Engineering of Science and Military Technologies
Volume (2) - Issue (3) - July 2018

ISSN: 2357-0954
DOI: 110.21608/ejmtc.2018.939.1041

Original Article

A two-stage rice-based lossless compression method for telemetry data

Mohamed A. Elshafey

Department Computer Engineering, Military Technical College, Cairo, Egypt

Keywords:
IRIG-106, lossless compression, rice coding, telemetry data, XOR Operation.

Corresponding Author:
Mohamed A. Elshafey, Department of Computer Engineering, Military Technical College, Cairo, Egypt,
Tel: (+20)1158546666
Email: m.shafey@mtc.edu.eg

Abstract
The article presents results of a study of a proposed lossless compression method for telemetry data. A two-stage scheme of lossless compression, consisting of decorrelation and entropy coding, is presented. In experiments, a bitwise XOR operation is implemented as a decorrelator and rice coding method as entropy coder. A comparison is performed between two compression methods: one is based only on rice coding method, while the other is based on the proposed method. The results are based on estimates of the gain in variance and entropy of the output signal from the decorrelator. In experiments, parameters of telemetry information of automatic control systems are studied, such as temperature, pressure and positioning data. Streams of telemetry frames, with different frame structures: one and two level of commutation, in IRIG-106 standard format, are formed and tested. Based on experimental results, conclusions are done for lossless telemetry data compression.

I. INTRODUCTION

Nowadays, telemetry systems are widely used for both control and gathering information at some remote locations and transmitting data to a convenient distant location to be displayed, analyzed and recorded. The transmission media may be air and space for satellite and airborne applications, or copper wire and fiber cable for static ground environment[1].

Telemetry data has a huge size, so it should be compressed at the telemetry source before transmission and/or archived in compressed form, at the receiving and registering station, for later processing.

Normal requirement for telemetry data compression algorithms is the ability to recover initial data without loss of information. for the analysis of any kind of abnormal events, recovery of bad sites within the telemetry data stream and for any other types of post- or real-time data processing[2–6].

In this paper, a proposed two-stage compression method, based on XOR method and Rice coding method, is presented. Rice coding algorithm is widely used for telemetry data compression without loss, but it may be not suitable if the data is highly deviated and has high variances. The proposed lossless compression method is implemented, tested and verified with real telemetry data, in standard format IRIG-106[7], in different frame structures.

II. TELEMETRY SYSTEMS OVERVIEW

A. The overall telemetry system

The generic telemetry system is shown in Fig. 1. In data acquisition subsystem, sensors measure the amount of a physical attribute and transform the measurement to an engineering unit value. The commutator measures values from multiple different sensors and outputs a single stream of pulses, each with a voltage relative to the respective measured value (Time Division Multiplexing, TDM). The analog multiplexed data is then transformed to a digital form, then modulated and transmitted via radio transmitter and antenna, coax cable, telephone line, etc. to a receiving station. At the receiving station, the received data stream is amplified and the decommutator then recognizes the synchronization pattern and returns the serial digital stream to parallel data[1].

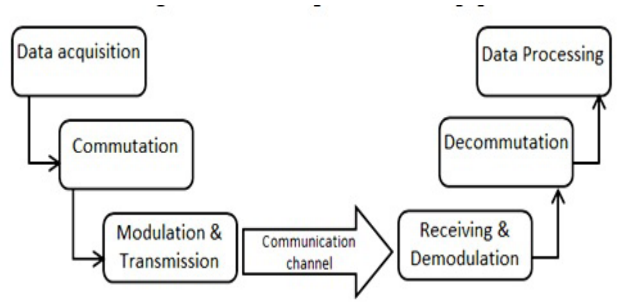
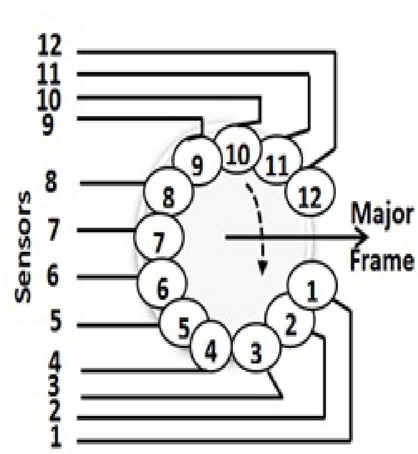


Fig. 1: Generic telemetry system

B. Telemetry frame structure

Telemetry stream is a sequentially transmitted telemetry frames. A complete scan by the commutator (one revolution) produces data words of the telemetry frame, which contains the value of each measured parameter. Every scan produces the same sequence of words. For the process of decommutation, a unique frame synchronization code is added to each frame to serve as a reference and again the values of the parameters are extracted into individual measured values. The length of the frame synchronization code is longer than usual data words to reduce the probability of actual data matching[1,8].

