

Maxwell demon and the correspondence between information and entropy

R. P. Poplavskii

Usp. Fiz. Nauk 128, 165-176 (May 1979)

Thermodynamic models are discussed for various methods of introducing order in a system (for carrying out a control process). The concept of entropic efficiency is introduced as a measure of the degree of irreversibility of various processes (the process approaches reversibility as this efficiency approaches unity from below). Although the efficiency of the Maxwell demon is much less than unity, it is comparable to the efficiency of the Carnot cycle. The efficiency of a model for a modern version of the Maxwell demon is estimated. This modern version is the information approach for producing a population inversion in a two-level maser. A thermodynamic model for the general case of control processes confirms the conclusion that the information acquisition and control processes are intrinsically irreversible.

PACS numbers: 05.70.Ln

CONTENTS

1. Introduction	371
2. Maxwell demon and its exorcism	372
3. Entropic efficiency of the Maxwell demon	373
4. The ammonia-beam maser: A twentieth-century Maxwell demon	375
5. The general case: Thermodynamic model for a transducer, the simplest unit of a control system	376
6. Information and entropy	378
References	379

1. INTRODUCTION

Recent developments in the physics of black holes have revealed profound analogies with thermodynamics. If the surface area of a black hole is assumed to be proportional to entropy, while the surface gravitation is taken to be proportional to temperature (these assumptions are confirmed in gedanken experiments on the interaction of a black hole with an incident object), then it is possible to formulate physical laws for black holes which are equivalent to the corresponding laws of thermodynamics.¹

In this paper we will not take up the many interesting questions raised by this analogy (in particular, the reasons why the third law does not hold in all the formulations used in thermodynamics,² and the analysis of a new mechanism for obtaining statistical laws from dynamics). Instead, we will be concerned only with that conclusion which can be drawn from this analogy which is important for the present topic: that, as in thermodynamics, entropy is related to the absence of detailed information on the internal structure of a system (a black hole).

It is this fact which has revived the interest of many physicists in the correspondence between information and entropy in classical statistical thermodynamics, and which has served as an impetus for me to return to a discussion of this question, which is a development of Ref. 3.

A natural starting point is an analysis of the activity of the Maxwell demon, which is the simplest thermo-

dynamic model of a control process, and apparently the first to appear in the literature.

After describing the Maxwell model itself we will briefly discuss the basic work which has been carried out to "exorcise" this demon. Then we examine the model in more detail, finding quantitative relationships between information and entropy, in accordance with the purpose of this paper. The useful effect of a control process is described by a measure of the order in the system—the decrease in its statistical entropy. This characteristic is a natural one for a thermodynamic treatment. Everywhere below we will use the term "negentropy" (negative entropy); the useful ordering effect in the system is called the "negentropic effect."¹

In Section 4 we will discuss a particular modern version of the Maxwell demon: the ordering of molecules in a two-level maser.

Finally, we will analyze a thermodynamic model for the control process in a general case, leaning heavily

¹To avoid confusion, we note that the same effect is called the "entropy defect" in Ref. 3. This has turned out not to be a good term, however, because it was used previously in Ref. 15 with a different meaning. It is the meaning of Ref. 15 which is used in Ref. 16. Incidentally, we note that in using the concept of "negentropy," introduced by Brillouin,⁸ we are nowhere using the "negentropic information principle," also formulated by Brillouin. Both the useful effect and the expenditure associated with the entropy increase are calculated directly for all the models.

on the results of Ref. 3.

For all the models considered for the control process we find that the negentropic effect is much weaker than the entropy increase caused by energy dissipation during the acquisition of information and the control step proper.

We thus confirm the conclusion reached in Ref. 3: that information processes are intrinsically irreversible.

2. MAXWELL DEMON AND ITS EXORCISM

In the penultimate section of his book *Theory of Heat* (in the chapter on the molecular theory of the structure of matter), Maxwell examined limitations of the second law of thermodynamics. It was in this section that he outlined a model which was to provoke a stream of papers which were still appearing a century later. After formulating the second law as the impossibility of creating a temperature or pressure difference without performing work, Maxwell gave a description of a model, which we will reproduce in its entirety here.⁴

"But if we conceive a being whose facilities are so sharpened that he can follow every molecule in its course, such a being, whose attributes are still as essentially finite as our own, would be able to do what is at present impossible to us. For we have seen that the molecules in a vessel full of air at uniform temperature are moving with velocities by no means uniform though the mean velocity of any great number of them, arbitrarily selected, is almost exactly uniform. Now let us suppose that such a vessel is divided into two portions *A* and *B*, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from *A* to *B*, and only the slower ones to pass from *B* to *A*. He will thus, without expenditure of work, raise the temperature of *B* and lower that of *A*, in contradiction of the second law of thermodynamics."

Maxwell went on to point out that this was only one of the examples which show that the conclusions we draw from experiments on objects consisting of huge numbers of molecules may prove inapplicable to more subtle observations and experiments, carried out under conditions such that it would be possible to distinguish and control the individual molecules with which we usually deal only in large numbers.

We wish to draw attention to two aspects of this model. On the one hand, we emphasize that fluctuations (in the velocities) are being used, and the objects of the ordering are the individual molecules; a macroscopic effect is built up as the result of a large number of microscopic control operations.

On the other hand, and we think this is a more important point, we have seen here a description of a process in which the measured values of one parameter (the velocity) are used to control another physical quantity (the position of the trap door which opens and closes the aperture), so that there is an ordering in the sys-

tem—a negentropic effect. As mentioned above, in this sense this model is the simplest particular example of a model for a control process.

Initially, only the first of these circumstances was pointed out in the literature. Various modifications of the perpetual motion machine of the second kind described by Maxwell were proposed, and these schemes made important use of fluctuation effects. [It was this time that Lord Kelvin (William Thomson) introduced the term "Maxwell demon"; Maxwell himself had talked in terms of a "being".]

A bibliography of this work was given by Smoluchowski,⁵ as part of a detailed and clear analysis of fluctuation phenomena carried out in an effort to prove the validity of kinetic theory. Here it was pointed out for the first time that the Brownian motion of the trap door itself would greatly hinder the demon's work and the operation of similar automatic devices. In particular, a ratchet drive mechanism with a toothed wheel and a pawl was discussed,⁶ and it was concluded from plausible arguments that it would be impossible to devise a machine which would operate for a long time through the use and ordering of fluctuations, specifically because of the Brownian motion of the control element itself. Smoluchowski stated⁵ that a rigorous proof could be carried out only by statistical mechanics (and this has been done, in particular, in Ref. 6).

