# Dynamics and information 

B B Kadomtsev

## Contents

1. Introduction ..... 425
2. Information ..... 426
3. Entropy ..... 427
4. Carnot cycle ..... 428
5. Irreversible processes ..... 429
6. Entropy and information ..... 429
7. Entropy encore! ..... 431
8. Control ..... 432
9. Gas dynamics ..... 432
10. Waves ..... 433
11. Correlation function ..... 435
12. The Schrodinger equation ..... 435
13. Particle in a thermostat ..... 436
14. Wave-particle dualism ..... 437
15. Radioactive decay ..... 439
16. Schrodinger's cat ..... 440
17. Irreversibility of our environment ..... 441
18. Who throws the dice? ..... 442
19. Brownian motion ..... 442
20. Microworld and macroworld ..... 444
21. Behaviour of a microparticle ..... 445
22. Perception ..... 447
Abstract. The dynamical and informational aspects of the behaviour of complex physical systems are considered. In classical physics the information coupling appears in the interaction of nonlinear systems with stochastic behaviour when a small external perturbation may greatly alter the paths of a classical system in phase space. In quantum systems the information coupling to the environment appears in measurement processes when the coherence of the wave function of a quantum object is destroyed and corresponding information appears in the external environment. These processes can be described in terms of the collapse of wave functions. Numerous examples of the collapse are considered, including those leading to the classical behaviour of macroscopic bodies with the information coupling to a nonequilibrium environment. The Einstein-Podolsky-Rosen paradox is discussed in

[^0]23. Fluctuations and irreversibility ..... 450
24. Measurement in quantum theory ..... 454
25. The Einstein-Podolsky-Rosen ..... 459
paradox
26. Bell's inequalities ..... 460
27. Quantum cryptography ..... 461
28. Random wave function ..... 461
29. Mesoworld ..... 464
30. Collapse of wave functions ..... 468
31. Second quantisation ..... 471
32. Molecular chaos ..... 473
33. Convection ..... 476
34. Self-organisation ..... 478
35. Approach to equilibrium and collapse ..... 480
36. Sokolov effect ..... 482
37. Quantum telegraph ..... 486
38. Superluminal communication ..... 486
39. Present, past, future ..... 489
40. Free will ..... 492
41. Hierarchical systems ..... 494
42. Summary ..... 496
References ..... 499
detail together with its possible role in the processes of information transfer over a distance.

## 1. Introduction

It is generally accepted that the main branches of physics are built on the principles of dynamics. The starting point is always the mechanics of a material point and Newton's laws, which introduce the principal dynamical concepts: mass, velocity, momentum, and force. Theoretical mechanics simply casts the elementary laws of mechanics in a more splendid form in terms of differential equations and variational principles. The simplest laws of motion of a material point provide the basis for more complex equations of motion of continuous media: gases, liquids, and elastic solids. Here we meet for the first time the continuous functions of the coordinates and time which play the role of fields, although it is usual to regard fields in vacuum - for example an electro-magnetic field - as the intrinsic fields. The field equations are again the equations of dynamics. Thermodynamics is a phenomenological science only at first sight. In reality it can be constructed on the basis of statistical physics, which is simply a special variety of dynamics. The fact that physics is based on the principles of dynamics is manifested also in the main units of measurements (for example, the centimetre, the gramme, and the second)
which are introduced initially in the mechanics of a material point and then are applied to other more complex branches of physics.

Something quite different from dynamics appears in quantum theory when the square of the wave function is interpreted as the corresponding probability. The probability then comes forward as an essential element of the theory and the discussion is still continuing about the meaning of the wave function and of the evolution of the probabilities of observation of one or another physical quantity described by this function. Following Albert Einstein, one would like to regard the quantum probability as representing an incomplete description of a microscopic object and assume that there might exist a more rigorous theory which could account for the randomness of the observed quantities on the basis of the dynamics of some hidden variables. However, a convincing proof has been provided recently that there is no local realism, no local hidden variables. Consequently, as stressed by Niels Bohr, the quantum probability has a deeper meaning and it imparts to the wave function certain characteristic features which carry information.

A mathematically rigorous definition of the amount of information had appeared together with the birth of cybernetics, which is the science of control and automation of dynamical processes. More specifically, if a physical quantity can a priori assume several values, the observation of one of these values immediately increases information about this physical quantity. If an automaton has a sufficient energy, then after receiving information it can alter the value of the physical quantity in the required direction.

The signal representing the value of a measured quantity can be as small as we please. The control signal can also be very small if it is applied to a sufficiently powerful amplifier.

It follows that the profile of a signal, i.e. its meaningful content, and not its magnitude is important in control. We can say that when automata interact with one another, the important aspect is not so much the exchange of energy as the exchange of symbols, i.e. signals.

However, this applies not only to deliberately built automata, but also to natural physical systems at the margin of their stability. Small signals acting on such systems can have major consequences. We are in fact speaking here of open physical systems through which large energy fluxes may pass. Such systems are far from equilibrium and their dynamical behaviour is complex. They are usually called just that: complex physical systems.

Both aspects, dynamical and informational, may play an equally important role in the description of complex physical systems. We are then faced with the problem of the simultaneous action of forces and information on a system under conditions far from thermodynamic equilibrium. The interplay of nonlinear dynamical processes in such systems very frequently results in self-organisation when both dynamical and informational aspects of the process are matched very accurately and combine to form a single 'organism'.

All these problems are topical in modern physics and they provide its strong foundations. I would like to acquaint the reader with the most interesting ideas and directions of research in this field. The subject will be presented deliberately at a popular level and very frequently the rigour of treatment will be sacrificed to a lively argument. The paper therefore looks more like a set of thoughts based on the physical ideas which themselves have been the subject of serious books, but which would be difficult to combine in a
unified text without deliberate major simplification of the presentation style. In addition to the many familiar facts and their theoretical treatments, this paper puts forward some new and currently unconventional ideas. The readers may regard them as unconvincing and even simply incorrect. However, they are asked not to make hasty judgements: some of the topics are discussed again and even several times from different points of view, and the relevant conclusions become more and more convincing. It may be that the readers will not agree with some of my ideas, but I shall be satisfied if this paper stimulates a desire to consider more deeply the problems discussed here and to arrive at one's own conclusions.

The references at the end of the paper are unfortunately in no sense complete: they list only those papers and books which have come up naturally in the course of this presentation. My excuse is only the circumstance that the presentation itself is not sufficiently fundamental.

## 2. Information

In our age, when we are flooded with information from all sides it would seem there is no need to explain what information is. However, this is not quite true. There is a rigorous mathematical definition of the amount of information and, for the benefit of those readers who have not encountered the concept, I shall try to explain what this concept means in an unhurried step-by-step manner.

The most usual form of information is the printed word. For example, the text that you are reading gives you an opportunity to acquire the information in this paper. This information is carried by symbols arranged in a line-by-line fashion: these symbols are letters, gaps between the words, and punctuation marks. It is perfectly obvious that the greater the number of text pages, the greater the amount of information carried by the text. However, in order to characterise the amount of information in the form of a specific number, it is necessary to begin with a very simple example. Let us assume that the text is transmitted by the Morse alphabet, in which each letter corresponds to a certain number of dots and dashes. Moreover, let us consider the simple case when the text is continuous without any gaps between letters and words. The result is one continuous sequence of dots and dashes. At each position there can be only one of two symbols: either a dot or a dash. When we are dealing with just one of two symbols, it is usual to assume that each cell carries one bit of information. The whole Morse sequence with $N$ symbols contains $N$ bits of information. We can say that such a sequence carries the 'memory' of a certain text and that in each of its $N$ 'memory cells' there is one bit of information. The total number of different texts which can be stored in a ribbon of $N$ cells is obviously $2^{N}$.

