UDC 681.391

MATHEMATICAL AND PHYSICAL NATURE OF CHANNEL CAPACITY

S. Rassomakhin

V. N. Karazin Kharkiv National University, Svobody sq., 4, Kharkov, 61022, Ukraine
rassomakhin@karazin.ua

Reviewer: Victor Krasnobayev, Dr., Full Professor, V.N. Karazin Kharkiv National University, Svobody sq., 4, Kharkov, 61022, Ukraine;
krasnobayev@karazin.ua

Received on November 2016

Abstract*. The classic methodological approaches to the determination of channel capacity have been considered. The contradiction between analytical and geometric definitions of maximum achievable transmission rate has been shown. Objectivity of maximum likelihood rule usage in low-quality channels with low signal/noise ratio has been analyzed. The correct formulation of the mathematical and physical content of channel capacity has been made. Invariance of capacity to a noise distribution in continuous channels has been proved. The main causes of the crisis in the development of information transmission theories have been indicated.*

Keywords*: differential entropy, channel capacity, maximum likelihood rule, uncertainty sphere, random encoding.*

Introduction and problem formulation

Currently, the definitions of fundamental limits of speed, reliability of data transmission, channel capacity, value of signal/noise ratio, as the key indicator of predicted communication quality, have become the most extensively used categories in communication theory and its applications. The works of Kotelnikov [1] and Shannon [2], published in 1946–1948, are considered to be the discovery of the fundamental laws of compression, data transmission and marks the birth of information theory in its modern sense. The theory based on the deep intersection with probability theory, statistics, computer science and other fields of knowledge was the basis for the development of communications, data storage and processing, and other information technologies.

This theory can be defined as a science dealing with the study and optimization of information encoding/decoding algorithms in order to create economical and reliable ways of its transmission through communication channels and its memory storage. The theory has arisen from the needs of radio, radar, telephone, television and computer technology, and is the theoretical base for the construction of communication systems. This theory focuses on the problem of optimal (in terms of speed, reliability and efficiency) usage of available technical devices for transmission, transformation, distribution and storage of information. At present, by the depth and amount of the researches, information theory can be matched with many branches of mathematical physics.

Undoubtedly, the main category of modern information theory is the concept of noisy channel capacity defined by Shannon [2,6]. According to his interpretation, capacity is a boundary of the data transmission rate, which cannot be exceeded with any encoding/decoding methods under any high level of transmission reliability, but it can be approached arbitrarily close to by choosing the proper methods of encoding and decoding. Channel capacity was expressed in statistical terms by introducing mathematical characteristic of the joint probability distribution of two random variables, called the amount of information. It is equal to the maximum amount of information in the signal at the channel output relative to the signal at its input, where the maximum is taken over all probability distributions of the input signal. The amount of information, in its turn, is expressed through another value, which has long been used in thermodynamics – the entropy, and represents the difference between the entropy of the channel output signal and the conditional entropy, if the input signal is known. Methodological role of capacity is extremely high in information theory, because it is not only the basis for the coding theorem stated by Shannon, but also is instrumental in proving the majority of other fundamental theorems and the existing limits.

Despite the undeniable achievements in information theory, it has been criticized recently. The reason for this is not only a lack of practicality and constructiveness in various statements of theorems but, moreover, the theory development crisis is manifesting. Visible technological progress in communication services cannot hide the absence of significant increase in specific efficiency of telecommunication equipment. The channel and physical layer protocols of information transmission system (ITS) are rather expensive. Error correcting codes, which have history of theoretical and experimental studies that amounts to more than 70 years, almost are not used in practice. The reason is not only the computational complexity of constructing and decoding cumbersome constructions in high-speed channels, but also the unacceptability of substantial residual amount of erroneous decoding probability for a transmission of data and program texts. It can be said without exaggeration that the specific efficiency of telecommunications has not changed since the twenties of the last century. The development of technique and communication technology is purely extensive. Performance improvement is achieved mostly by the development of transceiver technological base, as well as the bandwidth expansion and transmitter power (which, actually, determines the mathematical definition of capacity). It has negative moral, material and ecological effects. The problem of electromagnetic compatibility is becoming all the more essential. Overloaded traditional radio frequency ranges and a small bandwidth of metallic communication lines have forced switch to the optical range (however its potential is not infinite). As a result, ITS has become less reliable and more expensive. Mobile technology is not undergoing radical changes, but only extensive modifications. Geostationary orbit of communication satellites is approaching the saturation limit. The increase in demand rate for communication services is exceeding the rate of ITS performance increment. This serves as a testament of the explicit crisis in the theory and practice of data transmission system construction.

The purpose of this work is to reveal three main causes of the crisis – the errors embodied in the "base" of information theory which are the root cause of its evolution dead end, in particular:

- proving obvious methodological errors in the existing definitions of continuous channels capacity;
- justifying incorrectness of the statement that capacity is the limit of attainable rates for any continuous channel models;
- «debunking" the view that the decision-making based on maximum likelihood rule is the best way to estimate noisy channel output state at low signal/noise ratio.

Undoubtedly, the issues being considered can be seen as debatable, especially out of context of newly obtained scientific results which are not the subject of this work and are waiting for being published. This paper should be considered as the motivations for searching fundamentally new solutions in the mathematical theory of communication, which correspond to the true physical content of information transmission process.

1 The differential entropy of continuous distributions and analytical determination of Gaussian channel capacity

Considering the work's subject, at first let's pay attention to some well-known facts. The first definition of the capacity of discrete binary channel without memory with symmetric transition graph determined by the error probability p_0, is given in [2], and uses statistical measure of uncertainty of discrete choice, called entropy:

$$C = V \cdot \max_{P(X)} \left\{ H(X) - H(X|Y) \right\}, \tag{1}$$

where X, Y – the messages at the input and output of a noisy channel; $P(X) = \{p(0), p(1)\}$ – probability distribution of binary alphabet symbol; V – the number of binary symbols transmitted through the channel per second;

$$H(X) = -\left\{ p(0)\log\left[p(0)\right] + p(1)\log\left[p(1)\right] \right\} \tag{2}$$

– the entropy (uncertainty) of a binary message source (if information is measured in bits – logarithm base equals two);

$$H(X|Y) = -\{p_0 \log[p_0] + (1-p_0)\log[1-p_0]\} \quad (3)$$

– the channel unreliability – the entropy (uncertainty) of noise. If the channel quality specified by the parameter p_0, s known, maximum (1) is achieved with equiprobable source symbols $p(0) = p(1) = 1/2$ and amounts to:

$$C = V[1 - H(X|Y)]. \quad (4)$$

The definition $I_b = C/V$ is often used for calculating the average amount of information, which a single binary symbol on the output of a discrete noisy channel contains. It is particularly used in assessing the index of specific effectiveness of ITS [3].

The equation (4) has been generalized for the case of non-binary channel without memory (see, for example, [4,10]). By now, in addition to the above cases for discrete channel models, analytical definitions of channel capacity with erasing and some "exotic" examples of asymmetric transition graphs discussed by C. Shannon in his original paper [2] are known.

By itself, any discrete channel model is a kind of an add-on to the model of continuous (in time and level) channel. The equations (1)-(4) are objectively understandable, are clear from the physical and mathematical point of view and will not be discussed further. They need to be considered in order to keep track of the continuity of the methodological approach used by Shannon for an analytic derivation of the continuous channel capacity equation. The class of continuous channels with defined capacity is narrowed to "Gaussian" [2, 5, 7, 8] in the current paradigm. Its incorrectness will be shown below.