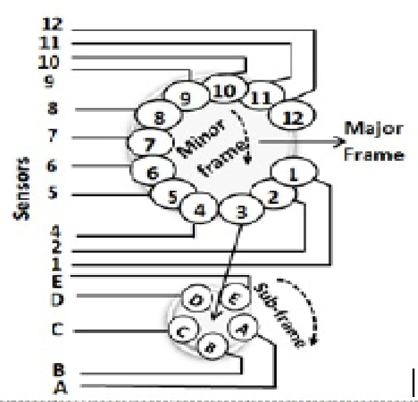
Fig. 2 shows a commutator (one level of commutation), which multiplexes (TDM) values of (12) sensors in a Major_Frame.



Frame Synchronization	Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	Sensor 6	Sensor 7	Sensor 8	Sensor 9	Sensor 10	Sensor 11	Sensor 12

Fig. 2: One level of commutation and the resulting Major_Frame structure.

In most of telemetry applications, measured values change at different rates, often by several orders of magnitude. The slowest changing data may not even require sampling once per frame. Multiple slow changing parameters values can share a single frame word (subcommutation), as shown in data word number (3) in Fig. 3. To distinguish between the shared positions in Minor_Frames, a Sub_Frame synchronization pattern is required (normally as a counter)[1,8].



Frame Synch.	Sensor 1	Sensor 2	Sensor A	Sensor 4	Sensor 5	Sensor 6	Sensor 7	Sensor 8	Sensor 9	Sensor 10	Sensor 11	Sensor 12
Frame Synch.	Sensor 1	Sensor 2	Sensor B	Sensor 4	Sensor 5	Sensor 6	Sensor 7	Sensor 8	Sensor 9	Sensor 10	Sensor 11	Sensor 12
Frame Synch.	Sensor 1	Sensor 2	Sensor C	Sensor 4	Sensor 5	Sensor 6	Sensor 7	Sensor 8	Sensor 9	Sensor 10	Sensor 11	Sensor 12
Frame Synch.	Sensor 1	Sensor 2	Sensor D	Sensor 4	Sensor 5	Sensor 6	Sensor 7	Sensor 8	Sensor 9	Sensor 10	Sensor 11	Sensor 12
Frame Synch.	Sensor 1	Sensor 2	Sensor E	Sensor 4	Sensor 5	Sensor 6	Sensor 7	Sensor 8	Sensor 9	Sensor 10	Sensor 11	Sensor 12

Fig. 3: Two levels of commutation and the resulting Major_Frame structure, which includes (5) Minor_Frames and a Sub_Frame transmitted at data word number (3) in the Major_Frame.

III. PROPOSED LOSSLESS COMPRESSION METHOD

The effectiveness of methods for lossless telemetry data compression is largely determined by the properties of the data to be compressed. Compression algorithms show better results if they can adapt with the characteristics of the input data, which are in most cases rapidly change[9–11]. One of the most important algorithms that are widely used for lossless compression of telemetry data is Rice coding algorithm[4,12].

Modern lossless compression algorithms can be described as two stages: The first stage: decorrelation stage, which exploits the redundancy between the neighbouring samples in the data sequence and The second stage is entropy coding, which takes advantage from decreasing variance and lowering entropy value of the data (entropy value is the smallest average number of bits needed to represent the source output symbol), resulting from the first stage[9–11] as in Fig. 4.

Efficiency of the decorrelation stage can be evaluated by two parameters[9–11]:

- Variance gain *Gain_Var*, which is the relation between variance σ_X^2 of the data input (X) to the decorrelator and variance σ_E^2 of data (E) outputting from it:

$$Gain\_Var = \sigma_X^2 / \sigma_E^2 \qquad (1)$$

- The entropy *H*:

$$H = -\sum_{i=1}^{L} P_i * \log_2(P_i) \quad \text{(bits / symbol)} \qquad (2)$$

Where L is the number of symbols in the encoded message (length of the alphabet) and Pi is the probability of each symbol (i) in the alphabet.

For evaluating the efficiency of the compression algorithm, the compression ratio (Cr) is checked:

$$C_r = V_{input} / V_{output} \qquad (3)$$

Where V_{input} is the size of the source data (X) at the input of the system, and V_{output} is the size compressed data (Y) at the system output (output of entropy coder).