This work by Smoluchowski⁵ is of considerable interest because it was apparently the first to point out so clearly the limiting role of fluctuations (the thermal motion of the molecules or noise) for both the measurements and the control step proper. The implication here is that energy must be expended on both steps, which are necessarily part of the control system.²⁾

In most of the work on this model, the effort has been concentrated on the first step of the control process: the acquisition of information. Szilard⁷ was the first to point out the relationship between entropy and information (and essentially used a quantitative measure of it, which corresponds to that introduced by Shannon). Analyzing in detail a simplified model which was an extension of that described by Maxwell, Szilard showed that for any ordering of the molecules information should be obtained on the coordinates (or velocities) of the molecules. In other words, measurements should be made. This acquisition of information involves an increase in entropy in the system, and this increase is at least no smaller than the entropy decrease resulting from the ordering of the molecules.

All the subsequent work (by Demers, Jacobson, Gabor, Brillouin, *et al.*), which is discussed in detail by Brillouin,⁸ has been devoted to estimating the energy expended on the measurement step in various modifications of the Maxwell demon model (including one

²⁾In a later book (1914), cited by Szilard (Ref. 7, p. 841), Smoluchowski mentions the need for an energy expenditure during control even more definitely, even mentioning the unavoidable energy dissipation in the operation of the sensor and motor regions in the cerebral cortex.

with a "pressure" demon rather than a "temperature" demon) and to comparing this energy expenditure with the useful effect.

If we define the entropic efficiency η of the process as the ratio of the negentropic effect $\Delta N = -\Delta S^{(-)}$ (the lowering of the entropy of the system as a measure of its order) to the entropic expenditure $\Delta S^{(+)}$ (the entropy increase due to the energy dissipation in the acquisition of information and the control), it follows from all this work that

$$\eta = \frac{\Delta N}{\Delta S^{(+)}} \leq 1. \quad (1)$$

The energy expended on the control step proper is, as a rule, ignored. This result means precisely that the demon has been exorcised,³ since it shows that the second law holds.³⁾

3. ENTROPIC EFFICIENCY OF THE MAXWELL DEMON

The estimate of the efficiency η in (1) can be improved by finding more accurate estimates of the energy expended and by taking into account the energy expended in all the steps of the process, i.e., in the direct control of the trap door as well as in the measurements.

Everywhere below we will express the temperature T in energy units, while the entropy S will be in relative units⁴⁾ (as, for example, in Ref. 9). Let us first calculate the negentropic effect $\Delta N = -\Delta S^{(-)}$.

We denote by T the original gas temperature in the vessel, while n is the total number of molecules (there are $n/2$ molecules in each half, A and B). We assume that as the demon is working $\Delta n/2$ "hot" molecules pass from A to B , while $\Delta n/2$ "cold" molecules pass from B to A . Then temperatures T_B and T_A are established in the respective parts of the vessel, and these temperatures satisfy

$$\frac{T_B - T_A}{T} = \frac{\Delta T}{T} = \Theta. \quad (2)$$

To relate Θ and $\Delta n/n$, we assume that ϵ_{hot} and ϵ_{cold} , the average energies of the "hot" and "cold" molecules, differ from the initial average energy $\epsilon_0 = c_V T$ (c_V is the specific heat at constant volume per molecule) by $+\Delta \epsilon_1$ and $-\Delta \epsilon_2$, respectively.

After one molecule with energy $\epsilon_0 + \Delta \epsilon_1$ has moved from A to B , while a different molecule, with an energy $\epsilon_0 - \Delta \epsilon_2$, has moved from B to A (the first step of the process), a temperature is established in B which corresponds to the average energy $\epsilon_1^B = \epsilon_0 + (2/n)(\Delta \epsilon_1 + \Delta \epsilon_2)$, while the temperature established in A corresponds to $(\epsilon_1^A = \epsilon_0 - (2/n)(\Delta \epsilon_1 + \Delta \epsilon_2))$.

Now considering the successive steps of the process, we obtain in the $(\Delta n/2)$ -th step,

$$\Theta = \frac{\Delta T}{T} = 2 \frac{\Delta \epsilon_1 + \Delta \epsilon_2}{\epsilon_0} \frac{\Delta n}{n}, \quad \frac{\Delta n}{n} = \kappa, \quad \beta = \frac{\Delta \epsilon_1 + \Delta \epsilon_2}{\epsilon_0}. \quad (3)$$

³⁾We have already mentioned³ the unconvincing attempts in Ref. 10 to prove that this exorcism is only apparent.

⁴⁾In transforming to the usual temperature units (degrees), we should divide T by the Boltzmann constant k and multiply S by the same factor.

The efficiency of this process increases with increasing β , as defined in (3), but it is clear from the relationship between the average energy and its dispersion that we have $\beta \approx 1$.

It is obvious from the very meaning of the model that we have $\kappa \leq \alpha$, where α is the fraction of molecules which meets the energy requirements. For maximum efficiency we should choose $\alpha = \alpha_{\text{max}} = 1/2$. In this case the "hot" molecules should be chosen as all those whose velocities exceed the rms value $\sqrt{\bar{v}^2}$ corresponding to the initial temperature T (with an analogous choice of "cold" molecules). Then using (3) we can write

$$\kappa = \frac{\Theta}{2\beta} < \kappa_{\text{max}} \leq \frac{1}{2}, \quad \beta \approx 1, \quad \Theta < 1. \quad (4)$$

Strictly speaking, α decreases in each step of the process in accordance with $(\alpha(i+1) = \alpha(i)[1 - (4/n)])$ [see the derivation of Eq. (3)] because of the temperature change, i.e., the shift of the average energy. Furthermore, there are many other factors which would lead to a stronger inequality.

$$\kappa \ll 1, \quad \Theta \ll 1. \quad (4')$$

However, since these other factors are of a technical, rather than fundamental, nature (as will be seen from the description of the scheme for selecting molecules), we will not use condition (4') below.

