26	29	31	34	36	45	51	55	45	36	26	19	11
$+[\log_2 m]$	12	17	21	25	29	32	35	37	40	42	51	57	61	51	42	32	25	17
$W(n)$	2	2	3	3	4	4	4	5	5	5	6	7	7	6	5	4	3	2
$k_a(n)$	9	10	12	13	15	16	17	19	20	21	27	33	42	55	59	62	63	64

3. Influence of Statistics of Number of Active Channels in a Frame on Frame Compression Factor

If we know the number n of active channels in a frame and the formula for formation of a compressed frame, we can determine the dependence of the frame compression factor on n for any system. However, we never know in advance the number of active channels in each frame. The distribution of probabilities of the number of active channels in active frames p(n)

is also unknown and it is difficult to make assumptions concerning its form.

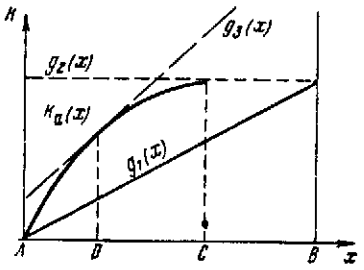


Figure 3. Determination of Extreme Values of Mean Frame Length

We must find a method of estimating the compression factor of an active frame with arbitrary forms of p(n).

Let us formulate the problem: find p(n) such that for each fixed mean value of the number of active channels in a frame $\bar{n}$, the greatest (least) value of mean compressed frame length k_a is achieved for an arbitrary compression system.

In other words, find p(n) providing

$$\sum_n k_a(n)\, p(n) = \begin{matrix}\min\\ \max\end{matrix}$$

where

$$\bar{n} = \sum_n np(n) = \text{const.}$$

Let us first study the task of finding the minimum. We use for convenience in place of n the continuous variable x and Figure 3.

For any $k_a(x)$, there is a line $g_1(x) \leq k_a(x)$. Since $p(x) \geq 0$, the following inequality is always fulfilled in [A, B]:

$$\int k_a(x)\, p(x)\, dx \geqslant \int g_1(x)\, p(x)\, dx = \text{const.}$$

Equality of the integrals, i.e., the minimum of the left integral, is achieved only where /236

$$p(x) = \alpha\delta(x - A) + \beta\delta(x - B), \tag{12}$$

where α and β are defined from the equation system

$$\left.\begin{aligned} \alpha + \beta &= 1 \\ A\alpha + B\beta &= \bar{x} \end{aligned}\right\}.$$

Where A = 1, B = m, $\bar{x} = \bar{n}$

$$\left.\begin{aligned} \alpha + \beta &= 1 \\ \alpha + m\beta &= \bar{n} \end{aligned}\right\}.$$

The solution of the system of equations yields the following values for α and β:

$$\alpha = \frac{m - \bar{n}}{m - 1}, \quad \beta = \frac{\bar{n} - 1}{m - 1}. \tag{13}$$

Then the minimum mean length of an active frame is

$$k_{a\ min} = k_a(1)\alpha + k\beta.$$

Substituting the values of α and β, we produce the formula for the maximum mean frame compression factor:

$$\bar{r}_{a\ max} = k_{a\ min}^{-\frac{k(m-1)}{k_a(1)(m-\bar{n}) + k(\bar{n}-1)}}. \tag{14}$$

We can see from (14) that $\bar{r}_{a\ max}$ with fixed $\bar{n}$ depends only on the length of a frame with one active channel $k_a(1)$, which is determined by the method of writing the address and time information in the compressed frame.

For the system example studied we produce

$$k_{a1}(1) = 3, \quad k_{a2}(1) = 8, \quad k_{a3}(1) = 14, \quad k_{a4}(1) = 9.$$

Substituting the values k = 64, m = 58 into (14), we produce

$$\bar{r}_{a1\,max} = \frac{3648}{110 + 61\bar{n}}; \qquad \bar{r}_{a3\,max} = \frac{1824}{374 + 25\bar{n}};$$
$$\bar{r}_{a2\,max} = \frac{456}{50 + 7\bar{n}}; \qquad \bar{r}_{a4\,max} = \frac{3648}{458 + 55\bar{n}}. \tag{15}$$

Similarly, using the supplementary lines $g_2(x)$ and $g_3(x)$, we can find the conditional maximum of the integral

$$\int k_a(x)\, p(x)\, dx.$$

The result of search for the maximum is as follows:

1) with $\bar{x}$ in the interval (A, C) and $k_a(x) = a + bx$, the maximum of the intergral yields a distribution of probabilities p(x) of any form falling completely in (A, C);

2) with $\bar{x}$ in the interval (C, B), when $k_a(x) \equiv$ const, the maximum of the integral yields a distribution of probabilities p(x) of any form falling fully in (C, B);

3) with $\bar{x}$ = C, the maximum of the integral yields a δ function at point C, i.e., $p(x) = \delta(x - C)$;

4) for all $\bar{x}$ from (A, B) and any $k_a(x)$, function $p(x) = \delta(x - \bar{x})$ also yields the maximum of the integral.

Returning to discrete values of the variable, we find that with fixed $\bar{n}$ the greatest length of a frame yields p(n) fully concentrated in point $\bar{n}$, i.e., $p(\bar{n}) = 1$.

Consequently, with fixed $\bar{n}$ the minimum value of the compression factor of a frame is /237

$$\bar{r}_{a.\,min} = \begin{cases} \dfrac{k}{k_a(\bar{n})} & \text{where} \quad \bar{n} \leqslant n_i, \\ 1 & \text{where} \quad \bar{n} > n_i. \end{cases}$$

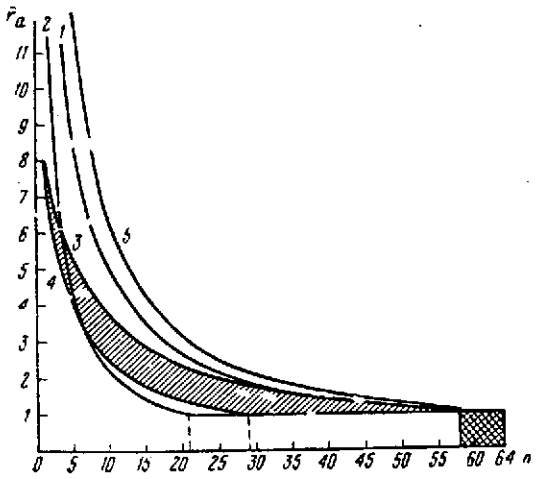


Figure 4. Extreme Compression Factors of an Active Frame. System 1: 1, $\bar{r}_a$ max; 2, $\bar{r}_a$ min; System 2: 3, $\bar{r}_a$ max; 4, $\bar{r}_a$ min; 5, $\bar{r}_u$

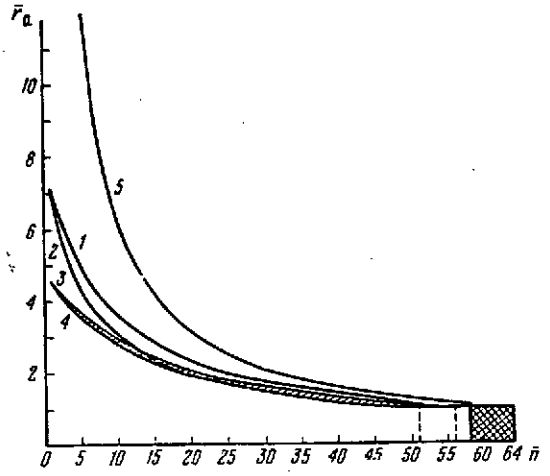


Figure 5. Extreme Compression Factors of an Active Frame. System 4: 1, $\bar{r}_a$ max; 2, $\bar{r}_a$ min; System 3: 3, $\bar{r}_a$ max; 4, $\bar{r}_a$ min; 5, $\bar{r}_u$

The graphs of the dependences $\bar{r}_a$ max $(\bar{n})$ and $\bar{r}_a$ min $(\bar{n})$ for systems 1 and 2 are presented in Figure 4, for systems 3 and 4 -- in Figure 5. The upper curve in these figures shows the greatest possible ideal frame compression factor (ignoring service information):

$$\bar{r}_u = \frac{k}{\bar{n}}.$$

Comparison of the curves shows that for systems 1 and 2 there is a significant area of values of $\bar{n}$ in which the minimum mean compression factor of an active frame is equal to 1. The dependences for systems 3 and 4 show almost complete insensitivity of the frame compression factor to the statistics of the number of active channels p(n). Here, /238 for almost all n, a frame compression factor of greater than 1 is guaranteed. System 4 always provides better compression than system 3. A guaranteed value of compression factor greater than 1 is achieved by decreasing its value with a small number of active channels in the frame.

4. Experimental Evaluation

A data compression system was modeled and the statistical characteristics of the output signal of the telemetry systems of a number of spacecraft in the "Venera" and "Zond" series launched in the USSR were studied. Statistical processing was performed on the assumption that the data were compressed by a zero order predictor.

The distributions of probabilities p(n) of the number of active channels in an active frame with various sensor interrogation rates, various comparison thresholds, etc. calculated in [3] allowed the compression factor of the actual telemetry signal to be calculated for various values of $\bar{n}$ and various formulas for recording a compressed frame.

The distribution of probabilities p(n) was found to be comparatively compact, concentrated around a mean value similar in form to distribution p(n), yielding the minimum value of compression factor with given values of $\bar{n}$. The calculated values of the actual compression factor yielded values differing little from the minimum compression factor for all systems of recording the supplementary service information.

Conclusions

The separation of the data compression factor of the collection and processing system of spacecraft into two parts used in this work has allowed us to determine the compression factor of an active frame, depending not only on the statistics of activity of channels in the telemetry frame, but also on the method of introduction of the additional address and time information to each frame.

The problem of the extreme distributions of probability of the number of active channels in a frame was stated and solved, allowing us to avoid discussion on the models of actual distributions. The extreme distributions of probabilities allow us to calculate the values of maximum and minimum mean frame compression factor with any values of mean number of active channels per frame for specific systems of entry of the service information to a compressed frame.

Modeling of the telemetry data processing systems of a number of long-range spacecraft has revealed an interesting feature of the actual statistical characteristics of the signal -- the practical similarity of actual frame compression factors to the minimum possible factors with the values of the mean number of active channels per frame produced.

REFERENCES

1. Lynch, T. G., Sequence Time Coding for Data Compression, Proc. of IEEE (letters), 1966, Vol. 54, October.
2. Kotov, P. A., Povysheniye dostovernosti peredachi informatsii [Increasing the Reliability of Transmission of Digital Information], Moscow, Svyaz' Press, 1966.
3. Khodarev, Yu. K., V. P. Yevdokimov, V. M. Pokras, "Statistical Analysis of Information from Long-Range Spacecraft," This collection, p. 239.

N75 11014

STATISTICAL ANALYSIS OF INFORMATION FROM LONG-RANGE SPACECRAFT

Yu. K. Khodarev, V. P. Yevdokimov
and V. M. Pokras

In order to determine the capabilities and effectiveness of the use of data compression systems, we must know some of their statistical characteristics. /239

As was shown in [1], we can make an a priori estimate of the quality of data compression systems on the basis of the statistical characteristics of the sensors used.

However, it is difficult to produce the statistical characteristics on the Earth for many types of sensors studying physical processes in space.

It therefore seems useful to make a qualitative estimate of the capabilities for using some system of compression based on statistical analysis of data produced from spacecraft which have already flown.

A priori ~~estimation of the quality~~ of a compression system by the method suggested in [1] requires that we know the channel activities of sensors, whose values form each telemetry frame. The evaluation can also be based on the distribution of the number of active channels per telemetry frame (a channel is considered active if the change of its value within the frame in question relative to the value in the previous frame satisfies a certain predetermined criterion of activity).