If we agree to measure the amount of information in bits, then the quantity of information $I_{B}$ can be described by the relationship [20]

$$
\begin{equation*}
I_{\mathrm{B}} \equiv N \equiv \log _{2} M_{N} \tag{1}
\end{equation*}
$$

Here, $M_{N}=2^{N}$ is the total number of different texts. According to Eqn (1) the amount of information is simply equal to the minimum number of binary cells by means of which this information can be written down.

The relationship (1) can be represented in a somewhat different form as follows. If there is a set of $M_{N}$ different texts, the probability $P_{N}$ that the text we are reading is
identical with that selected at random from $M_{N}$ different texts is obviously $P_{N}=1 / M_{N}$. Therefore, instead of Eqn (1) we can use the expression

$$
\begin{equation*}
I_{\mathrm{B}}=-\log _{2} P_{N} \tag{2}
\end{equation*}
$$

The higher the value of $N$, the lower the probability $P_{N}$ and the larger the amount of information $I_{\mathrm{B}}$ contained in a specific text.

Let us now return to an ordinary letter text. Let us assume that the number of letters in the alphabet is 32 (which is true, for example, of the Russian alphabet apart from the letter ë). The number can be represented as $32=2^{5}$ so that it is sufficient to have five binary cells in order to assign to each letter a unique combination of, for example, dots and dashes. If the lower-case letters are supplemented by the capitals, the number of letters is doubled to 64 and we need another bit of information, so that the amount of information per letter (lower-case or capital) becomes $I_{\mathrm{B}}=6$. The addition of the gaps between the words and of the punctuation marks increases further the amount of information carried by one text symbol.

However, such direct calculation of the amount of information per symbol is not quite correct. This is because the alphabet contains letters which are encountered very rarely in text. In the Morse alphabet one can 'expend' more dots and dashes on such letters but in the case of the frequently encountered letters one can economise by assigning shorter ribbon 'sections' of the sequence to these letters. A rigorous definition of the ammount of information has been provided by Shannon [21]. It appears as follows:

$$
\begin{equation*}
I=-\sum_{i} p_{i} \ln p_{i} . \tag{3}
\end{equation*}
$$

The summation is carried out over all the symbols and $p_{i}$ denotes the probability of the appearance of a symbol labelled with the number $i$. The general expression given by Eqn (3) applies both to the frequently employed letters and to those which are very unlikely to appear in a text. The natural logarithm is used in Eqn (3): it corresponds to a unit of information called a 'nat'.

It is known from probability theory that in the case of a random quantity $x_{i}$ we can introduce the definition of its average value or of the mathematical expectation in accordance with the expression

$$
\begin{equation*}
\left\langle x_{i}\right\rangle=\sum p_{i} x_{i} \tag{4}
\end{equation*}
$$

where the summation is carried out over all possible values of a random quantity $x_{i}$ with a probability $p_{i}$ of the $i$ th value. We can see that Eqn (3) can also be written in the form

$$
\begin{equation*}
I=-\left\langle\ln p_{i}\right\rangle \tag{5}
\end{equation*}
$$

If we consider the text again and assign $I$ to one cell of the text, i.e. to one symbol, then $p_{i}$ is the probability of the appearance of a symbol with the serial number $i$. For example, if this is the letter ' $a$ ', then the relevant probability can be found by calculating how many times the letter ' $a$ ' appears on one page and dividing the number obtained by the total number of symbols on one page.

If we know the amount of information (3) per symbol, then in the case of a text with $N$ symbols the amount of information should be increased simply by the factor $N$. Naturally, Eqn (3) can also be applied to the whole text.

Then the probabilities $p_{i}$ are much lower because the number of possible combinations rises steeply and the result remains the same: the information carried by one symbol should be multiplied by $N$ to obtain text information.

Logarithms with different bases are used in Eqns (2) and (3): in Eqn (2) this is the logarithm to the base 2 and in Eqn (3) the logarithm is natural. Since for any number $N$ we have $N=2^{\log N}=\mathrm{e}^{\ln N}$, it follows that

$$
\begin{equation*}
I_{\mathrm{B}}=\frac{I}{\ln 2} \cong 1.44 I \tag{6}
\end{equation*}
$$

In other words, the number of bits is almost 1.5 times greater than the number of nats.

It will be shown later that in dealing with physics (and not with computer technology) it is more convenient to use the Shannon definition of information given by Eqn (3), i.e. to measure it in nats. If desired, Eqn (6) can then be used to find the number of bits corresponding to a given value of $I$. The Shannon definition (3) can be used not only in the case of a text, but also in the case of any other discrete (digital) information. For example, a black-and-white image on a television screen can be expanded into a set of discrete black-and-white dots, as well as several intermediate grey shadings. Then, Eqn (3) gives the amount of information provided by a given instantaneous image on the screen. A similar definition applies also to a colour image on a television screen or to the paper output of a printer if we allow suitably for the information carried by colour shading.

It will be shown later that the information defined by Eqn (3) plays a major role in nonequilibrium physical processes. It should be stressed this applies to the numerical expression for the amount of information irrespective of its intellectual content.

## 3. Entropy

The concept of entropy is one of the fundamental ideas in physics. The reader obviously already knows what it means. Nevertheless, the continuity and consistency of the presentation requires that some time be spent in making clear this physical quantity. It is convenient to use the simplest physical object, which is an ideal gas.

Let us assume that a monatomic ideal gas with a particle density $n$ kept at a temperature $T$ occupies a volume $V$. The temperature $T$ is measured in energy units (for example, in ergs if the cgs system of units is used). Consequently, the Boltzmann constant will not appear in the relationship considered below. Each atom of this gas has an average kinetic energy of thermal motion amounting to $3 T / 2$. Therefore, the total thermal energy of the gas is

$$
\begin{equation*}
E=\frac{3}{2} T n V \tag{7}
\end{equation*}
$$

The gas presure is known to be $p=n T$. If the gas can exchange heat with the external medium, the law of conservation of the gas energy is

$$
\begin{equation*}
\mathrm{d} E=-p \mathrm{~d} V+\mathrm{d} Q \tag{8}
\end{equation*}
$$

Therefore, the internal energy of the gas may change both because of the work done by the gas and because of the deposition of a certain amount of heat $\mathrm{d} Q$ from outside. Eqn (8) is known to describe the first law of thermodynamics, i.e. the law of energy conservation. The gas is assumed to be in equilibrium, i.e. $p=$ const throughout the gas volume.

