Thermodynamic models of information processes

R. P. Poplavskiĭ

Usp. Fiz. Nauk 115, 465-501 (March 1975)

This review considers the thermodynamic aspects of the theory of information processes—measurement, transmission of information, and its processing. Considerable attention is given to an explanation of the limiting relations between the accuracy of a physical measurement and the ensuing energy degradation. Hence, in particular, a relation is established between physical entropy and quantity of information. These results are used to determine the energy cost of information transmission with allowance for the energy expenditures on decoding. The process of information processing (computational process) is treated as indirect measurement, and on the basis of this model the energetic complexity of the process is determined (the minimum value of the energy that must be degraded in order to realize the prescribed processing with the necessary accuracy).

PACS numbers: 89.70.

CONTENTS

1. FUNDAMENTAL PROBLEM OF THE THERMODYNAMICS OF INFORMATION PROCESSES

a) Introduction. The fundamental problem of the thermodynamics of information processes consists in establishing connections (limiting relations) between the most important thermodynamic characteristics (energy, entropy) and the informational (accuracy, quantity of information). In the application to the primary information process—measurement—one can distinguish three aspects of the problem:

—whether the obtaining of information is connected with a decrease of the statistical entropy of the system, and what the nature of this connection is;

—what sort of compensating increase of entropy there is in the system because of dissipation of energy;

—how it is possible after measurement to ensure additional decrease of entropy in the system by means of controlling action.

Despite the obvious importance of this problem, the literature devoted to it is scanty and is very inconsistent. Therefore the range of questions discussed in this review is limited to just this problem and, furthermore, to its application to classical (nonquantum) topics. The last limitation is due to the fact that, as will be evident from what follows, in order to solve the fundamental problem of the thermodynamics of information processes, as formulated above, it is necessary, and in the first approximation sufficient, to restrict oneself to investigation of the classical part of the measurement apparatus.

On the other hand, an attempt has been made to treat from a unified point of view all the basic information processes: the obtaining of information (measurement), the transmission of information, and the processing of information (the computational process).

b) History of the problem. After the introduction into thermodynamics, by Clausius, of the concept of entropy as a state function, Boltzmann gave entropy a statistical (informational) meaning. A few years before this (see, for example, [1]), Maxwell formulated the paradox of the demon who sorts molecules according to velocity and in consequence "violates" the second principle of thermodynamics. Precisely by investigation of the "Maxwell's demon" paradox, Szilard in 1929 [2] first pointed out the physical connection between information and entropy.

The mathematical definition of quantity of information, introduced in 1928 by Hartley [3] and then generalized in 1948 in the works of Wiener [4] and, especially, Shannon [5], coincides to within a constant multiplier with the statistical definition of entropy according to Boltzmann. This formal circumstance served as the basis for the introduction by Shannon of the concept of "entropy of a communication source." Shannon did not discuss the physical meaning of the formulas for quantity of information and for entropy. The usefulness of an investigation of the interesting meaning of the analogy between these concepts was pointed out by Wiener [6] and by Kolmogorov [7].

But information theory, beginning with the basic works of Shannon [5], developed predominantly as a mathematical discipline [31,32] with application to the process of information transmission. Kolmogorov [7] gave a rigorous presentation of the basic problems of this theory and considered the limits of its applicability.

The basic researches of 1944—1951, analyzing the connection between thermodynamics and information, related to the investigation of the "Maxwell's demon" problem from various points of view and developed the foundation-laying work of Szilard [2] already cited. These researches of Gabor, Demers, Jacobson, and finally Brillouin are set forth in detail in Brillouin's book [8]. It was shown that the important thing in Maxwell's paradox is not that the "demon" is a thinking being (he might be called a robot), but that information is used for purposeful control, leading to a decrease of entropy (ordering of the system either with respect to temperature or with respect to pressure—a "pressure demon" [8]). The obtaining of information is inevitably

connected with a degradation of energy, which leads to an increase of entropy exceeding its decrease in consequence of ordering of the molecules. It seems that the results presented in Brillouin's book [8] give a definitive solution of the "Maxwell's demon" paradox. It is difficult to agree with Shambadal' [9], who regards the "exorcism of the demon" as specious. His arguments are based on the different dimensions of entropy and of quantity of information, and also on the fact that a number of the works cited treat a one-molecule gas. It is well known that dimensionless energy (corresponding to measurement of the temperature in energy units) is often used[1)] in statistical physics [10]; and thermodynamics is still valid for a one-molecule gas in a thermostat (see [11], English transl., p. 399). But one can agree [9] that Maxwell's paradox has been solved only qualitatively: the problem of the quantitative relations between the information and energy characteristics cannot be considered solved in the works cited.

Natural, therefore, is the dissatisfaction expressed in 1955 by Kharkevich [12] because of the absence, in information theory, of relations of the conservation-law type. This can explain the well-known remarks of Gabor [13] in 1950 and of von Neumann [14] in 1951 on the necessity for developing a physical (thermodynamic) direction in information theory.

The application of thermodynamics to information theory found systematic development only in 1956, in the well-known book of Brillouin [8]. Here were formulated and treated a whole series of problems going far beyond the limits of the "Maxwell's demon" problem. In particular, a detailed analysis was made of the connection between energy and information characteristics in a whole series of thought experiments in measurement. Formulated in general form was the "negentropy principle of information," which sets a lower bound to the energy cost of a measurement. But neither from the examples treated nor from general considerations was it explained whether this lower bound is attainable; and if it is not, whether the estimate may not be made more accurate. Thus even on the basis of Brillouin's book [8], the limiting relations between information and energy characteristics cannot be considered definitively established. Therefore the importance of this problem is stressed even in the literature of the sixties (see, for example, [15]).

In a subsequent book, L. Brillouin [16] develops his ideas in a somewhat different direction: he discusses the problems of the information content of physical laws and theories, the significance of errors of measurement in the cognitive process, etc.

In recent years, papers have appeared that use the methods of information theory to analyze technical information systems [17,18,37] and that consider the formal analogies between thermodynamic and informational concepts [19,20]. Work in these directions, of course, has not posed the problem of giving an answer to the question of the physical connections between energy and information characteristics.

Further development of the physical theory of information proceeded principally in the direction of investigating the limits imposed by the quantum nature of matter on the process of information transmission. This aspect of physical information processes has great theoretical and practical importance and has been presented in a voluminous literature.

But only a few of the works in this direction have a relation to the problem posed above. The results of these works, in particular [21,22], and also ones directly related to the theme of the review, in particular [23-25], will be discussed in detail below.

c) The measurement process and its informational characteristics. In order to establish the limiting physical relations between energy and information characteristics, it seems natural to begin with an analysis of the primary information process—physical measurement.

By measurement is usually understood any single-valued transformation of a measurable physical quantity to some other physical quantity, called the recording parameter. But for thermodynamic analysis of measurement, this definition needs to be narrowed: it is necessary to describe a measurement as a thermodynamic process of transition of the system from one equilibrium state to another. We shall assume that:

1) the measurable physical quantity is an internal parameter of the system being investigated (SI);

2) the physical quantity chosen as recording parameter is one that characterizes the state of the measuring apparatus (MA) and therefore is capable of use as a control parameter;

3) in the process of interaction between SI and MA (that is, in the measurement process), there is established a new stationary value of the recording parameter, uniquely related to the measurable quantity.

According to this definition, for completion of a measurement it is sufficient to restrict oneself to a single step (and, correspondingly, to a single act of elementary interaction between SI and MA) only when:

a) the measurable quantity satisfies condition 1);

b) it can be directly transformed to a recording parameter satisfying 2) and 3).

In general this is not so, and the measurement may require a long chain of successive transformations. It appears that in order to explain the general relationships, it is sufficient to consider a chain of four steps:

$$l \to \lambda \to F \to y \to x; \qquad (1.1)$$

here l is the measurable parameter, which may be either classical or quantum; it by no means always satisfies condition 1) and therefore must be transformed to a physical quantity that does satisfy condition 1). In the course of this transformation, accomplished by a transducer, often (explicitly or implicitly) there is an intermediate step, in which l leaves a "trace" λ in some coded form. We shall illustrate what has been said by some examples of possible $l-\lambda$ pairs:

frequency of a signal being detected—position (or number) of a tunable resonator;

distance to an object—time interval recorded by the main and the reflected pulses;

difference of phase or of velocity between two electromagnetic signals—interference pattern in the place where they meet;

momentum of micro- or macroparticles—instantaneous value of the deflection of a test body; etc.

It is evident that the "trace" λ is a positional coding of the measurable quantity and is explicitly present in those measurements in which the presence of an obser-

ver is assumed. Therefore λ is often accepted as the final result of the measurement, as follows from the examples given. But if the presence of an observer is not assumed, then the step $l-\lambda$ may be present implicitly, since simultaneously with rearrangement of the apparatus according to the parameter l there occurs in the transducer a transformation of it to F. The parameter F has the character of a generalized force (in particular, an electromotive force) acting on MA. Thus SI (together with the transducer) has a certain supply of free energy that insures the interaction of SI and MA.

The action of the generalized force F on MA, leading to a corresponding change of a coordinate x, cannot always be accomplished directly. For example, if F is a measurable voltage in an electric circuit being studied, and if x is the angle of rotation of a pointer with a spring, then y is the value that becomes established, in the measurement process, of the current or charge of voltage in the instrument circuit that acts directly on the pointer. The parameter y and the device that accomplishes the step $F \to y$ in (1.1) we shall call "matching." This step is introduced into the consideration in order to estimate its effect on the thermodynamic characteristics of the measurement process. From later discussion it will be evident that the limiting relations are determined by the direct transformation $F \to x$, and that any additional step only worsens the estimates.

From what has been said, it is evident that for the thermodynamic characteristics of a measurement according to 1)—3), only the classical part of the MA is important. The definitions introduced above will be made more concrete as follows.

Every physical measurement is connected with an interaction of two systems, that under study and the experimental apparatus; the interaction is such that an exchange of energy occurs between SI and MA, and as a result there is a change in the value of a parameter x of the recording device in the experimental apparatus. Dissipative processes due to the energy exchange between SI and MA are the only possible reason for irreversibility of the physical measurement.

It is assumed that SI and MA are located in a thermostat with temperature T. Thermal fluctuations of the parameters F, y, and x are the only irremovable cause of measurement error. The measurement is based on the fact that there is an unambiguous correspondence between the mean values $\overline{F}$, $\overline{y}$, and $\overline{x}$ (averaged over the fluctuations that occur), and from the value of $\overline{x}$ a judgement is made regarding the measurable quantity F. A linear relation is assumed between F, y, and x.

Such a model of a measurement was considered in [23,26,27] and will be described in detail in Chap. 2.

We pass on to the definition of the fundamental informational characteristics of a measurement.

Let the true value of the scalar quantity being measured be l, the measured value l'; their difference

$$l - l' = \Delta l \tag{1.2}$$

is the error in the given measurement.

A measurement in a domain, or an estimate of a random quantity [28], is characterized by the fact that a domain (segment) L of length l_m is known in advance,

$$l_m = l_{max} - l_{min}, \quad l \in L \Leftrightarrow l_{min} \leqslant l \leqslant l_{max} \tag{1.3}$$

and the prior distribution $P(l)$ in L is given.

The error Δl of a given realization of the measurement of course cannot serve as its criterion of quality: for this purpose, one must choose some quantity averaged over the joint distribution $P(l', l) = P(l)P(l'/l)$. A physical measurement is usually characterized by the property of uniformity; that is, that $P(l'/l) = P(|l - l'|) = P(|\Delta l|)$ and is independent of l.

An average over all fluctuations (that is, over $P(|\Delta l|)$ we shall denote by a bar above, and an average over the prior distribution by triangular brackets, $\langle \ldots \rangle$.

Then the mean square of the fluctuations is

$$\langle \overline{\Delta l^2} \rangle = \overline{\Delta l^2}. \tag{1.4}$$

In addition, a measurement is often characterized [8,29] by an interval Δl_p that has the meaning of minimum interval of resolution or scale division. It must be emphasized that Δl_p, in contrast to $\sqrt{\overline{\Delta l^2}}$, is an ambiguous characteristic, since

$$P\left(|\Delta l| > \frac{\Delta l_p}{2}\right) = w > 0, \tag{1.5}$$

that is, there is a nonzero probability w of an error (of exceeding the interval Δl_p). In this connection, there arises the question of the relation of the error Δl_p to the reliability [29] of the experiment.

A measurement according to the scheme (1.1) is characterized by a linear transformation

$$\frac{x}{x_m} = \frac{l - l_0}{l_m} \tag{1.6}$$

of the segment L of (1.3) to a segment X of length x_m. Here, as is well known [30], the distributions of l and of x belong to the same type and differ only by the configuration parameters: a scale factor l_m/x_m and a centering parameter l_0.

We shall hereafter consider only distributions $P(l)$ symmetric with respect to $\langle l \rangle$, and therefore the cases of a symmetric measurement,

$$l_0 = \langle l \rangle, \quad -\frac{x_m}{2} \leqslant x \leqslant \frac{x_m}{2} \Leftrightarrow x \in X_+ \tag{1.7}$$

and an asymmetric measurement,

$$l_0 = l_{min}, \quad 0 \leqslant x \leqslant x_m \Leftrightarrow x \in X_-. \tag{1.8}$$

In simultaneously treating both cases (1.7) and (1.8), we shall write: $x \in X$, having in mind either X_+ or X_-. The transformation (1.6) is in general related to the fact that in the transducer there is used a certain standard generalized force F_{st}, which jointly with $F(l)$ gives

$$\begin{aligned} \langle F_{st} + F(l) \rangle &= 0 \quad \text{for (1.7)}, \\ F_{st} + F(l_{min}) &= 0 \quad \text{for (1.8)}. \end{aligned} \tag{1.9}$$

We notice that improvement of experiments is usually connected with an increase of their sensitivity; that is, discernibility of small differences $l_1 - l_2$. It is therefore natural that in these cases principal consideration is given to the lowering of the absolute error $\sqrt{\overline{\Delta l^2}}$ of (1.4). Here, when one speaks of an increase of accuracy, one has in mind a decrease of the error, although formally the dimensional quantity $(1/\sqrt{\overline{\Delta l^2}})$ can hardly be called the absolute accuracy.

In contrast to this, the relative accuracy can be formally defined as the reciprocal of the relative error. As will be evident from what follows, just this dimensionless quantity plays a decisive role in the thermodynamics of information processes.