For the reasons outlined above, we present in this work the results of analysis of these statistical characteristics, ignoring the physical nature of channel activity. Analysis of the distribution of the number of active channels and the channel activities was performed using data from scientific experiments conducted on the Zond-1 and Venera-4 spacecraft. These spacecraft were selected for the following reasons:

1) long-range spacecraft are quite promising from the standpoint of the use of data compression systems;

2) Zond-1 and Venera-4 had similar scientific apparatus functioning over the flight trajectory;

3) the spacecraft flew in different directions from the Earth in relationship to the sun.

The materials studied were produced in communications sessions with the spacecraft after approximately 1 month of flight by each spacecraft.

In order to eliminate the influence of errors in the communications channel, sections with high signal/noise ratios were selected. Parity checking in each telemetry word was also used to detect errors.

Each telemetry system frame consisted of 58 similar channels. The telemetry systems did not use data compression.

The criterion of channel activity in each frame was considered a change in the value of a channel since the last active frame by a quantity greater than a certain threshold z, i.e., a zero order prediction algorithm was used.

The data files to be analyzed contained an average of 50 telemetry channels.

The zero order predictor was modeled and all subsequent calculations performed on an Ural-11 computer.

Figures 1 and 2 show the change in distribution of probabilities of the number of active channels p(n), where n is the number of active channels. The distributions were determined for the two spacecraft with various thresholds z. The interrogation period was once each four hours. The distributions are constructed for thresholds varying from Q = 3% of full scale to 9Q.

Figure 3 shows the change in p(n) with z with an interrogation period equal to one interrogation each 8 seconds. The distributions were produced from the materials of one communications session with the Venera-4 spacecraft. /240

The dependences on Figures 1, 2 and 3 are based on the results of processing data of one, arbitrarily selected communications session. This was done because the distributions of probabilities of the number of active channels with fixed threshold z changed little from session to session (Figure 4). /241 The distributions presented in Figure 4 were produced in sessions separated in time by approximately one month.

The calculations of the mean values of number n as to distribution and corresponding compression factors were performed for two compression systems:

1) a system in which the value of each active channel is transmitted with an address word and a time code word. This system is represented by the formula 3n;

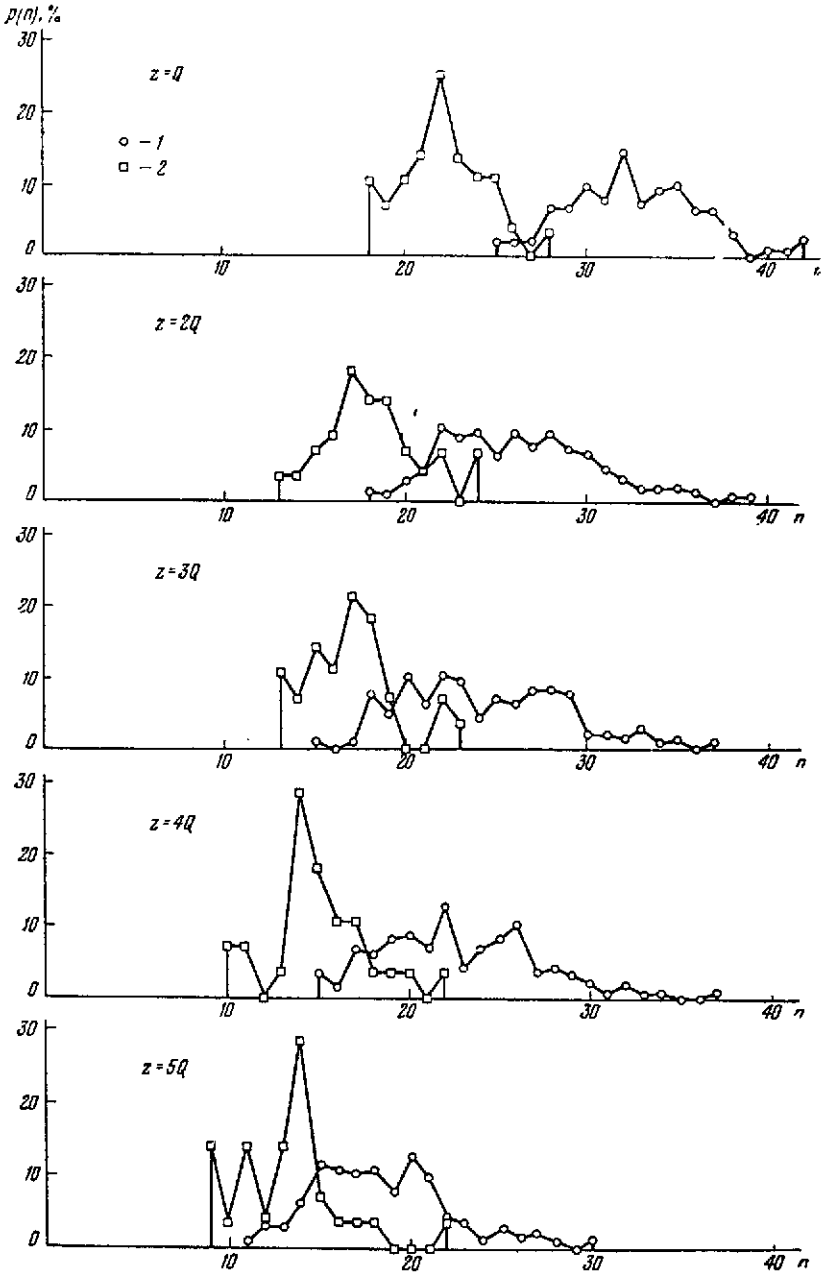


Figure 1. Distribution of Probability of Number of Active Channels p(n) for Threshold z = 1-5 Q (Q = 3%). Interrogation Period 1800 T. 1, Zond-1; 2, Venera-4

2) a system in which the value of each active channel is transmitted with an address word, while the time code is

included in a service group consisting of 6 words. This system is represented by the formula 6 + 2n.

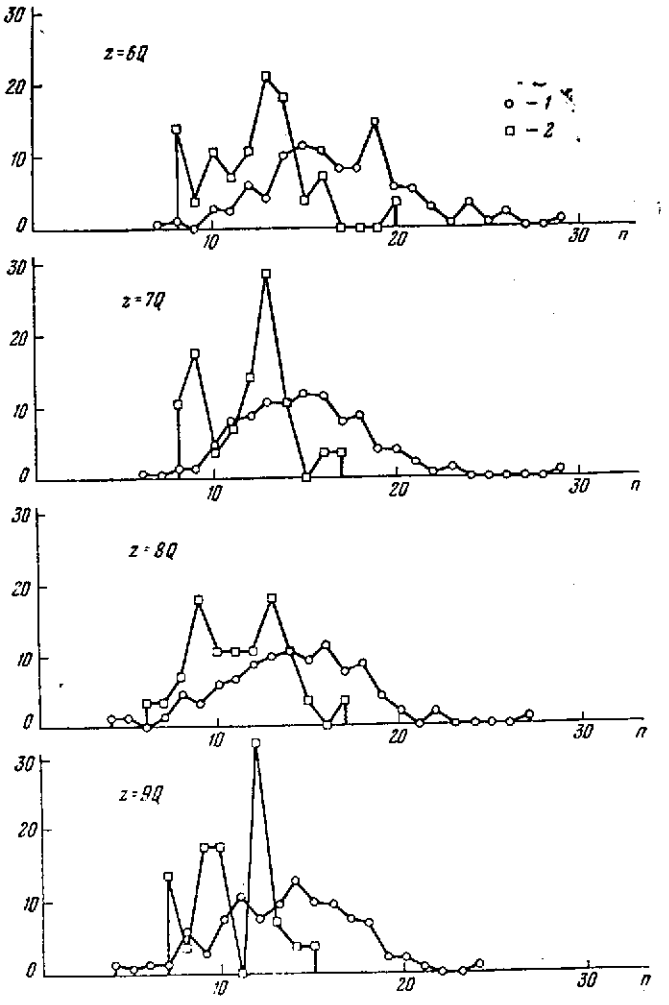


Figure 2. Distribution of Probability of Number of Active Channels p(n) for Threshod z = 6-9 Q (Q = 3%). Interrogation Period 1800 T. 1, Zond-1; 2, Venera-4

These compression systems are analyzed in detail in [2]. The results of calculations for the 3n and 6 + 2n systems with various z are presented in Table 1.

The values of compression factor of the system and mean frame length were calculated from the formulas /242

$$R = \frac{N \cdot k}{N_a \cdot \bar{k}_a}; \quad \bar{k}_a = \sum_{i=1}^{m} [k_a(n_i) \cdot p(n_i) \cdot \alpha + 64 \cdot p(n_i) \cdot (1 - \alpha)];$$

$$\alpha = 1, \quad \text{if} \quad k_a(n_i) < 64; \quad \alpha = 0, \quad \text{if} \quad k_a(n_i) \geqslant 64,$$

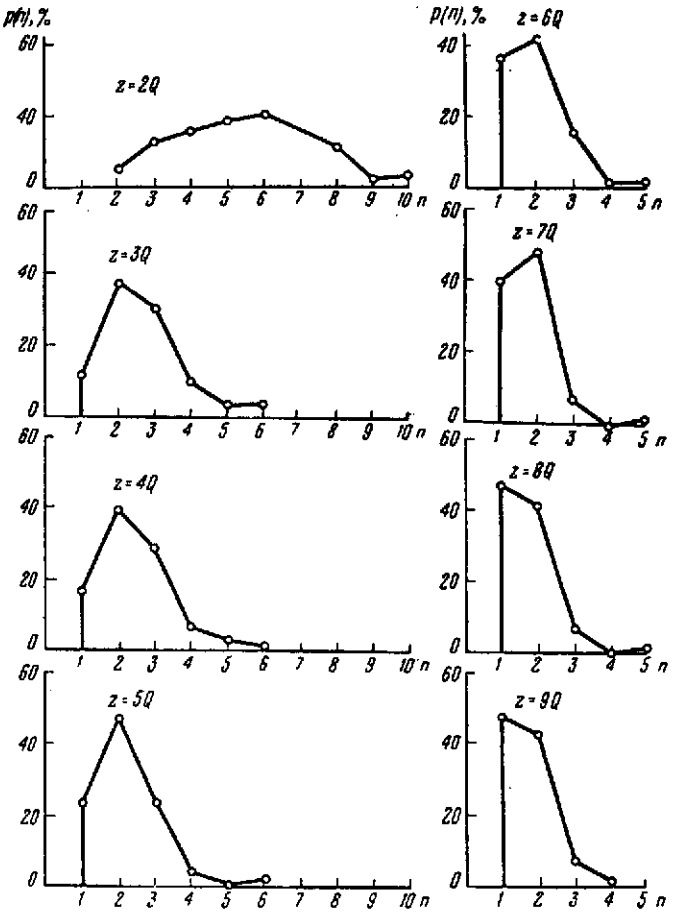


Figure 3. Distribution of Probability of Number of Active Channels p(n) for Thresholds z = 2-9 Q. Interrogation Time T. Venera-4 Spacecraft

where R is the frame compression factor considering empty frames; N is the total number of frames studied; N_a is the number of active frames; k is the number of information (m) and service channels in a frame (for the telemetry systems in question, K = 64); $k_a(n_i)$ is the length of a frame in the compression system with n_1 active channels.