For an ideal gas with n molecules, the entropy is⁹ $S = n(c_V \ln T + \text{const})$.

Adopting the notation $S_{1n} = 2S(n/2, T)$ and $S_{f1n} = S(n/2, T_B) + S(n/2, T_A)$, we can write $S_{1n} - S_{f1n} = -\Delta S^{(-)} = \Delta N = -(n/2)c_V \ln[1 + (\Delta T/2T)][1 - (\Delta T/2T)]$.

Using (3) and (4), we find the final equation for the negentropic effect:

$$\Delta N \approx \frac{n}{2} c_V \frac{\Theta^2}{4} = \frac{n}{2} c_V \beta^2 \kappa^2. \quad (5)$$

To determine the efficiency η in (1), we will estimate the entropy increase $\Delta S^{(+)} = \Delta S_1^{(+)} + \Delta S_2^{(+)}$ due to the energy dissipation during the acquisition of information ($\Delta S_1^{(+)}$) and during the step in which the trap door is actually moved ($\Delta S_2^{(+)}$).

For this purpose we consider a specific scheme for selecting molecules. We restrict the measurement of the molecule's velocity to an observation of this velocity in a specified volume element. The velocity is determined from the Doppler shift of the frequency ν of the signal used to prove the volume element.

In order to select all velocities $v \geq \sqrt{\bar{v}^2}$ (i.e., with $\alpha \approx 1/2$) and to separate accurately the "hot" molecules from the "cold" ones, we can use two measurement systems. In the first we use several detectors ($m \geq 5$), each of which detects molecules in a comparatively narrow velocity interval δv (or a narrow energy interval $\delta \epsilon < \Delta \epsilon_1, \Delta \epsilon_2$). In the second measurement system we use a wide-band detector with a sharp band edge.

In the first measurement system, the probability for each detector that a "hot" (or "cold") molecule will belong to the appropriate velocity interval is $q = \alpha/m \leq 0.1$. In the second system, this probability is $q = \alpha$

≈ 0.5 .

In both cases, however, the *a priori* probabilities for the detection of the appropriate molecule ("hot or "cold") are $\alpha \approx 0.5$. Then in each measurement it is, generally speaking, necessary to acquire an amount of information I_1 equal to $\ln 2$. Because of the nonzero probability w for false operations of the detector, caused by thermal noise, the amount of information acquired in the observation is $I_1 < \ln 2$, and this amount of information becomes smaller as w approaches 0.5 (at $w = 0.5$, $I_1 = 0$). The decrease in w , however, requires an increase in the energy from the probing signal, E .

Let us carry out a calculation for the first measurement system (that consisting of several narrow-band detectors). Our motivations for choosing this system are that, first, only in this case can we correctly take into account the delay between the observation time and the time in which the trap door is opened and, second, the noise level is higher in a wide-band detector (in the second measurement system).

We choose the characteristics of the gas and the dimensions of the trap door such that $\hat{\tau}$, the average time between collisions of molecules with the trap door, is equal to τ , the duration of the probing signal. This latter quantity is in turn equal to the time constant of the matched detector (and the time required to open the trap door). We note that if $\hat{\tau} < \tau$ then "harmful" molecules will be incident at the time the door is opened and will degrade the efficiency. If $\hat{\tau} > \tau$, on the other hand, the efficiency would not be affected, but the time required for the entire apparatus to work would be increased. Although a false operation in this case would allow the passage of a "harmful" molecule only with a probability $\tau/\hat{\tau}$, the number of such events would increase by a factor of $\hat{\tau}/\tau$.

For efficient operation of the device, the probability w for a false detection in each time interval τ must be low. Let us find this probability.

A matched detector which is tuned to the frequency ν (which corresponds to ε_{hot} or $\varepsilon_{\text{cold}}$) and which has a time constant $\tau = 1/\Delta\nu$, where $\Delta\nu$ is the passband, is a system with one oscillatory degree of freedom. For low frequencies ($R\nu \ll T$) this system is similar to a classical oscillator, for which the probability of exceeding the energy level E_t is known to be $w = \exp(-E_t/T)$. At the entrance to the detector, the signal energy E_s must thus exceed the threshold $E_t = T \ln(1/w)$. We can write an equation for the amount of information, $I_1(w)$:

$$I_1 = \ln 2 - w \ln \frac{1}{w} - (1-w) \ln \frac{1}{1-w}.$$

The values of w and thus E_s are found by maximizing the efficiency

$$\frac{I_1 T}{E_s} \rightarrow \max.$$

We find

$$w = w_{\text{opt}} \approx 0.04, \quad E_s > E_t = T \ln \frac{1}{w}, \quad E_s = CT, \quad C > 5. \quad (6)$$

The energy of the signal reaching the detector is dis-

sipated, and it increases the entropy of the heat reservoir. If the total number of detection events is $\Delta n = \kappa n$, the entropy increase is

$$\Delta S_1^{(*)} = C \kappa n, \quad C > 5. \quad (7)$$

The amount of information obtained is $I_{\Delta n} < \kappa n \ln 2$. It can thus be seen from (7) that the entropy cost of a unit of information during the detection step is higher than T (for the general case of measurements, the extent to which this cost exceeds T is much greater; see Ref. 3 and Section 5 of the present paper).

To determine η in (1) we must also take into account the energy expended on the control step proper: the opening of the trap door. Repeating all the arguments which we made regarding the detection, we can easily show that the energy which should be expended on controlling the trap door (to avoid false operations due to the fluctuations in the trap door itself) should be the same as the energy expended during the detection. In other words, the total entropy increase in the system is

$$\Delta S_1^{(*)} = 2\Delta S_1^{(*)} = 2C \kappa n > 10 \kappa n. \quad (8)$$

In this particular control model, the second step (the control step proper) increases the expenditure by a factor of only two (in comparison with the first step, i.e., the measurement step). In general, on the other hand, this increase can be much greater, so that we cannot ignore the second step in calculating the entropic efficiency of the control process (Section 5).

From (5), (8), (4), and (1) we finally obtain

$$\eta \leq \frac{\nu\beta}{4\Theta} \Theta = a\Theta, \quad a \leq 0.06 \quad (9)$$

for all real values of c_v and β .