If it is also assumed that the gas is in complete thermodynamic equilibrium ( $T=$ const $)$, then the relationship given by Eqn (8) can be considered to represent an elementary process of variation of the gas parameters at a very low rate, so that the thermodynamic equilibrium is not disturbed. It is for these processes that the concept of entropy $S$ is introduced by the relationship

$$
\begin{equation*}
\mathrm{d} S=\frac{\mathrm{d} Q}{T} \tag{9}
\end{equation*}
$$

In other words, an equilibrium gas not only has an internal energy, but also another important internal characteristic associated with the thermal motion of its atoms. It follows from Eqns (8) and (9) that at a constant volume, $\mathrm{d} V=0$, the change in energy is proportional to the change in entropy and in general we have

$$
\begin{equation*}
\mathrm{d} E=-p \mathrm{~d} V+T \mathrm{~d} S \tag{10}
\end{equation*}
$$

Since $p=(N / V) T$ and $E=\frac{3}{2} N T$, where $N=n V=$ const is the total number of atoms in the investigated gas, we can write Eqn (10) in the form

$$
\begin{equation*}
\mathrm{d} S=N\left(\frac{3}{2} \frac{\mathrm{~d} T}{T}+\frac{\mathrm{d} V}{V}\right) \tag{11}
\end{equation*}
$$

Hence, integration of Eqn (11) yields directly

$$
\begin{equation*}
S=N\left[\ln \left(V T^{3 / 2}\right)+\mathrm{const}\right] \equiv N s \tag{12}
\end{equation*}
$$

The constant of integration is retained so as to leave some additional freedom in the subsequent discussion. The entropy of the gas is proportional to the number of particles and the expression in the brackets, equal to $s$, is the entropy per particle.

If the volume and temperature of the gas vary in such a way that $V T^{3 / 2}$ remains constant, the entropy $S$ does not change. According to Eqn (9) this means that the gas does not exchange heat with the external medium, i.e. that the gas is separated from this medium by reliable heat-insulating walls. A process of this kind is known as adiabatic. In an adiabatic process the following relationship is valid:

$$
\begin{equation*}
p V^{\gamma}=\mathrm{const} \tag{13}
\end{equation*}
$$

where $\gamma=5 / 3$ is known as the adiabatic exponent. This relationship is obtained from the condition $V T^{3 / 2}=$ const and from $p=n T$. In the adiabatic process the temperature and pressure vary with the density as follows:

$$
\begin{equation*}
T=\mathrm{const} \times n^{\gamma-1}, \quad p=\mathrm{const} \times n^{\gamma} \tag{14}
\end{equation*}
$$

## 4. Carnot cycle

In the subsequent discussion it will be necessary to make recourse to what are known as the thought (gedanken) experiments involving different types of ideal processes and ideal devices. In thermodynamics one such ideal device is the Carnot heat engine, in which work is done at the expense of thermal energy.

Let us assume that there are two thermostats kept at different temperatures $T_{1}$ and $T_{2}$, such that $T_{1}>T_{2}$. The first thermostat may be called a heater and the second a cooler. According to Carnot, an ideal gas can be used as the working substance to produce work at the expense of the thermal energy. Let us postulate that at the temperature $T_{1}$ the volume of the gas is $V_{1}$ and its pressure is $p=p_{1}=\left(N / V_{1}\right) T_{1}$. Here, $N$ is the total number of particles. A reversible adiabatic process can cool the
working gas to the temperature $T_{2}$ because, according to Eqn (14), the gas temperature falls as it expands in accordance with $T \propto V^{-\gamma+1}$. For $\gamma=5 / 3$ we have $T \propto V^{-2 / 3}$. Therefore, for $V_{2} / V_{1}=\left(T_{2} / T_{1}\right)^{-3 / 2}$ the temperature of the gas in the first state is $T_{1}$ and in the second state is $T_{2}$. The transition from the first to the second state is adiabatic. The adiabatic expansion of the gas from the initial volume $V_{1}$ to the final volume $V_{2}$ performs the work $W=\int p \mathrm{~d} V$. Since according to Eqn (13) the pressure varies proportionally to $V^{-\gamma}=V^{-5 / 3}$, i.e.

$$
p=\frac{N T_{1}}{V_{1}}\left(\frac{V_{1}}{V}\right)^{5 / 3}
$$

this work is given by

$$
\begin{equation*}
W=\frac{3}{2} N T_{1}\left[1-\left(\frac{V_{1}}{V_{2}}\right)^{2 / 3}\right]=\frac{3}{2} N\left(T_{1}-T_{2}\right) \tag{15}
\end{equation*}
$$

It follows from the law of conservation of energy that this is exactly the amount by which the internal energy of the gas is reduced.

If the gas is compressed again from the volume $V_{2}$ to $V_{1}$, this requires work equal to that given by Eqn (15) and the energy of the gas is restored again. So far there is no cycle: these are simply forward and reverse processes which do not alter anything in the external world. A cycle appears when the working substance receives heat from the heater and gives it up to the cooler. At the heater this is an isothermal expansion of the gas at the temperature $T_{1}$ from its initial volume $V_{1}$ to some intermediate volume $V_{1}^{*}$. The gas then expands adiabatically to the final volume $V_{2}^{*}$ in such a way that the final temperature is $T_{2}$, i.e. $V_{2}^{*} / V_{1}^{*}=\left(T_{2} / T_{1}\right)^{-3 / 2}$. The gas is then compressed isothermally to the volume $V_{2}$ and converted adiabatically to the initial state. The forward and reverse work done in the adiabatic parts of the cycle balance each other exactly so that, in accordance with Eqn (15), the work $W$ is governed solely by the difference between the initial and final temperatures. It is in the isothermal parts of the cycle that the amounts of work done are different. In fact, if the heater supplies an amount of heat $Q_{1}$ to the body, then $Q_{1}=T_{1}\left(S_{*}-S\right)$, where $S$ is the initial entropy of the gas and $S_{*}$ is the value after heating. In the adiabatic parts of the cycle the entropy does not change so that the entropy in the cooler has to be reduced from $S_{*}$ to $S_{1}$ by transfer to the gas of a negative heat (i.e. by removal of heat) amounting to $Q_{2}=-T_{2}\left(S_{*}-S\right)$. Then, since in the isothermal parts the internal energy does not change, the difference between the amounts of work done $W_{1}$ and $W_{2}$ is $W=W_{1}-W_{2}=$ $\left(T_{1}-T_{2}\right)\left(S_{*}-S\right)$. We can now calculate the efficiency of the cycle, which is the ratio of the work done $W$ to the amount of heat $Q_{1}$ obtained from the heater:

$$
\begin{equation*}
\eta=\frac{W}{Q_{1}}=1-\frac{T_{2}}{T_{1}} \tag{16}
\end{equation*}
$$

The most important feature of the Carnot engine is its reversibility: if the machine is reversed, it acts as a cooler transferring heat from the hot to the cold body as a result of work done from outside.

It is this circumstance together with the second law of thermodynamics that leads to the conclusion the Eqn (16) represents the maximum possible efficiency of a heat engine.

The second law of thermodynamics itself is formulated as the impossibility of constructing a perpetual motion machine
of the second kind: it is not possible to realise a process in which work could be done at the expense of heat without any change in the external world.

Let us suppose for a moment that it is possible to construct an engine with an efficiency greater than that given by Eqn (16). Such an engine can then be brought into contact with a Carnot engine operating in the cooling regime. This combination of engines could simply pump out the heat energy and convert it into work, which is forbidden by the second law of thermodynamics. Consequently, Eqn (16) does represent the maximum possible efficiency of any heat engine. It is interesting to ask what happens in the external world during the Carnot cycle? In one cycle the Carnot engine receives the energy $Q_{1}$ from the heater at the temperature $T_{1}$ and at the same time removes the entropy $Q_{1} / T_{1}$ from the external world. It then transfers the heat $Q_{2}$ to the cooler at the temperature $T_{2}$ and transfers the entropy $Q_{2} / T_{2}$. Therefore, the total change in the entropy contributed by the Carnot engine is

$$
\begin{equation*}
\Delta S=\frac{Q_{1}}{T_{1}}-\frac{Q_{2}}{T_{2}}=0 \tag{17}
\end{equation*}
$$

provided we use the expressions for $Q_{1}$ and $Q_{2}$. As we can see, the Carnot engine is fully reversible and does not alter the entropy of the external world although it transforms part of the heat into mechanical work. Heat nevertheless leaks from the hot to the cold body.