We define the relative accuracy $1/\sigma_x$ of a single measurement as

$$\frac{1}{\sigma_x}=\frac{x}{\sqrt{\overline{\Delta x^2}}}=\frac{l-l_0}{\sqrt{\overline{\Delta l^2}}} \tag{1.10}$$

This quantity has meaning either as a characteristic of a single realization of a measurement in a domain, or in estimation of a constant quantity [28]. In the latter case, in speaking of the accuracy of a quantity l it is usual to assume that in (1.10) and (1.8) $l_0 = l_{min} = 0$. In the former case, for a characteristic of any process of measurement in a domain we shall define the mean relative accuracy $1/\sigma$ as the ratio of the prior root-mean-square error to the posterior:

$$\frac{1}{\sigma}=\sqrt{\frac{\langle(x-\langle x\rangle)^2\rangle}{\langle\overline{\Delta x^2}\rangle}}\equiv\sqrt{\frac{\langle\delta x^2\rangle}{\langle\overline{\Delta x^2}\rangle}}. \tag{1.11}$$

Often the relative (or, according to [37], reduced) error $\sqrt{\overline{\Delta\xi^2}}$ is introduced through the dimensionless variable

$$\xi=\frac{x}{x_m}=\frac{l-l_0}{l_m}, \tag{1.12}$$

then

$$\frac{1}{\sigma^2}=\frac{\langle\delta\xi^2\rangle}{\langle\overline{\Delta\xi^2}\rangle}\equiv\frac{\langle\delta\xi^2\rangle}{\sigma_0^2}. \tag{1.13}$$

We shall introduce also the relative (reduced) resolution interval

$$\varepsilon=\Delta\xi_p=\frac{\Delta x_p}{x_m},\quad P\left(\frac{|x-x'|}{x_m}>\frac{\varepsilon}{2}\right)=w>0. \tag{1.14}$$

We shall call $1/\epsilon$ the resolving power, $1/w$ the reliability (following [8]), and $1/\sigma$ the accuracy of the measurement.

It remains to determine the mean quantity of information $I(x, x')$ characterizing the process of measurement in a domain.

We have, according to [7,31],

$$I(x, x')=H(x)-MH(x/x'), \tag{1.15}$$

where $H(x)$ is the initial entropy, describing the indeterminacy of the prior distribution $P(x)$; $H(x/x')$ is the conditional entropy, describing the posterior indeterminacy of x given the measured value x' (that is, the indeterminacy of the distribution $P(x/x')$); and $MH(x/x')$ is the value of the conditional entropy $H(x/x')$ averaged over $P(x')$.

Supposing, in accordance with [32], that maximum entropy on a segment corresponds to a uniform distribution, whereas maximum entropy for given dispersion corresponds to a Gaussian distribution, we shall hereafter assume for the respective distributions

$$p(x)=\text{const}=\frac{1}{x_m},\quad x\in X, \tag{1.16}$$

$$p(x/x')=p(|x-x'|)=\frac{1}{\sqrt{2\pi}\,\sigma_0}e^{-\Delta x^2/2\sigma_0^2}\qquad(-\infty\leqslant x-x'\leqslant\infty). \tag{1.17}$$

Here, in agreement with formulas (1.13), (1.15), (1.16), (1.7), and (1.8), we have

$$\langle\delta x^2\rangle=\frac{1}{12}x_m^2,\quad \sigma_0^2=\frac{1}{12}\sigma^2. \tag{1.18}$$

If for $\sigma_0^2 \ll 1$ (accurate measurements) we neglect the difference of $p(x')$ from $p(x)$ and calculate (1.15) according to (1.16) and (1.17), we get

$$I(x, x')=\frac{1}{2}\ln\frac{1}{\sigma^2}+\frac{1}{2}\ln\left(\frac{6}{\pi e}\right)=\frac{1}{2}\ln\frac{1}{\sigma^2}-0.17\approx\ln\frac{1}{\sigma},\quad \sigma^2\ll 1. \tag{1.19}$$

Thus for high accuracy $1/\sigma$ (formula (1.11)) of the measurement, the quantity of information is determined by its logarithm[2)] ($\ln 1/\sigma$) to a sufficient accuracy (the ratio of variances was taken for such different distributions as the uniform and the normal).

If we calculate $p(x')$ on the basis of the expressions (1.16) and (1.17), we find that within the interval $P(x')$ does not differ from $P(x)$; only on the edges of the segment ($x_b \equiv x_{max}$, x_{min}) for $|x - x_b| \approx \sigma_0 x_m$ does the difference of $p(x')$ from $p(x)$ become appreciable. In the calculation of $I(x, x')$ this is equivalent to an effective decrease

$$\langle\overline{\Delta\xi^2}\rangle\equiv\sigma_0^2\approx\overline{\Delta\xi^2}(1-\sigma_0^2), \tag{1.20}$$

which is unimportant in formula (1.19) (for $\langle\delta\xi^2\rangle \gg \overline{\Delta\xi^2}$, $\sigma^2 \ll 1$), but which must be taken into account[3)] in the calculation of $I(x, x')$ for crude measurements. In this case, recalling the symmetry property of $I(x, x')$:

$$I(x, x')=I(x', x)=H(x')-MH(x'/x), \tag{1.21}$$

we note that the conditional entropy is exactly determined by the normal distribution, since the averaging in (1.21) is over $P(x)$, while $H(x') > H(x)$ because of the edge effect mentioned.

Calculations show that the edge effect can be taken into account with sufficient accuracy by supposing that the effective increase of $H(x')$ is equivalent to the fact that the initial variance is equal to the sum of the prior $\langle\delta x^2\rangle$ and the posterior $\overline{\Delta x^2}$. Then from the preceding formula we have the equality

$$I(x, x')\approx\frac{1}{2}\ln\left(\frac{\langle\delta x^2\rangle+\overline{\Delta x^2}}{\overline{\Delta x^2}}\right)=\frac{1}{2}\ln\left(\frac{1}{\sigma^2}+1\right), \tag{1.22}$$

which, in contrast to (1.19), is valid also for $\sigma^2 \geq 1$.

It is evident that for very crude measurements, when $1/\sigma^2 \ll 1$,

$$I(x, x')\approx\frac{1}{2\sigma^2}\quad\text{for}\quad\frac{1}{\sigma^2}\to 0. \tag{1.23}$$

Finally, we shall give expressions for $\langle\delta x^2\rangle$ of (1.11) and (1.13), which simplify in the case of the uniform distribution (1.16). For a symmetric measurement (1.7),

$$\langle\delta x^2\rangle=\langle x^2\rangle=\frac{1}{12}x_m^2, \tag{1.24}$$

and for an asymmetric measurement (1.8)

$$\langle\delta x^2\rangle=\frac{1}{4}\langle x^2\rangle=\frac{1}{12}x_m^2. \tag{1.25}$$

2. THERMODYNAMIC LIMITS OF ACCURACY OF A PHYSICAL MEASUREMENT

a) Entropy defect in a measurement. To begin the analysis of the characteristics of the measurement process, we shall isolate the principal questions of the basic problem, formulated above, of the thermodynamics of information processes (see Sec. a) of Chap. 1).

I) Does the coincidence of the formulas for quantity of information I and for change of entropy ΔH reflect the physical content of the process, or is this simply a formal analogy? More exactly: is the obtaining of information connected with a decrease of the statistical entropy (entropy defect $\Delta H_1 \leq 0$ [21]) of the system, and what is the nature of this connection?

II) What is the nature of the compensating increase of entropy $\Delta H_2 \geq 0$ in the system? That is to say, what is the minimum dissipation of energy in the measurement process, and how is it related to the informational char-

acteristics, i.e., what is the energy cost of the information? In other words: is the process of classical measurement reversible or irreversible, and what is the lower bound to its irreversibility?

III) By what means is it possible, in the obtaining of information, to produce a certain ordering in the system (of the type produced by "Maxwell's demon," and in the general case to accomplish control of the system by means of information); that is, to insure an additional decrease of entropy ΔH_3 ($\Delta H_3 \leq 0$)? In this section we shall consider the answers to question 1.

Until very recently, one encountered statements (see, for example, [9,33]) in which the authors insisted on the essentially different nature of physical and of informational entropy (in accordance with [33], "energetic" H_e and "cybernetic" H_c). In [9], as was mentioned above, the principal arguments were based on the difference in their dimensions. In [33], it was furthermore emphasized that H_e is defined only for an equilibrium state of a closed thermodynamic system, whereas H_c is defined for a stationary random process of delivery of messages by an information source and thus can be related to an arbitrary mathematical model of a source of messages, not connected with a state of a physical system. With this, of course, it is impossible not to agree; but it is clear that juxtaposition of H_e and H_c (more accurately, of I and ΔH_1 according to (I)) makes sense only in an analysis of a change of state of a real physical system in the process of obtaining information[4)] (analogous to the discussion in [33] with application to the control process for the other components: ΔH_2 (II) and ΔH_3 (III)).

The majority of authors identify physical and informational entropy, like Brillouin [8,16], restricting themselves to heuristic considerations (see, for example, [34-36]), which cannot serve as proof for this point of view. The "positive" (according to [21]) aspect of the negentropy principle of information, formulated by Brillouin [8,16], also consists precisely of the assertion of the equivalence of the obtaining of information (bound[4)]) to the decrease of the physical entropy of the system.

Here it is simply assumed that after the obtaining of the information, the number of possible microstates in the system is decreased by a factor $1/\epsilon$ (see (1.14)), and it is postulated that $I = |\Delta H_1| = \ln(1/\epsilon)$. It has been mentioned above (see footnote 2) that the definition of relative accuracy and of quantity of information in [8] is not rigorous. Here it must be emphasized, however, that it was Brillouin who first directed attention to the importance of the relative (and not the absolute) accuracy for establishing quantitative relations between the thermodynamic and the informational characteristics.

In [21] question I) was investigated more thoroughly. The authors consider the process of information transmission and relate the obtaining of information to a deviation of the system from the equilibrium state. It is hypothesized that each i-th signal transfers the system to the i-th macrostate. The concept of entropy defect ΔH_1 is introduced; it shows how much further from a state of thermodynamic equilibrium the state of the system is, on the average, under the action of a definite signal than under the action of a randomly chosen signal. It is asserted that the quantity of information I obtained by a physical system is equal to its entropy defect $|\Delta H_1|$; but here it is noted that this principle must be considered heuristic, since it has not been proved rigorously. It is further asserted [21] that the inequality $I \leq |\Delta H_1|$ always holds and that this is a consequence of the second principle of thermodynamics. The latter part of this assertion is erroneous, since the second principle can be applied only in estimations of the compensating increase of entropy ΔH_2 (II) in another part of the system; that is, when one is using the "negative" (according to [21]) aspect of Brillouin's negentropy principle, according to which $\Delta H_2 \geq |\Delta H_1|$. This assertion of Brillouin can be considered a consequence of the second principle of thermodynamics.

It appears that the answer to question I) follows from an analysis of the equilibrium thermodynamic model considered in reference [23]. According to [23,26,27] and the definition given in Sec. c) of Chap. 1, the measurement process reduces to an interaction of two bodies, the first of which, Y, contains the system being studied (together with matching apparatus), and the second is the recording device (for concreteness, mechanical) of the measuring instrument. It is assumed that before the measurement ($t < 0$) the bodies Y and X are not interacting and are in an initial state of equilibrium: $y(t \leq 0) = x(t \leq 0) = 0$. We shall use the name "equilibrium models" for those in which, during the measurement process ($t \geq 0$), there is established a new equilibrium state of the now coupled system Y + X, characterized by new equilibrium values of the parameters y_∞, x_∞ proportional to the measurable quantity F.

We shall first consider, in conformity with [23], models of the first class, in which the accuracy of the measurement is uniquely determined by the energy U transferred from body Y to body X.

The transfer process in the recording apparatus (body X) is described by the equation

$$\mu\ddot{x} + \eta\dot{x} + \beta x = \varkappa y \equiv f(t),\ x(0) = \dot{x}(0) = 0. \qquad (2.1)$$

The transfer process in body Y (SI with matching apparatus) is described either by the equation

$$M\dot{y} + Hy = F - \varkappa\dot{x},\quad y(0) = 0 \qquad (2.2)$$

(if y is a generalized velocity) or by the equation

$$Hy + B\dot{y} = F + \varkappa x,\quad y(0) = 0 \qquad (2.3)$$

(if y is a generalized coordinate). Here the following generalized coefficients[5)] have been introduced: inertias μ, M; resistances η, H; elasticities (stiffnesses) $\beta = \partial^2 U/\partial x^2$, B; $\varkappa$ is the coupling coefficient of the subsystems Y and X.

It is evident that in accordance with formula (1.10)

$$\sigma_x^2 = \frac{\overline{\Delta x^2}}{\overline{x_\infty^2}} = \frac{kT}{\beta x_\infty^2} = \frac{kT}{2U}. \qquad (2.4)$$

Considering a measurement in a domain (symmetric, in accordance with (1.7)), we get from (1.11) and (1.24)

$$\frac{1}{\sigma^2} = \frac{\langle x^2\rangle}{\overline{\Delta x^2}} = \frac{2\langle U\rangle}{kT}, \qquad (2.5)$$

where $\langle U\rangle$ is the value of the energy transferred from body Y to body X, averaged over P(x).

In the equilibrium model, it is easy to calculate directly the entropy defect ΔH_1. In fact, before the measurement

$$H_0 = H(Y) + H(X) = H(\alpha) + H(x) + H(y), \qquad (2.6)$$

where $H(\alpha)$ is the entropy determined by all the microscopic and macroscopic parameters except y in body Y and x in body X. After the measurement, there is established between the two (previously independent) param-

eters (y and x) a coupling, which leads to a decrease of entropy of the now coupled subsystems. We have

$$H_\infty = H(Y + X) = H(\alpha) + H(x) + MH(y/x). \qquad (2.7)$$

The entropy defect is

$$-\Delta H_1 = H_0 - H_\infty = H(y) - MH(y/x) = I(x, y). \qquad (2.8)$$

But the indeterminacy in y for known x is due only to the error $x - x' = \Delta x$ of the measurement. Recalling (1.15) and (1.22), we get, with use of (2.5):

$$I(x, y) = I(x, x') = \frac{1}{2}\ln\left(\frac{1}{\sigma^2} + 1\right) = \frac{1}{2}\ln\left(\frac{2\langle U\rangle}{kT} + 1\right). \qquad (2.9)$$

It is evident that the degree of coupling between subsystems Y and X is described by the quantity of information obtained, which is equal to the decrease of entropy, and the chief problem of analysis of the model consists in the determination of the minimum permissible dissipation of energy in the measurement process; that is, in the answer to question II).

But even before the carrying out of this analysis, the often expressed point of view that with unlimited measurement time it is possible to obtain arbitrary relative accuracy, expending as little energy as is desired, is easily refuted by the estimates presented above.