We emphasize that the efficiency of direct heating (and cooling) is much higher, even if the process is definitely irreversible. If we use heat reservoirs with temperatures T_B and T_A , respectively, we find

$$\Delta S_1^{(*)} = \frac{1}{2} n c_v \frac{\Delta T}{2} \left(\frac{1}{T_A} - \frac{1}{T_B} \right) \approx \frac{1}{2} n c_v \frac{\Theta^2}{2}. \quad (10)$$

Using this equation along with (5) we find (regardless of Θ)

$$\eta = 0.5. \quad (11)$$

In contrast, the efficiency of the Maxwell demon depends on Θ and does not exceed 0.06 [see (9)] even at the maximum possible values in (4) (these maximum possible values are essentially unattainable). We thus see that the efficiency of the Maxwell demon is in fact low [(9), (4), (4')], but comparable in order of magnitude (if $\Theta \ll 1$) to the thermodynamic efficiency of the Carnot cycle for heat engines:

$$\eta_C = \frac{\Theta}{1 + (\Theta/2)} \approx \Theta. \quad (12)$$

It is worthwhile to compare η with η_C , despite the very different physical meanings of these quantities: the former is a measure of the irreversibility of the control process (the creation of order in the system), while the latter is a measure of the efficiency of the reversible conversion of heat into work. In order to extract work from the system, however, we need both the first and second steps. A comparison of η with η_C shows that if the first step is implemented by an information method its efficiency will be the same (in order

of magnitude) as η_C , the maximum efficiency of the subsequent use of the negentropic effect.⁵⁾

To conclude this analysis of Maxwell's paradox, we note that equations analogous to (5) and (8)–(11) are found in the case of a "pressure demon,"⁸ in which case the role of $\Delta T/T$ is played by $\Delta P/P = \Delta n/n$, and the heat reservoirs are replaced by barostats [for a calculation analogous to (10)].

In summary, the example of the Maxwell demon shows that the entropic efficiency of the control process is low [(9)]. Before we analyze a general thermodynamic model for the control process, we will briefly discuss an example of the use of an information method to separate molecules in a maser, which is pertinent to our topic.

4. THE AMMONIA-BEAM MASER: A TWENTIETH-CENTURY MAXWELL DEMON

Certain methods used to create a population inversion for lasers operating at a high frequency ($h\nu \gg T$) may be called "high-power" methods.¹¹ For example, there can be direct pumping from the lower level to a higher level, which is the third energy level (the auxiliary pump frequency is $\nu_a > \nu$). This method makes use of a difference in the time constants of the thermal relaxation processes.

There is another method for creating a population inversion which is interesting from the thermodynamic standpoint: the direct conversion of thermal energy into coherent emission. Without going into the details of this (thermal) excitation method (see Ref. 12), we note that in this case an equilibrium distribution is initially established for all three levels. This distribution corresponds to the heater temperature. Then, at the temperature of the refrigerator, a population inversion is achieved between the lower and intermediate levels because the time for relaxation between the upper and intermediate levels (at the frequency of the idler transition) is much shorter than the times corresponding to the other two frequencies (the signal and auxiliary frequencies). Konyukhov and Prokhorov¹² have shown that the quantum efficiency here is less than the efficiency of a Carnot cycle: $\eta_q \leq \eta_C$.

For two-level lasers with a low signal frequency ($h\nu \ll T$), the number of excited molecules is comparable to the number of unexcited molecules. Therefore,

⁵⁾The step of the subsequent useful exploitation of the negentropic effect lies outside the control process under discussion here (and the object of the work lies outside the thermodynamic system under consideration here). Analogously, the source of the initial negentropy (deviation from equilibrium), whose presence is mandatory not only in the direct-heating method but also in the Maxwell demon model, lies outside the system under consideration. This source may be, in particular, some device which generates signals, whose energy (per degree of freedom) must exceed T [see the derivation of (7)]. Alternatively, it could be a radiation source with a temperature $T_R > T$ (since the frequency ν must satisfy the condition $h\nu > T$; Ref. 8). Furthermore, an energy source is of course required to actually move the trap door.

in such devices the most efficient way to create a population inversion is to separate the excited and unexcited molecules physically. This method is naturally called an "information" method. This was the operating principle of one of the first masers: the ammonia-beam maser.^{11,21,22}

Here the function of the Maxwell demon—separating the excited molecules from the unexcited molecules—is performed by an inhomogeneous electric field, which deflects the unexcited molecules toward a strong field, while the unexcited molecules are deflected toward a weak field.^{11,21,22}

If a sufficient number of excited molecules enter a high- Q resonator per unit time, self-excitation conditions will be satisfied, and the maser will operate continuously.

To estimate the entropic efficiency and the quantum efficiency¹² of this maser, we will calculate the separation negentropy $\Delta N = -\Delta S^{(-)}$ for two types of noninteracting molecules, differing only in energy. The separation entropy $\Delta S^{(-)}$ obviously differs from the mixing entropy⁶⁾ of the gases ($-\Delta S^{(-)}$) only in its sign.

We first note that the partial densities $n^{(+)} / n$ and $n^{(-)} / n$ of the two molecular species are completely governed by $h\nu$, the difference between the energies of the two levels, and that these partial densities vary monotonically as the parameter $x \equiv h\nu / T$ is varied. Here n is the total number of molecules, and $n^{(+)}$ and $n^{(-)}$ are the numbers of excited and unexcited molecules, respectively.

For an ammonia-beam maser at room temperature, we have

$$x \equiv \frac{h\nu}{T} \ll 1, \quad n^{(+)} = \frac{n}{2} \frac{1-x}{1-(x/2)} \approx \frac{n}{2}, \quad n^{(-)} = \frac{n}{2} \frac{1}{1-(x/2)} \approx \frac{n}{2}. \quad (13)$$

In the opposite limiting case, we have

$$x \gg 1, \quad n^{(+)} = \frac{ne^{-x}}{1+e^{-x}} \approx ne^{-x}, \quad n^{(-)} = \frac{n}{1+e^{-x}} \approx n. \quad (14)$$

Like the partial densities, the separation negentropy ΔN is a monotonic function of the parameter x .