## 5. Irreversible processes

The Carnot cycle is an ideal completely reversible thought process. The operation of all real heat engines involves irreversible processes and it is therefore desirable to understand the meaning of such irreversibility. It is convenient to start again with the Carnot engine.

The Carnot cycle has adiabatic and isothermal parts. The adiabatic process of compression or expansion of a gas can be quite readily represented as very close to reversible. It is sufficient to postulate the compression or expansion rates much lower than the velocity of sound, and to create a sufficiently good thermal insulation, which is automatically ensured if the dimensions of the gas-occupied volume are large. The situation is much more difficult in the case of the isothermal parts of the cycle: it is in these parts that heat has to be delivered or removed. If the isothermal process is accelerated, the result may be a considerable reduction in the efficiency because the gas can no longer reach the temperature $T_{1}$ during heating and the temperature $T_{2}$ during cooling. In the limit of very fast thermal oscillations it is possible to lose all the work and simply accelerate the flow of heat from the hot to the cold body.

Let us now consider in detail the irreversible process of the transfer of heat by conduction. Let us assume that there is a rod of length $L$ and the temperatures at its ends are $T_{1}$ and $T_{2}$. If the thermal conductivity of the $\operatorname{rod}$ is $\kappa$, then the heat flux along the rod is

$$
\begin{equation*}
q=\kappa \frac{T_{1}-T_{2}}{L} \tag{18}
\end{equation*}
$$

Therefore, across each transverse cross section of the rod the entropy flux is $q_{S}=q / T$. If the temperature is a linear function of the coordinate $x$, so that

$$
\begin{equation*}
T=T_{1}-\frac{\left(T_{1}-T_{2}\right) x}{L}, \tag{19}
\end{equation*}
$$

the entropy flux considered as a function of $x$ is

$$
\begin{equation*}
q_{S}=\frac{q L}{(L-x) T_{1}+T_{2} x} . \tag{20}
\end{equation*}
$$

This flux increases along the rod from the hot end where $T=T_{1}$ to the cold end where $T=T_{2}$. As we can see, at the cold end the entropy flux is $T_{1} / T_{2}$ times greater than at the hot end. In other words, an excess entropy is produced inside the rod and this entropy then flows out in the direction of the cooler. In addition to the entropy flux $q / T_{1}$ entering the rod per unit time, there is an entropy created inside the rod so that a larger flux $q / T_{2}$ is ejected outside.

We recall that the Carnot engine does the work $W=n q_{1}=\left[1-\left(T_{2} / T_{1}\right)\right] q_{1}$, where $q_{1}$ is the flux entering the engine. Consequently, the flux $q_{2}=q_{1}-W=$ ( $T_{2} / T_{1}$ ) $q_{1}$ emerging from the Carnot engine corresponds to the absence of internal entropy production. If the efficiency is less, the heat engine creates an entropy

$$
\begin{equation*}
\frac{q_{2}}{T_{2}}-\frac{q_{1}}{T_{1}}=\frac{1}{T_{2}}\left(q_{1}-\eta q_{1}\right)-\frac{q_{1}}{T_{1}}=\frac{q_{1}}{T_{2}}\left(\eta_{\mathrm{C}}-\eta\right) \tag{21}
\end{equation*}
$$

Here, $\eta_{\mathrm{C}}=1-\left(T_{2} / T_{1}\right)$ is the efficiency of the Carnot cycle and $\eta$ is the real efficiency. If $\eta=0$, then no work is done, $W=0$, and $q_{2}=q_{1}=q$. In this case the pure heat flux performs no work and simply thermally 'contaminates' the medium where it generates entropy at the rate

$$
q\left(\frac{1}{T_{2}}-\frac{1}{T_{1}}\right)
$$

It thus follows that the heat flux itself increases the entropy of the medium and this process continues until the temperatures of the hot and cold bodies become equal.

## 6. Entropy and information

In the spirit of ideal thought experiments let us now consider the case when the whole of our ideal gas consists of just one particle. It might seem that this is an absolutely absurd approach but let us not be too hasty in our judgement. If one particle is enclosed in a vessel of volume $V$ with walls kept at a temperature $T$, then sooner or later this particle comes into equilibrium with these walls. At each moment in time the particle is naturally at one definite point in space and has a very definite velocity. However, all the processes may be assumed to be so slow that the particle not only has a chance to fill on average the whole space of volume $V$, but it can repeatedly change the magnitude and direction of its velocity by inelastic collisions with the vessel walls. In this case we can speak of a particle which has a Maxwellian velocity distribution and on average fills the vessel containing it. A very important condition is that we do not need to know anything about the particle except that it collides with the walls and exerts on them an average pressure, and that its velocity distribution is Maxwellian with the temperature $T$.

If this particle is now, for example, compressed adiabatically, its average effect on the walls is exactly the same as that of $N$ particles, except that it is $N$ times smaller. The temperature of the particle can also be changed slowly by suitably altering the temperature of the walls of the vessel and thus allowing the particle to reach thermal equilibrium with the walls.

The average pressure exerted by one particle on a wall when $N=1$ is obviously equal to $p=T / V$ and the average density is then $n=1 / V$. For this particle we again can
organise a Carnot cycle and thus create an ideal heat microengine which is fully reversible and whose efficiency is $\eta=1-\left(T_{2} / T_{1}\right)$. Now with the aid of some ideal devices and by application of the second law of thermodynamics we shall try to establish a relationship between entropy and information. We shall begin with the simplest case of isothermal processes. We shall thus assume that $T=$ const. It follows from the first law of thermodynamics at $T=$ const and from the relationship $p=T / V$ that

$$
\begin{equation*}
T \mathrm{~d} S=p \mathrm{~d} V=T \frac{\mathrm{~d} V}{V} \tag{22}
\end{equation*}
$$

One should use here the lower-case symbol $s$ for the specific entropy, but $S$ is retained so as to go over later to more general relationships.

Hence, we find that the entropy change is independent of the value of $T$, so that we can assume

$$
\begin{equation*}
S=\ln \frac{V}{V_{0}} . \tag{23}
\end{equation*}
$$

Here, a small volume $V_{0} \ll V$ is introduced deliberately simply for the sake of normalisation: the value of $V_{0}$ should be considerably greater than the size of the particle in order not to violate the ideal-gas approximation.

The work done in an isothermal process is also described by a fairly simple expression:

$$
\begin{equation*}
W=\int p \mathrm{~d} V=T \int \frac{\mathrm{~d} V}{V}=T \ln \frac{V_{2}}{V_{1}} . \tag{24}
\end{equation*}
$$

We can see that $W=T\left(S_{2}-S_{1}\right)$, i.e. the work is expressed simply in terms of the difference between the corresponding entropies. We shall stress once again that we are dealing with a very slow, fully reversible process.