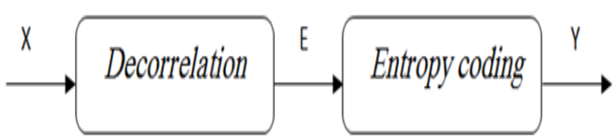


Fig. 4: Two-stage lossless compression scheme

This paper presents the results of research carried out in order to develop an effective method of two stage lossless compression for telemetry data stream. The presented method involves implementing XOR operation to telemetry data frames, and then performing entropy coding for the XORed data using Rice coding method.

A. Rice coding overview

Rice codes[12] are a special case of Golomb codes, which are effective for encoding processes, which match with the geometric distribution:

$$P(i) = (1-\zeta) * \zeta^i , i \in [0,1] \qquad (4)$$

The length of code word for symbol (*i*) is $L_i = \lfloor i/m \rfloor + \lceil \log_2(m) \rceil + 1$, Where (m) is an adjustable parameter for the coding algorithm. First $\lfloor i/m \rfloor$ bits of the code word are zeros. This is followed by a bit equal to unity, which divides the code word into two parts. The following $\lceil \log_2(m) \rceil$ bits are used to hold the remaining of (i/m).

If $m = 2^k$, then the code is called Rice code or Golomb-Rice code. This version of the Golomb code is used in several compression algorithms due to its simple hardware realization. Parameter *k* is linearly related to the variance of the data symbols. In this paper, the following formula[9,13,14] are used in experiments:

$$k = \log_2(\log_e(2) * E(|x|)) \qquad (5)$$

Where $E(|x|)$ is the average value of symbols of the encoded message.

A code word for an encoded number (v) by Rice coding algorithm is divided into two parts: $v = v_i + v_f$, where $v_i = \lfloor v/2^k \rfloor$ and $v_f = v \bmod 2^k$. The coded word is formed by value of (v_i +1), provided by unary code, and value of vf , represented by (*k*) bit binary code.

Rice codes are effective for coding values, which are close to , and are recommended for the encoding processes, which match the geometric distribution. Fig. 5 shows the relationship between lengths of some Rice codes and encoded values.

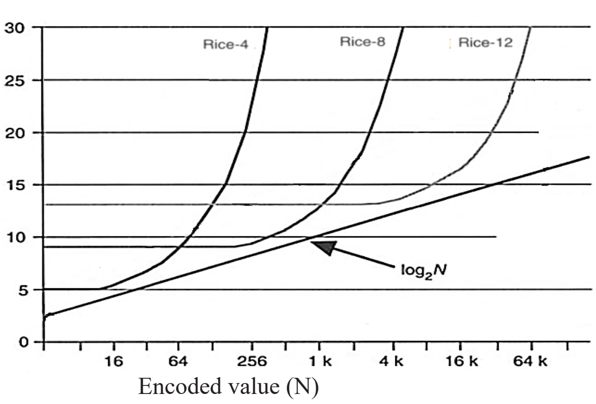


Fig. 5: Relationship between some Rice codes lengths and encoded values.

Efficiency of Rice codes depends on a controllable parameter (k), the value of which depends on the variances of symbols values in the message.

B. Probability model of encoded data

The encoded data should have a non-uniform probability distribution and centred on an average value. One of the distribution models corresponding to this situation is Laplacian distribution, which is quite picked at zero mean and with variance as defined by the formula[15]:

$$f(x) = \frac{1}{\sqrt{2\sigma^2}} e^{\frac{-\sqrt{2}|x|}{\sigma}} \qquad (6)$$

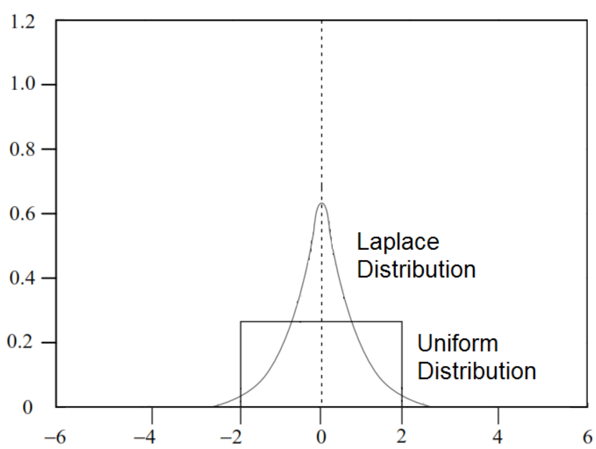


Fig. 6: Uniform distribution versus Laplacian.