It might seem that this point of view is in full agreement with the classical concept of reversible processes. But when we consider the process of obtaining information, we may not neglect, as is often done, the small[6)] but in this case decisive effect of the interaction of the two systems, which leads to the result that the entropy of the coupled system is in general less than the sum of the entropies of the previously noninteracting systems. And if this is so, then, according to the second law of thermodynamics, there must necessarily occur in the interaction process a dissipation of energy that compensates this decrease of entropy; that is, the rate of interaction must be nonzero.

In the following section, estimates of this rate will be obtained, and the physical meaning of the irreversibility of the measurement process will be explained.

b) Lower bound to the irreversibility of the process of physical measurement. Investigations made by J. von Neumann[11] and also by other authors[38-41] demonstrate an irreversibility of the measurement process that is due only to the quantum nature of the object of study.

Recently papers have appeared which investigate the quantum-mechanical irreversibility of the measurement process from the point of view of the concepts of information theory[22,42,43]: in the terms[21] discussed above, it is shown that $I < |\Delta H_1|$ if only the two operators corresponding to the parameters being measured do not commute. We note in this connection that the quantum-mechanical limitation on the sensitivity (absolute error Δl) applies only in simultaneous measurement of at least two noncommuting variables.

Limitations on the accuracy of measurement of a single quantity arise only in the classical macroscopic part MA and are due to thermal fluctuations. References 26 and 27 give corresponding estimates of the variance of the absolute error $\overline{\Delta x^2}$. Above, in accordance with[23], we gave estimates (2.4), (2.5) of the relative variance, relating it (and the relative accuracy $1/\sigma$) to the energy U of interaction of the system under study and the measuring instrument, and also to the entropy defect (2.8), (2.9). In order to answer question II), however, it is essential to relate the accuracy not to the energy U transferred but to the energy Q dissipated in the measurement process.

It is well known that the dissipated energy Q in the process (2.1) of transfer of energy U to the body X is essentially determined by the time of interaction. If the force increases rapidly (that is, the interaction time is limited by the time constant of the MA), then the work produced exceeds U by a factor 2: kinetic energy equal to the potential is dissipated in the system—transferred to the thermostat. On the other hand, it is well known[10] that in a reversible process the minimum work is $R_{min} \to U$ (that is, $Q \to 0$). It is not that the coefficient of friction is negligibly small (an explanation sometimes given): if we neglect the frictional forces, then the new equilibrium state of the coupled system will never be established. An approximation to a reversible process can be insured, independently of the value of the coefficient of friction η (with increase of η, only the time scale changes, since the time constant of MA changes) by slowing down the interaction process (slow increase of the force). The explanation of this is well known: the transferred energy U is determined by the displacement $x = vt$ (v = mean velocity, t = time of interaction), whereas the dissipated energy $Q \sim v^2 t$. Therefore by slowing down the transition process as much as is desired, that is by increasing t and decreasing v proportionally, it is possible, for given U, to insure $Q \to 0$.

We note that one usually does not investigate what it means to slow down the transition process: it is assumed that the experimenter slowly introduces a force (moves a piston, removes a resistance in an electric circuit, and so on). But to estimate how the work performed by the experimenter depends on the slowing down does not seem possible.

Therefore in the model considered, the role of the experimenter is reduced to a minimum: he only removes the subsystems from a state in which they do not interact to a state of interaction, in which the transitional process begins in the coupled system SI + MA. This action of the experimenter (advancing of a shutter, closing of a switch in an electric circuit, and so on) is a single act and not connected either with the energy being transferred or with energy stored in the shutter or the open switch. The work performed here by the experimenter must only exceed (in general by many times) kT: the energy barrier that prevents the subsystems from spontaneously reaching the state of interaction. Furthermore, the slowing mechanism considered in the models is itself realized not by the experimenter but by some regulator which (in contrast to the experimenter) can be investigated by physical methods. Here it turns out that since the regulator itself participates in the thermal motion, the realization of the slowing down is connected with a certain additional dissipation of energy that is usually not taken into account.

Returning to the model (2.1), (2.2) or (2.1), (2.3), we see that if $f(t) = \kappa y(t)$ increases rapidly ($\tau_y \ll \tau_x$, where τ_y and τ_x are the time constants of bodies Y and X respectively), then the dissipative losses in body X are $Q_x = U$. It is also evident that the matching apparatus can serve as a damper, slowing down the growth of the force in (2.1). Then, however, although it is possible to make $Q_x \to 0$, the dissipative losses in the body Y are $Q_y > U$. This result follows from the fact that the coup-

ling coefficient κ and the dimensionless coefficients proportional to it, $\delta = \kappa^2/\eta H$ in the system (2.1), (2.2) and $\Delta = \kappa^2/\beta B$ in the system (2.1), (2.3), are bounded from above. Decrease of δ and Δ below the upper bound is possible but is connected with an increase of Q_y. Those values are energetically optimal for which $\sigma_y^2 = \sigma_x^2$ (see (2.4)). Then

$$\begin{aligned} Q_\Sigma &= Q_y + Q_x = 2Q_x = 2U, \\ \sigma^2 &= \sigma_y^2 + \sigma_x^2 = 2\sigma_x^2, \end{aligned} \tag{2.10}$$

and we attain the lower bound of the estimate for $\sigma^2 Q_\Sigma$ in the case of two degrees of freedom (see[23])

$$\sigma^2 Q_\Sigma \geqslant 2kT. \tag{2.11}$$

This estimate, in consequence of (2.10), is four times worse than the limiting estimate for one degree of freedom, when the matching parameter y is absent and the force F acts directly on the recording device.

Then, in accordance with (2.4) and $Q_X = U$, we have

$$\sigma^2 Q_\Sigma \geqslant \frac{1}{2} kT. \tag{2.12}$$

Both estimates relate to an irreversible realization of the measurement process: either when the interaction time is limited by the time constant τ_X, (2.12), or when the growth of the force is slowed down in X, but as a result the losses in Y increase, (2.11).

Thus in order to decrease the dissipation during the measurement, it is insufficient to slow down the growth of the force in X: it is necessary also to slow down the growth of the force in Y. In the model under consideration, this can be done only by introduction of a third body Z—a regulator, which slowly changes the parameter H in (2.2) or B in (2.3). But since the regulator also participates in the thermal motion, it is then found that for given U (that is, accuracy $1/\sigma$) the decrease of Q_Σ is not unlimited[7].

Let the regulator insure the optimal law of growth of the force f(t) in (2.1):

$$f(t) = \begin{cases} \beta x(t) + \varepsilon & \text{for } 0 \leqslant t \leqslant t_0, \\ f_\infty = \beta x_\infty = \text{const} & \text{for } t \geqslant t_0. \end{cases} \tag{2.13}$$

Then the dissipation Q_1 connected with decrease of velocity decreases by a factor $n = t_0/4\tau_X$ (irreversible realization of the process corresponds to mean velocity $x_\infty/4\tau_X$). On the other hand, in consequence of (2.13), equation (2.1) is transformed to $\mu\ddot{x} + \eta\dot{x} = \epsilon$, when it is necessary to take account of a Brownian drift[8]

$$\overline{\Delta x^2}(t_0) = \frac{2kT}{\eta} t_0. \tag{2.14}$$

When $t = t_0$, so that $f(t) = f_\infty$, the body X falls into a potential well; the larger t_0, the more its energy exceeds the equilibrium value $\overline{\beta \Delta x_0^2}/2 = kT/2$. This energy is delivered to the thermostat; this corresponds to a second term in the dissipative losses, Q_2. From the condition

$$Q_\Sigma = Q_1 + Q_2 = 2U \frac{4\tau_x}{t_0} + 2kT \frac{t_0}{4\tau_x} \to \min \tag{2.15}$$

we get

$$Q_{\Sigma\min} = 4\sqrt{UkT}, \quad \left(\frac{t_0}{4\tau_x}\right)_{\text{opt}} \equiv n_{\text{opt}} = \sqrt{\frac{U}{kT}} = \frac{1}{\sqrt{2}\sigma_x} = \frac{1}{\sigma}, \tag{2.16}$$

whence follows at once the limiting estimate (see[23]) for the "reversible" realization[9] of the measurement process:

$$\sigma Q_\Sigma \geqslant 4kT. \tag{2.17}$$

For a model with one degree of freedom, analogous to the preceding, the estimate improves:

$$\sigma Q_\Sigma \geqslant 2kT, \tag{2.18}$$

but in this case not by a factor 4, like (2.12) and (2.11), but only by a factor 2.

The same result is obtained for a discrete realization of the slowing down, when the regulator fixes each i-th intermediate equilibrium state (i = 1, 2, ..., n); in the optimal case, $\Delta x_i = \Delta x_j = x_\infty/n$. Since the fixation of each equilibrium state requires an energy coupling of the regulator with the system exceeding kT, each i-th transition is connected with a dissipation $Q_{i\,\min} \geq kT$.

We have in this case

$$\begin{aligned} Q_1 &= \sum_{i=1}^{n} \Delta Q_{1i}, \quad Q_2 = \sum_{i=1}^{n} \Delta Q_{2i}, \quad U = \sum_{i=1}^{n} \Delta U_i, \\ \Delta Q_{1i} &= \Delta Q_{1j} \equiv \Delta Q_1 = \frac{U}{n^2}. \end{aligned} \tag{2.19}$$

On the other hand, on taking into account the coupling of the regulator with the system, we have

$$\Delta Q_{2i} = a^2 kT, \quad a \gg 1. \tag{2.20}$$

We write in analogy to (2.15):

$$Q_\Sigma = na^2 kT + \frac{U}{n} \to \min. \tag{2.21}$$

Hence, setting $a^2 = 1$ (weak restraints), we get directly (2.18).

In order to obtain estimates of the increase of entropy ΔH_2, it is necessary to go over from a single measurement to measurement in a domain. It is easy to see that the estimates (2.11) and (2.12) for an irreversible realization of the process are valid also for the mean values $\langle Q_\Sigma \rangle$ in a symmetric measurement (1.7) (in consequence of (1.24), in analogy to (2.4), (2.5)). For an asymmetric measurement (1.8), the estimates, in consequence of (1.25) worsen by a factor 4.

In an optimal realization of the process, however, the coefficient in the estimates (2.17), (2.18) is, strictly speaking, somewhat smaller (by a factor $\sqrt{3}/2$), because $1/\sigma \sim \sqrt{\langle x^2 \rangle}$ whereas $\langle Q \rangle \sim \langle \sqrt{x^2} \rangle$. For the same reason, for an asymmetric measurement the estimates (2.17), (2.18) worsen not by a factor 4 but by a factor 2.

Hereafter we shall neglect the coefficient $\sqrt{3}/2$ in (2.17), (2.18) and shall apply these estimates, as well as (2.11), (2.12), also to the case of a symmetric measurement in a domain[10], when $Q_\Sigma = \langle Q_\Sigma \rangle$ (we shall omit the brackets).

From formulas (2.12) and (2.18) we find that the minimal increase of entropy ΔH_2 resulting from dissipation is completely determined by the accuracy of the measurement and the mode of realization of it (irreversible or in the limit "reversible"):

$$\Delta H_2^{(\text{i})} \geqslant \frac{Q_\Sigma}{kT} = \frac{1}{2\sigma^2}, \quad \Delta H_2^{(\text{r})} \geqslant \frac{2}{\sigma} = 2\sqrt{2\Delta H_2^{(\text{i})}}. \tag{2.22}$$

In accordance with an earlier result (see (2.9),

$$|\Delta H_1| = I = \frac{1}{2} \ln\left(\frac{1}{\sigma^2} + 1\right) = \begin{cases} \ln(1/\sigma), & \sigma^2 \ll 1, \\ 1/2\sigma^2, & \sigma^2 \gg 1. \end{cases} \tag{2.23}$$

It was mentioned above that according to Brillouin[8,16]

$$\Delta H_2 \geqslant |\Delta H_1|, \tag{2.24}$$

the equality sign defines the lower bound of an irreversible measurement process. From (2.22) and (2.23) it is

evident that this lower bound is much too low and can be attained only in the limit $I \to 0$. Even this, however, is not altogether accurate, since formulas (2.22) are in general valid only for $\sigma^2 \leq 1$. In the contrary case (that is, when $U < kT$, (2.4)), in accordance with (2.20), $Q_\Sigma > kT > U$; that is, $\Delta H_2 > 1$ always. In this connection we remark that the possibility of attaining the lower bound (2.24) in the examples considered in [6] has not been proved. But if it is nevertheless attainable for $\sigma^2 < 1$, then from (2.23) and (2.5) there is obtained an extremely unnatural relation between the optimal speed of the transition process and the parameters of the measurement and the MA (in this connection, see footnote 9).

But the chief objection, in principle, to the possibility of the equality in (2.24) for $I > 0$ arises in the course of answering the question III) formulated above. From consideration of Maxwell's paradox it is known that the information obtained does not completely describe the result attained: it can be used later for the introduction of order into the system (for control); that is, for a decrease of entropy $\Delta H_3 < 0$.

To wit, a low value of the informational efficiency

$$r_{\text{inf max}} = \frac{I}{\Delta H_{2\min}} = \frac{\sigma}{2} \ln \frac{1}{\sigma} = \frac{1}{2} Ie^{-I}, \quad \sigma^2 \ll 1, \tag{2.25}$$

obtained from the relations (2.23) and (2.22), insures the possibility of effective control without violation of the second principle of thermodynamics:

$$\Delta H_2 - |\Delta H_1| = \Delta H_2 (1 - r_{\text{inf max}}) \geqslant |\Delta H_3| > 0. \tag{2.26}$$

In contrast to the informational, the mechanical efficiency r_{mech} completely describes the result attained:

$$r_{\text{mech max}} = \frac{U}{U + Q_{\Sigma \min}} \approx 1 - 2\sqrt{\frac{kT}{2U}}, \quad U \gg kT, \tag{2.27}$$

it differs less from unity, the greater the useful effect (the transferred energy U).

Thus the analysis of the slowing down of the transition process achieved by an ideal regulator (yet one that participates in the thermal motion) leads to the natural bounds (2.16) and (2.22) in the approximation to a reversible process.

All these results related to models of the first class, in which the transferred energy U completely determined the informational characteristics of the measurement process (see (2.4), (2.16), and (2.23)).

It is possible to consider models of a second class, in which the estimate (2.12), relating Q_Σ with σ, is obtained directly and cannot be improved; but then the value of the transferred energy U (which no longer determines σ) may appreciably exceed the bounds (2.16) and (2.27).

Let the measurable quantity F, in contrast to the model just considered, be not a lumped parameter, but one that slowly decreases with distance z. Then the role of regulator Z is played by a "truck" that slowly brings the body X closer to SI.