TABLE 1*

FRAME COMPRESSION FACTOR AS A FUNCTION OF THRESHOLD Z FOR VARIOUS INTERROGATION RATES

	Period		z=Q	2Q	3Q	4Q	5Q	6Q	7Q	8Q	9Q	System
Venera-4	1800 T	n	21,9	18,0	16,8	15,0	13,2	12,3	11,7	11,1	10,5	
		R1	1,04	1,20	1,27	1,42	1,82	1,73	1,82	1,92	2,03	3n
		R2	1,28	1,52	1,62	1,78	1,98	2,09	2,18	2,27	2,37	6+2n
	T	n	5,00	3,3	3,0	2,7	2,4	2,2	2,1	2,1	2,0	
		R1	4,25	6,5	7,1	7,9	8,9	9,7	10,3	10,3	10,7	3n
		R2	4,00	5,1	5,3	5,6	5,9	6,1	6,3	6,3	6,4	6+2n
Zond-1	1800 T	n	32,5	26,5	24,3	22,7	18,3	17,9	14,8	13,9	13,3	
		R1	1	1,02	1,03	1,05	1,19	1,29	1,45	1,53	1,60	3n
		R2	1,01	1,12	1,19	1,42	1,5	1,53	1,8	1,89	1,97	6+2n

*Translator's Note: Commas represent decimals.

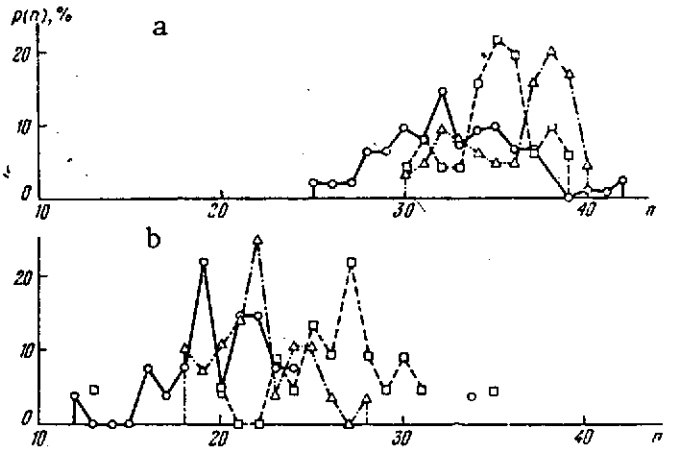


Figure 4. Distribution of Probability of Number of Active Channels p(n) for Threshold z = 1Q in Various Communications Sessions. a, Zond-1; b, Venera-4 /243

As we can see from these formulas, it was assumed in the calculation that if the length of a frame in the compression system is greater than k, the uncompressed frame of length k is transmitted.

Table 2 presents the results of calculation of mean values of channel activity as a function of interrogation rate. The calculations were performed for interrogation periods T, 2T, 3T, and 4T. For comparison, this same table presents the values of mean active frame length with interrogation period 18T (T = 8 sec).

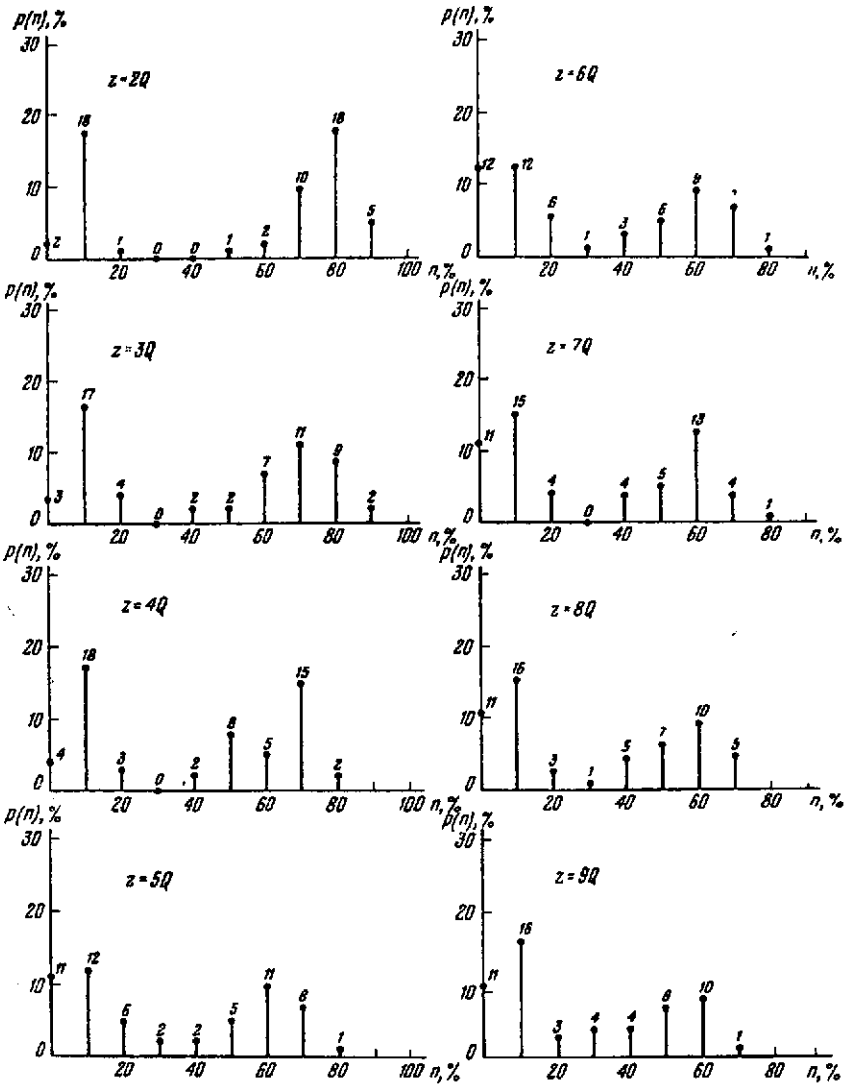


Figure 5. Histogram of Number of Channels as a Function of Activity p(n) for Threshold z = 2-9Q. Interrogation Period 1800 T. Venera-4 Spacecraft

Calculations of local activity A were performed by the formula $A_i = (a_i/N)\ 100\%$, where a_i is the number of changes in the ith channel exceeding the fixed threshold; N is the total number of values in the ith channel studied.

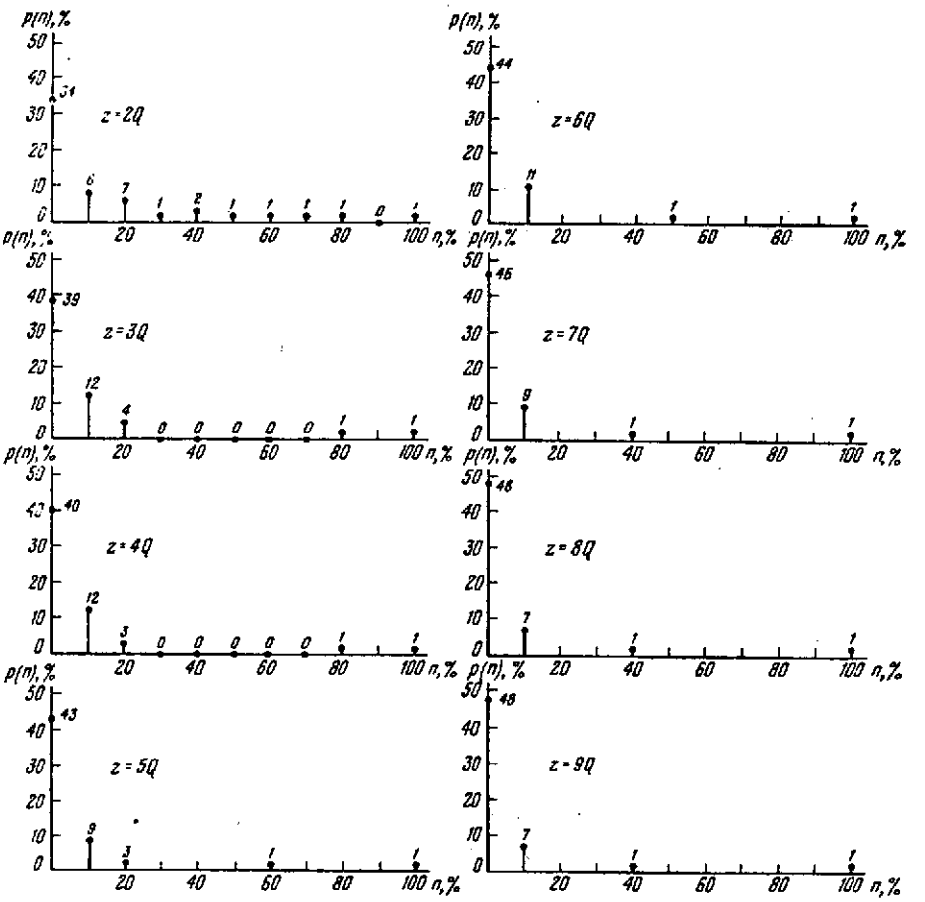


Figure 6. Histograms of the Number of Channels as a Function of their Activity p(n) for Threshold z = 2-9Q. Interrogation Period T. Venera-4 Spacecraft /244

Figures 5 and 6 present histograms of the number of channels as a function of activity. The entire range of channel activities is divided into ten sectors: from 0 to 10%, 10 to 20%, etc. The values at the lower boundaries of the sectors relate to the preceding sectors. Since all sectors are equal, in place of areas proportional to the number of channels falling within each sector, the ordinates at the lower boundaries of the sectors were constructed. For convenience of reading, the numerical values are written beside the ordinates.

Figure 5 shows that there are two compact groups of channels -- slowly changing and rapidly changing channels. The

relatively slight change in groups with various thresholds indicates that there are many channels with significant changes in values.

TABLE 2*

CHANGE IN MEAN NUMBER OF ACTIVE CHANNELS IN A FRAME WITH CHANGING INTERROGATION PERIOD T /245

	Q								
	1	2	3	4	5	6	7	8	9
1T	4,8	3,00	2,8	2,3	2,2	2,14	2,12	1,96	1,96
2T	5,0	3,00	2,8	2,5	2,2	1,9	1,9	1,8	1,8
3T	7,06	5,5	4,8	4,0	3,0	2,5	2,5	2,0	2,0
4T	8,00	4,9	4,8	4,1	3,1	2,6	2,6	2,4	2,2
1800T	21,9	18,0	16,8	15,0	13,2	12,3	11,7	11,1	10,5

*Translator's Note: Commas represent decimals.

Conclusions

1. The mean number of active channels changes little (by 10-20%) over long time intervals during flight.

2. The mean number of active channels changes significantly (from 4-5 to 20-30) when the interrogation frequency of the sensors is changed from T to 1800 T.

3. The mean number of active channels increases more slowly than the interrogation period. For example, when the interrogation period is increased by a factor of four, the mean number of active channels increases by 40%, while when the period is increased by 1800 times, the mean number of active channels increases by 5-6 times.

4. The mean number of active channels changes by a factor of approximately two as the threshold is changed from 3 to 25% of the full measurement scale.

5. The mean number of active channels for spacecraft carrying the same types of scientific apparatus flying in different directions from the Earth in relationship to the sun differs by approximately 1.5 times.

6. The distributions of probability of the number of active channels for different spacecraft and sessions are compact, concentrated about the mean value, indicating that

there are many independent or slightly dependent channels in the systems studied.

7. An increase in threshold from $z = Q$ to $z = 2Q$ causes a halving of the mean number of active channels. Further increases in threshold change the mean number of active channels only slightly.

8. Compression of information is achieved due to the small number of active channels in frames, since empty frames (without a single active channel) were almost never encountered.

9. The compression factor for $\bar{n} = 20$ types of distributions of the number of active channels for all compression systems studied was close to one.

REFERENCES

1. Khodarev, Yu. K., B. P. Yevdokimov, V. M. Pokras, "A Priori Estimate of the Quality of a Data Compression System Based on Statistical Characteristics of the Sensors Used," This collection, p. 379.
2. Khodarev, Yu. K., V. P. Yevdokimov, V. M. Pokras, "Potential Capabilities for Compression of Information of Certain Data Processing Systems," This collection, p. 353.