C. XOR operation as decorrelator

This paper presents a simple method for reducing correlational dependences in adjacent frames in telemetry stream, in which for a pair of telemetry words, located at the same position in adjacent Major_Frames, a bitwise "Exclusive OR" operation is carried out.

Fig. 7 shows how XOR operation can be applied to Major_Frames of one level of commutation in a telemetry stream, while Fig. 8 shows XOR operation for Major_ Frames, resulting of two level of commutation with a Sub_ Frame at the 3rd data word.

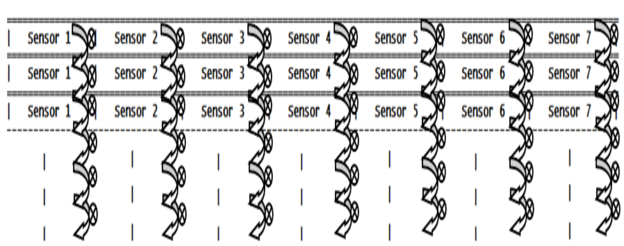


Fig. 7: XOR operation for one level of commutation frames.

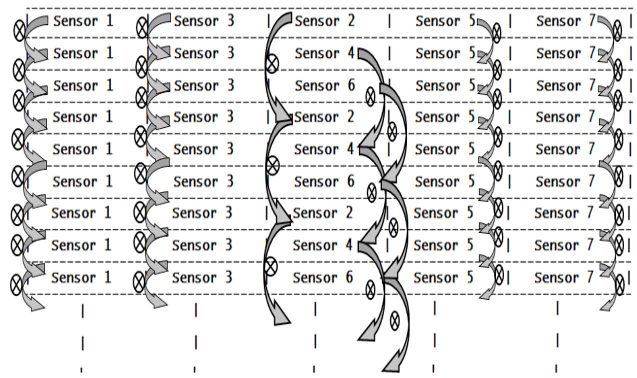


Fig. 8: XOR operation for two level of commutation frames.

IV. PREPARING DATA FOR EXPERIMENTS

Telemetry data samples, used in experiments, are represented in the IRIG-106 format [7]. It is an open standard of Telemetry Group (RCC), developed for use in aerospace industry, and is nowadays widely used in registering telemetry systems for various purposes. The information in the telemetry stream is transmitted in frames with a fixed length and a predetermined fixed internal structure. In frames, different readings of several sources of information; the digitized analog sensor readings and readings of digital devices, can be transmitted. Each data source is transmitted in a separate channel of telemetry data recording system. Current value of the channel is transmitted in a data word of telemetry frame at a given displacement from the beginning of the frame. All data words in telemetry frames are (8) bits wide, in which a quantized readout of a sensor signal is stored in a scale value of $0:255$.

The digitized data samples of analog signals, which represent typical telemetry parameters of automatic control system, such as temperature, pressure and positioning information, are obtained in the laboratory from physical real sources. Then, these data were inclosed in different telemetry frame structures in standard format IRIG-106. The telemetry stream itself is made by the telemetry simulation software described[16]. The software is a MATLAB–based simulator for generating different telemetry data streams, based on different telemetry frame structures, in IRIG–106 telemetry standard. The simulator can be considered a tool for research and design of different efficient algorithms for telemetry data processing, including data compression codes, error correcting based algorithms, and frame synchronization methods. Description of different simulated telemetry signals is provided, including test patterns, harmonic signals, arbitrary digital sequences, recorded data, and filtered signals. The simulator supports different telemetry frame structures till commutation based on three stage of circuit switching with different frame synchronization codes in IRIG-106 standard format. The simulator is able to simulate the effect of imposing the telemetry data stream to Additive White Gaussian Noise (AWGN) on communication channel during transmission, as well as the effect of the noise from the on-board sensors during acquisition phase.

For experiments, several commutation schemes are used and, based on these developed schemes, different telemetry data streams are generated[16]. Fig. 9 shows a timed portion of the acquired data samples.

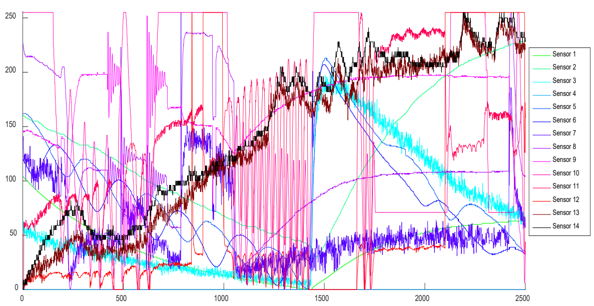


Fig. 9: A sample of the acquired data samples for (14) sensors.