In this case the optimal law of growth of the force is

$$f(t) = bx(t) + \varepsilon, \quad \beta - b = \beta_{\text{eff}} \ll \beta, \tag{2.28}$$

and in equation (2.1) β will be replaced by β_{eff}. Then the dissipation decreases by a factor β/β_{eff}, but the mean square of the error increases by the same factor, so that the estimate (2.12) does not improve. In this situation, however, the relation between the dissipated energy Q_Σ and the transferred energy U may be better than by (2.16), but the relation between the accuracy and U appreciably worse than by (2.4).

This kind of interaction is characterized by the fact that the test body is acted upon by the small difference of two large forces, produced by corresponding fields. Therefore there is a very weak coupling with the parameter x (the accuracy $1/\sigma$, which characterizes this coupling, is small), and a "strong" interaction occurs in equilibrium: an inappreciable fluctuation of the total energy leads to large fluctuations of the parameter x and consequently to large transfers of energy from one field to the other. As is well known [44], normality of the macroscopic variables is characterized by high stability of them ($1/\sigma \gg 1$). In the present case the energy of the system is a normal variable, but the energy of the subsystems is not; for their thermal fluctuations are large, and consequently the information that can be extracted by the test body (with recording parameter x) is small.

Thus only for such a transfer of energy U that the information extracted from the system is characterized in accuracy by (2.4) does the bound (2.16) occur in the approximation to a reversible process.

We shall now discuss the choice of the parameters of the recording apparatus.

From the preceding it is evident that for a given value of the directly measurable quantity $\kappa y_\infty = f_\infty$, an increase of the accuracy (that is, of U) can be attained by decrease of the elastic coefficient β. If, besides the energy limitations, we take into account also the limitations on the time of measurement, which in all cases (both irreversible and optimal) are determined by the relaxation time τ_x, it is easy to obtain the following recommendations for synthesis of the recording apparatus. It is known that the character of the transitional mode for equation (2.1) is completely determined by the parameter

$$\gamma = \frac{\eta}{2\sqrt{\mu\beta}}, \tag{2.29}$$

inverse to the figure of merit. It is easy to show that at the boundary ($\gamma = 1$) between oscillatory ($\gamma < 1$) and aperiodic ($\gamma > 1$) modes, a minimum of the relaxation time τ_x is attained (and with it, of the measurement time) for given value of β (that is, for given accuracy). We have

$$\tau_x = \frac{2\mu}{\eta} = \frac{\eta}{2\beta} \quad \text{for} \quad \gamma = 1. \tag{2.30}$$

From this it is evident why accurate measurements are characterized by small values of β and μ.

c) The energy cost of accuracy and of quantity of information. From the estimates obtained above for the measurement process, (2.12), (2.18), (2.22), and (2.23), there follow immediately limiting estimates of the energy cost $\hat{e}_\sigma$ of accuracy and $\hat{e}_I$ of unit quantity of information. We have from (2.18)

$$\hat{e}_\sigma = \frac{Q_{\Sigma \min}}{1/\sigma} = 2kT, \tag{2.31}$$

that is, for optimal slowing down of the process the cost of accuracy is constant.

But for irreversible realization of the process, the energy cost of accuracy

$$\hat{e}_\sigma^{(u)} = \frac{1}{2} kT \left(\frac{1}{\sigma}\right) \tag{2.32}$$

increases in proportion to the accuracy.

From (2.23) and the preceding formulas we have

$$\hat{e}_I = \frac{Q_{\Sigma\,\min}}{I} = 2kT\frac{1}{\sigma\ln(1/\sigma)} = 2kT\frac{e^I}{I}, \quad \sigma^2 \ll 1, \qquad (2.33)$$

$$\hat{e}_I^{(\Pi)} = \frac{kT}{\sigma^2\ln[(1/\sigma^2)+1]} = \begin{cases} (kT/2I)\,e^{2I}, & \sigma^2 \ll 1, \\ kT, & \sigma^2 \gg 1, \end{cases} \qquad (2.34)$$

that is, for $\sigma^2 \ll 1$, not only for irreversible but also for optimal realization of the process, the cost of unit quantity of information I increases exponentially (with respect to I). Only for $I \to 0$ ($\sigma^2 \gg 1$) does this cost approximate to kT, Brillouin's lower estimate, obtained also in [21].

One of the reasons for the obtaining of too low estimates for $\hat{e}_I$ is the fact that neither in [8] nor in [21] is there introduced an explicit bound on the energy (or, contrariwise, on the quantity of information or the accuracy)[11]; because of this, these estimates relate to the uninteresting limiting case $I \to 0$.

A second reason has a different origin: very often the energy expenditures are estimated only in the first stage of the measurement (see the scheme (1.1)), especially if this step is detection[12].

We shall consider briefly the limiting energy relations in detection as dependent on accuracy. The statistical theory of detection developed during approximately the same years as did information theory and has been presented in an extensive literature. Here we must select the pioneering papers of Kotel'nikov [45,46] and the later [47-49]. Without discussing the numerous results of this theory, we note an important basic fact important for the question under discussion: the threshold energy is determined first of all by the energy of the noise kT, and the threshold level can be found from the requirements on the probabilities of detection $(1 - w_+)$ and of false alarms w_-. It is known that for arbitrary relations between the duration $\hat{\tau}$ of the signal pulse and the time constant τ for relaxation of the receiver, the signal energy for sure detection $E_c > kT$.

If we are concerned with detection of a constant force of duration $\hat{\tau}$ (a video pulse) or of a pulse of sinusoidal oscillations of known frequency ν and of duration $\hat{\tau}$, then a matched receiver, as is well known [50], is characterized by a relaxation time $\tau \approx \hat{\tau}$: in this case it is optimal to use the band $\Delta\nu$ of the receiver, and the chief part of the signal energy is transferred to the receiver and is dissipated in it.

Use is often made of a signal with a large number $n \gg 1$ of degrees of freedom (wide-band, either repeated pulses or a constant force of long duration in comparison with τ). Then the signal energy in a single degree of freedom may be smaller than kT (the noise power exceeds the signal power). In these cases, as is well known [50], one uses the method of accumulation (for a sinusoidal signal, coherent accumulation). Then the signal/noise energy ratio at the output of the receiver increases by a factor $n = \hat{\tau}/\tau \gg 1$, but the total signal energy $E_c^{(n)}$ is $\sqrt{n}$ times as large as the energy $E_c^{(1)}$ of a signal with one degree of freedom that would insure the same w_+ and w_-.

Recently the contrary case of a time ratio $\tau/\hat{\tau} = m \gg 1$ has acquired great importance. It was shown by Braginskiĭ [51] that an increase of the sensitivity of a mechanical receiver—an oscillator, described by (2.1)—is inevitably connected with an increase of the relaxation time τ and a corresponding narrowing of the band of the receiver. Braginskiĭ [51] proposed an optimal strategy for obtaining a solution which, despite the so significant mismatch of the receiver with respect to the signal, insures detection of almost the limiting threshold signal. Here the portion of the energy U transferred to the receiver (or extracted from it) may be m times smaller than kT. It must be mentioned, however, that the threshold signal energy E_c is not decreased: only a fraction 1/m of this energy is transferred to the receiver (because of the fact that the receiver has a band m times as narrow as the spectrum of the signal), while the principal part of the signal energy is dispersed, not entering the receiver, and is absorbed in the thermostat. Precisely for this reason, the application of the described effect of apparent lowering of the threshold of detection to information transmission or operative radar is of course inexpedient, since this method leads not to a decrease of the transmitter energy but only to an incomplete use of it in the receiver.

But the methodology presented in [51] and developed in [52] has great importance in physical experiments on the detection of minute effects (search for quarks, gravitational waves, etc.—see [51]). Because the energy U transferred to the receiver can be m times smaller than kT, quantum limitations begin to show up significantly earlier (when $U = kT/m \approx h\nu$). These questions are considered in detail in [52].

We shall briefly discuss the successive steps of a measurement described in [51,52] and shall compare them with the steps of the scheme (1.1). The force being measured excites small nonstationary changes of amplitude of the oscillations of the mechanical oscillator (2.1)—the step ($l \to \lambda$) of (1.1). Then with the aid of a capacitive transducer the mechanical oscillations are transformed to electrical; this corresponds to the step $\lambda \to F$ in (1.1). In [52] it is shown that in this step it is advantageous to be in a nonstationary mode: then it is possible to attain minimum threshold signal (for high initial level of the oscillator energy). The next step is of especial interest: it is shown [52] that measurement of the energy of the electrical oscillator before and after action of the force can be accomplished practically without disturbance. The interaction of an electron beam with a resonator can occur almost completely elastically—without transfer of energy. Between the interactions in a phase and the opposite phase, part of the electrons fall into the recording device; this corresponds to the last steps in (1.1): $F \to y \to x$; here y is the current, and x is the value of the recording parameter. The last step, as always, is quasistationary and is connected with transfer and degradation of energy exceeding kT.

Thus in all cases of detection, the energy degraded in the receiver and the surrounding medium exceeds kT.

In considering the accuracy of measurement in detection, it is necessary to distinguish two problems.

On the one hand, it is possible in detection to retain amplitude information and thus not simply to detect, but also to measure the force. Then in the first two cases, when $\hat{\tau} \geq \tau$, the requirement 3) of Sec. c) of Chap. 1, on establishment of a stationary state, is fulfilled, and the energy estimates obtained above are valid for the accuracy of the measurement of amplitude. In the third case, as is evident from what was just said, only the final step in (1.1) is stationary; for it, the estimates obtained are also valid.

On the other hand, and this is especially characteristic of detection, what is of interest is the accuracy of the measurement not of amplitude, but of some parameter l

with respect to which the detector reorganizes itself. This type of detection also appears in the first step ($l \to \lambda$) of measurement by scheme (1.1), when λ is either the instantaneous position of the receiver on some scale, or the number of the detector with a positive response (in the case of a parallel many-channel receiver).

In order to insure accuracy $1/\sigma_0$ (see (1.10), (1.13), (1.14)) of the order of the resolving power $1/\epsilon$, it is necessary that the probability of error

$$w \approx \frac{1}{\epsilon} w_- \approx \epsilon, \quad w_- \approx \epsilon^2, \tag{2.35}$$

where w_- is the probability of a spurious blip in a single detection event (resolved element). Hence it is easy to find the energy E_0 of the threshold level for detection of a signal

$$E_0 \approx kT \ln \frac{1}{w_-} = 2kT \ln \frac{1}{\epsilon}$$

and the approximate value of the signal energy

$$E_c \approx 2E_0 = 4kT \ln \frac{1}{\epsilon} \approx 4kTI. \tag{2.36}$$

Thus it is evident that in detection (that is, in the first step of measurement), minimum cost of unit information is, in order of magnitude, attained; but this cost relates only to signal energy and does not take into account energy expenditures in the second stage, corresponding to measurement of time or distance. Although practically this (second) part of the energy expenditures is usually unimportant (see footnote 11), in principle it is it that determines the energy cost of accuracy and of quantity of information (2.31)—(2.34).

In closing this chapter, we shall discuss the reasons why thermodynamic limitations have not shown up practically in the improvement of physical experiments.

In physical experiments the aim is, as a rule, to achieve an increase of sensitivity[51-55]; that is, a decrease of the absolute root-mean-square error $\sqrt{\Delta l^2}$ (1.4) and, correspondingly, a decrease of the resolving interval Δl_p (1.5). In this, however, the relative accuracy of the experiment is artificially decreased by having recourse to the transformation (1.6) because of the use of standards (1.9).

What has been said relates to the majority of experiments—even the most delicate and accurate: to the classical experiments of Michelson and to experiments on the use of the Mössbauer effect[53]; to experiments on detection of gravitational waves[51] or, in general, "crucial experiments";[54] etc.

In all these experiments, despite the high sensitivity, the accuracy $1/\sigma$ did not exceed 10^3. Then even according to the worst estimate (2.32), at T = 300°K it is necessary in a single measurement event to degrade only 10^{-7} erg. If, however, in accordance with[56], high relative accuracy is not very necessary in usual experiments, the fundamental physical constants must be known with high accuracy. Nevertheless, all the fundamental constants have been determined with a relative accuracy not exceeding 10^6—10^7 (see, for example,[56,57]). A single measurement event with such accuracy requires the degradation of ~1 erg, and this is also less than the actual technical limitations.

Thermodynamic limitations acquire importance not only in principle but in practice in contemporary systems for the collection, storage, transfer, and processing of information. In the first place, these systems are characterized by transfer and processing of large blocks of information with a high speed of operation. In the second place, transfer of information is connected with large losses on attenuation of signals. In the third place, modern computers operate with numbers of very high accuracy (~ 10^9—10^{13}).

We shall therefore discuss the thermodynamic peculiarities of information transmission and processing that results from the limitations on the accuracy of measurement.

3. THERMODYNAMIC CHARACTERISTICS OF THE PROCESS OF INFORMATION TRANSMISSION

a) The energy meaning of encoding. In order to explain the thermodynamic peculiarities of the process, it is sufficient to restrict ourselves to the analysis of the transmission of one-dimensional, uncorrelated messages (numbers) $x \in X$. This case is directly related to the measurement process considered above, and it enables us to obtain energy estimates of the accuracy of various number representations, which will be necessary later for investigation of the process of information processing.

Suppose that it is necessary to transmit over a communication line the measured value of a scalar quantity in the form of a number. In analogy to the preceding, we shall require that at the receiving end, the quantity of interest to us shall also be represented in the form of a single scalar characteristic. It is obvious that the obtaining of messages in the form of a scalar quantity with the necessary accuracy requires no less energy expenditure than a measurement. In contrast to measurement, however, here it is important what part of the energy is consumed by the system under study (in the present case, the energy of the arriving signal), and what part by the measurement instrument (in the present case, the receiver that accomplishes the decoding). If the only characteristic of the signal that carries the information about the number is its amplitude (energy), then all the estimates obtained in the analysis of measurement apply directly to the signal energy absorbed in the receiver[24,25], and no additional expenditures in the receiver are required. This method of encoding we shall call the analog, or natural, representation of a number. This method of transmission corresponds to a remote measurement, very disadvantageous in the energy respect: in consequence of attenuation of the signal, the energy expenditures in the transmitter are very large. Therefore it is expedient to redistribute the energy necessary to insure the prescribed accuracy of transmission, in such a way as to decrease as much as possible the energy of the received (and therefore of the transmitted) signal, correspondingly increasing the energy expenditures in the decoding at the receiving end. Such redistribution of energy is accomplished by various methods of encoding, when the information about a number is contained not only in the amplitude of the signal but also in its temporal position (positional codes) or in the carrier frequency.

These (and other) methods of encoding are based on the use of a number of degrees of freedom instead of one: a scalar quantity is reduced to a unique corresponding vector, each of whose components requires appreciably less accuracy of representation than did the

original scalar. It was shown above that at high accuracy, the energy expenditures grow exponentially with increase of the quantity of information about a given scalar quantity. On the other hand, it is clear that with increase of the quantity of information not because of accuracy, but in consequence of an increase of the number of components, the energy expenditures are proportional to the increase of the quantity of information (that is, to the number of components). Herein lies the physical meaning of any methods of encoding, among them the optimal—according to Shannon.

b) Natural representation of a number. Before estimating more economical (because of encoding) methods, we shall first consider the energy cost of accuracy of the natural representation of a number.