N75 11015

A PRIORI ESTIMATE OF THE QUALITY OF A DATA COMPRESSION SYSTEM BASED ON STATISTICAL CHARACTERISTICS OF THE SENSORS USED

Yu. K. Khodarev, V. P. Yevdokimov
and V. M. Pokras

The requirements placed on the operating characteristics of /246
an on-board data collection and processing system in a spacecraft depend to a great extent on the scientific equipment carried by the vehicle. Knowledge of the composition and certain statistical characteristics of the output signals of the scientific instruments allows us to determine possible signal processing algorithms in the system and to evaluate the prospects for utilizing data reduction.

In ordinary telemetry systems without data processing, the output signals of the sensors are fed to the output of the system by the telemetry commutator, forming frames which also contain supplementary synchronization, recognition and other signals. Systems are widely used in which the output signals of the commutator are sent to a long-term memory device in the spacecraft for subsequent transmission to the Earth.

In collection systems with data processing, the output signals of the commutator are sent to the central processor or computer, from which they are sent to the on-board memory after processing according to various programs.

One processing algorithm might include data reduction. After the reduction process, the channels of the frame contain only a sample of signals, significant according to some criterion. The channels with significant samples at any given moment in time will be called active channels for this moment. The activity criteria for different channels may be different. In planning a telemetry system with data processing, it is desirable to estimate the approximate savings in the number of bits of information which must be transmitted to the Earth resulting from data reduction. Knowledge of this quantity allows us to determine the expediency and "cost" of introduction of additional complication to the on-board data collection and processing system.

This type of estimation becomes possible if we know some of the statistical characteristics of the sensors used. We will show that these characteristics include the mean activity of

telemetry channels, even if the mutual correlation of the sensor signals is not known.

Let us study the information mass formed by a system commutator in a certain time T. The columns of the mass are the time scans of the commutator channels, the rows are the frames formed. In principle, the mass can be sequentially analyzed by two methods -- by rows, as the commutator does, and by columns (Figure 1).

Let us define the ideal data reduction factor as the ratio of the total number of samples inspected to the number of words transmitted after processing to the long-term memory device:

$$R_i = I_0 / I_{ai}$$

With row-by-row (frame) recording of the mass, the ideal data reduction factor is

$$R_i = \frac{Nk}{\sum_{i=1}^{N} n_i} = \frac{N}{N_a} \frac{k}{\bar{n}} = R_a \bar{r}_i \tag{1}$$

Here N is the number of frames inspected; N_a is the number of active frames of N (containing at least 1 active channel); N_i is the number of active channels in the ith frame; n is the mean number of active channels in an active frame (empirical); k is the total number of channels in a frame (including service channels); N_a/N is the frequency of appearance of active frames; $R_a = N/N_a$ is the data reduction factor resulting from the presence of empty frames (with no active channels). /247

Let us study the recording of a mass by columns. We represent by a_i the number of active samples in time T in channel i and write the number of active samples for the m information channels

$$I_{at} = a_1 + a_2 + \ldots + a_m = \sum_{i=1}^{m} a_i.$$

The ideal data reduction factor for the system is

$$R_i = \frac{Nk}{\sum_{i=1}^{m} a_i} = \frac{k}{\sum_{i=1}^{m} \frac{a_i}{N}}, \tag{2}$$

where a_i/N is the activity, i.e., the frequency of active samples in channel i. Combining formulas (1) and (2), we produce

$$R_i = \frac{N}{N_a}\frac{k}{\bar{n}} = \frac{k}{\sum_{i=1}^{m} \frac{a_i}{N}}. \tag{3}$$

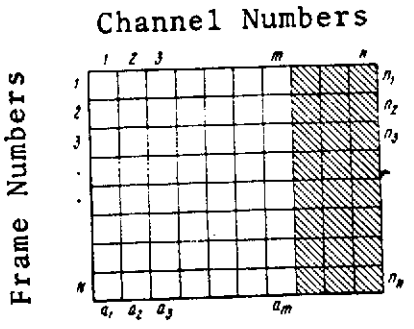


Figure 1. An Information Mass: N, Number of Frames Inspected; n_i, Number of Active Channels in Frame i; a_i, Relative Share of Active Samples in Channel i

In many practical cases, it is possible to determine the a priori reduction factor for a channel N/a_i using modeling or past experience.

Let us study the possibilities opened in this case when we know the values of a_i/N for all channels of the system.

Using formula (2) and the fixed a_i/N, we can determine the ideal reduction factor of the system R_i. We must now find a method of estimating the share of active frames N_a/N and the mean number of active frames in a channel $\bar{N}$, using this value of reduction factor. These quantities are necessary to calculate the mean length of an active frame considering the addition of service information $\overline{k_a(n)}$, necessary for recognition of channels after the reduction process, and the actual system reduction factor [1]:

$$R = \frac{N}{N_a}\frac{k}{\overline{k_a(n)}} = R_a \bar{r}_a. \tag{4}$$

Since the mutual correlation of sensor signals is unknown, in order to estimate $R_a = N/N_a$ let us study the case of the maximum and minimum time coincidence of activity of neighboring channels.

1. Total Absence of Empty Frames Possible /248

Suppose after calculating R_i we find that

$$R_i \leq k \tag{5}$$

This means that the activity of channels can be rather evenly distributed among frames, so that they will all be active in time T, i.e.,

$$R_{a\ min} = N/N_a = 1. \tag{6}$$

Time coincidence of channel activities (activity correlation) is minimal in this case. Then

$$R_i = k/\bar{n}, \tag{7}$$

from which

$$\bar{n}_{min} = k/R_i. \tag{8}$$

Knowing the mean number of active channels $\bar{n}$ in an active frame, we can determine the minimum and maximum values of frame reduction factor, found in [1], for a specific system for recording of a reduced frame $k_a(n)$:

$$\bar{r}_{a.\,max} = \frac{k(m-1)}{k_a(1)(m-\bar{n}) + k(\bar{n}-1)} = R_1;$$
$$\bar{r}_{a.\,min} = \frac{k}{k_a(\bar{n})} = R_2. \tag{9}$$

Thus, the actual reduction factor of the system falls within the limits

$$R_2 \leqslant R \leqslant R_1. \tag{10}$$

However, the activity of neighboring channels may be distributed unevenly over successive frames, so that "empty" frames appear (containing no active information channels).

In this case, we can estimate only the share of active frames relative to the total number with the maximum coincidence of activity, i.e., the greatest mutual correlation of activity. It is not difficult to show that the share of active frames is determined by the channel with the greatest share of active samples, i.e.

$$\frac{N_a}{N} \geqslant \frac{a_{i\,max}}{N}; \quad R_{a.\,max} = \frac{N}{a_{i\,max}}. \tag{11}$$

Expression (3) can be transformed to:

$$R_i = \frac{N}{a_{i\,max}} \frac{k}{\sum_{i=1}^{m} \frac{a_i}{a_{i\,max}}}. \tag{12}$$

We can see from (12) that with the maximum coincidence of channel activities, the mean number of active channels in a frame increases and becomes /249

$$\bar{n}_{max} = \sum_{i=1}^{m} \frac{a_i}{a_{i\,max}} = \frac{k}{R_i} \frac{N}{a_{i\,max}}. \tag{13}$$

For this value of n, we find $\bar{r}_{a\ max}$ and $\bar{r}_{a\ min}$ using formula (9).

Then, the values of actual reduction factor of the system can be found by multiplying the corresponding values of R_a and $\bar{R}_a$

$$\bar{n}_{max} = \sum_{i=1}^{m} \frac{a_i}{a_{i\,max}} = \frac{k}{R_K} \frac{N}{a_{i\,max}}. \tag{14}$$

The actual compression factor of the system falls within the limits

$$R_i \leq R \leq R_3. \tag{15}$$

From inequalities (10) and (15), we can find the maximum and minimum values, within which the actual system reduction factor will lie with any level of coincidence of activity in neighboring channels, since the boundary values are defined by these inequalities.

2. Empty Frames Must be Present

After calculating the ideal system reduction factor using the known values of a_i/N with formula (3) we find

$$R_i > k$$

This means that with any possible coincidence of activities in neighboring channels, among all the frames studied there must be some empty frames. The greatest possible number of active frames of all frames studied, i.e., the value of $R_{a\ min}$ can be estimated by expanding R_i into two cofactors, one of which is equal to k. Then the second cofactor gives us the value of $R_{a\ min}$.

$$R_i = R_{a\ min} k \tag{16}$$

Since we have taken the value $\bar{r}_i$ in expression (16) equal to k, this means that in all active frames the minimum number of active frames is equal to 1. Using the value of $\bar{n} = 1$ and the fixed formula for recording a reduced frame $k_a(n)$, we can find [1] the frame reduction factor $\bar{r}_{a\ max}$ and calculate the first value of our estimate of the actual reduction factor of the system (where $\bar{n} = 1$, $\bar{r}_{a\ max} = \bar{r}_{a\ min}$)

$$R_1 = R_{a\ min} \bar{r}_{a\ max}$$

C5

The resultant value of R_1 represents the estimate of the actual system reduction factor with the minimum possible coincidence of channel activities.

Let us now estimate the values of R_a and $\bar{r}_a$ with maximum coincidence of channel activities. This means that the share of active channels in the total number of channels drops. The value of $R_a = N/N_a$ increases, and its maximum value for a given R_i is

$$R_{a\ max}=N/a_{i\ max}$$

The mean number of active channels in an active frame, the maximum possible for the given R_i, can be found from the formula

$$\bar{n}_{max}=(k/R_i)(N/a_{i\ max}) \qquad (17)$$

Using this value of $\bar{n}$, we can determine the limiting values of reduction factor for an active frame for a given form of recording of the reduced frame $k_a(n)$ using formula (9): /250

$$R_2=R_{a\ max}\bar{r}_{a2\ max},$$

$$R_3=R_{a\ max}\bar{r}_{a3\ min}.$$

Selecting the least and greatest of the values of R_1, R_2 and R_3 found, we find the area of values of the actual system reduction factor with any channel activity coincidence

$$R_{min}\leq R\leq R_{max},$$

giving us our estimate of its possible value.

Thus, this method can estimate the compression factor of an on-board data collection and processing system with known mean activities of the sensors used and unknown mutual correlation of their output signals.

REFERENCE

1. Khodarev, Yu. K., V. P. Yevdokimov, V. M. Pokras, "Potential Capabilities for Data Reduction of Certain Data Processing Systems, This collection, p. 353.

N75 11016

METHODS OF SYNCHRONIZATION OF TELEMETRY FRAMES USING DISTRIBUTED AND COMBINED FRAME PAUSES

Yu. M. Shtar'kov

Introduction

Telemetry sensors are usually connected to the inputs of a commutator for successive cyclical interrogation. The operating mode of the commutator fully determines the structure of the stream of data at its output. It is natural to divide this stream into sectors corresponding to one operating cycle of the commutator, usually called frames. If we know the beginning of each frame, it is not difficult to divide it in groups of characters (messages) corresponding to the indications of the various sensors.

Thus, in order to utilize all the advantages of the frame structure of data, we need only know the beginning of each frame or, in other words, we must have frame synchronization. One very simple but effective means of frame synchronization is the use of the so-called frame pause. A frame pause is a fixed group of characters inserted before each frame. Recognition of this group allows the beginning of each subsequent frame to be detected.

The simplicity and effectiveness of frame synchronization using the frame pause have made this method quite popular. Many works have been published on the selection of the frame pause (for example, see [1-3]) and frame synchronization algorithms (the latest works in this area include [4, 5]).