Let us continue our ideal thought experiments. Let us assume that there are ideal barriers which can be used to divide a vessel into parts without loss of energy or creation of new entropy. Let us divide the vessel into two equal parts each of volume $V / 2$. The particle then remains in one of the halves, but we do not yet know which of them. Let us now assume that there is some means or measuring instrument which allows us to determine where the particle is located. For example, it might be detected in one of the halves by means of a spring balance in the gravitational field or simply from the disappearance of the pressure on the barrier exerted by the empty half-volume. If this is true, then out of the initial symmetric distribution with a $50 \%$ chance of finding the particle in either of the two halves, we now have the probability of $100 \%$ for one of the halves. The distribution of probabilities seems to 'shrink' or 'collapse'. The new entropy $S_{2}=\ln \left(V / 2 V_{0}\right)$ is correspondingly less than the initial entropy by an amount $\Delta S=\ln 2$. Because of a reduction in the entropy, mechanical work can be done. It is sufficient to move the barrier in the direction of the empty part until it disappears completely so that the particle occupies again the full volume. The corresponding work done is $W=T \Delta S=T \ln 2$. If nothing changes in the external world, then repetition of these cycles should make it possible to construct a perpetual motion machine of the second kind. However, since the second law of thermodynamics forbids the creation of work directly at the expense of heat, some event should occur in the external wall. What is that event?

Detection of the particle in one of the halves alters the information about the particle. Specifically, out of two
possible halves, there is now only one where the particle is located. This knowledge corresponds exactly to one bit of information. The process of measurement reduces the entropy of the particle and increases by exactly the same amount the information provided by the measuring device. If the halves are divided again into quarters, eighths, etc., the entropy gradually decreases and the information increases. In other words, we have

$$
\begin{equation*}
S+I=\text { const } \tag{25}
\end{equation*}
$$

The more is known about the particle or, in the more general case, about the physical system, the less its entropy.

However, we can now draw the conclusion that the appearance of information in the external world (or in external devices) is impossible without an increase in the entropy of the external environment by an amount not less than $\Delta I$. Otherwise our reversible heat microengine could produce work directly at the expense of heat. In other words, information which is a definite portion of order, can be acquired by external devices, automata, or simply by the external world only at the expense of an additional disorder (thermal motion) in the external environment. Information production is accompanied by an increase in the entropy by an amount no less than the information that has been gained.

An external device or environment which acquires information and can utilise it in the subsequent events can be called Maxwell'sdaemon. Maxwell postulated his daemon for a precisely similar situation: if the daemon can distinguish between hot and cold particles, then by closing an aperture with a shutter it can transfer heat from the cold to the hot end. The above discussion is analogous to the Maxwell's daemon concept. It is based on the universal second law of thermodynamics. It is this second law that requires the daemon himself to increase the entropy in the process of 'heavy work' in recognition of the information $I$.

It follows from Eqn (25) that the sum of $S$ and $I$ is constant. If in our model a particle is placed in an elementary cell of volume $V_{0}$, then $S=0$ and the information reaches its maximum value $I_{\max }=-\ln p_{\min }=\ln V / V_{0}$, since the probability $p_{\text {min }}$ of finding the particle in a given cell is equal to the volume ratio $p_{\text {min }}=V_{0} / V$. A particle can be located in an elementary volume, i.e. the information $I=\ln V / V_{0}$ can be acquired, if at least this amount of entropy is produced inside or outside it. If subsequently the particle begins to expand (on average) over a large volume, information is gradually lost and the entropy of the particle increases.

It should be stressed once again that one has to 'pay' for information by an increase in the entropy $S_{\mathrm{e}}$ of the external systems and the increase is such that $\Delta S_{\mathrm{e}} \geqslant I$. Indeed, if the acquisition of one bit of information could increase the entropy of the instrument by an amount $\Delta S_{\mathrm{e}}$ which is less than one bit, a heat engine can be reversed. In the opposite case, by expanding the half-volume occupied by the particle we would increase its entropy by $\ln 2$ and thus do the work $T \ln 2$, whereas the final total entropy of the particle and the instrument would decrease. But this is forbidden by the second law of thermodynamics.

We have considered so far the entropy associated with the spatial localisation of a particle. In fact, this applies also to the velocity measurement. It is convenient to consider the case of just one dimension, for example, the motion only along the $x$ axis.

Let us assume that we have an instrument which can be used to measure the $x$ component of the velocity $v$ of particles with an error $\Delta v$. It is convenient to assume that $\Delta v$ is proportional to $v$, which ensures a certain unity in the subsequent estimates. A particle enters a tube of length $L$ at a velocity $v$ along the tube axis $x$, and on entry its velocity is measured with an error $\Delta v$. After a time $t=L / v$, a barrier is introduced at a distance $\Delta L \approx L \Delta v / v$ from the closed end and imprisons the particle there. The measure of localisation is $\Delta L / L$, which is approximately the same for all the particles with the Maxwellian distribution. Such localisation again changes the entropy by $\ln (L / \Delta L)$ and, in accordance with the second law of thermodynamics, must be accompanied by an increase in the entropy of the instrument that measures the particle velocity. In other words, any measurement that increases the information about a particle must be accompa-nied by an increase in the entropy of the instrument or of the environment. This is the characteristic cost of information (discussed in greater detail in Refs [2225]).

## 7. Entropy encore!

The relationship of Eqn (25) suggests that the entropy of a physical system is related in some way to the distribution of velocities. This is indeed true, as described in detail in monographs and text books on statistical physics. I shall acquaint the reader briefly with this problem and once again shall rely on the simplest example for an ideal gas.

The famous Boltzmann formula applies to statistical physics: the entropy is equal to the logarithm of the number of possible states $\Gamma$ of a given physical system:

$$
\begin{equation*}
S=\ln \Gamma \tag{26}
\end{equation*}
$$

It must be pointed out once again that we are using energy units to measure temperature and, therefore, the Boltzmann constant does not occur in the above expression. Here, $\Gamma$ is the number of possible microscopic states which correspond to the same macroscopic state. It is assumed that in the course of thermal motion the system can assume all possible states with approximately the same probability. Consequently, the prob-ability $p_{i}$ of encountering a particular state is approximately $p_{i} \approx 1 / \Gamma$. If the probabilities of individual states are different, a more accurate definition of the entropy is

$$
\begin{equation*}
S=-\sum_{i} p_{i} \ln p_{i}=-\left\langle\ln p_{i}\right\rangle \tag{27}
\end{equation*}
$$

The angular brackets here denote a mathematical expectation, i.e. the average value.

For the ideal gas the probability is $p_{i}$ split into the configuration and velocity parts, i.e. into the probabilities of the distribution in space and over the velocities. The number of possible states in space with a volume $V$ and $N \gg 1$ particles can be found by introducing an elementary volume $V_{0}$. It seems that the number of such states is simply equal to $\left(V / V_{0}\right)^{N}$. However, this is not quite true: the atoms in the gas are identical and, therefore, their transpositions do not create a new state. Consequently, the above number should be divided once again by $N!\approx N^{N}$, where only the principal term is retained in the Stirling formula for $N!$. The velocity part of $\Gamma$ is simply equal to $-N\left\langle\ln f_{0}\right\rangle$, where $f_{0}$ is the Maxwellian func-tion for the distribution of the velocity of one of the atoms:

$$
\begin{equation*}
f_{0}(v)=\left(\frac{m}{2 \pi T}\right)^{3 / 2} \exp \left(-\frac{m|v|^{2}}{2 T}\right) \tag{28}
\end{equation*}
$$

Here, $m$ is the mass of an atom and $v$ is the velocity vector. It therefore follows from Eqn (27) that the entropy of an ideal gas is

$$
\begin{equation*}
S=N\left[\ln \left(\frac{V}{N V_{0}}\right)+\ln \left(c_{0} T\right)^{3 / 2}\right] \tag{29}
\end{equation*}
$$

here, $c_{0}=2 \pi \mathrm{e} / m$ is simply a constant factor which appears as a result of the averaging of $\ln f_{0}$; e is the base of natural logarithms. In contrast to the thermodynamic expression (12) obtained earlier, Eqn (29) includes an explicit expression for the integration constant, namely const $=\ln \left(c_{0} / N V_{0}\right)$ in Eqn (12).