In the present case ΔN can be calculated easily if we

⁶⁾A huge literature has grown up on the mixing paradoxes (Gibbs *et al.*). Here we will consider only some recent papers. Bazarov¹³ points out that there is a discontinuity in the partial density of a gas at the transition from a mixing of this gas with some other gas, with an arbitrarily small difference in properties, to mixing with an identical gas. Gel'fer *et al.*¹⁴ correctly point out that this explanation (or any similar explanation) fails to resolve the paradox, whose essence is the discontinuity (here the discontinuity is not eliminated, but simply shifted from one concept to another) (Gel'fer *et al.*¹⁴ also give an extensive bibliography on the question). In accordance with the general interference principles of quantum mechanics, a continuous parameter of the proximity of the two particles is introduced¹⁴; the scalar product of the Φ functions corresponding to the internal states of the gases being mixed. The mixing entropy turns out to be a continuous function of the degree of nonorthogonality of the internal states of gases, and there is a continuous transition from the complete identity ($\Phi = 1$) to complete dissimilarity ($\Phi = 0$).

use the familiar expression for the entropy in terms of the free energy⁹ and write this free energy in terms of the partition function Z , singling out in this function the one term in which we are interested (per molecule):

$$Z = 1 + e^{-h\nu/T},$$

$$\Delta N = -\Delta S^{(*)} = \ln Z + \frac{T}{Z} \frac{\partial Z}{\partial T}. \quad (15)$$

Then for arbitrary $h\nu/T \equiv x$ we find

$$\Delta N = \ln(1 + e^{-x}) + \frac{xe^{-x}}{1 + e^{-x}}.$$

In the two limiting cases we have

$$\Delta N = \Delta I = \begin{cases} \ln 2 - \frac{3}{8}x^2 \approx \ln 2, & x \ll 1, \\ xe^{-x} + e^{-x} \approx xe^{-x}, & x \gg 1, \end{cases} \quad (16)$$

where ΔI is the average amount of information obtained in the separation of a single molecule.

To calculate the entropy expended here we should estimate Q , the energy dissipated (in the heat reservoir) during the separation. During separation in an inhomogeneous field, each molecule experiences a decrease in potential energy by some amount U . This energy is then dissipated; i.e., $Q = U$. The value of U is easily related to the reliability of the separation. For reliable separation, we must obviously have

$$\frac{U}{T} = \frac{Q}{T} \equiv C > 1, \quad (17)$$

since there is always a probability w for a Brownian transition of an unexcited molecule into the region intended for the excited molecule (and vice versa). Using $w > 0$, we see that the negentropic effect in (16) decreases by an amount

$$\Delta I^* = w \ln \frac{1}{w} + (1-w) \ln \frac{1}{1-w}, \quad (18)$$

which is equal to the loss of information due to the thermal motion.

For the second limiting case ($x \gg 1$) we can easily find an estimate for C in (17), since the condition ($w = e^{-U/T} \leq e^{-x} = n^{(*)}/n$) must hold [see (14) and (16)]. From (16) and (17) we find

$$\Delta S^{(*)} \geq x \gg 1, \quad \eta \leq e^{-x} \ll 1; \quad (19)$$

in other words, the entropic efficiency is extremely low in this case. The case in which we are primarily interested is the first case [$x \ll 1$ in (16)], which corresponds to the ammonia-beam maser for which (13) holds.

Using $U = Q = T \ln(1/w)$ [see (6)], and taking the actual negentropic effect from (16) and (18), $\Delta N^* = \Delta N - \Delta I^*$, we can easily find the range of w corresponding to the maximum efficiency:

$$\eta = \eta_{\max} = \left(\frac{\Delta N^*}{\Delta S^{(*)}} \right)_{\max} = \frac{1}{C^*}, \quad C^* \approx 6, \quad (20)$$

$$w \approx 0.04, \quad \Delta S^{(*)} = C \approx 5.$$

However, in the ammonia-beam maser, the separation negentropy is by no means completely exploited: all that we can count toward the "useful effect" is the energy stored in the excited molecules at the frequency ν . Using (13), (17), and $\Delta S^{(*)}$ from (20), we find the quantum efficiency to be

$$\eta_q = \frac{x}{2C} < x\eta \quad x \ll 1, \quad C \approx 5. \quad (21)$$

In other words, and as usual (see the preceding sec-

tion), the subsequent use of the negentropic effect is related to the low efficiency, $\eta_q \ll \eta$ in (20).

In the other limiting case ($x \gg 1$) we have $\eta_q = \eta$ in (19), but here the entropic efficiency of the separation is itself extremely low. This circumstance may be one reason why this limiting case is not achieved in practice.

5. THE GENERAL CASE: THERMODYNAMIC MODEL FOR A TRANSDUCER, THE SIMPLEST UNIT OF A CONTROL SYSTEM

Any control system necessarily includes a transducer, a very simple unit whose operation can be described as follows: when a certain value of a continuous scalar physical quantity l is established at the input of the transducer, the transducer produces at its output a value of some other physical quantity r which is in a one-to-one correspondence with the value at the input. Let us assume that l and r are internal parameters characterizing the states of two respective systems: the system being studied and the system being controlled. We assume that both systems are in a constant-temperature chamber with a temperature T and that the systems do not interact before the conversion process. In this conversion process the system being studied and the system being controlled are put in an interaction state such that energy is exchanged between these systems, with the result that the particular value of r in one-to-one correspondence with the particular value of l is established with the necessary accuracy. Since the function $r(l)$ is single-valued, it is obvious³ that the manifold of states of the system being studied must be no smaller than the manifold of states of the system being controlled: $\sigma_l \leq \sigma_r$. Here the relative errors are

$$\sigma_l = \frac{V \overline{\Delta l^2}}{\delta l}, \quad \sigma_r = \frac{V \overline{\Delta r^2}}{\delta r}, \quad \sigma_r \leq \sigma_l \ll 1, \quad (22)$$

where $\overline{\Delta r^2}$ and $\overline{\Delta l^2}$ are the mean square thermal fluctuations of the corresponding parameters;

$$\delta r \equiv r_{\max} - r_{\min}, \quad r_{\min} \leq r \leq r_{\max},$$

$$\delta l \equiv l_{\max} - l_{\min}, \quad l_{\min} \leq l \leq l_{\max}; \quad (23)$$

and σ is the given relative error of the transformation. The relative accuracy $1/\sigma_r$ is a measure of the effective number (manifold) of different values of the controlled parameter.⁷⁾

The thermodynamic model for the transformation describes the elementary control event (in the simplest case, this is a single event, if there is only a single transducer in the control system). On the other hand, this model allows us to determine all the basic aspects of the control process (from the thermodynamic standpoint) and to determine the principal characteristics of this process: ΔN , $\Delta S^{(*)}$, and η . For brevity we will thus call this model the "thermodynamic model of the control process."⁸⁾

⁷⁾The controlled parameter r is simultaneously the controlling parameter for some object (or process) lying outside this model [see also footnote 5].