In this work, two modifications of the method of frame synchronization by means of a frame pause are studied. The first does not require that all characters in the frame pause follow each other immediately. This allows us to obtain a number of additional advantages. The second modification is /251 based on the use of one particular type of redundancy within the frame to achieve frame synchronization without setting aside a special frame pause.

For simplicity, let us limit ourselves to analysis of a stream of binary data. However, the methods and calculations of this article can be extended without difficulty to any other case.

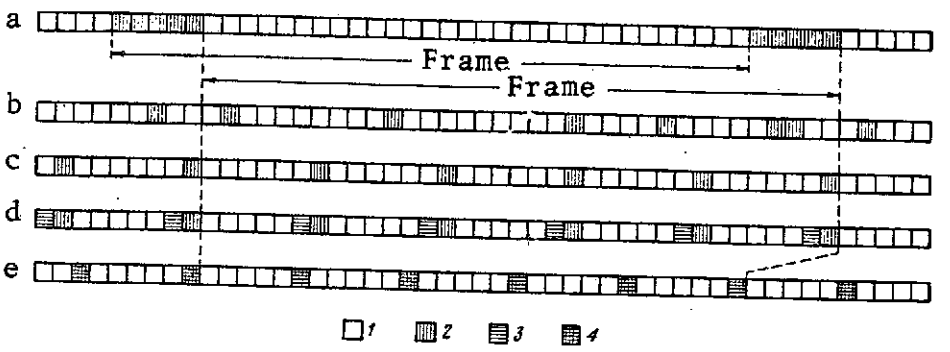


Figure 1. Structure of the Telemetry Channel. a, Using Concentrated Frame Pause; b, Using Unevenly Distributed Frame Pause; c, Using Evenly Distributed Frame Pause; d, With Parity Checking and Evenly Distributed Frame Pause; e, With Combined Frame Pause; 1, Data Characters; 2, Frame Pause Characters; 3, Parity-Check Characters; 4, Combined Frame Pause and Parity-Check Characters

In order to make our analysis more complete and comparison of various methods more convenient, let us first note the basic peculiarities of frame synchronization using an ordinary frame pause.

1. Concentrated Frame Pause

As we have noted, the frame pause is usually formed of a group of s characters placed at the beginning (or, which amounts to the same thing, at the end) of each frame (see Figure 1a). This frame pause will be called concentrated, number s -- the length of the frame pause. When this type of pause is used, the structure of the frame is insignificant. The only significant parameter of the frame is its length N -- the number of frame symbols (in addition to the frame pause), which determines the probability of proper synchronization.

Regardless of the frame pause recognition algorithm and the method of achieving and maintaining synchronization, we can note the following main peculiarities related to the use of a concentrated frame pause of length s.

1. If we ignore the initial portion of the sequence of data, which may not begin with the frame pause, it is sufficient for recognition of a frame pause to analyze the last s characters of any sequence of data received. Therefore, a memory of s bits capacity must be available (according to [6], if

the simplest recognition algorithm is used, the required memory capacity can sometimes be reduced. The first attempts at synchronization begin after reception of the first s characters of the data stream.

2. The sign of synchronism, developed immediately after completion of the frame pause, corresponds to the beginning of the frame. Therefore, the data received can be immediately divided into individual messages corresponding to the various sensors. Thus, synchronization requires no delay in utilization /252 of the data received.

3. If a certain sector of the data stream contains no frame pause, the sequence cannot be divided into individual messages. Furthermore, if the division into individual messages is somehow achieved, it is impossible to determine which sensor corresponds to which message (i.e., the main task of synchronization cannot be performed).

4. When a sequence of symbols is received, there is always a finite, generally very slight, probability of dropout or insertion of a character, i.e., recording of one extra character or failure to record one of the characters actually transmitted. This event can be detected, since the frame will contain not N, but N + 1 or N - 1 characters. However, it is impossible to localize the point of dropout or insertion. Considering that the division of the frame into individual messages in this case is incorrect, in this situation we must discard (not use) the frame.

5. If groups of errors may arise upon transmission of data, the danger arises that a significant number of characters in the frame pause may be distorted and synchronization performed incorrectly.

6. The probability of proper synchronization depends on the values of s and N, the probabilities of distortion of one character ε and the specific form of the frame pause. In selecting a frame pause, only its aperiodic autocorrelation function is significant.

Thus, the fact that the characters of a concentrated frame pause follow each other yields two advantages -- a small memory volume is required and no delay in synchronization occurs (delay in use of the data received). However, at the same time, concentration of the characters of the frame pause leads to three significant defects. In order to attempt to eliminate these defects, one must not place the characters of the frame pause one after another.

2. Distributed Frame Pause

Let us now assume that the characters of the frame pause are distributed in an arbitrary but fixed manner throughout the frame. The distribution may be uneven (see Figure 1b) or even (Figure 1c). From the technical standpoint, even distribution is more convenient. Furthermore, it can be shown that even distribution of the characters of a frame pause is somewhat better than uneven distribution from the standpoint of the probability of proper synchronization achieved. Therefore, we will study only evenly distributed frame pauses.

Let us represent the characters of an evenly distributed frame pause by $b_1, b_2, \ldots, b_s$. Then after each character of the frame pause there are $r = N/s$ data characters, then the next frame pause character, another r data characters, etc. The sets of $r + 1$ characters will be called "sectors" of the frame, and will be numbered from one through s. If we know the position of the individual messages within the frame, it is not difficult to determine the position of the messages and their association with any given sensor within the limits of each sector. Therefore, frame synchronization can be reduced to subdivision of the stream of data into individual sectors and determination of their ordinal numbers.

When an evenly distributed frame pause is used, the system shown on Figure 2 can be used for frame synchronization. At the moment when one frame is fully written in the shift register, the characters of the frame pause occupy the positions used for recognition in a strictly determined sequence. As for a concentrated frame pause, there is no need to specify the recognition algorithm. /253

Let us now note the main advantages related to the use of an evenly distributed frame pause of length s.

1. As Figure 2 shows, recognition of an evenly distributed frame pause uses a memory capacity (shift register) of $N + s = (r + 1)s$ bits. The memory volume can be somewhat reduced by discarding the first r locations in the shift register. In this case, we will always be studying the first $(r + 2)$th, $(2r + 3)$th, ..., $[(s - 1)r + s]$th characters received (numbering beginning with the last character received). However, the decrease in memory volume achieved is slight, and the recording of individual frames becomes more complex. However, it is difficult to expect a more radical reduction in memory volume.

Since the characters of the frame pause are evenly distributed throughout the frame, the first attempt at synchronization can be undertaken only after reception of $(s - 1)r + s$

characters. More precisely, these attempts can be undertaken earlier, but will utilize only a portion of the characters of the frame pause and the probability of proper solution will be lower than after the entire register is filled (see Figure 2).

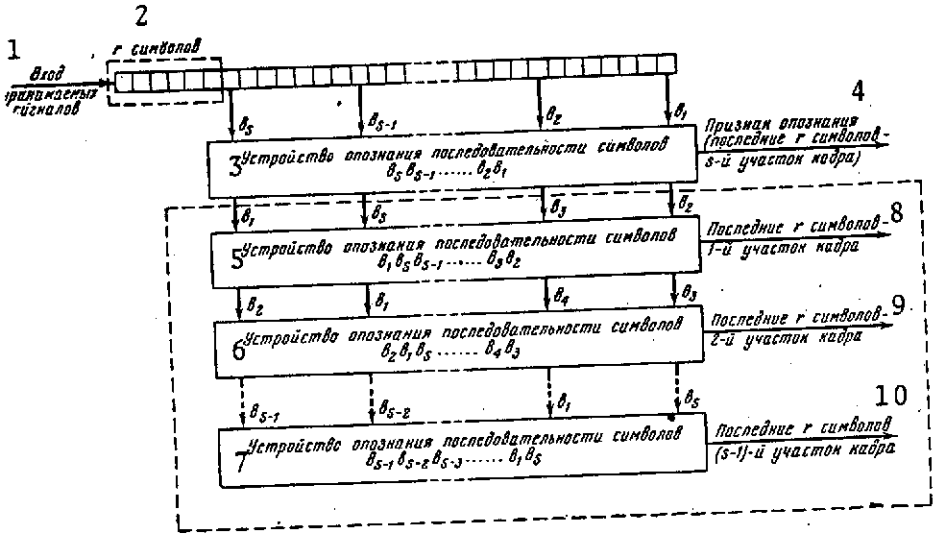


Figure 2. Synchronization Using Evenly Distributed Frame Pause. Key: 1, Input of Signal Received; 2, r Characters; 3, Device Recognizing Character Sequence $B_sB_{s-1} \ldots B_2B_1$; 4, Recognition Sign (Last r Characters of sth Frame Sector); 5, Device Recognizing Character Sequence $B_1B_sB_{s-1} \ldots B_3B_2$; 6, Device Recognizing Character Sequence $B_2B_1B_s \ldots B_4B_3$; 7, Device Recognizing Character Sequence $B_{s-1}B_{s-2}B_{s-3} \ldots B_1B_s$; 8, Last r Characters of 1st Frame Sector; 9, Last r Characters of 2nd Frame Sector; 10, Last r Characters of (s-1)th Frame Sector

2. The synchronization indicator is also produced immediately after completion of the reception of a frame pause. However, the last symbol of the frame pause is located at the very end of the frame. Therefore, there is a delay in synchronization -- any message in the frame can be used only after reception of all or almost all the frame, when frame synchronization and subsequent division of the frame into messages can be performed.

Synchronization delays can be avoided if we increase the number of recognition devices to s (Figure 2). In this case, synchronization of the first frame is performed with the same delay, since the recognition devices should not be put in operation until the first (s - 1)r + s characters are received (otherwise, the probability of proper synchronization would be

lower). However, after reception of the first sector of the next frame, recognition can be performed by the second device, meaning that the last r + 1 characters received are the first sector of the next frame (assuming that the shift register locations surrounded by the dotted line on Figure 2 are absent). Recognition by the third device means that the last r + 1 characters received are the "second" sector of the frame, etc. Generally, recognition by the jth device (1 < j < s) means that the last r + 1 characters received are the (j - 1)th frame sector (where j = 1 - sth frame sector). As a result, the delay in synchronization decreases from N + s = (r + 1)s characters to r + 1 characters (this is not true of the first frame). /254

Of course, the introduction of additional hardware for recognition increases the probability of improper decision and the volume of the synchronization circuit. However, the probability of improper decision increases only slightly, and the increase in circuit volume is compensated by the increase in potential of the synchronization circuit. In addition to the decrease in delay time, this expanded circuit has other advantages which we will describe below.

3. Let us assume that the data stream received begins in the middle of a certain frame. For example, the first sector received is the jth sector of a frame, followed by the (j + 1)th, etc. The recognition devices are turned on only after reception of (r + 1)s characters. At this moment, the jth device performs recognition, meaning that the last r + 1 characters correspond to the (j - 1)th frame sector, the preceding r + 1 characters -- to the (j - 2)th frame sector, etc. Thus, synchronization of the first N + s characters will be performed independently of the initial sector of the data stream. Subsequent synchronization is performed in the ordinary manner. For example, after reception of the next r + 1 characters, the (j + 1)th circuit performs recognition (jth sector), etc.

Synchronization can be performed similarly when for some reason the entire sequence received contains (r + 1)q < (r + 1)s characters. This requires that the unfilled portion of the register be disconnected (see Figure 2) along with those elements of the recognition circuit connected to the unfilled locations in the shift register. In other words, it is as if a shorter frame were studied, with a shorter frame pause, containing q characters, and the form of this shorter frame pause is not precisely known (q characters in a row of a cyclically repeating distributed frame pause can be selected by s different methods). Nevertheless, during the process of shifting of the symbols received in the shift register, in not more than r

cycles, recognition should be performed by one of the s devices, thus achieving synchronization. Of course, the smaller q, the lower the probability of proper solution. Furthermore, with sufficiently small q there may be not one but several equally probable solutions, and unambiguous synchronization may be impossible. However, this is quite natural (for example where $q = 1$, there may be very many solutions). The value of q at which ambiguity begins depends on the specific form of the distributed frame pause.