Above, an estimate of the cost of accuracy was obtained for a single measurement, and then these results were extended by simple considerations to measurement in a domain. Strictly speaking, however, the formulations of the corresponding extremal problems are different. We shall study this problem for the transmission process.

Following [5,7], we shall consider transmission of messages according to the scheme

$$x \to y \to y' \to x', \tag{3.1}$$

where x and x′ are the initial and the received messages, respectively, and y and y′ are the signals at the input and output of the transmission line.

The formulations of the extremal problems considered below, in contrast to the traditional ones [5,7,31], are simplified in that they relate only to uncorrelated messages (numbers) but have an important distinctive property related to the aim of the investigations. Whereas satisfaction of the "quality (accuracy) requirement" of transmission [7] is usually limited by the transmission capacity of the channel, below the role of limiting quantity is assumed by the energy[13]. Therefore only such limitations on the quality of the transmission are permissible as do not require infinitely large energy. Thus, for example, the limitation used in the usual mathematical formulations

$$P\left[\frac{|x-x'|}{x_m} \leqslant \frac{\varepsilon}{2}\right] = 1, \tag{3.2}$$

is not permissible.

This requirement (3.2) on the accuracy originates from the classical theory of approximation of functions, where x is the exact assignment of the function and x′ is the approximate. In this case, in which x and x′ are abstract mathematical items, condition (3.2) is completely correct. But when physical signals participate in the transition x → x′ (y → y′), the condition (3.2) cannot be attained if the energy is bounded.

For any physically encoded quantity, the assumption of zero probability w of an error that exceeds ϵ (if ϵ is less than the a priori interval) is inevitably connected with the assumption of a signal energy as large as may be desired (at nonzero temperature T). This follows from the classical statistical theory of fluctuations [10,26,27].

Thus with allowance for the finiteness of the energy, it would be necessary, in analogy with (1.5) and (1.14), to rewrite condition (3.2) thus:

$$P\left[\frac{|x-x'|}{x_m} \leqslant \frac{\varepsilon}{2}\right] = 1-w, \quad w>0. \tag{3.3}$$

But such a requirement on the transmission, in contrast to (3.2), does not describe its quality and does not permit comparison of different dependences y(x): for different y(x), the same values of ϵ and w can lead to appreciably different accuracy. The peculiarity stated shows up graphically, in particular, in a comparison of the natural representation of a number in analog systems with the digital representation in numerical ones.

Therefore we shall consider hereafter those requirements on the quality of transmission that are expressed through a single scalar characteristic—the mean relative accuracy $1/\sigma$.

As we did for measurement, we suppose, in analogy to (1.8) and (1.16), that the numbers x are distributed uniformly over a segment:

$$p(x) = \frac{1}{x_m}, \quad 0 \leqslant x \leqslant x_m \Leftrightarrow x \in X. \tag{3.4}$$

We shall consider the most natural case of the natural representation of a number—the linear amplitudinal:

$$y = Bx. \tag{3.5}$$

Assuming that the same processes occur in the receiver that were considered in the analysis of measurement, that is that the position x of the recording element is completely determined by the energy $U(x) = \beta x^2/2$ transferred to this element (in the present case a part of the signal energy), and neglecting noise in the channel, we give the conditional distribution $P(x'/x)$ of its density

$$p(x'/x) = p(|x-x'|) = \sqrt{\frac{\beta}{2\pi kT}} \exp\left[-\frac{\beta}{kT}\frac{(x-x')^2}{2}\right]. \tag{3.6}$$

Since the energy expenditures are determined by the relative error σ (or σ_0, in accordance with (1.13), (1.12), (1.18)), we rewrite (3.6), using the facts that $U_{max} = \beta x_m^2/2$ and $p(x) = 1/x_m$ by (3.4):

$$p(x'/x) = p(x)\sqrt{\frac{U_{max}}{\pi kT}} \exp\left[-\frac{U_{max}}{kT}\frac{(x-x')^2}{x_m^2}\right]. \tag{3.7}$$

The physical meaning of the coefficient B in relation (3.5) becomes clear if we take into account that the signal energy is

$$\widetilde{E}_c = y^2 = (1+\alpha)\frac{U_{max}}{kT}\frac{x^2}{x_m^2}, \tag{3.8}$$

where α defines the fraction that the dissipated energy Q is of the transferred energy U.

From the preceding relations (2.16)—(2.18) it is known that

$$\alpha = \begin{cases} \alpha_i = 1, \\ \alpha_r = 2\sqrt{\dfrac{kT}{U}}, \end{cases} \tag{3.9}$$

where the index "i" corresponds to irreversible and the index "r" to reversible, i.e. optimal, according to Sec. b) of Chap. 2, realization of the process of decoding in the receiver. From (3.8) it is evident that, in contrast to measurement, where going over to the optimal mode gives an appreciable power advantage (proportional to the accuracy $1/\sigma$—see (2.18), (2.11)), in the case of transmission this is not so (the signal energy (3.8), which determines the energy of the transmitter, changes by less than a factor two). This is also clear if one takes into account that, independently of the later use of the transferred energy U (useful or degraded in the receiver), at the transmitting end it is all the same necessary to insure that after attenuation, $E_c > U$.

We shall measure energy in relative units, using the notation

$$\frac{U_{\max}}{kT} \equiv A, \quad \frac{\langle U \rangle}{kT} \equiv S, \quad B = \sqrt{(1+\alpha)A}\,\frac{1}{x_m}. \tag{3.10}$$

From (3.4), remembering formulas (1.8) and (1.25), we get

$$S = \frac{A}{3}. \tag{3.11}$$

In transmission of numbers x assigned in the domain (3.4), there occurs a discretization (quantization) of them and of the corresponding signal amplitudes y (3.5), (3.10), in accordance with (1.14): $\Delta x_p / x_m = \epsilon$. Any choice of ϵ in accordance with (3.3) leads to a nonzero probability of error w. Both these quantities separately affect the accuracy $1/\sigma$ and the choice of the mean energy S of (3.11) for transmission of a single message. To determine the total energy S_Σ on transmission of all the messages x_i ($i = 1, 2, \ldots, 1/\epsilon$), it is necessary to multiply the mean energy S by the number of readings $1/\epsilon$:

$$S_\Sigma = S\,\frac{1}{\varepsilon}. \tag{3.12}$$

We write the formulation of the extremal problems[24]:

$$\sigma(\varepsilon, w) \to \min, \quad S(\varepsilon, w) \leqslant C_1 = \text{const}, \tag{3.13}$$

$$\sigma(\varepsilon, w) \to \min, \quad S_\Sigma(\varepsilon, w) \leqslant C_2 = \text{const}, \tag{3.14}$$

each of which corresponds to an energy optimization of the transmission parameters—either for a single message (3.13) or for the whole domain (3.14).

Solution of the problem (3.13) shows that $\sigma_{\min}$ (and correspondingly $I_{\max}$) is attained under the conditions

$$\varepsilon \to 0, \quad w \to 1$$

so that

$$\frac{1-w}{\varepsilon} = \sqrt{\frac{3}{\pi}S}, \tag{3.15}$$

that is, as was to be expected, for a single transmission (or measurement) discretization is not required ($\epsilon \to 0$).

We have in this case (taking account of (3.4), (1.8), (1.25))

$$\sigma^2 = \frac{2}{S}, \quad I \approx \frac{1}{2}\ln S, \quad S \gg 1. \tag{3.16}$$

It is evident that even for a continuous message ($\epsilon \to 0$), the quantity of information is finite and is determined by the released mean energy. We note that when symmetric encoding is used (that is, if in analogy to (1.7) we set $-x_m/2 \leq x \leq x_m/2$ in (3.4)), the energy cost of accuracy decreases by a factor 4:

$$\sigma^2 = \frac{1}{2S}, \quad -\frac{x_m}{2} \leqslant x \leqslant \frac{x_m}{2}. \tag{3.17}$$

Solution of the problem (3.14) for the symmetric domain (3.17) gives energetically optimal relations between the number of readings $1/\epsilon$ (or the resolving power for a measurement), the accuracy $1/\sigma$, and the reliability $1/w$:

$$w = 0{,}4; \quad \varepsilon^2 = \frac{0.6}{(S_\Sigma)^{2/3}} = \frac{0.47}{S},$$

$$\sigma^2 = 3\sigma_0^2 = \varepsilon^2 = \frac{5}{8S}. \tag{3.18}$$

On comparing the values of the minimum errors (σ, σ_0) with the corresponding values (3.17) (where, also, $\sigma^2 = 3\sigma_0^2$) obtained from problem (3.13), we see that the errors of discretization lead to an increase of the energy cost of accuracy by a factor 5/4. For the symmetric encoding (3.4), in correspondence with (1.18),

$$\sigma^2 = 12\sigma_0^2 = 4\varepsilon^2 = \frac{5}{2S}, \tag{3.19}$$

that is, the loss as compared with (3.16) is also in the ratio 5/4.

We point out that with the energetically optimal discretization (3.18) and symmetric encoding, $\epsilon = \sigma$; that is, in conformity with (3.16) and (1.19)

$$I \approx \ln\frac{1}{\sigma} = \ln\frac{1}{\varepsilon}, \quad \varepsilon^2 \ll 1. \tag{3.20}$$

This means that one can speak of a determination of energetically optimal ϵ-entropy in a solution of the extremal problem (3.14). The conditions (3.18) determine a unique value of ϵ, when for given S_Σ the entropy $H_\epsilon = \ln 1/\epsilon$ (despite the fact that $w > 0$ in (3.3)).

From formulas (3.20) and (3.16) it is evident that the quantity of information depends logarithmically on the energy (when $S \gg 1$); and only in the limiting case when $S \ll 1$ do we arrive at the linear dependence $I \approx S$.

A more accurate expression (as compared with (3.16)), valid for arbitrary S, has, in accordance with (2.23) and (1.22), the form

$$I = \frac{1}{2}\ln(2S+1). \tag{3.21}$$

We compare this expression with Shannon's[5] formula for the limiting mean over $\Delta\tau$ of the transmission capacity C (in a frequency band $\Delta\nu$):

$$C = \Delta\nu \ln\left(\frac{P_s}{P_n} + 1\right). \tag{3.22}$$

Since the noise power $P_n = kT\Delta\nu$ and the signal power $P_s = E_c/\Delta\tau$, and since $I = C\Delta\tau$, we get

$$I = \Delta\tau\Delta\nu \ln\left(\frac{E_s}{kT\Delta\nu\Delta\tau} + 1\right). \tag{3.23}$$

In accordance with[5, 8, 12], we suppose that for the quantity of information in one degree of freedom (which is given by formula (3.21)) it is appropriate to set $\Delta\nu\Delta\tau = 1/2$; then, remembering (3.10), (3.9), and (3.8), we arrive at formula (3.21) if in (3.9) $\alpha = \alpha_r = o(1)$. If $\alpha = \alpha_i = 1$, then $2S = \widehat{E}_s$ of (3.8), and the quantity of information decreases (quite inappreciably: by less than unity). Application of optimal slowing down for transmission is inexpedient not only because of the smallness of the positive effect, but also because it requires use of many degrees of freedom (of accumulation) for the transmission of a single reading. But here there is more advantage in the use of an encoding in which a signal with many degrees of freedom (vector) is placed in correspondence to a scalar quantity.

c) Positional methods of encoding. It was shown above that the limited energy resources lead to the necessity for discretization of continuous messages. Here the number of readings $1/\epsilon$ determines the maximum number of resolvable values of the scalar message (number) being transmitted. The transition from the usual limitations on power to limitations on signal energy shows that wide-band signals are more economical energetically (but not as regards power) only in consequence of the use of many degrees of freedom, but only within definite limits. In the discussion of detection it was explained that certainty of detection always requires that the signal energy $E_S > kT$, and various methods of accumulation do not change this limitation. Therefore increase of the number of degrees of freedom M of the signal to more than $1/\epsilon$ is meaningless.

The most economical method of encoding a number, from the point of view of transmitter energy, is the single-pulse positional, in which

$$M = M_{\max} \approx 1/\varepsilon. \tag{3.24}$$

Here use is made of a special case of the vector representation—a unit vector in which all components except one are zero. In consequence of (3.24), this method of encoding is the least economical from the point of view of bandwidth (or duration); that is, it requires an appreciable increase of the transmission capacity of the channel as compared with the natural representation of a number.

Another method of positional encoding, connected with the digital representation of a number, is the most widespread. This method is somewhat less economical with respect to power, but more economical with respect to the number of degrees of freedom:

$$M_a = (\ln a)^{-1} \ln \frac{1}{\varepsilon},$$

where a is the base of the positional numeration system. From the preceding it is known that it is energetically advantageous to take the minimum possible value $a_{\min} = 2$. Then

$$M_2 = (\ln 2)^{-1} \ln \frac{1}{\varepsilon}, \tag{3.25}$$

and the energy of the signal, as likewise for (3.24), is determined by the requirements for its certain detection. Because of the irregular (anomalous[12]) law of distribution of the errors (as compared with the natural amplitude encoding, (3.7)), the relations between σ_0, ϵ, and w are significantly different from (3.18).

On neglecting coefficients of order unity, we find at once:

$$w \cong \varepsilon \cong \sigma_0. \tag{3.26}$$

Hence it is easy to find the requirements on the probability w_- of false alarms—of a spurious blip in some position j different from the i-th (of the transmitted message).