⁸⁾Generally speaking, there are multichannel systems of dis-

If the function $r(l)$ is linear, this model for the control process corresponds precisely to the measurement-process model studied in detail in Refs. 3 and 17. The only distinction is that here the control element corresponds to the detecting element of the measuring instrument, while the controlled parameter r corresponds to the detection element φ .

Following Ref. 3, we assume that the *a priori* distributions of l and r are uniform on the intervals δl and δr in (23). This is a natural assumption for a study of the extreme values of the entropic efficiency η , since this assumption maximizes the negentropic control effect (in comparison with the effect corresponding to other possible distributions over the intervals).

It follows from Refs. 3 and 17 that in order to achieve the specified control accuracy, (22), the interaction energy U (the energy transferred from the system being studied to the system being controlled or vice versa) must increase in proportion to $1/\sigma^2$. The energy dissipated in the control process is $Q = U$ if the control is carried out promptly (within a limited time), or

$$Q \geq Q_{\min} = 2 \sqrt{UT} = \frac{2T}{\sigma}, \quad \frac{1}{\sigma^2} \geq \Delta S^{(+)} \geq \Delta S_{\min}^{(+)} = \frac{2}{\sigma} \quad (24)$$

where Q_{\min} and $\Delta S_{\min}^{(+)}$ correspond to the optimum retardation of the switching process (by a factor $1/\sigma$).

If the function $r(l)$ (r is the parameter being controlled, and l is the one being measured) is monotonic but not linear, then this functional transformation is on the average energetically equivalent to a linear transformation. If, on the other hand, $r(l)$ oscillates over the entire range of values, then the maximum energy expenditure increases in proportion to m , the number of regions within which the function is monotonic,³ since $\sigma_l = \sigma_r/m$.

The negentropic effect is estimated in Ref. 3; for this particular model, this effect is a measure of the result of the interaction between the two systems (in this case, the system being studied and that being controlled):

$$-\Delta S^{(-)} = \Delta N = \ln \frac{1}{\sigma} \approx \Delta I. \quad (25)$$

The entropic efficiency of the control process is thus

$$\eta \leq \eta_{\max} = \frac{\Delta N}{\Delta S_{\min}^{(+)}} = \frac{\sigma}{2} \ln \frac{1}{\sigma} \ll 1, \quad \sigma \ll 1. \quad (26)$$

We note that, while the control process necessarily involves a negentropic effect, a negentropic effect appears in the measurement process only if the result of the measurement is put in the form of a scalar physical quantity φ (for example, the angular deflection of a spring-loaded needle). Then this quantity can be used for the control, either directly (if the control element is on a common axis with the needle, $\varphi = r$) or after a

crete (relay) control, rather than continuous control, in which the control accuracy is governed by the number of channels, and the transducer which we are talking about here is not present. The thermodynamic model for such a system is less natural and more complicated to analyze. Without going into the details of the discrete-control model, we can point out that the entropic efficiency of such systems is no better than the extreme value of η achieved in continuous systems.

transformation $r(\varphi)$ (if there is a potentiometric pickup at this axis). The efficiency is of the same order of magnitude in these two cases [but a coefficient of $1/2$ appears in (26) during the transformation]. It is this model of the measurements (which are naturally called "active") which is the basis for the analysis of the measurement process in Ref. 3. It is pointed out in Ref. 3 that frequently only the first step of the process described in Ref. 3 is said to be a measurement (e.g., a detection), when the result of the measurement must be transformed by some device into a scalar physical quantity in order to be used for control purposes.

If this transformation [with $\Delta S^{(+)} \sim 1/\sigma$; see (24)] is assigned to the control step proper, then the energy cost of the measurement in the detection step (a "passive" measurement) is much lower. It can be shown for the general case¹⁹ that if $\sigma \ll 1$ then

$$\sigma = w, \quad p = w\sigma = \sigma^2, \quad E_t = 2T \ln \frac{1}{\sigma}, \quad (27)$$

$$\Delta S^{(+)} = C_I \ln \frac{1}{\sigma} = C_I I, \quad C_I > 2$$

(p is the probability for a false detection in one element). In this case, however, there is no negentropic effect in (25), so that it is meaningless to talk of the entropic efficiency of a "passive" measurement.

To conclude this section, we will briefly discuss some generalizations of this model for the control process in two directions: in the case in which the parameter being controlled depends on several physical quantities [$r(l_1, l_2, \dots)$] and in the case in which the object being controlled is separated from the system being measured in space or time. In both these cases, the system being controlled interacts with some image (model) of the system being measured, rather than directly with this system. Then it is necessary to expand the control model, incorporating in it the following steps (processes): information processing, information transfer, and information storage. These processes of course require a separate detailed analysis.^{3, 18, 19} Here we will simply discuss briefly the physical meaning (from the energy standpoint) of the information-representation methods which are used in the transfer, storage, and processing of the information.

We first note that these processes are not directly related to the negentropic effect—only to the transfer of information from one point to another, a duplication of this information, and its conversion. In these steps we can thus save a lot of energy by putting the scalar case aside and adopting a positional representation of numbers (only in the last step, the control step proper, do we have to transform back to the scalar representation of the controlled parameter r). In the case of a one-to-one transformation of a scalar quantity into a vector quantity, each component of the vector need be specified much less precisely; in the extreme case it may be sufficient to distinguish between two states: the presence or absence of a signal. Although an increase in the number of components (in the dimensionality of the vector space) is accompanied by a more stringent requirement on the reliability $1/w$ (i.e., on the probability w for a switch of the signal in any of the components into another digitization interval), the en-

ergy cost of an accurate representation of the entire number decreases.

We emphasize that all the known methods for reducing the energy cost of an accurate representation of numbers necessarily involve an increase in the time.