Thus, when a distributed frame pause is used, synchronization can be performed in all cases when the total volume of the sequence received is no less than a certain critical value, significantly less than $(r + 1)s$.

4. The advantages of a distributed frame pause are also seen when dropouts or insertions of characters occur. Let us assume, for example, that within the jth sector of a certain frame, a character has been inserted. Then at a certain moment in time $j = 1$, the characters of this sequence, used for recognition, are equal to $b_1, b_2, \ldots, b_{j-1}$, but the remaining $s - j + 1$ characters are equal to $b_j, b_{j+1}, \ldots, b_s$ only in a random manner with a probability of one half. When the characters in the register are shifted, the reverse situation arises -- the first $j - 1$ characters are equal to $b_1, b_2, \ldots, b_{j-1}$ with a probability of one half, while the last $s - j + 1$ characters are precisely equal to $b_j, b_{j+1}, \ldots, b_s$. Thus, when an insertion occurs in the jth sector at a certain moment in time, the first $j - 1$ characters studied (recognized) are the first characters of the frame pause (the remainder are random characters), and in the next cycle the $s - j + 1$ last characters correspond to the proper characters for a frame pause. This fact can be used not only for detection of the fact of insertion itself (which is not difficult), but also to localize the point of the insertion with some accuracy. A similar picture occurs when characters drop out, but in the reverse sequence. /255

Consequently, the use of a distributed frame pause allows not only detection, but localization of the insertion or dropout of a character with some accuracy. The greater the probability ε of distortion of one character during transmission, the less accurate will be localization. However, in any case, the error in localization will be less than a frame, i.e., better than when a concentrated frame pause is used.

5. In order for j characters of a frame pause to be included in a group of errors, the length of the group of errors must be at least $j(r + 1) - r$ characters if the frame pause is distributed, or at least j symbols if it is concentrated. Therefore, the distribution of the characters of a frame

pause causes a significant increase in interference stability against groups of errors.

6. As in the case of a concentrated frame pause, the probability of proper synchronization depends on the values of s and N, the probability of distortion of one character ε and the specific form of the frame pause. However, now in selecting a frame pause, its periodic (not its aperiodic) autocorrelation function is significant. This fact simplifies construction of a frame pause with a predetermined autocorrelation function. Furthermore, the use of a distributed frame pause yields one additional advantage, allowing some increase in the probability of proper synchronization.

Comparing now the properties of distributed and concentrated frame pauses, we can draw the following conclusions.

Concentration of the characters of a frame pause has the advantage that it allows extreme simplification of recognition and minimization of the required memory volume. However, at the same time, this concentration does not allow the quality of reception to be checked in the intervals between frame pauses, i.e., within each frame. Without such additional redundancy, it is quite impossible to localize the position of a dropout or insertion of a character, error burst within a frame, etc. Furthermore, even short error bursts may cause incorrect synchronization, since they may disrupt a significant portion or even all of the characters of the frame pause. Finally, it is impossible in principle to perform frame synchronization if the sector of the data sequence studied contains no frame pause. Thus, concentration of the characters of a frame pause has one advantage but, unfortunately, numerous disadvantages.

The reverse picture is true when a distributed frame pause is used. The only shortcoming is the requirement of a rather large memory volume. At the same time, distribution of the characters of the frame pause allows even checking of the entire frame, which has a number of advantages listed in detail earlier and eliminates the shortcomings of the concentrated frame pause.

In conclusion, we should note that the use of various properties of the distributed frame pause requires that s parallel-operating recognition devices be used. Of course, this makes the hardware more cumbersome, but it should be emphasized that the introduction of these additional circuits on Figure 2 was done primarily to make explanation more convenient and the illustration of the properties of the distributed frame pause clearer. There is no doubt that much simpler hardware could be /256
suggested to replace the s parallel recognition devices. The

task is still simpler if the frame synchronization algorithm is run on a digital computer.

3. Combined Frame Pause

One widely used and simple test of the quality of data within a frame is the parity check. The entire frame is divided into m groups of k characters (so that N = mk) and a (k + 1)th character is added to each group, equal to the mod-two sum of all k characters of the group. This character is then called the parity check character, and the groups of n = k + 1 bits are called words. Let us represent the characters in the first word of the frame as $a_{1,1}$, $a_{1,2}$, ..., $a_{1,n-1}$, $a_{1,n}$, in the second word as $a_{2,1}$, $a_{2,2}$, ..., $a_{2,n-1}$, $a_{2,n}$, ..., the mth word -- as $a_{m,1}$, $a_{m,2}$, ..., $a_{m,n-1}$, $a_{m,n}$, considering that the parity check characters are $a_{1,n}$, $a_{2,n}$, ..., $a_{m,n}$. By definition,

$$a_{i,1} \oplus a_{i,2} \oplus \ldots \oplus a_{i,n-1} = a_{i,n}, \quad 1 \leqslant i \leqslant m \tag{1}$$

or

$$a_{i,1} \oplus a_{i,2} \oplus \ldots \oplus a_{i,n-1} \oplus a_{i,n} = 0, \quad 1 \leqslant i \leqslant m, \tag{2}$$

where the $\oplus$ represents mod-two addition. Relationships such as (2) are usually called parity checks.

Introduction of parity bits increases the number of characters in a frame from N = mk to N' = (k + 1)m = N + m. However, it allows a very simple check of the quality of reception data within a frame.

After transmission through a communications channel and reception of all characters of any word, it is not difficult to check to see whether parity has been maintained. If so, this means that there either were no errors in transmission, or that an even number of errors occurred. Nonfulfillment of the parity check indicates presence of an odd number of errors. It is therefore usually assumed that fulfillment of the parity check means proper reception of all characters of a word (the probability of proper reception is significantly greater than the probability of appearance of two, four, etc. errors). If the parity check is not fulfilled, the corresponding word is either rejected or, if possible, special additional measures are taken to correct one error (the probability of which is

significantly greater than the probability of appearance of three, five, etc. errors). Thus, in the overwhelming majority of cases, the presence of the parity check bit allows distortion to be localized within n characters of the corresponding word. Furthermore, certain possibilities appear for localization of dropouts and insertions of characters within a frame and for detection of groups of errors of significant length.

The introduction of the parity check means that the data are somewhat redundant. This redundance can be used for purposes of synchronization. For example, the method of using parity checks to synchronize words (subdivide sequences of characters into individual words) is widely known. This method utilizes the fact that when characters are divided properly into words, parity checks are fulfilled with a probability near unity, while if they are improperly divided, the probability of proper parity is one half. Therefore, after analyzing a sufficiently large number of characters, we can synchronize the words properly with high probability.

This method can be used in converted form for frame synchronization. Let us assume that a distributed frame pause of length s = m is used, or, in other words, the number of characters in the frame pause is equal to the number of words in a telemetry frame. Then the s characters of the frame pause may be located immediately following the s parity check characters (see Figure 1d). The basic idea of the methods suggested consists in combining the solution of the problem of parity checking and the neighboring frame pause character into a single character used both for frame synchronization and for error detection. The structure of a frame in this case is illustrated by Figure 1e. The character which ends each word is the character which combines these two functions. We will call the implicit distributed frame pause a combined frame pause, meaning that it allows us to combine two functions. /257

Suppose $b_1, b_2, \ldots, b_m$ are the characters of the distributed frame pause of length m, and

$$c_i = a_{i,n} \oplus b_i, \quad 1 \leqslant i \leqslant m. \tag{3}$$

To perform the task at hand, we replace the parity check character $a_{i,n}$ and the subsequent frame pause character b_i (see Figure 1d) with a single character -- c_i. This changes the structure of the frame, corresponding to the change from Figure 1d to Figure 1e. Now the ith word ($1 \leq i \leq m$) contains the characters $a_{i,1}, a_{i,2}, \ldots, a_{i,n-1}, c_i$ and, according to (2) and (3),

$$a_{i,1} \oplus a_{i,2} \oplus \ldots \oplus a_{i,n-1} \oplus c_i = b_i, \quad 1 \leqslant i \leqslant m. \tag{4}$$

Thus, by replacing the character $a_{i,n}$ with the character c_i, the sum of all characters in the ith word is no longer equal to zero, but rather to the corresponding character of the frame pause. This property is the basis of the method of frame synchronization here being suggested.

Let us represent by d_{ij} the sum of n successive characters:

$$a_{i,j} \oplus a_{i,j+1} \oplus \ldots \oplus a_{i,n-1} \oplus c_i \oplus a_{i+1,1} \oplus a_{i+1,2} \oplus \ldots \oplus a_{i+1,j-1} = d_{ij}. \tag{5}$$

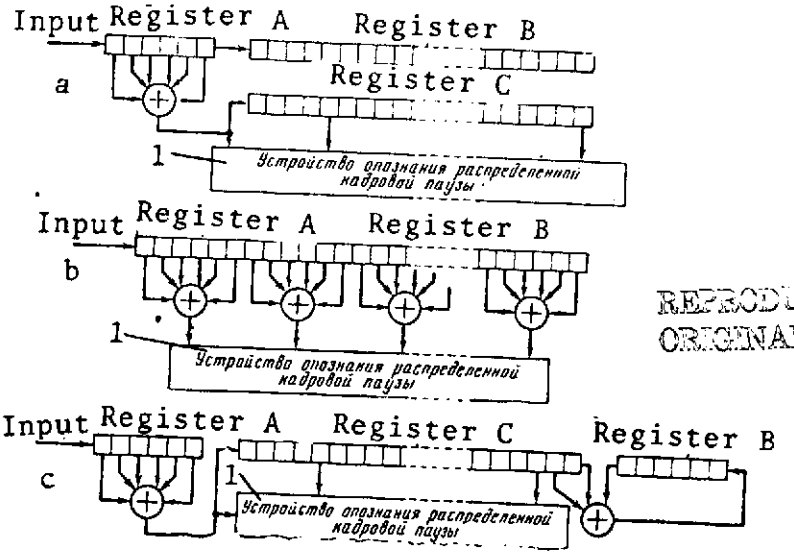
According to (4) and (5), $d_{i,1} = c_i$. However, if the data characters are independent and take on values of 0 and 1 with equal probability, for all $j \neq 1$ all d_{ij} are independent random quantities, taking on the values of 0 and 1 with probability one half.

In order to perform frame synchronization, we must continually calculate the modulo two sum of the last n characters received at the receiving end of the communications system. According to (5), the values of these sums successively produced (for an arbitrarily selected initial moment) will be

$$d_{ij}, d_{i,j+1}, \ldots, d_{i,n-1}, d_{i,n}, d_{i+1,1}, d_{i+1,2}, \ldots, d_{i+1,n-1}, d_{i+1,n}. \tag{6}$$

The sums $d_{1,1}, d_{2,1}, \ldots, d_{m,1}$, equal to the characters of the distributed frame pause, will recur in this sequence at interval n, and the sums d_{ij} ($j \neq 1$), the statistical characteristics of which are equal to the statistical properties of the data characters in a frame (these sums are independent and take on values of 0 and 1 with equal probability) will fall between them. But then the sequence of sums (6) has precisely the same structure as an ordinary telemetry frame with a distributed frame pause (see Figure 1, c).