Fig. 10 and Fig. 11 show two different types of telemetry frame structures, which are used in experiments (data word length for each of them equals to 8 bits long). The first type (Type-1) of examined frame structure, Fig. 10, is a telemetry frame structure, which consists of 14 channels in a Major_Frame without any Sub_Frames (which means one level of commutation). This frame further also includes additional information (SYNC_F), which provides frame synchronization in the telemetry stream.

SYNC_F	DW1	DW2	DW3	DW4	DW5	DW6	DW7	DW8	DW9	DW10	DW11	DW12	DW13	DW14
SYNC16	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14

Fig. 10: Telemetry frame structure without Sub_Frames (one level of commutation), Type-1.

The second type (Type-2) of the examined frame structures, Fig. 11, is a telemetry frame structure, which contains a Sub_Frame, resulting from a second level of commutation. Fig. 11 shows 5 Minor_Frames in IRIG-106 standard format, in which the Major_Frame contains one Sub_Frame of length (5) and this Sub_Frame is attached to the 6th data word. In that 6th channel of the main commutator, values are passed to the second commutation level. Thus, a complete cycle, which reads a complete set of telemetry parameters takes 5 frames of the main commutator (first level of commutation). During this cycle, channels of the main commutator are sampled 5 times.

SYNC_F	SYNC_SF1	DW1	DW2	DW3	DW4	DW5	DW6	DW7	DW8	DW9	DW10
SYNC16	C1	S2	S4	S6	S8	S9	S1	S10	S11	S13	S14
SYNC16	C2	S2	S4	S6	S8	S9	S3	S10	S11	S13	S14
SYNC16	C3	S2	S4	S6	S8	S9	S5	S10	S11	S13	S14
SYNC16	C4	S2	S4	S6	S8	S9	S7	S10	S11	S13	S14
SYNC16	C5	S2	S4	S6	S8	S9	S12	S10	S11	S13	S14

Fig. 11. Telemetry frame structure with a Sub_Frame of length (5) at the 6th data word (two level of commutation), Type-2.

V. EXPERIMENTAL TEST AND ANALYSIS

Fig. 12 shows the probability distribution of telemetry data samples in the studied telemetry streams. The graph shows that the values are distributed over the entire range of values $\overline{0:255}$. The data samples has entropy value of H_x= 7.37 (bits/symbol) and variance σ_X^2 = 5765.8 .

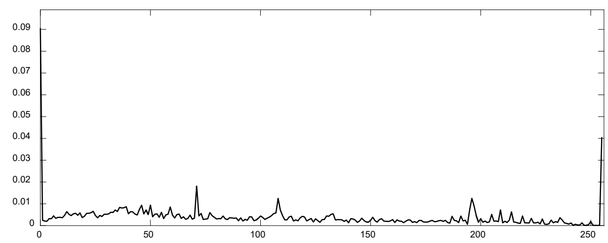


Fig. 12: Probability distribution of data samples in a telemetry stream.

Carrying out one-stage compression, based only on Rice coding method (without decorrelation stage), produced no compression gain (compression ratio C_r = 0.96) for estimated parameter k = 7 using (5).

Experiments, using the presented two stage compression method, which applied on generated telemetry streams, of telemetry frames Type-1 and Type-2 separately, generate the following results:

-The probability distribution of the decorrelated data (E) based on XOR operation is shown in Fig. 13, in which the graph has an expressed peak at zero, allowing qualitatively to evaluate effectiveness of XOR operation as a decorrelator.

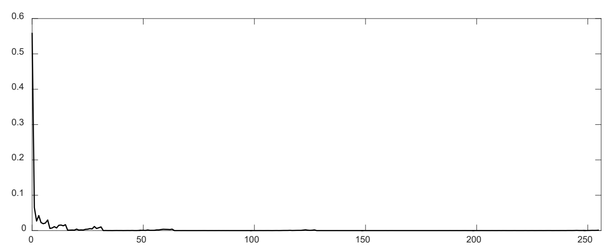


Fig. 13: Probability distribution of decorrelated data samples in a telemetry stream using XOR operation.

The entropy of the decorrelated data (E) is H_E = 3.1 (bits /symbol) and its variance is σ_E^2 = 694.24, which results in variance gain of $Gain\_Var$ = 8.30 .