For a continuous source, when the messages are selected from the infinite set, Shannon, following the logic of (1)-(4), introduces the concept of the entropy of a continuous distribution (often referred to as the differential entropy):

$$H(X) = -\int_{-\infty}^{\infty} f(x)\log[f(x)]dx, \quad (5)$$

where $f(x)$ – the probability density function (PDF) of continuous random variable x. Accordingly, the joint and conditional entropy of two statistically related random arguments which determine the input and output of a continuous channel are given by:

$$H(X,Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log[f(x,y)]dx\,dy; \quad (6)$$

$$H(Y|X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log\left[\frac{f(x,y)}{f(y)}\right]dx\,dy. \quad (7)$$

The main properties of the entropy of the continuous case (5) include the following:

1) for a given constraint on the average power σ^2 of the continuous process centered relatively to zero, the entropy (5) is maximal if this process is Gaussian, i.e.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x^2}{2\sigma^2}\right), \quad (8)$$

in this case

$$\max_{f(x)}(H(x)) = \log\sqrt{2\pi e\sigma^2}; \quad (9)$$

2) unlike Shannon's discrete definition (5) – (9) [see. 2] the differential entropy measurement is relative to the given coordinate system, i.e., it is not absolute. This means that when the argument of the logarithm after calculating the integrals is less than unity, the differential entropy (5) – (8) can take on negative values! Such computing subjectivism has no a sensible physical interpretation till

now, and therefore, in most cases, simply is suppressed. Although Shannon tried to justify this fact asserting that, the possibility of negative differential entropy notwithstanding, the sum or the difference between two definitions of entropy is always positive [2]. However, such justification does not prevent the collapse, which will be shown below in the analytical determination of capacity by average mutual information (ratio of differential entropy).

In a continuous channel, the input source signals $x(t)$ are continuous functions of time, and the output signals – $y(t)=x(t)+\xi(t)$ are their implementations distorted by summing them with noise. The noise implementations $\xi(t)$ are also a continuous function of time. Continuous channel capacity is defined in [2] as the maximum (over all possible input distributions) of the function which essentially similar to the expression (1):

$$C=\frac{1}{T}\left(2FT\cdot\max_{f(x)}\left\{H(Y)-H(Y|X)\right\}\right), \tag{10}$$

where F – the frequency band which restricts the channel; T – duration of channel output observation; 2FT – number of degrees of freedom, defined on the duration T, as the number of independent measurements of function with a limited spectrum, defined by the sampling theorem [1,2]. In the formula (10) H(Y) – denotes the channel output entropy, and conditional entropy H(Y| X) defined by the expression (7). The difference, the maximum of which is sought in (10), is usually referred to as the average mutual information between the input and output per one channel usage:

$$I(X,Y)=H(Y)-H(Y|X)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\log\frac{f(x,y)}{f(x)f(y)}. \tag{11}$$

Then for one channel usage:

$$C=\max_{f(x)}\left\{I(X,Y)\right\}. \tag{12}$$

It is convenient to consider the relationship of Shannon's information definitions for a continuous channel using the Venn diagram, shown in Fig. 1.

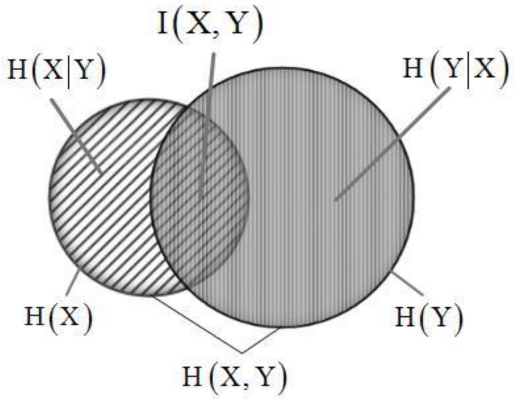

Fig. 1 – Relationship definitions of entropy for continuous channel

Therefore, the capacity of a continuous channel where noise is additive and not statistically associated with the signal, per one dimension equals the maximum of average mutual information for all variants of input distributions. [2,4,7,8] state that

$$C=2F\cdot\max_{f(x)}\left\{H(Y)-H(\xi)\right\}, \tag{13}$$

where $H(\xi)$ – the noise entropy. The theorem 16 in [2] postulates that when noise and signal are independent and additively interactive, the data transmission rate per one channel usage equals the difference between the channel output entropy and noise entropy:

$$R = H(Y) - H(\xi), \tag{14}$$

accordingly

$$C = \max_{f(x)} \{R\}. \tag{15}$$

The formulas (10), (12), (13) and (15) represent, in fact, various ways of defining the same physical magnitude for different types of measurements (total or per one channel usage). Let's continue the reasoning for the Gaussian channel in accordance with the logic of the presentation in [2], which is traditionally used in textbooks and monographs on information theory:

$$H(\xi) = \log\sqrt{2\pi e N}, \tag{16}$$

where N – the noise power. To maximize the rate, based on the properties (9), it is necessary to require that the source distribution is to be also Gaussian with the power S:

$$H(X) = \log\sqrt{2\pi e S}. \tag{17}$$

Since signal and noise are not linked statistically, due to the stability of normal distribution to the composition of any number of summable random variables [9], the distribution of their sum will be also normal with a total power which equals $(S+N)$

$$H(Y) = \log\sqrt{2\pi e(S+N)}. \tag{18}$$

As a result, we arrive at the well-known formula

$$C = F\left[\log\left(2\pi e(S+N)\right) - \log\left(2\pi e N\right)\right] \tag{19}$$

or

$$C = F\log\left(\frac{S+N}{N}\right). \tag{20}$$

It should be noted that the distribution of channel output is to be normal in the only case, when both *signal and noise are Gaussian.* The formula (20) being derived, only the signal and noise probability density functions has been used, and the methods of information receiving have not been mentioned, thereby this formula is referred to as *"Capacity of Gaussian channel"* [2–8].

Now let's focus on a strange behavior of the component analytical determination (19). To do this, we should recall that the minuend is the channel output entropy $H(Y)$, and the subtrahend is the noise entropy $H(\xi)$. What happens to the value C in case of the noise power decrease? It follows from (20) that if $F > 0$ then $\lim_{N\to 0} C = \infty$. At the same time the formula (19) shows that the capacity increases indefinitely not due to the growth channel output entropy (which, on the contrary, decreases), but due to the fact that the noise entropy (the subtrahend in (19)) tends to minus infinity:

$$\lim_{N\to 0} H(\xi) = -\infty. \tag{21}$$

This observation contradicts the physical meaning which is inherent in the definition of the difference (14). This change in the sign and adding of the subtrahend to the output entropy occurs already at "weak" noise: $N \le (2\pi e)^{-1}$. It is difficult to understand the physical meaning of this phenomenon. Although in the form (20) the capacity formula shows the monotonicity of the function $C(N)$ at $N \to 0$, that allows to explain this phenomenon by the difference between determining differential and discrete entropy, noted earlier. However, due to the lack of a clear physical interpretation of this phenomenon, correctness of the analytical derivation of capacity by using the concepts of the differential entropy and the average mutual information is doubtful.

As we will see later, attributing to this formula the ability to determine the upper limit of data transmission rates for the Gaussian channel is even more doubtful.

2 Geometric definition of capacity

After the publication of his work [2], a year later Shannon published a paper [6], which provides another method for determining capacity based on multi-dimensional geometric construction of the signal and noise space, represented in the "flat" image in Fig. 2. Any implementation of a continuous random signal, which has duration T, and which frequency spectrum is limited to F, is represented as a point in n = 2FT - dimensional space. If the transmission system is “good”, those points – S_i are uniformly distributed within the hyper sphere with the radius determined by the average signal power and the dimension of the space

$$r_S \approx \sqrt{nS} \tag{22}$$

and volume

$$V_S \approx \frac{\sqrt{\pi}^n}{\Gamma\left(n/2+1\right)}\left(\sqrt{nS}\right)^n, \tag{23}$$

where Γ(n/2+1) – gamma function. For uniform distribution of signal points, an arbitrary choice of n coordinates – random variables with zero mean and variance, which equals S can be used. Providing the dimensions of space n increase unlimitedly, the distribution of points will monotonously approach the uniform. This asymptotic property of uniformity is the basis for the construction of random codes, almost any of which is "good" [7]. The random signal realization is a channel form of a codeword of a random code and can be obtained by two following ways:

$$S(t) = \sum_{i=0}^{2FT-1} s_i \frac{\sin(2\pi F\left(t - i\cdot\Delta t\right))}{2\pi F\left(t - i\cdot\Delta t\right)}, \quad \Delta t = 1/\left(2F\right); \tag{24}$$

$$S(t) = \sum_{i=1}^{2FT} \left\{ s_{2(i-1)} \sin\left(2\pi F\frac{i}{T}\right) + s_{2(i-1)+1} \cos\left(2\pi F\frac{i}{T}\right) \right\}. \tag{25}$$

The formula (24) is an expansion of a random signal in the basis of the sinc-functions and has a continuous spectrum effectively bounded by the frequency F.