Formally the expressions (3) for information and (27) for entropy are identical. However, their meaning is completely different. Information of Eqn (3) corresponds to just one sampling from an enormous (for example $\Gamma$ ) number of possible states. The measure of this information is $I=\ln \Gamma$. The entropy, however, corresponds to the possibility of finding a system with a certain probability $1 / \Gamma$ in each of the available states. The quantity $S=\ln \Gamma$ corresponds to the maximum 'occupancy' of all the states. The quantities $I$ and $S$ are formally equal because $I$ corresponds to the maximum information for just one state and $S$ is defined for the set of all the states.

Let us assume that, for example, the amount of information $I$ represents the text of the present paper. The entropy of this text is zero since there is only one fixed sequence of letters and other typographical symbols. Consequently, $\Gamma=1$. Let us now assume that the whole text is set in 'thermal motion' as a result of which the letters begin to jump rapidly and become transposed. Very soon all the information is totally lost, but in the course of such thermal motion all possible states from the total number $\Gamma$ are assumed, i.e. $S=\ln \Gamma$. In an intermediate variant when a part of the text is conserved and another goes over to a completely chaotic 'thermal motion', we have Eqn (25): $S+I=$ const. The entropy and information of a closed system seem to have a mutual relationship to one another: the 'forgetting' of information automatically increases entropy.

According to the second law of thermodynamics the entropy of a closed system cannot decrease with time. In the example under discussion this means that the boundary between the 'thermal' and 'information' parts of the text can naturally move in just one direction, namely in the direction of forgetting information. A new portion of information, i.e. the deliberate 'freezing' of part of the text which is in thermal motion, can occur only at the expense of additional erasure of some part of the main text if this process occurs without introducing additional information from outside. Order can only appear because of the destruction of a different order (in a closed system).

It follows from this discussion that the problems of analysis of order and of creation of new information require, as a rule, going beyond a certain isolated physical system. Therefore, open nonequilibrium systems will be basic to our discussion and we shall deal with them step by step from simple to complex.

## 8. Control

To continue our ideal thought experiments, we shall consider a different formulation of the problem. Let us assume that our particle executes Brownian motion in a plane $(x, y)$. The probability of finding it in this plane obeys the Fokker Planck equation, which in the simplest variant looks like the diffusion equation:

$$
\begin{equation*}
\frac{\partial p}{\partial t}=D \nabla^{2} p \tag{30}
\end{equation*}
$$

Here, $p(x, y, t)$ is the probability density for the position of the particle in the plane, $\nabla^{2}$ is the two-dimensional Laplacian, and $D$ is the diffusion coefficient. The solution in the form of a point-source function yields

$$
\begin{equation*}
p=\frac{1}{4 \pi D t} \exp \left(-\frac{x^{2}+y^{2}}{4 D t}\right) \tag{31}
\end{equation*}
$$

This expression can be used to find the expression for the specific entropy:

$$
s=\int p \ln p \mathrm{~d} \boldsymbol{r}=\ln (4 \pi \mathrm{e} D t)
$$

Here, we are allowing for the fact that $\left\langle x^{2}+y^{2}\right\rangle=4 D t$.
We shall now select a certain value $t=t_{0}$ corresponding to the localisation of a particle described by the average value $\left\langle x^{2}+y^{2}\right\rangle=4 D t_{0}$. We shall now wait a certain time $\Delta t$ such that $\Delta t=t-t_{0}$. In this time the entropy rises by $\Delta s=\ln 2$. We assume that there is a device which can automatically move the cell where the Brownian particle is located. Such a device could centre the position of the particle and restore its previous localisation. However, this requires measurement of the position of the particle with an error not exceeding such localisation. Acquisition of this information and its subsequent use in the movement of the cell requires an increase in the entropy of the device by at least $\ln 2$.

Therefore, if this device is used in the controlled localisation of a particle, we can stop the rise in the entropy of the particle itself, but only at the expense of constant creation of the entropy inside or outside this device at the rate $\dot{s}$, which is at least $\ln \left(2 / t_{0}\right)$. The dissipation and the corresponding irreversibility seem to be transferred outside the cell containing the Brownian particle.

They are transferred to the system of control of the position of the particle. The particle itself is converted into an open system from which an excess entropy 'slag', created by dissipation, is continuously removed. The total entropy of the particle and of the control device naturally increases, but this increase occurs somewhere far from the object of interest to us. This example demonstrates that entropy creation need not occur strictly locally, but may proceed somewhere far from the system in question.

## 9. Gas dynamics

At first sight it seems that dissipation should always and inexorably bring a physical system close to thermodynamic equilibrium. In some sense this is true, but the real processes may be much more complex and this is clear from the simplest example of gas dynamics.

In general, the behaviour of a low-density gas is described well by the Boltzmann kinetic equation

$$
\begin{equation*}
\frac{\partial f}{\partial t}+v \cdot \nabla f=\operatorname{St}(f) \tag{32}
\end{equation*}
$$

Here, $f$ is the local function representing the distribution of the particle velocities and $\operatorname{St}(f)$ is the collision term. It is collisions that are responsible for the approach to local equilibrium, i.e. for dissipation. If the collisions are sufficiently frequent, the distribution function becomes Maxwellian:

$$
\begin{equation*}
f_{0}=n\left(\frac{m}{2 \pi T}\right)^{3 / 2} \exp \left[-\frac{m|v-u|^{2}}{2 T}\right] \tag{33}
\end{equation*}
$$

Here, $n, T$, and $\boldsymbol{u}$ are the local values of the density, temperature, and average velocity. For the Maxwellian distribution function the local equilibrium is complete, i.e. $\operatorname{St}(f)=0$. However, if $n, T$, and $\boldsymbol{u}$ are functions of the coordinates and time, the left-hand side of Eqn (32) does not vanish automatically on substitution of $f=f_{0}$. This means that a complete thermodynamic equilibrium is not achieved although the collision term exerts a strong tendency towards equilibrium. We can readily see that changes in the quantities $n, T$, and $\boldsymbol{u}$ with time should obey certain constraints imposed by the nature of Eqn (32) itself. This is because the collision term is constructed in such a way that it conserves the number of particles, their total momentum, and the their total energy. Therefore, the left-hand side of Eqn (32) should also be subject to these constraints.