For single-positional encoding, in consequence of (3.24) and (3.26), we get in analogy to (2.35)

$$w = M_{\max} w_-, \quad w_- = \varepsilon^2. \tag{3.27}$$

For the binary digital representation (3.25), we have

$$w = M_2 w_-, \quad w_- = \frac{\varepsilon \ln 2}{\ln (1/\varepsilon)}. \tag{3.28}$$

Hence it is easy to find the threshold energy for both cases:

$$E_{\text{th}} \approx kT \ln \frac{1}{w_-}, \tag{3.29}$$

and on setting $w_- = w_+$ (the probability of nondetection), we get for the energy of a signal that represents a number by the single-positional code (3.27):

$$E_s^{(\Phi)} \approx 2E_{\text{th}}^{(\Phi)} = 4kT \ln \frac{1}{\varepsilon} \approx 4kTI. \tag{3.30}$$

For the binary digital representation, the signals (ones) can be in all M_2 positions (3.25). On taking into account that on the average a number contains half of unity, we get

$$E_p = \frac{1}{2} M_2 E_s^{(p)} = M_2 E_{\text{th}}^{(p)} = kT \ln \frac{1}{\varepsilon} (\ln 2)^{-1} \left(\ln \frac{1}{\varepsilon} + \ln \ln \frac{1}{\varepsilon} + \ln \frac{1}{\ln 2} \right),$$

$$E_p \approx kT \left(\ln \frac{1}{\varepsilon} \right)^2 = kT (I)^2. \tag{3.31}$$

It is evident that with the single-positional encoding (3.30), the energy cost of unit information,

$$\hat{e}_0 = \frac{E_s^{(\Phi)}}{I} \approx 4kT, \tag{3.32}$$

as for detection (2.36), agrees in order of magnitude with the minimal estimate according to[8,21]. But as has already been mentioned in Sec. a) of Chap. 3 and in Sec. b) of Chapter 2, in the decoding of the signal after detection a measurement is made with accuracy $\sim 1/\epsilon$, when all the energy estimates of Secs. b), Chap. 2 and b), Chap. 3 are valid. Here it is important that the energy expended in the decoding at the receiving end does not affect the choice of the energy of the signal (which attenuates in the transmission channel).

The energy cost of unit information in the digital representation,

$$e_p = \frac{E_p}{I} \approx kT \ln \frac{1}{\varepsilon} \approx kTI \tag{3.33}$$

increases with increase of I (the number of digits $\sim \ln 1/\epsilon$). As for single-positional coding, for the transition to a scalar quantity, that is in decoding (deciphering), here again the energy estimates of Secs. b), Chap. 2 and b), Chap. 3 are valid.

4. THERMODYNAMIC MODELS OF THE PROCESS OF INFORMATION PROCESSING

a) Complexity criteria of information processing. During the last decade, considerable development has occurred in the mathematical theory of complexity; its various directions have been quite fully elucidated in the monograph and review literature[58-64]. In[65], the applied aspects of the theory of complexity were considered from the point of view of the problems of control theory. In this connection, a unified approach was proposed[65] to the various directions of complexity theory, and a classification of these directions was carried out. We shall discuss two basic classes of complexity criteria[65,63] that have a direct relation to information processing.

By "information processing" we shall understand a single-valued transformation Γ of a set of elements $x \in X$ to a set of elements $\Gamma x = f \in X_f$, when both sets X and X_f are subsets of some set of numbers. Informally—in the spirit of the preceding sections—we shall represent information processing as indirect measurement. Let it be necessary to determine the value of some physical quantity f, direct measurement of which is impossible, but somehow a unique relation is known between it and another physical quantity x that is accessible for direct measurement. Then measurement of x and subsequent transformation of the physical quantity corresponding to the value of x into a value of the function f(x), represented in the same form, constitutes indirect measurement (information processing).

Consequently it is necessary, first to prescribe somehow the transformation Γ, and second to realize it; that is, to perform the corresponding transformation.

The first class of complexity criteria[65,63], in the application to information processing, characterizes the complexity of the prescription of the transformation Γ. This may be complexity of the text, of the algorithm (number of symbols), or of the scheme (number of elements), in which the rules are described for going from the objects $x \in X$ to the objects $f \in X_f$. In particular, this may be the length of the binary word (program) by which

some automaton (machine) recovers a word f from a word x. In the papers of Markov [66] and Kolmogorov [67,68], a general theory of complexity criteria of this class was developed—with application to any calculable functions. It is important that in [67] a proof was given of the invariance (to within an additive constant) of these complexity criteria with respect to the automaton that accomplishes the deciphering of the program.

We note that the criteria described characterize the complexity of an algorithm (its transcription), which in general is unrelated to the complexity of its execution (that is, for example, the duration of the calculation). The latter aspect of the complexity of information processing, that is the laboriousness of the calculation by a given algorithm, relates to the second class of criteria [63,65] and characterizes the complexity of the processing process itself. Here are introduced signalling functions [63]: temporal or capacitive, characterizing respectively the resources of time (for example, the number of elementary operations) or of memory necessary for realization of the given transformation Γ. It is important that these estimates depend both on the algorithm and, in general, on the machine (the mathematical model of the automaton) on which the algorithm is realized. A satisfactory machine-independent theory of criteria of this class does not exist at present [63,65].

Along with this, it would be desirable to estimate computational complexity (complexity of the process of information processing) in a manner invariant with respect to the algorithm and the machine, and based only on very general rules. It appears that in order to determine such an invariant characteristic of the complexity of the processing process, it is necessary to go over from mathematical models to physical. In the preceding chapters it has been shown that the complexity of the processes of measurement and of information transfer are uniquely characterized by the work performed and the energy thereby degraded. It would be tempting to represent the process of information processing also as a thermodynamic process and to characterize its complexity by the energy degraded[14]. We shall set forth some considerations on this score, which were briefly projected in [24,25].

b) Energy criterion for the complexity of information processing. The essential difference between mathematical and physical models of information processing is that the former operate from the very beginning only with discrete objects (most often with binary words), while the latter consider macroscopic physical parameters having the character of continuous quantities. On the other hand, the physical models explicitly introduce limitations on the energy resources; and it is purely in consequence of this, according to Sec. b), Chap. 3, that there arises a necessity for discretization (quantization) of the objects of study. Mathematical models, on the contrary, start from the hypothesis of potential realizability [64], which according to [69] consists of "an abstraction from the real limits to our constructive possibilities, imposed by the finiteness of our existence in space and in time." Thus practical realizability [64] is determined precisely by physical limitations; and if it is a question of thermodynamic limitations, then the practical limitations determined by them depend to a significantly smaller degree on the level of development of technology than do, for example, limitations with respect to speed of operation.

We shall now discuss the fundamental difference between the processes of measurement and of information transfer and the process of information processing. Whereas for the former, an energy criterion of complexity was determined solely by a single quality criterion—accuracy—of these processes, for processing this is not so. Processing of information (like any processing, for example of materials) cannot be characterized by a single scalar quantity, in contrast to transmission of information (or transportation of materials). Here there is a significant change of the quality (value) of the information (or of the other product being processed). If it is necessary to obtain a piece not only of given weight but also of a definite form, then the complexity of such a piece is determined not only by the weight of the original billet but also by the process of treating it. Analogously, the complexity of a function is characterized not only by its accuracy but also by its form, and correspondingly is determined not only by the accuracy of the original information but also by the process of treating it. Of course the complexity of the processing of a piece depends on the technology, so that we must take as an unconditional characteristic the minimum (over all imaginable technologies) complexity. Analogously, the complexity of a process of information processing must be characterized by the minimum (over all imaginable models of the processing) energy expenditures. It appears that it is possible to approximate such estimates by considering only two thermodynamic models of information processing [24].

Returning to the treatment of the process of information processing as an indirect measurement[15], we represent it according to the scheme

$$x \to y \xrightarrow{\Gamma} z \to f. \tag{4.1}$$

Here x is the physical quantity being measured (or a number corresponding to it), y is the signal representing it, and z is the signal, represented in the same form (as y), corresponding to the value of the function—the number f(x). In (4.1) the symbols correspond to (3.1) (and not (1.1)); therefore, as in (3.5) and (3.10), y and z describe the amplitude of the signals, and y^2 and z^2 their energy (see (3.8) and (3.9)).

Because of the insufficiency of a scalar characteristic for estimation of the quality of information processing, and in accordance with the generally accepted method of vector representation of functions with integrable square, we shall represent them in the form of points (vectors) in a Hilbert space L_2. It is necessary to choose a space L_2 because, as was shown above, the mean energy expenditures are proportional to[16]

$$\langle f^2 \rangle = \int_X p(x) f^2(x)\, dx = \| f \|^2. \tag{4.2}$$

Here $\|f\|$ is the norm of the function in L_2; the distance $\rho(f, g)$ between functions f and g is defined thus:

$$\rho^2(f, g) = \| f - g \|^2 = \int_X p(x) [f(x) - g(x)]^2\, dx. \tag{4.3}$$

As earlier, we shall suppose that a requirement on the accuracy of realization of f(x) is prescribed. From the requirement on the accuracy $1/\sigma_{0f}$ (see (1.13), (1.18), (1.24), (1.25)) it is easy to find requirements on the accuracy of representation of the original variable (the directly measurable quantity x):

$$\sigma_{0f}^2 = \sigma_{0x}^2 \int_X p(x) \left(\frac{df}{dx}\right)^2 dx = \sigma_{0x}^2 \left\| \frac{df}{dx} \right\|^2. \tag{4.4}$$

It is evident that, in general, for p(x) = const

$$\frac{1}{\sigma_{0x}} \geqslant \frac{1}{\sigma_{0f}}, \tag{4.5}$$

this corresponds fully to the intuitive idea that quantitatively there must be no less of the original product than of the final (in the present case, as regards quantity of information $I \sim \ln(1/\sigma)$). The inequality (4.5) is the stronger (see (4.4)), the more f(x) oscillates.

From (4.2) and the preceding it is evident that the norm of the function is determined entirely by the requirements on the accuracy $1/\sigma_{0f}$ (or $1/\sigma_f$) and is proportional to it. We shall require that the encoding of the original information (3.5) shall be the same as of that sought; that is, in (4.1)

$$z = Bf. \tag{4.6}$$

This means that, in consequence of (4.4) and (4.5), the original variable must be so normalized as to insure higher accuracy for the same energy coefficient A in (3.10). This reduces to an increase of the range x_m of the measurement of x; that is, for the original signal we must set

$$\begin{gathered} y = Bu, \quad u = cx, \quad c > 1, \\ 0 \leqslant x \leqslant x_m, \quad 0 \leqslant u \leqslant u_m = cx_m. \end{gathered} \tag{4.7}$$

The simplest method of realization of information processing according to the scheme (4.1) consists in measurement of x with the accuracy necessary according to (4.4) and subsequent functional transformation of the measured quantity to the f(x) sought. If this second step is realized with a functional transformer (for example, a potentiometer), then it corresponds completely to a second measurement. This step can also correspond to information transmission (with subsequent decoding), if the signal y = Bcx is used to interrogate a table for the values of f.

The last case illustrates especially graphically the need for discretization of x (and y) and quantization of f (and z). According to (3.18) and (3.19), we may choose the energetically optimal values

$$\varepsilon_x = \sqrt{3}\,\sigma_{0x}, \quad \varepsilon_f = \sqrt{3}\,\sigma_{0f} \geqslant \varepsilon_x. \tag{4.8}$$

Thus it follows from this model of information processing that its energetic complexity is

$$\begin{gathered} S_x^f = S_f + S_x, \\ S_x^f \approx C \| f \|^2 \left(1 + \left\| \frac{df}{dx} \right\|^2\right), \end{gathered} \tag{4.9}$$

where the coefficient of proportionality C is determined on the basis of the preceding chapters. It is easy to see that (4.9) does not always determine the minimum necessary energy expenditures: thus even if f(x) is very close to x ($f \to x$, $df/dx \to 1$), the model considered requires double the expenditures as compared with direct measurement. The closeness of the functions (4.7) and (4.6) is characterized by the distance $\rho(f, u)$ of (4.3):

$$\rho^2(f, u) = \int_X p(x)\,[f(x) - cx]^2\,dx. \tag{4.10}$$

With the notation

$$f(x) - cx \equiv \varphi(x), \tag{4.11}$$

we get

$$\rho^2(f, u) = \| \varphi(x) \|^2 \tag{4.12}$$

and go over to the description of the second model [24]. It seems obvious that for $\|\varphi\| \ll \|f\|$ it is expedient in the information processing device (IPD) to use the energy of each i-th original signal y_i by slightly changing its value; that is, to execute energy jumps

$$\zeta_i^2 = B^2\varphi_i^2 = B^2(f_i - cx_i)^2. \tag{4.13}$$

It is easy to understand, however, that this energy is insufficient for processing: it is still necessary to know the value of the jump φ_i corresponding to each x_i at the input. In the general case it is possible to interrogate a table for the values φ_i, using part of the energy of the input signal for interrogation of the memory (as in the first model), and to sum another part with the jump energy ζ_i^2 of (4.13). It is evident that in such a model (the second), a continuous transition is insured from information processing to measurement for $\|\varphi\| = \rho(f, u) \to 0$, and its energetic complexity is

$$\begin{gathered} S_x^\varphi = S_\varphi + S_x, \\ S_x^\varphi \approx C \| f \|^2 \left(1 + \frac{\| \varphi \|^2 + \| d\varphi/dx \|^2}{\| f \|^2}\right). \end{gathered} \tag{4.14}$$

A general definition of the energetic complexity $S_{x,f}$ of a realization of a function f(x) is, in accordance with the preceding and with (4.13) and (4.14):

$$S_{x,f} = \min \{S_x^f, S_x^\varphi\}. \tag{4.15}$$

The fact that the complexity of the transition y → z in (4.1) is actually characterized by the distance $\rho(f, u)$ of (4.10)–(4.12) and that ρ^2 is proportional to the mean energy jump (4.13) follows from the calculation model [24], which constitutes the central part of the second processing model.

Let the IPD be replaced by a "black box," into whose input enters a signal y = Bcx, (4.7). If the "black box" is located in a thermostat, then on the signal is superposed additive noise with mean energy kT determined by the temperature T of the thermostat. For $T > 0$, there is a nonzero probability that an arbitrary value of the input will be converted to an arbitrary other value, including that prescribed by the processing z = Bf of (4.6). The higher this probability, then, by hypothesis, the simpler the processing.

We write the expression for the probability density w(f/u) of transition of input values to output, starting from a Gaussian distribution law for the fluctuations (3.6), (3.7), and (4.13):

$$w(f/u) = \sqrt{\frac{B^2}{\pi}}\, e^{-B^2(f-cx)^2} = \sqrt{\frac{B^2}{\pi}}\, e^{-B^2\varphi^2(x)}$$

and, taking into account the necessity for discretization of the input and output values, and also the finite accuracy of reproduction of a function, we go over from probability density to the finite values of the probability of transition of the point cx_i to an ϵ_f-segment:

$$W_i(f_i/u_i) = \frac{B}{\sqrt{\pi}}\, e^{-B^2\varphi_i^2(x)}\varepsilon_f. \tag{4.16}$$

We furthermore write the mean (over P(x) value of the logarithm of this probability:

$$\langle -\ln W(f/u) \rangle = \sum_{i=1}^{1/\varepsilon} p_i \left[B^2\varphi_i^2(x) + \ln \frac{\sqrt{\pi}}{B\varepsilon_f}\right], \tag{4.17}$$

where, in accordance with (3.10) and (3.18), $B\epsilon$ = const (of order unity). Taking the ratio of (4.17) to the logarithm of the probability of spontaneous appearance of unity from zero (that is, eliminating the accuracy of reproduction of the coefficient B), and letting T → 0, we get:

$$\lim \frac{\langle -\ln W(f/u)\rangle}{\langle -\ln W(1/0)\rangle} = \rho^2(f, u) = \| \varphi_i'(x) \|^2. \qquad (4.18)$$

It is well known [10] that the argument of the exponential function that describes the probability of the corresponding fluctuation is equal to the minimum work that must be performed on the system for transition of it from an initial equilibrium state to another, determined by the fluctuation. Thus the calculation model described gives an expression for the energy that it is necessary to transfer from one body to another. And this energy, as follows from the preceding, completely determines the value of the dissipated energy (in the irreversible case, is simply equal to it). Thus formula (4.18) represents the value of the mean energy jump for the prescribed processing. The mean value of the increase of entropy in the system is proportional to the square of the distance $\rho^2(f, u)$ (in the irreversible case) or to the distance $\rho(f, u)$ (for optimally slowed-down realization of the calculation).