This increase in the time, however, reduces the energy cost only down to certain limits. Although these limits are different for different information processes, they are always finite: the energy cost of the information cannot be made arbitrarily small, even if an arbitrarily long time (or an arbitrarily wide frequency band) is available.

In the measurements, the conversion from the behavior of $\Delta S^{(+)} \sim 1/\sigma^2$ to the behavior $\Delta S^{(+)} \sim 1/\sigma$ is associated with the optimum retardation of the switching process³ (by a factor of $1/\sigma$).

By dropping the scalar representation and adopting the vector representation, we obtain the following behavior: $\Delta S \sim \ln^2(1/\sigma)$ for a positional coding method, in which case the time Δt (or the frequency band) is increased by a factor $\sim \ln(1/\sigma)$, or $\Delta S^{(+)} \sim \ln(1/\sigma)$, in the case of single-position coding,⁹⁾ with $\Delta t \sim 1/\sigma$ (Ref. 19).

The Shannon redundant coding method, which uses a positional representation with auxiliary (check) positions, is asymptotically the best method (in terms of the product $\Delta S^{(+)} \Delta t$). In this case a further increase in the dimensionality of the vector space makes it possible to detect and correct errors of a certain multiplicity, so that the requirements on the probability p for a distortion (and the requirements on the signal-to-noise ratio) in a single position are relaxed. In this case we have $\Delta S^{(+)} \sim \ln(1/\sigma)$, and the corresponding increase in the time is by a factor of only $\ln(1/\sigma)$. Even with the best coding methods, however, the energy cost per unit information²⁸ is greater than T , i.e., $\Delta S^{(+)} > I$, and the entropic efficiency is $\eta \ll 1$ [see (26)] (since a negentropic effect is given only by those information processes which involve control of an "active" measurement, i.e., which involve a scalar representation of numbers).

6. INFORMATION AND ENTROPY

A distinction should be made between two aspects of the correspondence between information and entropy.

One aspect is related to the statistical interpretation of entropy (which goes back to Boltzmann and Gibbs), as a characteristic of the incompleteness (indefiniteness) of the information on the internal structure of the system when its macroscopic parameters are specified (see, for example, Ref. 2). Here it is pertinent to go back to the analogy with black holes,¹ brought up in the Introduction. Bekenstein²⁰ has formulated and studied on certain examples a generalized second law of thermodynamics for the physics of black holes: the sum of the entropy of the black holes and of the entropy of the matter outside them can never decrease. It is

⁹⁾In this case (as in the case of detection), the number is represented by a pulse in one of the $1/\sigma$ possible positions.

important to note that the increase in the entropy of the black hole as it captures some external object is directly characterized by a loss of information about this object. Working in this manner, and using the example of the capture of an elementary particle without any internal structure (so that the information loss is one bit), Bekenstein²⁰ defines a dimensionless factor which relates the entropy of the black hole to its surface area. Then, as in thermodynamics, the decrease in the information ($-\Delta I$) about the internal structure naturally corresponds, by definition, to an increase in the entropy ($+\Delta S$) of the system. The fundamental distinction is that in thermodynamics there is the possibility of again compressing or cooling the gas, i.e., of reducing its entropy ($-\Delta S^{(-)} = \Delta N$), while the entropy of an object in a black hole is at the maximum for the given mass and cannot be reduced: information on the internal structure of the black hole is strictly inaccessible to an external observer. We note that the negentropic effect in a gas which we mentioned,

$$-\Delta S^{(-)} = \Delta N = \Delta I, \quad (28)$$

is achieved by a "noninformation" method: upon compression (or cooling), there is a decrease in the uncertainty of the coordinates (or velocities) of all the molecules simultaneously.

Then by comparing two analogous thermodynamic systems with different entropies ($S_1 - S_2 = \Delta S^{(-)}$), we are justified in asserting, in accordance with (28), that this difference is governed precisely by the different amounts of information about the internal structures of these systems. Furthermore, if the gas molecules are acted upon collectively it is possible to approach a reversible process arbitrarily closely, i.e., to achieve $\eta = 1$. This statement means that in these cases (and only in these cases) equal signs apply in Brillouin's negentropic information principle,⁸ which in our notation is:

$$\Delta S^{(+)} \geq \Delta I \geq \Delta N. \quad (29)$$

A completely different aspect of the correspondence between information and entropy is the acquisition of information on the macroscopic states of a thermodynamic system and generally all ordering (or control) processes which are brought about by information methods. It has been shown for the models discussed above that all the information processes are inherently irreversible. Making the assumption that, taken together, these models give a fairly complete picture of the ordering (or control) methods which are possible, we will point out their common and distinctive features.

From (29) we see two basic factors which lead to $\eta < 1$. In the first place, regardless of the method used to acquire the information, there is a strict inequality $\Delta S^{(+)} > \Delta I$ on the left side of (29). In other words, the energy cost of a unit of information is definitely higher than T .

Second, if the transition $\Delta I \rightarrow \Delta N$ is taken into account accurately, then the negentropic effect is smaller than the information effect in the case of ordering also; in other words, a strict inequality also holds on the right side of (29).

A distinctive feature of the control model with a Maxwell demon is that the measurement step can be limited to a detection—without the subsequent conversion of the result of the measurement into a scalar quantity. This circumstance arises when the quantity to be measured is a time—not a time interval but a particular instant; only in this particular case can the result of the observation be used directly for control. The energy cost per microscopic control step is thus governed by (6), rather than by (24).

In the subsequent steps, however, the inefficiency of the control of individual molecules becomes apparent: even the temperature effect is smaller than the information effect (for a single step, $\delta T/T = 4/n$, and $I \approx \ln 2$), and the negentropic effect is much smaller yet [see (4) and (5)]. The reason for the second order of smallness, $\Delta N \sim \Theta^2$, is that the effect is of a difference nature¹⁰⁾ (T and S decrease in only one half of the vessel; they increase in the other half). In this case, however, the costs are summed [see (7) and (8)]. The ultimate result is a value $\eta \ll 1$ in (9).

The heat-reservoir method for cooling (and heating) is much more efficient, as mentioned above, because of the collective (noninformational) nature of the effect on all the molecules. Even in an irreversible process of this type, we would have $\eta \approx 0.5$ [see, in particular, (11)].