Let us integrate Eqn (32) with respect to the velocities. If $f$ is assumed to be identical with $f_{0}$, the result is

$$
\begin{equation*}
\frac{\partial n}{\partial t}+\operatorname{div}(n u)=0 \tag{34}
\end{equation*}
$$

This equation of continuity automatically guarantees that the collision term does not annihilate or create particles. If we now multiply Eqn (32) by $m v$ and integrate again with respect to $v$ allowing for $f=f_{0}$, then the result is - subject to Eqn (34) - the Euler equation

$$
\begin{equation*}
m n \frac{\mathrm{~d} \boldsymbol{u}}{\mathrm{~d} t}+\nabla p=0 \tag{35}
\end{equation*}
$$

where the operator is

$$
\frac{\mathrm{d}}{\mathrm{~d} t}=\frac{\partial}{\partial t}+(\boldsymbol{u} \cdot \nabla)
$$

and the quantity $p=n T$ is the local gas pressure.
Similarly, multiplication of Eqn (33) by $\frac{1}{2} m v^{2}$ and integra-tion with respect to the velocities on the assumption that $f=f_{0}$, subject to Eqns (34) and (35) gives the relationship

$$
\begin{equation*}
\frac{\mathrm{d} p}{\mathrm{~d} t}+\frac{5}{3} p \operatorname{div} \boldsymbol{u}=0 \tag{36}
\end{equation*}
$$

where $p=n T$.
Therefore, instead of the Boltzmann equation we now have the system (34)-(36) for two scalar and one vector quantity, and instead of a six-dimensional phase space we can use the usual configuration space.

The equations of ideal gas dynamics become more precise as $\operatorname{St}(f)$ increases, i.e. as the collisions between the atoms become more frequent. Mathematically this means that the equations of hydrodynamics represent an asymptotic form of the Boltzmann equation in the limit $\mathrm{St} \rightarrow \infty$. The collisions themselves select a set of quantities $n, T$, and $\boldsymbol{u}$, which are transformed into dynamical variables. They can be called the
order parameters because they represent the main characteristics of a local thermodynamic equilibrium.

The variables $n, T$, and $\boldsymbol{u}$ may be considered as smooth functions of the coordinates. Their behaviour is governed by the equations of gas dynamics. Naturally, against the background of this dynamics there are small thermodynamic fluctuations due to the discrete structure of the atomic gas, but at this stage they are of no interest to us. The equations of gas dynamics, which are nonlinear in their structure, provide natural means for the description of such processes when the changes in space and time of the relevant dynamical variables $n, T$, and $\boldsymbol{u}$ are much greater than the thermodynamic background of the usual fluctuations. Consequently, we can speak of states and processes which are very far from thermodynamic equilibrium.

The equations of gas dynamics represent only the simplest example of the description of physical systems far from equilibrium. In other nonequilibrium physical systems one frequently encounters a situation that certain order parameters arise spontaneously and these begin to play the role of dynamical variables. Our task will be to provide a qualitative description of such dynamical systems subject to the allowance for the role of informational processes.

As demonstrated earlier, informational processes are very closely intertwined with irreversible dissipative processes in systems far from thermodynamic equilibrium. As far as possible, we shall consider these processes step by step going from simple to complex.

## 10. Waves

Flow or oscillations of a gas at speeds much less than that of sound represent a very wide class of gas-dynamic processes. They include low-amplitude sound itself as well as incompressible flow with $\operatorname{div} \boldsymbol{u}=0$. Such flow will be discussed later and here I shall deal with sound. I shall assume that the amplitude of oscillations is small so that deviations of the density $\tilde{n}=n-n_{0}$ and pressure $\tilde{p}=p-p_{0}$ from their equilibrium values $n_{0}$ and $p_{0}$ are small. If the velocity $\boldsymbol{u}$ is also assumed to be small, the equations of gas dynamics may be linearised. The solutions of linear homogeneous equations can always be regarded as composed of elementary solutions of the plane-wave type $\exp (-\omega+i \boldsymbol{k} \cdot \boldsymbol{r})$, where $\boldsymbol{k}$ is the wave vector. For a plane wave, Eqns (34)-(36) become

$$
\begin{equation*}
\omega \tilde{n}=n_{0} \boldsymbol{k} \cdot \boldsymbol{u}, \quad m n_{0} \omega \boldsymbol{u}=\boldsymbol{k} \tilde{p}, \quad \omega \tilde{p}=\gamma p_{0} \boldsymbol{k} \cdot \boldsymbol{u} \tag{37}
\end{equation*}
$$

Here, $\gamma=5 / 3$ is the adiabatic exponent. It follows from the last two equations in the system (37) that

$$
\begin{equation*}
\omega^{2}-c_{\mathrm{s}}^{2} k^{2}=0 \tag{38}
\end{equation*}
$$

where $c_{\mathrm{s}}=\left(\gamma p_{0} / m n_{0}\right)^{1 / 2}$ is the speed of sound.
The relationship of Eqn (38) between the velocity and wave number is known as the dispersion relation. According to the second of the equations in system (37) the displacement of a medium occurs in the direction of the wave vector. A comparison of the first and third equations in system (37) shows that

$$
\tilde{p} / p_{0}=\frac{5}{3} \tilde{n} / n_{0}
$$

Since $p=n T$, it follows that the propagation of sound perturbs its temperature, so that $\widetilde{T} / T_{0}=\frac{2}{3} \tilde{n} / n_{0}=(\gamma-1) \tilde{n} / n_{0}$. This is to be expected because in the case of ideal acoustic vibrations, the pressure varies in accordance with the adiabatic law.

We shall now find the expression for the momentum density $P$ and the energy density $\varepsilon$ in an acoustic wave. If we use the angular brackets for the averaging in space, in the second approximation with respect to the amplitude of the oscillations we obtain

$$
\begin{equation*}
\boldsymbol{P}=\langle m n \boldsymbol{u}\rangle=\langle m \tilde{n} \boldsymbol{u}\rangle=\frac{\boldsymbol{k}}{\omega} m n_{0}\left\langle u^{2}\right\rangle \tag{39}
\end{equation*}
$$

where $u$ is the component of the velocity along $\boldsymbol{k}$.
We shall use here the first equation from system (37) to express $\tilde{n}$ in terms of $u$. The energy of the oscillations consists of the kinetic and potential components. The potential energy is readily found if the last two equations in system (37) are written as a simpler relationship,

$$
\begin{equation*}
\ddot{\xi}=-\alpha \xi \tag{40}
\end{equation*}
$$

where $\alpha=k^{2} c_{\mathrm{s}}^{2}$ and $\xi$ is the displacement of the medium, i.e. $\dot{\xi}=u$, and $u$ is the component of the velocity $\boldsymbol{u}$ along the direction of the wave vector (this is the only component that exhibits vibrations). F or a wave with a given wave number we simply have a harmonic oscillator and its average kinetic energy is equal to the average potential energy. We then have simply

$$
\begin{equation*}
\varepsilon=m n_{0}\left\langle u^{2}\right\rangle \tag{41}
\end{equation*}
$$

A comparison of the expressions given by Eqns (39) and (41) shows that we have the remarkable relationship

$$
\begin{equation*}
\boldsymbol{P}=\frac{\boldsymbol{k}}{\omega} \varepsilon=\frac{\boldsymbol{k}}{k} \frac{1}{v_{\varphi}} \varepsilon \tag{42}
\end{equation*}
$$

where $v_{\varphi}=\omega / k$ is the phase velocity of the wave. This is an important relationship. It is of much more general validity and applies not only to an acoustic wave, but also to any other small-amplitude waves in isotropic media. We shall now consider what this relationship means.

Any longitudinal wave can be excited by a force applied along the wave vector in resonance with the wave. Let us assume that this force is $F$, calculated per unit volume of the medium, and that it moves at the phase velocity $v_{\varphi}$. In the time $t$ the externally applied force supplies a momentum $P=F t$. In this time the force does work $W=F s$, where $s=v_{\varphi} t$ is the path traversed by the point on a wave with same resonant phase. However, it follows from the law of conservation of energy that $W=\varepsilon$, so that $\varepsilon=v_{\varphi} P$.