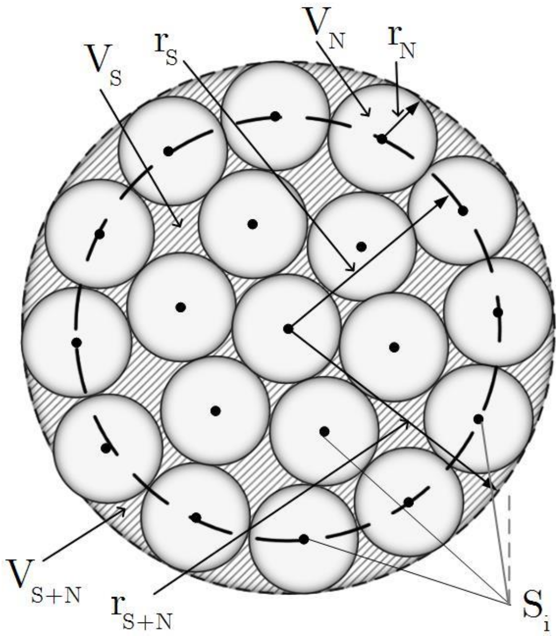

Fig. 2 – The geometric representation of the information transmission system space

For (25) the Fourier expansion in the orthogonal (on the interval T), harmonic basis is used. Thereby the realization S(t) is periodic on T, and if it is repeated indefinitely, it will have a discrete spectrum bounded above by F, non-zero measurements of which are arranged with a frequency step of $1/T$.

Both methods (24) and (25) may be used in the description of capacity attainment by means of coding given by Shannon in [2] (quoting 1): «…Let $m = 2^k$ samples of white noise be constructed, each of duration T. These are assigned binary numbers from 0 to m-1. At the transmitter, the message sequences are broken up into groups of k and for each group the corresponding noise sample is transmitted as the signal. At the receiver the m samples are known and the actual received signal (perturbed by noise) is compared with each of them. The sample which has the least R.M.S. discrepancy from the received signal is chosen as the transmitted signal and the corresponding binary number reconstructed. This process amounts to choosing the most probable (a posteriori) signal…». The formulas (24) and (25) in conjunction with the above quote is a description of the process of construction and decoding of a random code, where the decoding is performed according to the rule, which is traditionally called the *Rule of Maximum Likelihood* (MLR). With an unlimited increase in the length of the code block (synchronous increase parameters k and $n = 2TF$), if a noise is not too large, the probability of an error in the received codeword can be arbitrarily small. Thus, the geometric definition of capacity is the highest attainable rate of an arbitrary code which is decoded with the MLR and an arbitrarily low unreliability is provided.

In the geometric interpretation of the best code (Fig. 2), the point son channel output, $S_i, i \in [0, m-1]$, which correspond to the transmitted code words, are displaced under the influence of Gaussian noise within the spheres of uncertainty with the radius

$$r_N \approx \sqrt{nN} \tag{26}$$

and the volume

$$V_N \approx \frac{\sqrt{\pi}^n}{\Gamma(n/2+1)} \left(\sqrt{nN}\right)^n. \tag{27}$$

In accordance with the law of large numbers, when n increases, the probability of finding the displaced points outside the sphere with the radius $r_N + \varepsilon/\sqrt{n}$ tends to zero (ε – an arbitrary small value). Spheres of uncertainty become more delineated. Shannon compares them with regular billiard balls [2,6,8]. Since the signals of codewords and noise do not depend on each other, the total radius of hyper spherical space, which contains m spheres of uncertainty, is characterized by the radius and volume:

$$r_{S+N} \approx \sqrt{n(S+N)}, \tag{28}$$

$$V_{S+N} \approx \frac{\sqrt{\pi}^n}{\Gamma(n/2+1)} \left(\sqrt{n(S+N)}\right)^n. \tag{29}$$

With $n \to \infty$ and $\varepsilon/\sqrt{n} \to 0$ we can determine the maximum amount of non-overlapping spheres, which can be packed in the volume V_{S+N}, in such a way that there is practically no empty space between them:

$$m_C = V_{S+N}/V_N = \sqrt{\frac{S+N}{N}}^n = \sqrt{\frac{S+N}{N}}^{2FT}. \tag{30}$$

Let's recall that, if codewords are constructed in accordance with the rules (24) or (25), 2FT = n – the dimension of geometrical code space. Finding the logarithm (30) and averaging over the time T gives the maximum achievable code rate, or (*according to the modern information theory*) – channel capacity:

$$C = \frac{1}{T}\log m_C = F\log\left(\frac{S+N}{N}\right). \quad (31)$$

The results of (20) and (31) are the same apparently; allegedly it confirms the definition of C as the *maximum achievable information rate* in a channel with an additive noise and arbitrarily small unreliability. However, it should be noted, that according to the logic of the formulae derivation (31), (and of the quote discussed above as well), the value C is the *limit rate of the best code*, *when the Maximum Likelihood Rule is used in decoding*. If this was not true, and the receiver would not need to store samples of the signal realization segments in its memory in order to use them in the MLR comparisons (as it is described in the above quote from [2]) and when $N > S$, it would be sufficient to switch to the noise receiving (would there be any difference which of these two processes could be reliably distinguished from their mix?), in order to compensate the noise in the output mixture of the channel. We can refer to [8] or other works, which consider the physical and mathematical meaning of capacity, and see, that the value C, in the theorems proved by the author, is strictly an upper limit of rates for the codes in the Gaussian channel when the MLR is used, but not for the Gaussian channel itself, transmission and signal processing method notwithstanding. In the prevailing views on information theory there is no difference between these two concepts, because in the scheme of ITS, introduced by Shannon [2], the channel's encoder and decoder are present by default. The possibility to build an effective ITS, which does not use coding, is not considered at all! This contradicts the practical observation, noted in the introduction, that the error-correction codes are hardly used in the systems where mistakes are not allowed. To answer the question: *what the value C denotes: just the upper limit rate of information transmission over a channel with additive noise or the upper limit data transmission rate over the channel when the MLR is used for encoding and decoding* (unreliability is arbitrarily low in any case), let's turn to the analysis of mathematical and logical correctness of the reasoning used in the derivation of the formulas (20) and (31). For an objective analysis we need to change the conditions for which the formulas (20) and (31) have been obtained, i.e. consider the channel models different from the Gaussian one.

3 Comparison of the analytical and geometric definitions of capacity for non-Gaussian channel

Let's consider the following model of a continuous channel with the bandwidth limited value F, (where F - the frequency band which restricts the channel)and additive, stationary and signal-independent noise. Let the signal be Gaussian process with the probability density function:

$$f_1(x) = \frac{1}{\sqrt{2\pi S}}\exp\left(-\frac{x^2}{2S}\right), \quad (32)$$

with the mathematical expectation and variance

$$M[x] = 0, \quad D[x] = S. \quad (33)$$

The entropy of the signal is determined by the expression (17). The noise in the channel adds a random error to any signal measurement. This error has a uniform probability density in the range of $\left[-\frac{a}{2}, \frac{a}{2}\right],\ a > 0$:

$$f_2(y) = \begin{cases} 1/a, & \text{при } y \in [-a/2, a/2]; \\ 0, & \text{при } |y| > a/2. \end{cases} \quad (34)$$

The corresponding numeric characteristics of distribution (34) are:

$$M(y) = 0, \quad D[y] = N = a^2/12. \quad (35)$$

The entropy of the noise is defined by the value:

$$H(N) = \log a. \tag{36}$$

In some cases, the exposure of the quantizer of level signal when it is measured with the values of the sampling interval $\Delta t = 1/2F$ and the limited (greater than zero) value a (the quantization step) can be described with such a noise model [3].

By the theorem 18 in [2] Shannon defines the limits of the capacity value for arbitrary non-Gaussian channel in the following form:

$$F \log \frac{S + N_1}{N_1} \le C \le F \log \frac{S + N}{N_1}, \tag{37}$$

Where N_1 – an entropy power, i.e., the power of equivalent Gaussian noise which has the same entropy as the original non-Gaussian noise do. For this model, we can calculate the entropy power by equating the values (16) and (36):

$$N_1 = a^2 / (2\pi e) = \frac{12}{2\pi e} N. \tag{38}$$

Now let's calculate the capacity of the channel described, using an analytical approach (11) – (14). The channel output entropy, in this case, is the differential entropy of the process, which obtained by adding two independent processes:

– normal (signal) - with a mathematical expectation and variance (33);

– uniform (noise) - with a mathematical expectation and variance (35).