Returning to the determination of the energetic complexity $S_{x,f}$ of processing as indirect measurement (4.15), (4.14), (4.9), we note that always

$$S_{x,f} = S_f \frac{1}{r}, \quad r < 1, \qquad (4.19)$$

where S_f is the energetic complexity of a direct measurement. Thus the coefficient 1/r characterizes the relative complexity of the processing: the increase of complexity of reproduction of f(x) according to indirect information about x, accessible for direct measurement.

Incidentally, it is possible to introduce also the concept "value of initial information" [24], which it is natural to describe by the coefficient $r < 1$, the reciprocal of the coefficient of relative complexity of the processing[17].

As follows from the above considerations, the complexity of a calculation is greater (and the value of the initial information less), the greater the distance $\rho(x, f)$ between the vectors of the initial and the sought functions.

This treatment of the complexity of a calculation and of the value of initial information makes it possible, it seems to us, to dispel misunderstandings that have arisen in discussion (in terms of a scalar representation of information) of the useful effect of the work of a computer (see, for example, [8], p. 295 of original English edition), and the related problem of comparing the information obtained during its solution. A computer in the process of solving a problem of course does not increase the quantity of information, but decreases it, simultaneously increasing the value of the information.

c) **Effective methods of calculation.** In practical calculations, functional transformers (or the tabular method of realization of functions), corresponding to the first processing model of Sec. b), Chap. 4, are applied quite rarely. Besides the nonuniversality of such devices, here it is essential to use amplitude encoding, in which the energetic complexity is proportional to the square of the accuracy (in actual cases of limitation on the processing time). At $T = 300°K$, a number with accuracy $10^9 \approx 2^{30}$ requires energy 10^5 erg = 10^{-2} J. This is why analog methods are not used for accurate calculations. It has already been mentioned that accuracy this high (and higher) in the natural representation of numbers not only is not required and is practically not realizable from energy considerations, but also does not correspond to anything physical, since even the accuracy of the fundamental constants is considerably less (apropos of this, see also other fundamental considerations in [74]).

The requirement of high accuracy arises in many-dimensional problems, when the number of distinguishable initial situations increases exponentially with increase of the number n of measurements of the initial domain $X^{(n)}$ of definition of the function. In this case, use of the first processing model is difficult. In fact, for this purpose it is necessary to bring the dimensionality n of the domain $X^{(n)}$ of the input variables into agreement with the dimensionality μ of the memory. Here the energy expenditures are $S^{(n)} \sim nq^{n/\mu}$, where $q = 1/\epsilon$ is the number of distinguishable situations for each of the variables. Since $\mu \leq 3$ (practically, even $\mu \leq 2$), for large n the increase of $S^{(n)}$ is appreciable.

In the second model of information processing, the execution of the energy jumps in the transition $y \to z$ of (4.1) is not necessarily connected with interrogation of a table for the values φ_i. Properly, a calculation is usually characterized by a sequence of certain transformations of the argument x and of the corresponding signal y (for example, multiplications and additions in the realization of a polynomial). In this case the transition from argument to function proceeds not along a straight line connecting the corresponding vectors $\{x_i\}$ and $\{f_i\}$, but along some broken line, which is closer to straight, the broader the set of elementary operations. Nevertheless, this transition along a broken line may prove energetically advantageous, since it does not require turning to a table. In particular, in this case it is possible to use (as is done in a digital computer) a binary digital representation of numbers, which energetically is significantly cheaper than the natural representation[18] (see (3.31) and (3.33)).

In digital coding, the energy of each given number is on the average the same (in contrast to amplitude coding, where the energy is proportional to the square of the value of x) and depends only on the digit capacity (n). Therefore all the elementary operations entail about the same energy expenditures (of the order of n unit jumps (3.33)), and the energetic complexity of calculations is, in a natural manner, characterized sufficiently well by the number of elementary operations. As before, the distance $\rho(f, u)$ determines the complexity, but in consequence of (3.31) the energy expenditures are proportional to $(\ln \rho)^2$, and not to ρ^2 as in amplitude coding.

We note that the definitions and model of information processing considered above, with application to realization of a function f(x), can be easily extended to the solution of some problem (in particular, the concepts of complexity of the solution, value of the initial information, etc.). The effectiveness of one or another method of solution of a given problem—under given requirements on the accuracy of the solution—is characterized sufficiently well by the number of elementary operations. We recall that the effectiveness of the second model of processing as compared with a general table method (first model) was determined by a lucky choice of the zero-order approximation to the desired function. In the present case, the independent variable itself was used as the zero-order approximation, and the expediency of such a method was determined by the nearness according to (4.10); this corresponded to a change of requirements on relative accuracy for the second stage (interrogration of tables for values of φ_i instead of f_i).

In connection with the facts just noted, we shall make

a few qualitative remarks about constructive methods in computational mathematics. Application of good zero-order approximations is peculiarly effective in many-dimensional problems (as follows, in particular, from the models considered above). Therefore it is important to have some general approaches to the search for such zero-order approximations. If we consider some class of problems, then for "almost all" problems of this class, some general and, as a rule, computationally very cumbersome method seems the only possible one. In particular, for example, for the solution of boundary problems described by equations of elliptic type, the method of grids serves as such a method.

On the other hand, as a rule, in each class of problems there exist "singular points" in a whole domain of this class, for which special but very effective methods of solving them are known (or perhaps proposed). In particular, for the class of boundary problems mentioned, a computationally simple method has been proposed for regions whose boundaries are described by equations of the second order (see [75,76]).

If we use solutions obtained for special problems as zero-order approximations (principal part [75]) for solution of a broader class of problems, in some sense "close" to these special problems, then it is natural to expect a significant economy in the computational work. Thus, for example, for the boundary problems [75,76] criteria of closeness have been given. If in the transition thereafter from the zero-order approximation to subsequent ones use is made, for example, of the same general grid method, it is found that the grid step can be chosen appreciably larger. In [77] it is shown that the effectiveness of this kind of zero-order approximation is connected with an appreciable lowering of the requirements on the relative accuracy of the residual part of the solution.

5. CONCLUSION

A whole series of physical problems requires the carrying out of computational jobs that cannot be provided by contemporary computational technology [78]. Similar difficulties arise also in problems of planning, control, and projection (see [79,80]). It is interesting to have estimates that limit in principle the possibilities of computational technology. It is desirable that such estimates not be too high, like those cited in the literature [81-83] and characterized by a number of units of information to be processed $\sim 10^{100}$. This figure was obtained either from a calculation of the number of nucleons in the universe and of the number of events at the atomic level from the instant of formation of the earth's crust [81], or from an estimate of the energy of the universe equivalent to its mass and used for the recording of information [82,83].

From thermodynamic considerations, the energy cost of accuracy was determined above for various forms of number representation; on the basis of this, it is possible to obtain more realistic estimates of the limiting possibilities of computational technology.

The requirements for high accuracy are realized in digital (in particular, binary) representation of numbers, where the energy is proportional to the square of the logarithm of the accuracy. Here high accuracy is connected with great depth of the calculations and leads to higher requirements on reliability [14,84-86]. In this case, for a thirty-digit binary number, if we allow for the requirement on the reliability, the energy cost of a binary unit at T = 300°K has the order 10^{-11} erg. Analysis of data on the logical schemes of contemporary technical devices (electronic [87] and optical [88]) shows that over a wide range of frequencies (from 10^4 to 10^9 Hz) the energy of a single switching event is approximately the same and has the order 10^{-3} erg. Thus the limiting value differs by 8 orders of magnitude from that actually attained. Assuming that this discrepancy can in principle be overcome, we get the following estimate.

If all the energy produced on the earth in the course of a year ($\sim 10^{14}$ kWh) is expended on information processing, then there will be enough of it for the performance of 10^{36} (300/T) elementary operations (such as addition of thirty-digit binary numbers). We emphasize that, despite the fact that this estimate takes account only of thermal noise, it is valid in all cases. In fact, if for $kT \gg h\nu$ this is an accurate estimate, then under the necessity for taking account of quantum noise it can be taken as a top estimate (as regards number of operations). This estimate is not as excessive as those cited above, and it shows that the energy resources actually and quite strongly limit the possibilities of digital computers of the contemporary structure.

In order to make this assertion concrete, we shall cite an example [78]. For a quantum-mechanical calculation of the methane molecule, it is necessary to perform calculations by the grid method at 10^{42} points. If we suppose that at each point it is necessary to perform in all 10 elementary operations, and if we assume that all the calculations are carried out at a superlow temperature ($T = 3 \times 10^{-3}$ °K), then even under these conditions the calculation of the methane molecule requires expenditure of the energy produced on the Earth in about a century.

Similar requirements arise in the problem of calculation of plasma traps [78] and in numerous problems of control and of planning [79,80]. Therefore, along with increase of the productivity of digital computers and building up of the stock of them, it seems urgent to search for other (nonarithmetic) methods of solving complicated problems.

The author expresses his thanks to B. B. Braginskiĭ and L. A. Rivlin for useful discussions, which contributed to the improvement of this article.

[1)] Throughout the present paper also, dimensionless entropy is used and is denoted by the symbol H. This same notation has been accepted in information theory, beginning with Shannon's paper. In the physical literature, however, this notation occurs much earlier: in Boltzmann and in Gibbs.

[2)] For the special case in which both the prior and the posterior distributions are normal, relations of the type of l.19) and the later (1.22) were given [37]. It must be emphasized that until recently one encountered definitions of accuracy according to Brillouin [8, 16] as (in our notation (1.14) $1/\epsilon$, but of quantity of information as $I \approx \ln(1/\epsilon)$. Because of the condition $w>0$ of (1.14), such a definition is of course ambiguous and therefore incorrect. This problem will be discussed in detail below (Chap.3) in connection with energy estimates.

[3)] It is possible to derive (1.20) formally if in the definition of the error $\rho(x, x')$ we take into account the prior data:

$$\rho(x, x') = \begin{cases} |x-x'| & \text{for} \quad x' \in X, \\ x_{max}-x & \text{for} \quad x' > x_{max}, \\ x-x_{min} & \text{for} \quad x' < x_{min}. \end{cases}$$

In this case, of course, equalities of the type (1.4) become incorrect.

[4)] Brillouin [8] therefore introduces a distinction between free and bound information. The latter arises when the possible cases (events) can be described as microstates of a physical system.

[5)] In the particular case when Y is an electric circuit, the system (2.1), (2.2) describes the transfer process in a galvanometer (see [27]), the system (2.1, (2.3) in an electrometer.

[6)] We note that this effect $|\Delta H_1|$ is small only in comparison with the enetropy of the whole system (H_0 or H_∞ in (2.6) or (2.7); but for $H_0 - H(\alpha)$, that is for the entropy determined by the parameters y and x, the effect $|\Delta H_1|$ is always significant. This is important, since the exchange of energy U occurs precisely through these degrees of freedom (y and x).

[7)] It can be shown that results analogous to those given below also hold if the regulator slowly changes, instead of the parameter H or B, the parameter β in (2.1).

[8)] That is, with decrease of the force difference ϵ, there is an increase of t_0, and with it of the indeterminate (Brownian) mismatch of the forces.

[9)] The quotation marks point out that the optimal, from the point of view of the connection between σ and Q_Σ, realization of the measurement process, as is evident from (2.16), corresponds to a finite speed of the process—of the order of the thermal. Not only an increase but also a decrease of the speed, relative to the "natural" (thermal) speed, is connected with an increase of the dissipative losses.

[10)] Because the relations (2.11), (2.12), (2.17). and (2.18) and also the subsequent (2.22) are valid both for measurement in a domain (estimation of a random quantity) and for a single measurement (estimation of a constant parameter), it seems natural to extend formula (2.23) and the concept of quantity of information (defined only for a ran. om quantity) also to a single measurement.

[11)] The corresponding natural conditional extremal problems are discussed in the next chapter.

[12)] Practically, this is sometimes completely justified: for example, in a radar measurement of distance all that is calculated is the energy of the signal necessary for certain detection (this corresponds to the step $l \to \lambda$ in (1.1)), and there is no interest in the energy expended in the scanning device or the meter. Along with this, it is precisely in this step ($F \to x$ in (1.1)) that a physical quantity is formed that is proportional to the distance, and it is to it that the estimates obtained above relate. From (2.32) it is evident that for the accuracy usually realized ($\sim 10^3$), the expenditures in the second stage are negligibly small in comparison with the energy of the signal.

[13)] It is important that in this formulation the results obtained always lead to a finite value of the transmission capacity. On the other hand, some of the results in the usual formulation actually are valid only with an infinitely large energy. We remark that the limitations on the transmission capacity introduced in the usual formulations, and connected with a limitation of the mean signal power, are not equivalent to limitations on the mean energy.

[14)] We note in this connection that even in formal mathematical formulations, the term "computational work" is used [63]—not, of course, in the thermodynamic sense.

[15)] A treatment of the computational process similar to that given in Sec. a, Chap. 4, is contained in [70], where consideration is given to mathematical problems of the theory of (indirect) measurements, related to approximate solution of certain integral equations on computers.

[16)] See (3.4) and (3.8) - (3.11) for the signal energy. The degraded energy Q, in accordance with (2.5), (2.11), and (2.12), is also proportional to the square of the accuracy in an irreversible (operative) realization of the processes. The transition to estimates of Q for optimal slowing down of the processes is easily accomplished by formulas (2.16) - (2.18) and (2.22); we shall as a rule not do this.

[17)] The concept "value of information," outside the connection with the computational process, was considered in [33, 77-73] and other papers, where this concept assumed a more general case of use of the information obtained: for attainment of some purpose (control). In the case considered here, the purpose consists entirely in calculation of a prescribed function.

[18)] If in a calculation with digital numbers it is necessary to use a table, then a transition to a natural number is required (for example, in a decoder), and this entails an increase of the energy expenditures.

[1] Ya. M. Gel'fer, Istoriya i metodologiya termodinamiki i statisticheskoĭ fiziki (History and Methodology of Thermodynamics and Statistical Physics), M., Vysshaya shkola, 1969.