This means that "frame synchronization" can be performed for the sequence of sums containing the distributed frame pause. All of the characteristic features of this synchronization noted in section two except the last one remain in force. However, synchronization of the sequence of sums also means synchronization of the initial sequence of characters, since the moment of proper recognition of the distributed frame pause in the sequence of sums corresponds to completion of a frame of initial data.



REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

Figure 3. Methods of Construction of Synchronization Systems Using a Combined Frame Pause. Key: 1, Distributed Frame Pause Recognition Device

Thus, in spite of the fact that the initial sequence does not contain the frame pause in explicit form, linear processing of the arriving data can separate the distributed frame pause and allow it to be utilized for frame synchronization of the initial data.

After frame synchronization, the sector of the sequence is divided into words, the position of each word in the frame being known. This allows error detection. If relationship (5) is fulfilled for the ith word, all characters in the word have been properly received (or, significantly less probably, an even number of errors have occurred). If relationship (5) is not fulfilled, there are errors in the word. Thus, the same characters (c_i) are first used for frame synchronization, then for error detection.

Completing our analysis of the combined frame pause, let us note two peculiarities related to its application and distinguishing it from the distributed frame pause.

The first comment is related to the volume of memory required for synchronization using a combined frame pause.

The simplest case is represented on Figure 3a. Register A, containing n locations, is used to form the sums d_{ij}. Register B contains the N' - n characters of the sequence received (total volume of registers A and B equal to the number of

characters in a frame). However, frame synchronization is performed using sums $d_{i,j}$, the circuit contains register C for the N' - n sums $d_{i,j}$, which is filled simultaneously with register B and connected to the device for recognition of the distributed frame pause. The set of data in registers A and B at the moment of recognition corresponds to one full frame of telemetry data.

Thus, the circuit of Figure 3a requires about 2N' memory locations, i.e., twice as much as when the distributed frame pause is used. The simplest and most obvious plans for reducing the required memory volume are as shown in Figures 3b and 3c.

The circuit of Figure 3b contains no C register. It utilizes the fact that only m different sums d_{ij} arrive at the recognition circuit simultaneously. These sums are calculated using n-input modulo two adders. /259

An n-input adder is equivalent to n - 1 two-input adders. Using these devices, the circuit contains m(n - 1) = N'm two-input adders. Thus, eliminating N' - n memory locations (in comparison to the circuit of Figure 3a), we have been forced to introduce N' - m - n + 1 two-input modulo two adders. Consequently the comparative complexity of the circuits of Figures 3a and 3b is determined by the comparative complexity of a shift register location and a two-input modulo two adder.

Register C is eliminated in the circuit of Figure 3b because registers B and C contain the same data in different forms. This same fact is used in the circuit of Figure 3c, where register B is almost completely eliminated, replaced by register B', containing only n locations.

Let us assume that at a certain moment in time register C contains the values of sums $d_{i,j}$, $d_{i,j+1}$, $d_{i,j+2}$, etc. written from right to left; register B' contains $a_{i,j}$, $a_{i,j+1}$, ..., $a_{i+1,j-1}$, written from left to right. We note that the modulo two sum of all characters in register B' is equal to sum $d_{i,j}$ written in the far right location of register C.

The connection of registers C and B' is shown on Figure 3c. If we consider that, similarly to (5),

$$d_{i,j+1} = a_{i,j+1} \oplus a_{i,j+2} \oplus \ldots \oplus a_{i+1,j-1} \oplus a_{i+1,j},$$

then, using this relationship (5) and the fact that $a \oplus a = 0$, we find that the input of register B' receives the quantity

$$d_{i,j} + d_{i,j+1} \oplus a_{i,j} = a_{i+1,j}. \quad (7)$$

Thus, as a result of each successive shift in the registers, register C will contain (from right to left) the sums $d_{i,j+1}$, $d_{i,j+2}$, ..., while register B' (from left to right) will contain $a_{i,j+1}$, $a_{i,j+2}$, ..., $a_{j+1,i}$. As before, the sum of all characters in shift register B' is equal to the sum written in the far right location of register C. This means that in the process of operation of the circuit the input of register B' successively receives the values of all characters in the sequence received, beginning with a_{ij}. Thus, the circuit of Figure 3c contains no register B, but register B' allows us to successively restore the initial sequence of characters and divide it into frames, performing synchronization.

The process of initial filling of registers C and B' represents no serious difficulties, and its analysis will be omitted here.

The circuit of Figure 3c contains N' + n shift register locations and two adders -- one with n inputs and one with three inputs. Consequently, this circuit has approximately the same complexity as the synchronization circuit using a distributed frame pause (see Figure 2). However, we should note that the elements of this circuit must meet higher requirements for reliability. Actually, if one of the locations of register C or B' drops a character, the calculations performed according to formula (7) will produce improper results and the "track" resulting from the single dropout will continue without limit. This dropout has little or no influence on the synchronization process, but an error is introduced to register B', and will continue to circulate there, causing a deterioration in the quality of the data restored.

The circuits of Figure 3 can operate with either one or m recognition devices (see Figure 2). When one recognition device is used, any one of these circuits synchronizes the first full frame received, i.e., only the initial portion of the sequence of data, not forming a full frame, is lost. The introduction of m recognition devices either eliminates the delay of synchronization or eliminates the loss of the initial portion of the data (allowing synchronization relative to the first word, but with delay). /260

If we use m synchronization devices and allow the initial portion of the data, varying in volume from N' - n to N' characters, to be lost, we can use the circuit of Figure 3a without register B. Actually, the recognition device is switched on after reception of N' characters, then the recognition

indication appears at the output of one of the m devices at an interval of n characters, meaning that the n characters written in register A correspond to one word of the frame. The total memory volume is equal to N', the circuit contains one n-input adder, and no severe requirements are placed on the elements of the circuit. Furthermore, delays in synchronization are completely eliminated.

Even this rather superficial analysis shows us that when a combined frame pause is used, it is quite possible to construct a circuit containing about N' memory locations and practically equal in complexity to the circuit used with a distributed frame pause.

One second peculiarity distinguishing a combined frame pause from a distributed or concentrated frame pause is the probability of proper synchronization. If the probability of distortion of one character during transmission is $\varepsilon << 1$, the value of the sum $a_{i,1} \oplus a_{i,2} \oplus \ldots \oplus a_{i,n-1} \oplus c_i = b_i$ will be calculated improperly with probability $\varepsilon' \approx n\varepsilon$. Since when a combined frame pause is used, synchronization is performed on the basis of analysis of sums $d_{i,j}$, replacement of the distributed frame pause with a combined frame pause is equivalent to increasing the probability of distortion of one "character" by n times. Obviously, with otherwise equivalent conditions, the resulting probability of proper synchronization is lower.

This statement is correct if the lengths of the distributed (or concentrated) and combined frame pauses (s and m) are equal. However, it was assumed above that s = m only for greater clarity of the explanation. Generally speaking, this is not necessary at all. The value of s is determined by the required probability of proper synchronization, the value of m by the structure of a frame (number of words in a frame). In many cases, m is significantly greater than s. Therefore, using a combined frame pause of length $m > s$, we can produce an equal or better result than when we use a distributed or concentrated frame pause of length s. We will not present here the results of comparison as a function of ε, s, m and n, since the result will also depend on the synchronization algorithm and the correlation functions of the frame pauses used. However, in each specific case it is easy to perform the corresponding calculations.

Finally, we should note that it is not difficult to use combined and distributed (or concentrated) frame pauses simultaneously. For example, suppose s is the length of a distributed or concentrated frame pause providing the required probability of proper synchronization, $m \leq s$. If we now introduce an

additional combined frame pause of length m (not requiring the introduction of additional characters) and perform synchronization by both (distributed or concentrated and combined) simultaneously, the probability of proper synchronization is increased. We can then decrease s to a certain value s', for which combined application of the two frame pauses still allows the required probability of proper synchronization to be /261 achieved. Depending on the values of ε, s, m, n, etc. the value will be somewhat less than s. If at the same time m > s, it may occur that s' simply vanishes and the combined frame pause remains. This case was mentioned above.

Conclusions

1. Concentration of the characters of a frame pause allows simple synchronization circuits to be used. However, at the same time, this frame pause does not allow localization of a dropout or insertion of a character, is not protected from group errors, etc.

2. Even distribution of the frame pause characters within a frame eliminates these shortcomings but increases the required memory volume in the synchronizing device to the size of a frame.

3. If a message within a frame is divided into words of equal length, each of which contains one parity-check bit, this can be used for purposes of frame synchronization. The parity-check characters are used both for frame synchronization and for ordinary quality control of the data received. The frame pause, implicitly introduced by means of the parity-check characters, is called a combined frame pause. It has all advantages of a distributed frame pause and requires a synchronization circuit of approximately the same complexity.

4. Simultaneous use of concentrated or distributed and combined frame pauses is possible. The use of a combined frame pause, not requiring the addition of extra characters, can reduce the length of the concentrated or distributed frame pause or simply eliminate it.

REFERENCES

1. Good, G. E., J. L. Phillips, Optimum PCM Frame Synchronization Codes and Correlation Detection, Proc. National Telemetering Conf., 1961, pp. 11-13.
2. Maury, J. L., F. J. Styles, Development of Optimum Frame Synchronization Codes for Goddard Space Flight Center PCM Telemetry Standards, Proc. National Telemetering Conf., 1964, pt. 3-1.
3. Meshkovskiy, K. A., N. Ye. Kirillov, Kodirovaniye v tekhnike svyazi [Coding in Communications Technology], Svyas' Press, 1966.
4. Hawkes, T. A., J. Dupraz, PCM Frame Synchronization with a Self-Varying Threshold, IEEE Trans. on Aerospace and Electronic Systems, 1967, AES-3, March, No. 2, pp. 158-164.
5. Hawkes, T. A., J. Dupraz, Girault, J. H., Anglade, P., Construction and Performance of a PCM Frame Synchronizer with Self-Varying Threshold, IEEE Trans. of Communication Technology, 1968, COM-16, February, No. 1, pp. 142-148.
6. Marvin, P., The Synthesis of Binary Sequence Detectors, IEEE Trans. on Computers, 1968, Vol. 17, No. 9, pp. 873-880.

ABSTRACTS OF ARTICLES

Determination of Landing Point and Orientation of a Spacecraft on Mars, G. A. Mersov (p. 1)

This article presents formulas to calculate the landing point of a spacecraft on Mars. The areographic latitude is taken as the angle between a plumb line at the landing point and the plane perpendicular to the axis of rotation of Mars relative to the sun. The ephemeral time of the upper culmination of the sun at the landing point is used to determine areographic longitude. A coupled system and a topocentric system of coordinates are introduced to develop the first series of equations. Both errors which are independent random quantities and systematic errors are considered, as well as the influence of the distance of the landing point from the boundary of the area of polar night.

Low-Noise Parametric Amplifiers for Radio Astronomy Observations at 18-21 cm Wavelength, B. Z. Kanevskiy, V. M Veselow, I. A. Strukov, V. S. Etkin (p. 23)

The principle characteristics and use of SHF parametric amplifiers for radiometer input devices are explored. Balanced parametric amplifiers (BPA) are considered as the SHF signal amplifiers allowing production of the amplifier circuit without a special filter to achieve decoupling. Formulas to calculate the basic parameters of a BPA are given. A modulator based on coaxial lines is discussed as the input element of the SHF. Results of laboratory tests of the receiver section and long-term stability studies of the SHF sector are presented.

Problems of the Design of Low-Noise Input Devices, V. M. Manokhin, Yu. A. Nemlikher, I. A. Strukov, Ye. A. Sharkov (p. 28)

This article presents an analysis of the requirements placed on the elements of parametric centimeter waveband amplifiers for achievement of minimal noise temperatures. A low-noise semiconductor parametric amplifier using germanium parametric diodes for a receiver operating in the 4 GHz band was developed and tested confirming the possibility of satisfying all requirements.