To calculate the entropy of the output channel H(Y) it is necessary to define the probability density function of the overall process f(y). The function, in this case, will be a composition of two distributions [9]:

$$f(z) = \int_{-\infty}^{\infty} f_1(w) f_2(z - w) dw. \tag{39}$$

Using (32) and (34) in (39) makes it possible to write:

$$f(z) = \int_{-a/2}^{a/2} \frac{1}{a} \left[(2\pi S)^{-1/2} \exp\left(\frac{-(z - w)^2}{2 \cdot S} \right) \right] dw = \frac{1}{2a} \left[\operatorname{erf}\left(\frac{a + 2z}{\sqrt{8 \cdot S}} \right) + \operatorname{erf}\left(\frac{a - 2z}{\sqrt{8 \cdot S}} \right) \right], \tag{40}$$

where $\operatorname{erf}(A) = \frac{2}{\sqrt{\pi}} \int_0^A e^{-t} dt$.

Due to the independence of those two processes, numerical characteristics of the composition (40) are:

$$M[z] = M[x] + M[y] = 0; \quad D[z] = D[x] + D[y] = S + N. \tag{41}$$

The distribution (40) is not Gaussian, although is very similar to it. To make a comparison, Fig. 3 shows the probability density function (PDF) (40) and the similar PDF of the equipotent centered Gaussian process with the normalized dispersions $a = 2\sqrt{3}; \quad S = N = 1$.

Naturally, values of the differential entropy computed for PDF composition of two normal processes (formula (18)) and for the PDF composition of the case considered, are very similar as well. For example, for the values of numerical characteristics shown in Fig. 3 in (18) we have:

$$H(Y) = \log \sqrt{2\pi e \cdot 2} = 2{,}547.$$

Calculation of the entropy of the distribution (40) yields:

$$H'(Y) = -\int_{-\infty}^{\infty} f(z)\log f(z)\,dz = 2{,}544\,,$$

i.e., the entropy of the channel output with a uniform noise almost coincides with similar entropy of the Gaussian channel but it remains a bit smaller

$$H(Y) \approx H'(Y)\,. \tag{42}$$

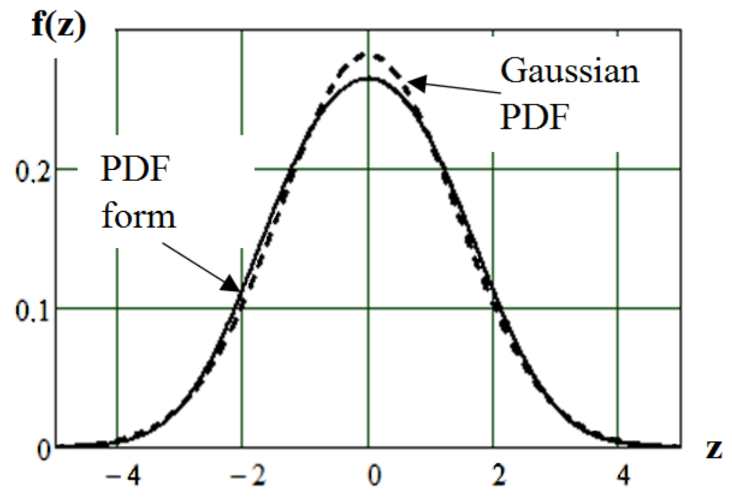

Fig. 3 – Comparison of Gaussian and composite PDFs

This result is a natural consequence of the central limit theorem of the probability theory [9]. We can write the expression for the analytical calculation of the capacity per one usage of this channel which has a uniform PDF of noise in the following form:

$$C' = H'(Y) - \log\sqrt{12\cdot N}\,, \tag{43}$$

where value N determined by the formula (35).

The comparison of value (43) with the capacity per one usage of Gaussian channel, derived from (19), under the same energy conditions, gives

$$\frac{C'}{C} = \frac{H'(Y) - \log\sqrt{12\cdot N}}{\log\sqrt{2\pi e(S+N)} - \log\sqrt{2\pi eN}}\,. \tag{44}$$

Example 1 For the case of equipotent signal and noise S=N=1, considered for the PDF in Fig. 3, we have:

– the entropy power determined in (38) $N_1 \approx 0{,}703$;

– the boundaries (37) defined by Shannon $0{,}638 \le C' \le 0{,}755$;

– the actual value calculated from (43) $C' \approx 0{,}751$;

– the Gaussian channel capacity, defined by the expression (19) under equivalent energy conditions

$$C = 0{,}5\,;$$

– the ratio of capacities, which defined by (44)

$$C'/C \approx 1{,}502\,.$$

The conclusion: the results of an analytical entropic definition of channel capacity with uniformly distributed noise lead to the following statement:

The channel capacity with uniformly distributed noise ***half as much again*** as the Gaussian channel capacity calculated for equipotent signal and noise!	(45)

Now let's use the geometric method, discussed in Sec. 2, to determine channel capacity with uniformly distributed noise. To that end, we compare the geometric representation and the characteristics of uncertainty spheres of (Fig. 2), within which the signal points are shifted by the action of normal and uniform noise. Let's introduce the concept of normalized (to the dimension of the signal space n) displacement of a signal point under the influence of noise:

– for Gaussian noise
$$r_n = \left(\frac{1}{n}\sum_{i=1}^{n}\xi n_i^{\,2}\right)^{1/2}; \tag{46}$$

– for uniformly distributed noise
$$r_u = \left(\frac{1}{n}\sum_{i=1}^{n}\xi u_i^{\,2}\right)^{1/2}; \tag{47}$$

where $\xi n_i,\ \xi u_i,\ i\in[1,n]$ – random value i -th coordinates of additive noise for a normal and uniform noise respectively. The probability distribution densities of these quantities are determined by the formulas

$$f(\xi n) = \frac{1}{\sqrt{2\pi N}}\exp\left(-\frac{\xi n^2}{2N}\right); \tag{48}$$

$$f(\xi u) = \begin{cases} 1/\sqrt{12\cdot N}, \ \text{при}\ \ \xi u\in\left[-\sqrt{3\cdot N},\sqrt{3\cdot N}\right]; \\ 0, \quad \text{при}\ \ |\xi n| > \sqrt{3\cdot N}\,. \end{cases} \tag{49}$$

Normalized radii of uncertainty spheres $\overline{r_n}$ and $\overline{r_u}$ for two noise distributions, under consideration, are determined by the mathematical expectation of random variables (46) and (47), which are the functions of random summands having the PDF (48) and (49), and "delineation degree" of the spheres determined by their dispersion $D[r_n]$ и $D[r_u]$. The analytical result for normal noise is known [9]:

$$\overline{r_n} = M[r_n] = \sqrt{\frac{2N}{n}}\left[\Gamma\left(\frac{n+1}{2}\right)\Big/\Gamma\left(\frac{n}{2}\right)\right]; \tag{50}$$

$$D[r_n] = N\left\{1-\frac{2}{n}\left[\Gamma\left(\frac{n+1}{2}\right)\Big/\Gamma\left(\frac{n}{2}\right)\right]^2\right\}. \tag{51}$$

Analytical calculation of similar numerical characteristics for the uniformly distributed noise $\overline{r_u}$ and $D[r_u]$ is difficult because a multidimensional compositional PDF of a random variable (47) is a discontinuous (piecewise-linear) function. Therefore, these characteristics have been calculated by a statistical model. The results of the analytical and statistical research of the characteristics of uncertainty spheres for a normal and uniform distribution of the noise coordinates are illustrated in Fig. 4-6. The results are expectable due to the law of large numbers. Fig. 4 shows the "virtual" cross-sections by the plane of the multidimensional picture of the displacement points for a normal (left) and uniform (right) distribution of noise coordinates, and calculated for three different values of space dimension at the number of tests equal 10^6. The images of the cross-sections of the spheres are obtained under the even noise power (equals 1) and are normalized per space dimension for convenient comparison.