At this point it is appropriate to move on to the next model and to note a distinction feature of the collective effect during the separation of two molecular species which is different for the two species (i.e., an informational effect).

In these cases the control (separation) occurs simultaneously with measurement (of the detection type), and the negentropic effect is equal to the informational effect. In this case both the expenditure [see (6) and (27)] and the loss in the negentropic effect [see (18) and (16)] are completely governed by the separation reliability $1/w$. The requirements on this reliability depend on the *a priori* probabilities $\pi_1, \pi_2 = 1 - \pi_1$ for observing (for the presence of) two molecular species. The best separation conditions correspond to the ammonia-beam maser:

$$x \ll 1, \pi_1 \approx \pi_2 = 1/2$$

[see (13)]. Maximization of η leads to comparatively high values [$\sim 0.1-0.2$; see (20)]. In general, on the other hand, we have the requirement

$$w < \min\{\pi_1, \pi_2\},$$

and in the case $x \gg 1$ [see (14)] this requirement leads to an extremely low efficiency in (19).

Finally, for the general case of control on the basis of a macroscopic parameter we have $\Delta N = (0.5-1)\Delta I$, depending on the expenditure on the transformation in the switch from measurement to control [see (25) and (26) and the following paragraph]. The low efficiency in (26) with a high control accuracy $1/\sigma$ can be attrib-

¹⁰⁾In the Carnot cycle, the useful effect is again of a difference nature, so that we have $\eta_c \approx \Theta$ in (12).

uted to the extremely unfavorable expenditure of energy [which increases exponentially with increasing ΔI ; see (24) and (25)] on the “active” measurement and the control or the representation of the information in terms of one physical quantity (in one degree of freedom of the system). It is for this reason that always whenever a scalar representation of the information (numbers) is not required, various positional (vector) methods are used to represent the numbers in accordance with the preceding section. These positional methods correspond to an energy distribution with respect to many degrees of freedom.

The entropy thus supplements [see (28)] only the information on the microscopic states (on the internal structure) of the system.

However, all the methods for acquiring information on the macroscopic states of the system and on its order are strictly irreversible ($\eta < 1$).

I wish to thank V. L. Ginzburg and the participants of a seminar which he led, particularly D. A. Kirzhnits, for an interest in these questions which stimulated this paper. I would also like to thank I. A. Zaidenman, L. A. Rivlin, and L. I. Rozonoér for useful discussions.

- ¹V. P. Frolov, Usp. Fiz. Nauk. 118, 473 (1976) [Sov. Phys. Usp. 19, 244 (1976)].
- ²R. P. Terletskii, Statisticheskaya fizika (Statistical Physics), Vysshaya shkola, Moscow, 1973.
- ³R. P. Poplavskii, Usp. Fiz. Nauk. 115, 465 (1975) [Sov. Phys. Usp. 18, 222 (1975)].
- ⁴J. C. Maxwell, Theory of Heat, London, 1871, p. 308 (Russ. transl. Kiev, 1888, p. 288).
- ⁵M. von Smoluchowski, Z. Phys. 13, 1069 (1912) (Russ. transl.: A. Einstein, and S. Smoluchowski, Brownian Motion, ONTI, Moscow-Leningrad, 1936, p. 166).
- ⁶R. P. Feynman, R. B. Leighton, and M. Sands, The Feynman Lectures on Physics, Addison-Wesley, Reading, Mass. (Russ. Transl. Issue No. 4, Mir, Moscow, 1965).
- ⁷L. Szilard, Z. Phys. 53, 840 (1929).
- ⁸L. Brillouin, Science and Information Theory, Academic Press, New York, 1962 (Russ. transl. of earlier edition, Fizmatgiz, Moscow, 1960).
- ⁹L. D. Landau and E. M. Lifshitz, Statisticheskaya fizika, Nauka, Moscow, 1976 (Statistical Physics, Addison-Wesley, Reading, Mass.).
- ¹⁰P. Chambadal, Évolution et applications du concept d'entropie, Dunod, Paris, 1963 (Russ. Transl. Nauka, Moscow, 1967).
- ¹¹J. R. Singer, Masers, Wiley, New York, 1959 (Russ. transl. IL, Moscow, 1961); G. J. Troup, Masers, Wiley Press, New York, 1959 (Russ. transl. IL, Moscow, 1961).
- ¹²V. K. Konyukhov and A. M. Prokhorov, Usp. Fiz. Nauk. 119, 541 (1976) [Sov. Phys. Usp. 19, 618 (1976)].
- ¹³I. P. Bazarov, Usp. Fiz. Nauk. 118, 539 (1976) [Sov. Phys. Usp. 18, 271 (1976)].
- ¹⁴Ya. M. Gel'fer, V. L. Lyuboshits, and M. I. Podgoretskii, Paradoks Gibbsa i tozhdestvennost' chastits v kvantovoi mekhanike (The Gibbs Paradox and the Identity of Particles in Quantum Mechanics), Nauka, Moscow, 1975.
- ¹⁵D. S. Lebedev and L. B. Levitin, in: Teoriya peredachi in-

formatsii (Theory of Information Transfer), Nauka, Moscow, 1964, p. 5.

¹⁶V. V. Mityugov, Fizicheskie osnovy teorii informatsii (Basic Physics of Information Theory), Sov. radio, Moscow, 1976.

¹⁷R. P. Poplavskii, Dokl. Akad. Nauk SSSR 202, 562 (1972) [Sov. Phys. Dokl. 17, 52 (1972)].

¹⁸R. P. Poplavskii, Izv. Akad. Nauk SSSR, Ser. Tekh. kibernetika No. 3 (1979).

¹⁹R. P. Poplavskii, Izv. Akad. Nauk SSSR, Ser. Tekh. kibernetika No. 4 (1979).

²⁰J. D. Bekenstein, Phys. Rev. D7, 2333 (1973); D9, 3292 (1974).

²¹R. P. Feynman, R. B. Leighton, and M. Sands, The Feynman Lectures on Physics, Addison-Wesley, Reading, Mass. (Russ. Transl. Issue No. 8, Mir, Moscow, 1966).

²²A. N. Oraevskii, Molekulyarnye generatory (Molecular Generators), Nauka, Moscow, 1964.

Translated by Dave Parsons