Photogrammetric Method of Determination of Deflection of Light Beams by Porthole Glass Under Space Flight Conditions, Ya. L. Ziman (p. 36)

It is suggested that a measuring camera be used to determine the deflection of light rays through the glass portholes of hermetically sealed spacecraft. It is desirable that the camera have the greatest possible lens aperture diameter to increase the number and steller magnitude of stars photographed. The article describes comparison of photographs within and outside the spacecraft and recommends similar tests in flight using oriented spacecraft.

Development of the Design of a Between-the-Lens Band-Type Photographic Shutter, A. S. Dorofeyev (p. 42)

When photographic equipment is miniaturized, in order to conserve weight and size for spacecraft use, by using wide angle, short focus lenses, a band-type central between-the-lens shutter must be used. The rectangular slot shape was the form with the highest optical efficiency. To overcome the problem of burrs caused in the punching out of the slots in a carbon steel band, a shape limited by arcs with radius equal to half the slot width is preferred.

Method of Radar Studies of the Moon and the Planets from Spacecraft, N. N. Krupenio (p. 48)

Radar studies of the moon and planets from the Earth have been valuable but limited in comparison to radar studies from spacecraft. Autonomous and non-autonomous methods of radar research from space satellites and probes are considered. Results of both the Luna and Surveyor series of lunar spacecraft are presented. It is concluded that use of radar in spacecraft allows valuable, high resolution measurement even near the limb and on the back side of the moon.

Multiple-Slit Shadow Device for Sensing the Direction to a Shining Object, A. V. Bruns (p. 65)

A multi-slit shadow direction sensor of jalousie type (JDS) was developed to provide a simple, highly reliable tracking system to guide a rotating mirror toward the sun so that the reflected light would illuminate the input slit of a

photoelectric spectrometer. The operating principle and stability are considered as well as determination of initial signal level at the output of each photosensor. A procedure is recommended for calculation of a JDS. Both one- and two-coordinate sun direction sensors are described and considered for use in the tracking system of a photoelectric spectrometer.

Computers and Prospects for On-Board Spacecraft Electronic Equipment, V. V. Andreyanov (p. 88)

This is a study of how telemetry and transceiving devices can be changed in digital computers are available. The ways computers may allow selection of the most informative of the total number of measurements are discussed along with development of compatible transceiver equipment.

The Effectiveness of Correcting Codes in Reception in the Whole in Additive Normal White Noise, Yu. M. Shtar'kov (p. 97)

Some possible criteria for estimating the effectiveness of correcting codes are presented, and the energy effectiveness of correcting codes is studied for symbol-by-symbol reception. Expressions for the energetic effectiveness of binary correcting codes for reception in the whole are produced. Asymptotic energetic effectiveness and finite signal/noise ratio cases are considered.

Selection of an Optimal Plan for Construction of a Spacecraft Antenna System, B. A. Prigoda, M. B. Faynshteyn, V. A. Gur'yev, A. S. Markin (p. 107)

This article presents an attempt to summarise the experience of the design of antennas for various types of spacecraft: fully oriented, partially oriented and with arbitrary orientation. Several examples of the construction of antenna systems and their characteristics are noted. The optimal relationship between weight, dimension and power characteristics of on-board devices allows selection of the most acceptable versions. Questions are presented for consideration in design of optimal spacecraft antenna-feeder systems.

Study of the Redundancy of Telemetry Information of the Luna-9 and Luna-13 Automatic Lunar Probes, A. V. Kantor, T. A. Tolmadzheva (p. 120)

Knowledge of the statistical characteristics of natural telemetry information is very important for the development of principles and methods of reduction of redundancy of telemetry messages as well as planning of on-board information compression devices. The main purpose of this article is to determine these characteristics for the telemetry information of the Luna-9 and Luna-13 automatic lunar probes, to allow us to estimate the redundancy of the information. Histograms as well as numerical characteristics are used.

Digital Computer Modeling of the Process of Reduction of Redundancy in Multichannel Telemetry Information by the Method of Adaptive Discretization with Associative Sorting, T. A. Tolmadzheva, A. V. Kantor, L. V. Rozhkovskiy (p. 136)

The present work is dedicated to digital computer modeling of the process of adaptive discretization with associative sorting (ADAS) of actual multichannel telemetry information. The main task in modeling is production of initial data for determination of dependences describing the operation of the on-board information compression device. Conclusions are presented including the shortcomings of telemetric information used in modeling.

Circularly Polarized Antennas with Controlled Radiation Patterns, B. A. Prigoda, V. A. Gur'yev, N. S. Ul'yanin, N. N. Kalugin (p. 163)

Log-helix conical antennas are examined as a way to satisfy the requirement for circular polarization of antennas for spacecraft use. It is concluded that a multiple-helix antenna which allows the helix lead angle to be changed can be used when the radiation and directivity in the direction of the reception point must be changed during flight.

Reduction in Redundancy of Multichannel Telemetric Information by the Method of Adaptive Discretization with Associative Sorting, A. V. Kantor, V. G. Timonin, Yu. S. Azarova (p. 168)

The method of adaptive discretization is the most promising for elimination of redundancy from telemetry messages characterized by signal shape. This article considers adaptive discretization with associative sorting as a way to avoir the shortcomings of adaptive discretization with buffer smoothing and adaptive discretization with logical switching in on-board information compression devices (OICD) in spacecraft. Mathematical investigations of OICD are presented.

The Question of the Possibility of Using Associative Computer Devices (ACD) for Parallel Adaptive Discretization of Multichannel Telemetry Information, A. V. Kantor, S. M. Perevertkin, T. S. Shcherbakova (p. 195)

The method of parallel adaptive discretization of data is considered the most promising and allows the effective compression algorithms to be used for high information-capacity radio telemetry systems. An associative computer device (ACD), i.e., parallel computers based on associative memory units (AMU), are recommended for realization of this method. A detailed discussion of the problems of application of AMU is followed by description of a particular ACD, and its use is recommended.

High Frequency Switching Devices, B. A. Prigoda, Ye. A. Vlasov, A. S. Markin, E. I. Yakovlev (p. 217)

This article investigates use of switching devices using various circuits of electronic or mechanical elements depending on the operating conditions and purpose of the switching.

SHF Filters in Spacecraft Antenna-Feeder Systems, Ye. A. Vlasov, B. A. Prigoda (p. 222)

This article investigates use of SHF-band filters in spacecraft antenna-feeder systems. These filters, depending on specific applications and requirements, may be constructed both of elements with lumped constants and of elements with distributed parameters.

Checking the Parameters of AFS in Spacecraft in Flight, Ye. A. Vlasov, B. A. Prigoda, A. S. Markin (p. 228)

The use of elements for remote testing of matching of on-board spacecraft antennas, required by the need to receive timely telemetry information on the quality of operation of antenna-feeder systems (AFS) during long flights, is discussed in this article. Information on the traveling-wave ratio deficiency to the AFS allowed improvement of quality and reliability of communications with spacecraft and developed telemetric information on the matching of on-board antennas, the efficiency of the feeder system and power of transmitters during long flights.

The Accuracy of Automatic Tracking, V. V. Kastrov (p. 235)

It has been generally assumed that tracking accuracy changes similarly to the rate of change of the curve of the measurement conversion. This article further takes into account the problem that internal noise increases along with the signals processed by the tracking device and that tracking accuracy thus drops. The main prerequisite for solution is consideration of the dependences of the output signal of the tracking device sensor not only on the measured parameter but on the signal itself.

Communications by Light with Long-Range Spacecraft, M. Z. Novikov (p. 246)

This article considers light contact with long-range spacecraft flying between the Earth and the moon. The possibility of creating on-board light signaling devices and the problems of construction of light signal receiving devices are studied. It is concluded that signaling using constant burning lamps is most favorable for observation of the spacecraft and transmission of information from the spacecraft.

Data Compression, V. F. Babkin, A. B. Kryukov, Yu. M. Shtar'kov (p. 264)

An approach to data compression is discussed in which the effect achieved by compression is evaluated by the closeness of the approach to the minimum possible volume. The authors attempt to systematize the known results on data compression. The review contains: description of methods of data compression based on statistical coding and information

theory; application of methods of interpolation and extrapolation; a specific compression method (related to description of the histogram of a sample by means of a set of quantiles); some criteria of effectiveness and methods of service information representation; discussion of models suggested for theoretical analysis.

Calculation of the Effectiveness of a Gamma Telescope, Yu. A. Rylov (p. 319)

This work contains a calculation of the effectiveness of recording of gamma quanta by a gamma telescope. Consideration of the isotropic nature of the stream of gamma quanta reaching the gamma telescope revealed that low energy gamma quanta are recorded relatively more effectively, since they are recorded from a wider solid angle than higher energy gamma quanta. The problem is solved of determining the effectiveness of recording of gamma quanta and determination of their energies by a gamma telescope.

The Problem of Controlling the Directivity of Low-Directivity Antennas on Spacecraft in Flight, V. I. Mashkov, B. A. Prigoda (p. 335)

Interest exists in the creation of antennas allowing adjustment of the radiation pattern and directivity in flight without using complex switching and matching circuits. This report presents the results of certain experimental tests performed with log-periodic conical helical antennas to determine the possibilities of regulation of their radiation patterns without changing the structure of the current-carrying elements, supply connections, etc.

Some Problems of the Design of Highly Directional Spacecraft Antennas, B. A. Prigoda (p. 339)

Problems of optimization and selection of the most expedient forms of design of directional antenna systems encountered in spacecraft design form the subject of this article. Selection of a given type of antenna depends on its characteristic size, weight, potential and other data considered here.

Potential Capabilities for Compression of Information of Certain Data Processing Systems, Yu. K. Khodarev, V. P. Yevdokimov, V. M. Pokras (p. 353)

This article undertakes to study a generalized block diagram of a data collection and processing system of a spacecraft in which a number of sensors or outputs of scientific instruments are cyclically interrogated by a commutator, methods of writing the supplementary information in a frame on the example of a certain hypothetical telemetry system and the influence of statistics of number of active channels in a frame on frame compression factor. The separation of the data compression factor of the collection and processing system of spacecraft into two parts used in this work allows determination of the compression factor of an active frame depending not only on the statistics of activity of channels in the telemetry frame, but also on the method of introduction of the additional address and time information to each frame.

Statistical Analysis of Information from Long-Range Spacecraft, Yu. K. Khodarev, V. P. Yevdokimov, V. M. Pokras (p. 368)

This article presents the results of analysis of statistical characteristics of information from spacecraft data compression systems, ignoring the nature of channel activity. Analysis of the distribution of the number of active channels and the channel activities was performed using data from scientific experiments conducted on the Zond-1 and Venera-4 spacecraft.

A Priori Estimate of the Quality of a Data Compression System Based on Statistical Characteristics of the Sensors Used, Yu. K. Khodarev, V. P. Yevdokimov, V. M. Pokras (p. 379)

Knowledge of the composition and certain statistical characteristics of the output signals of the scientific instruments on board a space vehicle allows determination of possible signal processing algorithms in the system and to evaluate the prospects for utilization of data reduction. A method is presented in this article to estimate the compression factor of an on-board data collection and processing system with known mean activities of the sensors used and unknown mutual correlation of their output signals.

Methods of Synchronization of Telemetry Frames Using Distributed and Combined Frame Pauses, Yu. M. Shtar'kov (p. 387)

In this work, two modifications of the method of frame synchronization by means of a frame pause are studied. The first does not require that all characters in the frame pause follow each other immediately. The second modification is based on the use of one particular type of redundancy within the frame to achieve frame synchronization without setting aside a special frame pause. Concentrated, distributed and combined frame pause are described and their uses compared.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1974—635-044/33