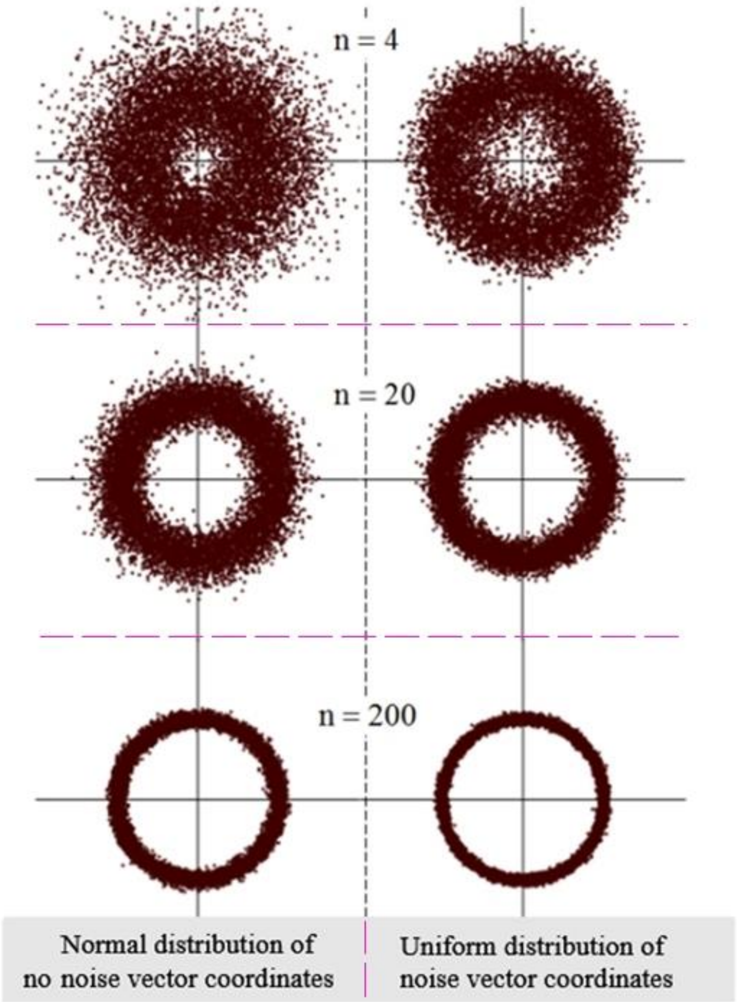

Fig. 4 – The projections of the normalized distributions of the vectors of Gaussian and uniform noise on the plane

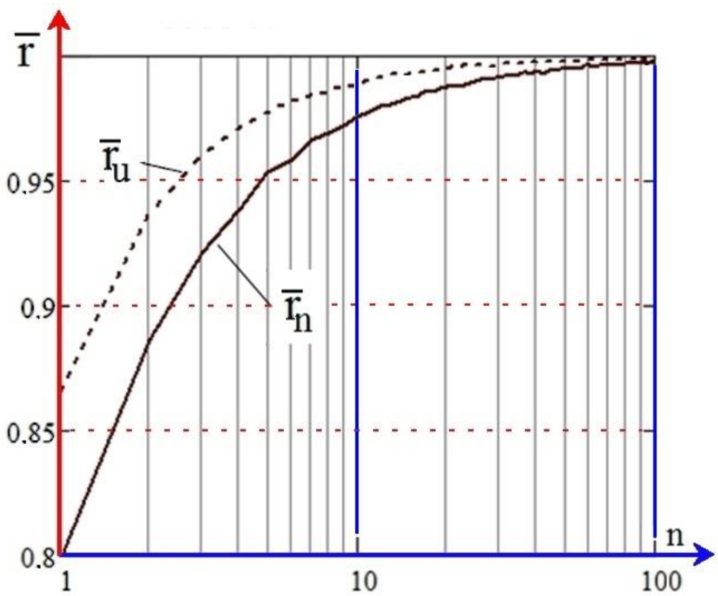

Fig. 5 – Dependence of the function $\overline{r_u}$ and $\overline{r_n}$ from the space dimension n for normal and uniform noise

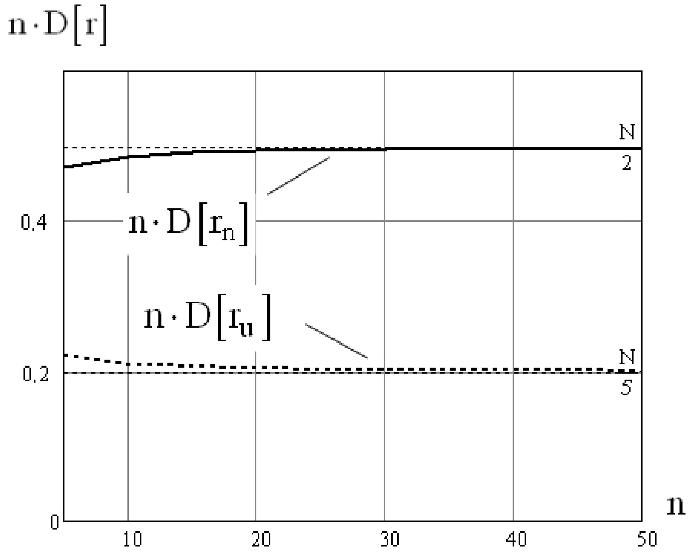

Fig. 6 – The dispersion of the radii of uncertainty spheres for the normal and uniform noise as a function of the dimension of space

The main conclusions from the results of the statistical experiment and analysis of the illustrations are following:

1) the spheres of normal and uniform noise have approximately equal average radii, while the value $\overline{r_u}$ tends to limiting and normalized value slightly faster than $\overline{r_n}$. This phenomenon is illustrated in Fig. 5;

2) the dispersion of scattering of the radius values for the sphere of uniform noise is smaller than the dispersion for the sphere of normal noise (the contours of the spheres on the right is more defined), and calculations yield the following limit ratio:

$$\lim_{n\to\infty}\left(D[r_n]/D[r_u]\right) = 2,5, \tag{52}$$

i.e., the effective width of the "ring" of scattering for uniformly distributed noise less, on average, in $\sqrt{2,5}$ times(practically, in any space dimension n), than the same parameter for normal noise. Limiting absolute values of the dispersions for the radii of the spheres are:

$$\lim_{n\to\infty}\{n\cdot D[r_n]\} = N/2, \quad \lim_{n\to\infty}\{n\cdot D[r_u]\} = N/5. \tag{53}$$

This phenomenon is illustrated by the graphs in Fig. 6 for $N = 1$. For the normally distributed noise, the absolute dispersion of the radius increases and tends to the limit value "from below", but for uniformly distributed noise it decreases and tends to the limit value (53) "from above". Finally, we can draw the main and obvious conclusion:

3) the average radii of the uncertainty spheres for the types of the noise PDF under consideration coincide asymptotically:

$$\lim_{n\to\infty}\overline{r_n} = \lim_{n\to\infty}\overline{r_u} = \sqrt{n\cdot N}. \tag{54}$$

This result is a consequence of the law of large numbers. Of course, it can be generalized for any kind of centered PDF of signal and noise, i.e. for any continuous channel with additive noise, which are not statistically associated with signal. The parameters of the geometrical representations of ITS at $n\to\infty$ are affected only by the average power values of continuous signal and noise, but not the type of their distribution!

For similar reasons, the radius of the hyper sphere on non-Gaussian channel output space also coincides with the value determined by the expression (29), then using (27), (29) and (30) we arrive at the same value of channel capacity with uniformly distributed noise: $C' = C = F\log\frac{S+N}{N}$, which contradicts the definition (43) and the statement (45). Thus, two Shannon's works [2] and [6] published at one year interval contradict each other when being applied to a non-Gaussian channel. To the question "which of two methods of determining the capacity, the analytical (entropic) or geometrical, is correct?" – there is the definite answer: the geometrical one. The correctness of the geometrical approach can easily be verified by the statistical modeling of a random code [12]. Analytical method gives the result which coincides with the result of the geometrical method in the only case when signal and noise are Gaussian processes. It's just a coincidence which can be explained by the properties of normal distribution, which has a special significance in the theory of probability and stochastic processes. Due to the mentioned reasons, the methodology of using "entropic power" and the boundaries defined by (37) are not correct.

The results of the analysis for both non-Gaussian and Gaussian channels (as, indeed, for any other model) have shown that these channels have the same capacity $C = C'$, the value of which depends only on the signal/noise ratio and the channel bandwidth. Therefore, the definition of C (20) as the limit of information transmission rate in a Gaussian channel with additive noise is not correct, to say the least.