[2] L. Szilard, Z. Phys. **53**, 840 (1929).

[3] R. V. L. Hartley, Bell Syst. Tech. J. **7**, 535 (1928) (Russ. transl. in: Teoriya informatsii i ee prilozheniya (Information Theory and Its Applications), M., Fizmatgiz, 1959, p. 5).

[4] N. Wiener, Cybernetics, Technology, Press-Wiley, 1948 (Russ. transl., Sov. radio, 1968).

[5] C. E. Shannon, transl. in: Raboty po teorii informatsii i kibernetike (Works on Information Theory and Cybernetics), M., IL, 1958).

[6] N. Wiener, The Human Use of Human Beings: Cybernetics and Society, Houghton Mifflin, 1950; Avon Books, 1967 (Russ. transl., M., IL, 1958).

[7] A. N. Kolmogorov, in: Sessiya AN SSSR po nauchnym problemam avtomatizatsii proizvodstva. Plenarnye zasedaniya (Session of the Academy of Sciences, USSR, on Scientific Problems of the Automatization of Production. Plenary Meeting), M., Izd-vo Akad. Nauk SSSR, 1957, p. 66.

[8] L. Brillouin, Science and Information Theory, Academic Press, 1956 (Russ. transl., M., Fizmatgiz, 1960).

[9] P. Chambadal, Development and Application of the Concept of Entropy, (Russ. transl.) Nauka, 1967.

[10] L. D. Landau and E. M. Lifshitz, Statisticheskaya fizika (Statistical Physics), M., Nauka, 1964 (transl., Pergamon, Press-Addison Wesley, 1969).

[11] J. von Neumann, Mathematische Grundlagen der Quantenmechanik, Springer, 1943, 1968 (transl., Mathematical Foundations of Quantum Mechanics, Princeton, 1955; Russ. transl., M., Nauka, 1964).

[12] A. A. Kharkevich, Ocherki obshcheĭ teorii svyazi (Outlines of a General Theory of Communication), M., Gostekhizdat, 1955.

[13] D. Gabor, Phil. Mag. **41**, 1161 (1950).

[14] J. von Neumann, Theory of Self-Reproducing Automata (ed. A. W. Burks), Univ. of Ill. Press, Urbana, 1966 (Russ. transl., M., Mir, 1971).

[15] Yu. A. Shreĭder, in: Informatsiya i kibernetika (Information and Cybernetics), M., Sov. radio, 1967, p. 15.

[16] L. Brillouin, Scientific Uncertainty and Information, Academic Press, 1964 (Russ. transl., M., Mir, 1966).

[17] P. V. Novitskiĭ, Osnovy informatsionnoĭ teorii izmeritel'nykh ustroĭstv (Foundations of an Informational Theory of Measuring Devices), M., Energiya, 1968.

[18] V. I. Rabinovich and M. P. Tsapenko, Informatsionnye kharakteristiki sredstv izmereniya i kontrolya (Informational Characteristics of Measurement and Control Media), M., Energiya, 1968.

[19] V. Yu. Urbakh, Problemy Kibernetiki, no. 10, M., Fizmatgiz, 1963, p. 99.

[20] H. Reiss, J. Stat. Phys. 1, 107 (1969).

[21] D. S. Lebedev and L. B. Levitin, in: Teoriya peredachi informatsii (Theory of Information Transmission), M., Nauka, 1964, p. 5.

[22] L. B. Levitin, in: IV konferentsiya po teorii peredachi i kodirovaniya informatsii. Sektsiya II (Fourth Conference on the Theory of Information Transmission and Coding. Section II), Moskva-Tashkent, Akad. Nauk SSSR, 1969, p. 111.

[23] R. P. Poplavskiĭ, Dokl. Akad. Nauk SSSR **202**, 562 (1972) [Sov. Phys. Doklady **17**, 52 (1972)].

[24] R. P. Poplavskiĭ, ref. 22, Sec. I, p. 209.

[25] R. P. Poplavskiĭ, in: V Vsesoyuznyĭ simpozium po kibernetike (Fifth All-Union Symposium on Cybernetics), Tbilisi, Akad. Nauk SSSR, 1970, p. 45.

[26] Ya. I. Frenkel', Statisticheskaya fizika (Statistical Physics), M.-L., Izd-vo Akad. Nauk SSSR, 1948.

[27] M. A. Leontovich, Statisticheskaya fizika (Statistical Physics), M.-L., Gostekhizdat, 1944.

[28] H. Cramér, Mathematical Methods of Statistics, Princeton, 1946 (Russ. transl., M., IL, 1948).

[29] M. I. Kornfel'd, Usp. Fiz. Nauk **85**, 533 (1965) [Sov. Phys.-Uspekhi **8**, 299 (1965)].

[30] W. Feller, Introduction to Probability Theory and Its Applications, vol. 2, Wiley, 1966 (Russ. transl., M., Mir, 1967).
[31] R. Fano, Transmission of Information, MIT Press-Wiley, 1961 (Russ. transl., M., Mir, 1965).
[32] S. Goldman, Information Theory, Prentice-Hall, 1953 (Russ. Transl., M., IL, 1957).
[33] I. A. Poletaev, in: Issledovaniya po kibernetike (Studies in Cybernetics), M., Sov. radio, 1970, p. 228.
[34] R. C. Raymond, Amer. Scientist **38**, 273 (1950).
[35] D. A. Bell, Amer. Scientist **40**, 682 (1952).
[36] M. V. Vol'kenshtein, Molekuly i zhizn' (Molecules and Life), M., Nauka, 1965 (Transl., Plenum Press, 1970, 1973).
[37] G. I. Kavalerov and S. M. Mandel'shtam, Vvedenie v informatsionnuyu teoriyu izmereniĭ (Introduction to the Informational Theory of Measurements), M., Energiya, 1974.
[38] L. I. Mandel'shtam, Poln. sobr. trudov (Complete Collected Works), vol. V, M.-L., Izd-vo Akad. Nauk SSSR, 1950.
[39] D. Bohm, Quantum Theory, Prentice-Hall, 1951, 1959 (Russ. transl., M., Fizmatgiz, 1961).
[40] D. I. Blokhintsev, Printsipial'nye voprosy kvantovoĭ mekhaniki (Fundamental Problems of Quantum Mechanics), M., Nauka, 1966.
[41] D. I. Blokhintsev, Usp. Fiz. Nauk **95**, 75 (1968) [Sov. Phys.-Uspekhi **11**, 320 (1969)].
[42] A. S. Kholevo, Problemy peredachi informatsii **9** (2), 31 (1973).
[43] A. S. Kholevo, Problemy peredachi informatsii **9** (3), 3 (1973).
[44] G. Uhlenbeck, Fundamental Problems of Statistical Mechanics: Lectures delivered 12–18 May 1968 (Russ. transl., Usp. Fiz. Nauk **103**, 275 (1971).
[45] V. A. Kotel'nikov, Problemy pomekhoustoĭchivosti radiosvyazi (Radiotekhnicheskiĭ sbornik) (Problems of Freedom from Interference in Radiocommunication (Radio-Engineering Collection)), M., Gosénergoizdat, 1947.
[46] V. A. Kotel'nikov, Teoriya potentsial'noĭ pomekhoustoĭchivosti (Theory of Potential Freedom from Interference), M., Gosénergoizdat, 1956 (Transl., Theory of Optimum Noise Immunity, McGraw-Hill, 1960).
[47] P. M. Woodward and I. L. Davies, (Russ. transl. in: Teoriya peredachi elektricheskikh signalov pri nalichii pomekh (Theory of Transmission of Electric Signals in the Presence of Interference), M., IL, 1953, p. 239).
[48] W. W. Peterson, T. G. Birdsall, and W. C. Fox, Trans. IRE Prof. Group on Inform. Theory, unnumbered volume, no. PGIT-4 (Sept. 1954), 171 (Russ. transl. in: Teoriya informatsii i ee prilozheniya (Information Theory and Its Applications), M., Fizmatgiz, 1959, p. 210).
[49] D. Van Meter and D. Middleton, Russ. transl. in: Priem signalov pri nalichii shuma (Signal Reception in the Presence of Noise), M., IL, 1960, p. 57).
[50] A. A. Kharkevich, Bor'ba s pomekhami (Interference Prevention), M., Nauka, 1965; S. E. Fal'kevich, Otsenka parametrov signala (Estimation of Signal Parameters), M., Sov. radio, 1970.
[51] V. B. Braginskiĭ, Fizicheskie eksperimenty s probnymi telami (Physical Experiments with Test Bodies), M., Nauka, 1970.
[52] V. B. Braginskiĭ and Yu. I. Vorontsov, Usp. Fiz. Nauk **114**, 41 (1974) [Sov. Phys.-Uspekhi **17** (5), (1975)].
[53] R. L. Mössbauer, Usp. Fiz. Nauk **72**, 658 (1960) [Sov. Phys.-Uspekhi **4**, 866 (1961)].
[54] G. L. Trigg, Crucial Experiments in Modern Physics, Van Nostrand Reinhold, 1971 (Russ. transl., M., Mir, 1974).
[55] G. C. Wick, Nuovo Cim. Suppl. **4**, 309 (1966) (Russ. transl., Usp. Fiz. Nauk **101**, 303 (1970)).
[56] B. N. Taylor, D. N. Langenberg, and W. H. Parker, Sci. Am. **223** (4), 62 (1970) (Russ. transl., Usp. Fiz. Nauk **105**, 575 (1971)).
[57] D. H. Menzel (ed.), Fundamental Formulas of Physics, Constable, 1955; (2 vols.) Dover, 1960 (Russ. transl. (2 vols.), M., IL, 1957).
[58] S. V. Yablonskiĭ, in: Informatsionnye materialy (Informational Materials), no. 5 (42), M., Izd. Nauchnogo soveta AN SSSR po kompleksnoĭ probleme "Kibernetika", 1970, p. 5.
[59] O. B. Lutsanov, ref. 58, p. 16.
[60] B. A. Trakhtenbrot, ref. 58, p. 61.
[61] A. K. Zvonkin and L. A. Levin, Usp. Matemat. Nauk **25** (6), 85 (1970) [Russ. Math. Surveys **25** (6), 83 (1970)].
[62] A. G. Vitushkin, Otsenka slozhnosti zadachi tabulirovaniya (Estimation of the Complexity of a Tabulation Problem), M., Fizmatgiz, 1959 (Transl., Theory of Transmission and Processing of Information, Pergamon, 1961).
[63] B. A. Trakhtenbrot, Slozhnost' algoritmov i vychisleniĭ (Complexity of Algorithms and Calculations), Novosibirsk, Izd-vo NGU, 1967.
[64] B. A. Trakhtenbrot, Algoritmy i vychislitel'nye avtomaty (Algorithms and Automatic Computing Machines), M., Sov. radio, 1974 (Transl. of 1960 ed., Heath, 1963).
[65] D. B. Yudin and A. P. Goryashko, Izv. Akad. Nauk SSSR Tekh. Kibernetika No. **3**, 34 (1974) [Engineering Cybernetics No. 3, (1974)].
[66] A. A. Markov, Dokl. Akad. Nauk SSSR **157**, 262 (1964) [Sov. Math. **5**, 922 (1964)].
[67] A. N. Kolmogorov, Problemy peredachi informatsii **1** (1), 3 (1965).
[68] A. N. Kolmogorov, Problemy peredachi informatsii **5** (3), 3 (1969).
[69] A. A. Markov, Tr. Matemat. Inst. im. V. A. Steklova (Akad. Nauk SSSR), No. 42 (1954) [Proc. Steklov Inst. Math., No. 42 (1954)].
[70] V. Ya. Arsenin, Tr. Leningr. otdeleniya MIAN SSSR **83**, 33 (1973) [Seminars in Mathematics].
[71] A. A. Kharkevich, Problemy Kibernetiki, no. 4, 1960, p. 53 [Problems of Cybernetics IV, p. 1193 (1962)].
[72] M. M. Bongard, Problemy Kibernetiki, no. 9, 1963, p. 71 [Problems of Cybernetics, IX].
[73] R. L. Stratonovich, Izv. Akad. Nauk SSSR Tekh. Kibernetika No. 5, 3 (1965) [Engineering Cybernetics No. 5, 1 (1965)].
[74] P. K. Rashevskiĭ, Usp. Mat. Nauk **28** (4), 243 (1973) [Russ. Math. Surveys **28** (4), (1973)].
[75] R. P. Poplavskiĭ, in: Vychislitel'naya matematika (Computational Mathematics), no. 6, M., Izd-vo Akad. Nauk SSSR, 1960, p. 63.
[76] R. P. Poplavskiĭ, Prikl. Mat. i Mekh. **24**, 387 (1960) [J. Appl. Math. and Mech. **24**, 562 (1960)].
[77] R. P. Poplavskiĭ, in: Issledovaniya po sovremennym problemam konstruktivnoĭ teorii funktsiĭ (Investigations in Modern Problems of Constructive Theory of Functions), M., Fizmatgiz, 1961, p. 358.
[78] A. Schlüter, Phys. Blätter **26** (8), 343 (1970) (Russ. transl., Usp. Fiz. Nauk **106**, 119 (1972)].
[79] D. B. Yudin, Avtomatika i telemakhanika **33** (5), 47 (1972) [Automation and Remote Control **33**, 747 (1972)].
[80] D. B. Yudin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **15**,

957 (1972) [Radiophys. and Quantum Electronics 15, 727 (1972)].
[81] W. Ashby, Russ. transl. in: Obshchaya teoriya sistem (General Theory of Systems), M., Mir, 1966, p. 171).
[82] H. J. Bremermann, in Proc. Conf. on Self-Organizing Systems, Washington, Spartan, 1962, p. 93.
[83] R. Landauer, Phys. Today 23 (7), 22 (1970) (Russ. transl., Usp. Fiz. Nauk 106, 125 (1972)).
[84] J. von Neumann, Russ. transl. in: Avtomaty (Automata), M., IL, 1956, p. 68.
[85] J. von Neumann, Russ. transl. in transl. of book by Turing, "Can a Machine Think?", M., Fizmatgiz, 1960, p. 59).
[86] J. von Neumann, Russ. transl. in Kiberneticheskiĭ sbornik (Cybernetic Collection), no. 1, M., IL, 1960, p. 11).
[87] J. R. Dailey and H. C. Kuntzleman, Russ. transl. in Zarub. radioélektron (Foreign Radioelectronics), no. 6, 39 (1971)).
[88] N. G. Vasov, V. V. Nikitin, and A. S. Semenov, Usp. Fiz. Nauk 97, 561 (1969) [Sov. Phys.-Uspekhi 12, 219 (1969)].

Translated by W. F. Brown, Jr.