After performing entropy coding for the decorrelated data (E), the evaluated parameter (k) for Rice coding method using (5) is (3) and the compression ratio between the size of original telemetry data V_{input} and compressed data V_{output} is C_r = 1.64.

VI. CONCLUSION AND FUTURE WORK

Applying XOR operation as a decorrelator is sufficient enough to transform probability distribution model of the encoded data samples into a model, which is more efficient, when implementing entropy coding.

In[9,10] it is described that using liner prediction as a decorrelator is an efficient method, especially when order of the filter used in prediction is assigned a value, which is equal to the length of the major telemetry frame, which means more hardware complexity and high computation in cases of long Major_Frames. On the other side, XOR operation is characterized by low hardware requirements and simple implementation.

In addition, decorrelation-based operation XOR simplifies calculations, since no need of calculation of additional parameters as in case of linear prediction. To implement the XOR operation, only a buffer or shift register is required, the size of which is sufficient to hold data of last received Major_Frame. It is required to perform XOR operation with the next input Major_Frame. Next, the last input Major_Frame is hold in place of the previous and so on.

XOR operation can be added to any adaptive two-stage compression method, which implements different decorrelators and select one output of them, which has the best results, to be entropy coded using Rice codes as in[4].

VII. REFERENCES

[1] Carden F., Jedlicka R., Henry R, Telemetry Systems Engineering. London (UK): Artech House, 2002, 622 p.

[2] J.P. Arcangeli, *et al.*, Compression for an Effective Management of Telemetry Data, NASA Technical Documents: nasa_ techdoc_19940019483 , 1993, pp. 823– 830.

[3] Martinez-Heras J.A., David E., Rainer T., "Housekeeping Telemetry Compression: When, How And Why Bother," 1st International Conference IEEE On Advances In Satellite And Space Communications, 2009, pp. 35 – 40.

[4] Lossless Data Compression. – Report Concerning Space Data System Standard, Informational Report CCSDS 120.0-G-3 Green Book. 2013.

[5] Staudinger P., *et al.*, "Lossless Compression For Archiving Satellite Telemetry Data," IEEE Aerospace Conference Proceedings, 2000, vol. 2, pp. 299 – 304.

[6] Maluf D.A., Tran P.B., Tran D, "Effective Data Representation and Compression in Ground Data Systems," IEEE Aerospace Conference, 2008. pp. 1 – 7.

[7] Telemetry Standards, IRIG Standard 106-13. Secretariat Range Commanders Council US Army White Sand Missile Range. New Mexico (USA), 2013. http://www.irig106.org/docs/106-13/.

[8] Horan S., Introduction to PCM Telemetering Systems. 2nd Edition. USA: CRC Press, 2002, 448 p.

[9] M.A. Elshafey, "Research of methods for forming a unified medium of telemeasurements data in a multichannel registering system of telemetry information," PhD dissertation. Bauman Moscow State Technical University (BMSTU), Moscow, Russia, 2015. 157 p.

[10] M.A. Elshafey and I.M. Sidyakin, A.C Kharitanof., D.C. Varhichef, "Research reversible compression methods of telemetry data," Herald Bauman Moscow State Technical University, Russia. Series: instrumentation, 2014, no. 3 (96), pp. 92 – 104.

[11] M.A. Elshafey and I.M. Sidyakin, "Lossless Compression of Telemetry Information using Adaptive Linear Prediction," Science and Education: Electronic scientific and technical journal, Russia, 2014, no. 4, pp. 354 –366. DOI: 10.7463/0414.0707364 .

[12] Rice R. F. Some Practical Universal Noiseless Coding Techniques – Technical Report, Jet Propulsion Laboratory Jpl-79-22. 1979. 130 p.

[13] Hans M., Schafer R.W, "Lossless Compression of Digital Audio," IEEE Signal Processing Magazine, 2001, vol. 18, no. 4, pp. 21 – 32.

[14] Kim J, "Lossless Wideband Audio Compression: Prediction and Transform," PhD dissertation, Institute for Communication Sciences, Technical University Berlin, Germany, 2004, 190 p.

[15] Sayood K. Lossless compression handbook. California (USA): Academic Press, 2003. 488 p.

[16] M.A. Elshafey and I.M. Sidyakin, "Simulation of telemetry information transmission over a noisy channel," Engineering Herald. Electronic Scientific and Technical Journal, Russia, 2014, no. 1, pp. 38–51.