The true physical meaning of capacity in the geometric derivation is to determine the *maximum of information transmission rate through a channel with any kind of additive noise* ***when the channel encoding and the maximum likelihood rule in decoding are used***.

Consequently, capacity is not a channel characteristic, it is the *natural limit which arises for any continuous channel model, as soon as we decide to use the encoding of information* (in the sense of making the decision according to the results of the comparison between the channel output and the known samples of valid signal realization). As a result, it is necessary to partition the signal space at the channel output into the fields of "similarity" which, in fact, are the spheres of uncertainty in the geometric representation in Fig. 2. These fields will not overlap as long as the noise power at a fixed transmitter power budget does not exceed a permissible value. This value does define the so-called capacity (actually, the limit rate of the best achievable code). The dominant axiomatic inevitability of code usage and the decision-making process based on the "the greatest similarity" principle are the source of fundamental limitations in the existing information theory paradigm. In other words, the scant achievements of the modern information transmission theory are the consequence of invariable usage of the so-called *maximum likelihood rule*.

In conclusion of this section we'd like to present some considerations as an additional argument for proving the incorrectness of the existing analytical definition of capacity as the maximum average mutual information, considered in Sec. 1. In the quotation from [2] (see Sec. 2), decoding is considered as the process of comparing the noise sample with one of $M = 2^k$ combinations of the source symbols. Therefore, obviously, the entropy of that sample can be defined correctly not by the formula (17), but as the uncertainty of discrete choice (according to the principle (2)), i.e.

$$H(X) = \log 2^k = k . \tag{55}$$

Therefore, it is this definition that should be used in the calculations (17) – (20). This leads to the another collapse, because in the same expression two different definitions of the entropy (for the discrete and the continuous choice) will be present which, according to Shannon, exist in different measurement systems.

4 The rule of maximum likelihood

> *Cramer Theorem (1740):*
> *"There is no other method of treatment of the experimental results, which would give a better approximation to the truth than the maximum likelihood method."*

The name of the rule (method) - the Maximum Likelihood Rule (MLR) is appropriate to its role in the statistical estimation of the random experience realizations and the decision-making processes under conditions of multiple-hypothesis. Modern information transmission paradigm in all known practical applications deals with the decision-making process concerning the noisy channel output state under the conditions of equiprobable hypotheses, i.e. all the source messages are assumed to be equally probable, and the effect of noise in the channel on them is assumed to be same (symmetric). This explains why other statistical methods and decision-making criteria are no alternative to the MLR. Without much exaggeration we can say that the rule of maximum likelihood came to the statistical theory of communication from our life experience. We always try to hear the phrase in a disturbing noise or to recognize the object in low visibility conditions, subconsciously using the algorithm: "what (known to us) does it most look like?" This explains why the usage of the MLR in all standard applications of the information transmission theory is axiomatic.

The quotation from [2], which has been already referred to (see Sec. 2 of this paper), reflects the justifiable (taking into consideration our physiological experience) opinion of Shannon that the decoder on the channel output has to make a decision on the received codeword (signal) by comparing the proximity (in the mean square sense) of the received sample of a random process at the channel output with the samples available to the receiver.

The same approach can be observed in the description of the ideal (according to Kotelnikov) receiver for the non-coded modulation [1] (quotation 2): «… we assume that, depending on the total oscillation y(t), which affects the receiver input, it is certain to reproduce one of the possible message values $S_1(t),\ldots,S_m(t)$. … Obviously … full range of possible values y(t) can be divided into m non-overlapping areas. … The correct messages will be reproduced more or less frequently according to the configuration of the areas determined by the receiver. … We will call the receiver the ideal one when it is characterized by such (correctly selected) areas and thereby gives the minimum number of incorrectly reproduced messages when noise is applied».

Consequently, the basic postulate of the modern theory of potential noise immunity [1], as well as the error-correcting coding theory [2], is the rule of processing noisy signals (codes) based on the maximum likelihood (or the maximum similarity), which is used by the authors as the foundation for the further theories.

If the values of apriority probabilities of source messages are the same, the mathematical formulation of the MLR in the selection of k-th hypothesis from m alternatives is following:

$$\frac{f\left(S_k|y\right)}{f\left(S_i|y\right)} > 1, \text{ for all } i \in [1,m], \; i \neq k, \tag{56}$$

where $f\left(S_i|y\right)$ – the likelihood function recorded for message S_i. The problem of finding the most reliable solution comes to maximizing the likelihood function, and, in some cases, may have an analytical (non-exhaustive search) resolution based on methods of finding the extremum known from the mathematical analysis. In cases for a continuous channel (see the quotations 1 and 2 above), the likelihood function for the message S_i on the duration T can be expressed via the Euclidean (Hilbert) distance:

$$f\left(S_i|y\right) = \left\{\int_T \left[S_i(t) - y(t)\right]^2 dt\right\}^{-1/2}. \tag{57}$$

In accordance with the maximum likelihood (maximum similarity) principle, the hypothesis, which has maximum of the function (57), is considered to be true [1,2]. Resorting to such a rule, we automatically introduce a limit on the permissible intensity of noise, i.e. *we limit from below S/N ratio at which the output signal point will not be outside its own area of similarity*. This process originates all the basic statements and, so-called, the fundamental limits of information transmission theory. These limits (the most important of which is, undoubtedly, channel capacity) are extremely rigid, unfortunately, and that is the reason for the scant achievements of the information transmission theory.

What is the value of probability P, which describes the similarity of the process at the channel output to the true transmitted message at the low S/N ratio? The answer is obvious – it is very small. Let assume that the channel alphabet allows you to send m different messages that may appear with an equal regularity. Then, for the fixed signal power S and increasing of noise power N it is true that:

$$\lim_{N\to\infty} P = m^{-1}; \quad \lim_{m\to\infty} P = 0. \tag{58}$$

With any heavy noise (if the rate is higher than channel capacity) the process at a channel output with high probability is not similar to the true transmitted message, since its representing point is equally likely to be in the area of similarity of almost any of the m possible messages. When signal points in n-dimensional space are packed most densely [11], the number of uncertainty spheres, which are adjacent to the similarity sphere of the true transmitted signal may be too large. It does not allow to create a multi-dimensional ordered manipulation codes (such as Gray code), which minimize the number of distorted binary symbols at errors of the true message transformation to the nearest to it in the ITS space. For example, at $n = 24$ there is the densest packing based on the Leech lattice and built with the Golay binary code [10, 11], in which the surface of one sphere is adjoined by 196560 surrounding spheres. If on the basis of this lattice any redundant $(24, k)$-codes with $k = 1, 2, \ldots, 18$, are constructed, it will be possible to provide mutual equidistance between all signal (code) points. Even if channel capacity is exceeded insignificantly (small overlapping of the uncertainty spheres), reception of any codeword on the channel output on the basis of MLR is almost equiprobable and practically independent from the transmitted word (message). In such conditions maximum likelihood rule usage certainly leads to an error in the reception. Therefore, there is a paradox and contradiction: on the one hand, MLR is the best way to receive, which minimizes the probability of errors at a low noise; on the other hand – the rule itself is the cause of limitations on the permissible rate and/or noise power. Can the decision-making rule be modified when we use encoding and probabilistic estimation of the channel output state?

5 Can codes work without Maximum Likelihood Rule?

It is convenient to estimate the possibility of changing the decision rule, when the true message is not considered to be the closest one to the realization on the channel output, with the help of the presentation of ITS space by Poisson field of points [12]. A random or ordered algebraic code being constructed, its codebook (a plurality of signal points) forms a random (Poisson) field of the points in a n-dimensional space, as following conditions are always satisfied:

1) at a fixed average power budget of the transmitter all the points of code words are placed in the limited volume of the multidimensional space, and with increasing n this placement asymptotically approaches a uniform (for random code) one, i.e. the density of the field of points is constant throughout the volume of code space;

2) the probability of occurrence of an arbitrary number of points in any volume of space does not depend on the quantity of points falling into any volumes which do not intersect the chosen one;

3) the probability of two or more points falling into the elementary volume is negligible in comparison with the probability of one point falling into it.

Let's assume that the transmission rate in an arbitrary Gaussian channel exceeds its capacity. In the geometrical representation it will lead to the mutual crossing of the uncertainty spheres which is shown for the fragment of channel output space in Fig. 7. To simulate the situation let's use the known [9] analytical description of PDF $\varphi(\Delta)$ of the random variable of the displacement under the noise influence $\Delta = \sqrt{n} \cdot r_n$ (here r_n is determined by the formula (46)):

$$\varphi(\Delta) = \frac{2\Delta^{n-1}}{\Gamma\left(\frac{n}{2}\right)\sqrt{2N}^{n}} \exp\left(-\frac{\Delta^2}{2N}\right). \quad (59)$$

The numerical characteristics $\varphi(\Delta)$ are derived from (50), (51) as follows:

$$M[\Delta] = \sqrt{n} \cdot \overline{r_n}, \; D[\Delta] = n \cdot D[r_n]. \quad (60)$$

Let the message, which corresponds to the point 1,be transmitted over the channel under the noise. Displacement caused by the noise is such that the point 2 is available for the receiver to observe at the channel output. Let's also assume that value of displacement is $\Delta = (M[\Delta] + \delta)$. Using the MLR in this situation will identify the point 3 (which is the closest one to the received point 2) as true transmitted, which, obviously, leads to the error.

Let's modify the decision-making rule as follows: taking the point 2 observed at the channel output as the center and let's reconstruct the surface of the sphere with the radius M[Δ] around it. Then, checking all codebook points one by one we can identify the point which is the closest to this surface. This point will be considered as true transmitted. In accordance with the rule described in Fig. 7, the point 1 located at a distance δ from the surface of the auxiliary sphere is the true transmitted. This corresponds to the error-free receiver solution in this example. Let's name this decision-making rule the "Uncertainty Sphere Rule" (USR). According to this rule, not the message, which is the most similar one to the observed channel output realization, is considered to be the true, but the message, which is the nearest one to the surface of the sphere with the radius M[Δ] drawn around the observed output point.

Likelihood function of an arbitrary signal S_i for the USR can be written in the form:

$$f\left(S_i|y\right)=\left|\int_T\left[S_i\left(t\right)-y\left(t\right)\right]^2dt-M\left[\Delta\right]^2\right|^{-1/2}. \quad (61)$$

The signal having a maximum value of the function (61) is considered to be truly transmitted. The described rule will lead to the error-free decision only on condition that the auxiliary sphere around the received point (on the Fig. 7 – a point 2) has a radius which is exactly equal to the magnitude of the noise displacement of the transmitted point, i.e. if the noise power added to the transmitted signal (codeword) in a particular realization of the observed channel output is known precisely.

However, since it is impossible to know the exact power of the noise component in the particular received realization of a signal-noise mixture, then the auxiliary sphere can be outlined only by a radius which is equal to its mathematical expectation M[Δ]. This can lead to a wrong decision if any other codebook point will occur in the layer (with the thickness of δ) between two concentric spheres with radii $\Delta=\left(M\left[\Delta\right]+\delta\right)$ and M[Δ] (the hatched ring in Fig. 7).

On the basis of the Poisson field properties, the probability of a wrong decision can be calculated as a function of $S,\ N,\ n$. The occurrence of at least one code point within the space between two concentric spheres will lead to the error. For the Poisson field, the probability of this event is:

$$P\left(\lambda,\Delta\right)=1-\exp\left(-\lambda\left(m\right)\cdot V\left(\Delta\right)\right), \quad (62)$$

where λ – a field density, containing m points:

$$\lambda\left(m\right)=m/V_{S+N}\ ; \quad (63)$$

the value V_{S+N} is defined by (29);

V(Δ) – the volume of a concentric layer around the auxiliary sphere:

$$V\left(\Delta\right)=\frac{\sqrt{\pi}^n}{\Gamma\left(\frac{n}{2}+1\right)}\begin{cases}\left\{M\left[\Delta\right]^n-\Delta^n\right\}, \text{ when } 0\le\Delta\le M\left[\Delta\right];\\ \left\{\Delta^n-M\left[\Delta\right]^n\right\}, \text{ when } \Delta>M\left[\Delta\right].\end{cases} \quad (64)$$

Using (62) - (64) and averaging the result in accordance with the distribution (59) we can calculate the probability of error in decoding by USR:

$$P_{er}=\int_0^{\infty}\varphi\left(\Delta\right)P\left(\lambda,\Delta\right)d\Delta\,. \quad (65)$$

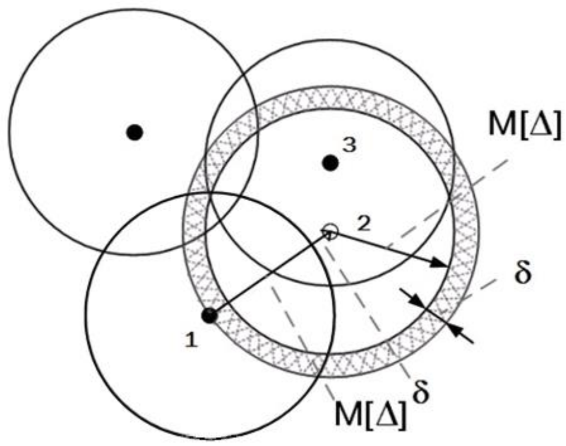

Fig. 7 – Geometric illustration the of uncertainty sphere rules (USR)

With the set values S, N and n, the number of signal points m_C in the code space which corresponds to the capacity is derived from (30), and the density of the field of points $\lambda(m_C)$ is calculated by the formula (63). Introducing the coefficient α changes of the transmission rate per one channel usage, we can model the situations, when the rate R exceeds the capacity C, which leads to intercrossing uncertainty spheres:

$$R > C \to R = \alpha \cdot C \to \alpha > 1 \Rightarrow m = (m_C)^{\alpha}; \quad (66)$$

or, by contrast, does not reach the channel capacity (uncertainty spheres do not intersect having a certain margin):

$$R < C \to R = \alpha \cdot C \to \alpha < 1 \Rightarrow m = (m_C)^{\alpha}. \quad (67)$$

For these expressions the argument, that regulates the simulated rate, is the number of points of different signals (codewords) for a fixed volume of signal space. For $m > m_C$ the channel capacity is exceeded, and for $m < m_C$ – the transmission rate does not reach the channel capacity. The coefficient α in (66) and (67) is located in the exponent because the transmission rate is measured by the logarithm of m.

The results of the calculation of a wrong decision probability (65) for USR with different values of the coefficient α and $S = N = 1$ are shownin Fig. 8.

Alas, the main conclusion from the analysis of the curves in Fig. 8 is disappointing – the USR (so attractive in the case of Fig. 7) leads to the same result as the MLR does! For $R > C$ the probability of error, when the space dimension (the number of degrees of freedom or the random code block length) increases, tends to unity monotonically.

When $R < C$ – the probability of error can became arbitrarily small with the corresponding increase of n. This result can be explained by the properties of multidimensional spheres, namely, almost all their volume is concentrated in a small area adjacent to the surface. In this area surrounding immediately the surface of the auxiliary sphere, PDF (defined by the expression (59)) reaches its highest values. Therefore, when the dimension of the space increases, the effective volume of layers around of the auxiliary sphere grows faster than the density of the field of points decreases. Certainly, we can try to formulate other criteria and decision-making rules, but, obviously, the results will be no better than the result of the MLR (*Cramer's theorem*). On the basis of the conducted simulation, it is possible to make an unambiguous conclusion – the MLR is the best and the only acceptable rule of statistical solutions for the codes – there are no alternatives to it.

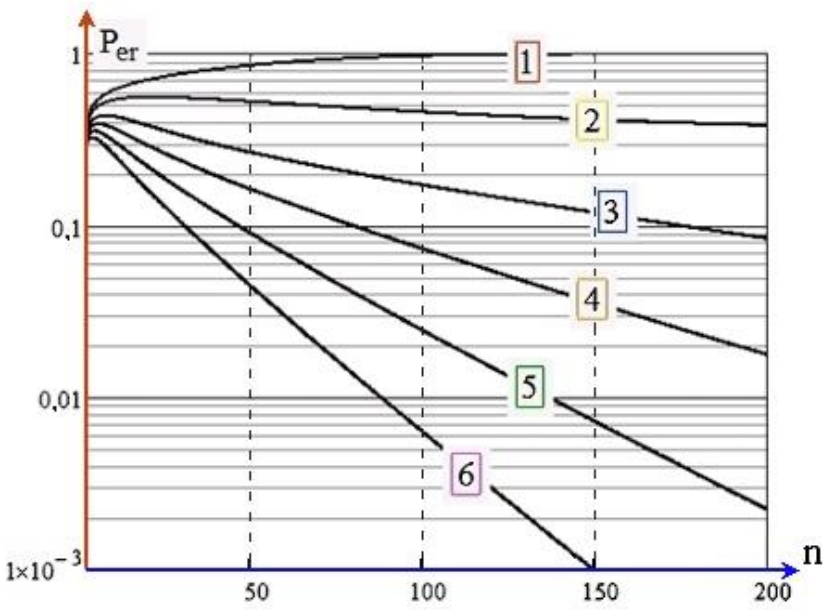

Fig. 8 – The dependencies $P_{er}(n)$ when using RSU and different rates
(1 → R = 1,05·C; 2 → R = 0,95·C; 3 → R = 0,8·C;
4 → R = 0,7·C; 5 → R = 0,6·C; 6 → R = 0,5·C)

6 Discussion of results and conclusions

The main conclusions from the results of the analysis of mathematical and physical nature of channel capacity, as well as from the modern information transmission theory contradictions are following:

1. The probability-entropy approach to the analytical determination of capacity of continuous channels, which uses the concept of the average mutual information between input and output (5) – (15), can be considered as the correct one only in case when the distribution of the source and the noise is Gaussian (16), (17). Since the usage of this approach for non-Gaussian models of continuous channels leads to the erroneous results (39) – (45), then the unjustified conclusion about the impossibility of analytical determining the capacity for such models has been made in many published works.

2. The mathematical definition (31) describes correctly the channel capacity value for any continuous channel where noise is a stationary random process. The value of channel capacity is not affected by a noise distribution type and is determined only by the signal/noise ratio and channel bandwidth. Different noise distributions manifest only in changes in the speed of approaching to the capacity when the duration of the samples of random noise code sequences increases.

3. The correct geometric definition of channel capacity determines its physical nature as the limit of in-formation transmission rate in a channel with any kind of additive noise, when the coding/decoding is used and the maximum likelihood rule is applied in decoding. Channel capacity is the physical limit only for systems, which use the maximum likelihood method.

4. The maximum likelihood rule is the best and only decision-making rule for the decoding. At the same time capacity is an indirect determination of the lower boundary of the signal/noise ratio when the noise displacement of message points in the multidimensional space of the output channel is not outside of the fixed "are-as of similarity". The existence of these areas is defined by the maximum likelihood method nature. Thus, on the one hand, the maximum likelihood rule is the best rule of the statistical decision-making, and on the other hand, it causes the appearance of the physical limit – channel capacity. Abandoning the MLR usage, which causes the appearance of the physical limit of data rates, in case when the work of receiver consists in solving the probabilistic and statistical problem, is impossible!

5. These considerations give rise to the following logical causal chain of factors that has led to the crisis in the information transmission theory development:

Representation of the signal receiving process under the noise influence as probabilistic and statistical problem;
↓
The usage of coding as the best realization of the maximum likelihood method;
↓
Appearance of physical limit rates – channel capacity.

As shown in this paper, we cannot break the chain represented above:

– it is impossible to exceed the capacity without abandoning the maximum likelihood method;

– it is impossible to abandon the maximum likelihood method when the work of receiver consists in solving the probabilistic problem.

Thus, the root of the considered problems is a probabilistic approach to receiving signals under the noise influence and even the most ambitious assumptions give no alternatives to it. However, it is not so. We do use not all the opportunities, which the nature offers for handling noisy digital signals in continuous channels. However, this will be discussed in the further publications.

References

[1] Kotel'nikov V. A. Teoriya potentsial'noi pomekhoustoichivosti / V. A.Kotel'nikov. – Moskva: Gosenergoizdat, 1956. – 152 s.
[2] Shannon C. E. A Mathematical Theory of Communication / C. E. Shannon // Bell Syst. Tech . – 1948, July-October. – Vol. 27. – P. 379 – 423; 623 – 656.
[3] Pomekhoustoichivost' i effektivnost' sistem peredachi informatsii / A. G. Zyuko, A. I. Fal'ko, I. P. Panfilov, V. L. Banket, P. V. Ivashchenko; pod red. A. G. Zyuko. – Moskva: Radio i svyaz', 1985. – 272 s.
[4] Gallager R. Teoriya informatsii i nadezhnaya svyaz' / R. Gallager; [per. s angl.; pod red. M. S. Pinskera i B. S. Tsybakova]. − Moskva: Sov. radio, 1974. − 720 s.
[5] Dobrushin R. L. Teoriya peredachi informatsii / R. L. Dobrushin, B. S.Tsybakov // Vestnik AN SSSR. – 1998. – № 6. – S. 76–81.
[6] Shannon C. E. Communication in the presence of noise / C. E. Shannon // Proc. IRE. – 1949, January. – Vol. 37. – P. 10 – 21.
[7] Verdu S. Fifty Years of Shannon Theory / S. Verdu // IEEE Transactions on Information Theory. – 1998, October. – Vol. 44. – № 6. – P. 2057 – 2078.
[8] Verdu S. A General Formula for Channel Capacity / S.Verdu, Te Sun Han // IEEE Transactions on Information Theory. – 1994, July. – Vol. 40. – № 4. – P. 1147 – 1157.
[9] Venttsel' E.S. Teoriya veroyatnostei i ee inzhenernye prilozheniya / E.S. Venttsel', L.A.Ovcharov. – Moskva: Vysshaya shkola, 2000. – 480 s.
[10] Khemming R.V. Teoriya kodirovaniya i teoriya informatsii / R.V.Khemming; per. s angl. – Moskva: Radio i svyaz', 1983. – 176 s.
[11] Konvei Dzh. Upakovki sharov, reshetki i gruppy / Dzh.Konvei, N.Sloen: [v 2 tomakh]. – Moskva: Mir, 1990.
[12] Rassomakhin S.G. Tekhnologiya psevdosluchainogo kodirovaniya v setevykh kommunikatsionnykh protokolakh kanal'nogo urovnya / S. G. Rassomakhin // Systemy obrobky informacii'. – 2012. – T.2. – Vyp. 3(101). – S. 206 – 211.

Рецензент: Віктор Краснобаєв, д.т.н., проф., Харківський національний університет імені В.Н. Каразіна, Харків, Україна.
E-mail: krasnobayev@karazin.ua

Надійшло: Листопад 2016.

Автори: Сергій Рассомахін, д.т.н., проф., Харківський національний університет імені В.Н. Каразіна, Харків, Україна.
E-mail: rassomakhin@karazin.ua

Математична і фізична природа пропускної здатності каналів.

Анотація: Розглянуто класичні методичні підходи до визначення пропускної здатності каналу (зв'язку). Показано протиріччя між аналітичним і геометричним визначенням максимально досяжної швидкості передачі. Аналізується об'єктивність правила максимальної правдоподібності при його використанні для каналів з низьким відношенням сигнал/шум. Виконано коректне визначення математичної і фізичної сутності пропускної здатності каналу. Доведена інваріантність величини пропускної здатності до виду розподілу шуму в неперервних каналах. Наведено основні причини кризи в розвитку теорії передачі інформації.

Ключові слова: диференціальна ентропія, пропускна здатність каналу, правило максимальної правдоподібності, сфера невизначеності, випадкове кодування

Рецензент: Виктор Краснобаев, д.т.н., проф., Харьковский национальный университет имени В.Н. Каразина, Харьков, Украина.
E-mail: krasnobayev@karazin.ua

Поступила: Ноябрь 2016.

Авторы: Сергей Рассомахин, д.т.н., проф., Харьковский национальный университет имени В.Н. Каразина, Харьков, Украина.
E-mail: rassomakhin@karazin.ua

Математическая и физическая природа пропускной способности каналов.

Аннотация: Рассмотрены классические методические подходы к определению пропускной способности канала (связи). Показаны противоречия между аналитическим и геометрическим определением максимально достижимой скорости передачи. Анализируется объективность правила максимального правдоподобия при его использовании для каналов с низким отношением сигнал/шум. Выполнено корректное определение математической и физической сущности пропускной способности канала. Доказана инвариантность величины пропускной способности к виду распределения шума в непрерывных каналах. Приведены основные причины кризиса в развитии теории передачи информации.

Ключевые слова: дифференциальная энтропия, пропускная способность канала, правило максимального правдоподобия, сфера неопределенности, случайное кодирование