It is interesting to consider what happens to a wave during the motion of a medium and, for example, at a velocity $v_{0}$ along the $z$ axis. In the expression $\exp (-i \omega t+i k \cdot r)$, describing the propagation of a wave in a medium at rest, we should replace $\boldsymbol{r}$ with $\boldsymbol{r}+v_{0} t$ and we then obtain a new exponential function $\exp \left(-\mathrm{i} \omega^{\prime} t+\mathrm{i} \boldsymbol{k} \cdot \boldsymbol{r}\right)$, where

$$
\begin{equation*}
\omega^{\prime}=\omega-\boldsymbol{k}_{0} \cdot v_{0}=\omega-k_{z} v_{0} \tag{43}
\end{equation*}
$$

This is the well-known Doppler frequency shift. In addition to this frequency shift, there is also a change in the energy density

$$
\begin{equation*}
\varepsilon^{\prime}=\varepsilon-\boldsymbol{P} \cdot v_{0}=\frac{\omega^{\prime}}{k} P \tag{44}
\end{equation*}
$$

Here, $\varepsilon$ is the energy density in a wave in the coordinate system moving with the medium and $\varepsilon^{\prime}$ is the energy density in the laboratory coordinate system. According to Eqn (44) the energy density $\varepsilon^{\prime}$ vanishes at $\omega^{\prime}=0$. Such a wave can be excited without energy loss. For an acoustic wave of
frequency $\omega=k c_{\mathrm{s}}$ the condition $\omega^{\prime}=0$ deduced from Eqn (43) is

$$
\begin{equation*}
\cos \theta \equiv \frac{k_{z}}{k_{0}}=\frac{c_{\mathrm{s}}}{v_{0}} \tag{45}
\end{equation*}
$$

This is the well-known condition for the emission of Cherenkov radiation. In the case of an acoustic wave, condition (45) governs the vertex of the Mach cone.

When the velocity of the medium is sufficiently high, the frequency $\omega^{\prime}$ may prove negative. The energy of the wave in the laboratory coordinate system also becomes negative: the excitation of this wave requires the removal of energy from the medium and not the supply of energy. In this case it is usual to speak of the anomalous Doppler effect.

The relationships describing the energy and momentum of a wave can be conveniently represented in a somewhat different form. Let us assume, for the sake of simplicity, that the wave propagates along the $x$ axis. The velocity of vibrations of a medium for a wave with a given wave number $k$ can be represented in the form
$u=\frac{1}{\sqrt{2}} u_{k} \exp \left(-\mathrm{i} \omega_{k} t+\mathrm{i} k x\right)+\frac{1}{\sqrt{2}} u_{-k} \exp \left(-\mathrm{i} \omega_{-k} t-\mathrm{i} k x\right)$,
where $u_{-k}$ and $\omega_{-k}$ are simply the corresponding values of the amplitude and frequency for the factor with the negative wave number $-k$. However, since $u$ should be a real quantity, it follows that $u_{-k}=u_{k}^{*}$ and $\omega_{-k}=-\omega_{k}$. We can easily see that the mean-square value of the velocity squared along the $x$ axis is

$$
\left\langle u^{2}\right\rangle=u_{k} u_{k}^{*}=\left|u_{k}\right|^{2}
$$

We shall now select some definite point on the $x$ axis, for example, $x=0$. The velocity at this point can also be represented in the form

$$
\begin{equation*}
u=\left(\frac{\omega_{k}}{2 m n_{0}}\right)^{1 / 2} a_{k}(f)+\left(\frac{\omega_{k}}{2 m n_{0}}\right)^{1 / 2} a_{k}^{*}(t) \tag{46}
\end{equation*}
$$

for the amplitudes $a_{k}$ and $a_{k}^{*}$ (known as phasors) are given by

$$
\begin{align*}
& a_{k}=\left(\frac{m n_{0}}{\omega_{k}}\right)^{1 / 2} u_{k} \exp \left(-\mathrm{i} \omega_{k} t\right) \\
& a_{k}^{*}=\left(\frac{m n_{0}}{\omega_{k}}\right)^{1 / 2} u_{k}^{*} \exp \left(-\mathrm{i} \omega_{k} t\right) \tag{47}
\end{align*}
$$

The energy density of the wave can be expressed very simply in terms of the amplitudes $a_{k}$ and $a_{k}^{*}: \varepsilon_{k}=m n_{0}\left\langle u^{2}\right\rangle=$ $\omega_{k} a_{k} a_{k}^{*}$. The velocity $u$ at the point $x=0$ can also be represented in the form

$$
u_{k}=\left(\frac{\omega_{k}}{2 m n_{0}}\right)^{1 / 2}\left(X_{k} \cos \omega_{k} t+Y_{k} \sin \omega_{k} t\right)
$$

The coefficients $X_{k}$ and $Y_{k}$ are known as the quadrature components because the difference between the phases of the cosine and sine is $\pi / 2$. We can easily show that

$$
X_{k}=\frac{1}{2}\left(a_{k}+a_{k}^{*}\right), \quad Y_{k}=\frac{1}{2 i}\left(a_{k}-a_{k}^{*}\right)
$$

Since the momentum density $P_{k}$ can be expressed in terms of the energy density,

$$
P_{k}=\frac{k}{\omega_{k}} \varepsilon_{k}
$$

the amplitudes $a_{k}$ and $a_{k}^{*}$ yield a simple expression for $P_{k}$ : $P_{k}=k a_{k} a_{k}^{*}$. We can now introduce an auxiliary function $\psi(x, t)$ defined by the relationship

$$
\begin{equation*}
\psi=\left(\frac{m n_{0}}{\omega_{k}}\right)^{1 / 2} u_{k} \exp \left(-\mathrm{i} \omega_{k} t+\mathrm{i} k x\right) \tag{48}
\end{equation*}
$$

> We can easily see that

$$
\begin{equation*}
\varepsilon=\omega_{k}|\psi|^{2}, \quad \boldsymbol{P}=\boldsymbol{k}|\psi|^{2} . \tag{49}
\end{equation*}
$$

In other words, the quantity $|\psi|^{2}$ can be interpreted as a certain arbitrary 'number of waves' per unit volume. The energy and momentum densities are obtained by a simple multiplication of the 'elementary energy' $\omega_{k}$ and of the 'elementary momentum' $\boldsymbol{k}$ by the arbitrary 'density of waves'.

The expressions in Eqn (49) are readily generalised to wave packets when the quantity $|\psi|^{2}$ is localised in a certain relatively wide region of space and vanishes outside this region.

In this case Eqn (49) can be integrated in space, and values of the total energy and total momentum can be found for a wave packet.

The relationships in Eqn (49) can be generalised to any small-amplitude waves because the expression $\boldsymbol{P}=\left(\boldsymbol{k} / \omega_{k}\right) \boldsymbol{\varepsilon}$ is universal. Consequently, the natural frequency $\omega_{k}$ in Eqn (49) need not be equal to $c_{\mathrm{s}} k$, as is true of an acoustic wave, but may be governed by the appropriate dispersion relation for any homogeneous medium.

According to the dispersion relation (39), for each value of the wave vector $\boldsymbol{k}$ we can find two frequencies: $\omega=c_{\mathrm{s}} k$ and $\omega=-c_{\mathrm{s}} k$. Consequently, two waves may exist: one propa-gating along $\boldsymbol{k}$ and the other in the opposite direction. For one of these waves the dispersion relation (provided the root is not multiple) can be written in the form $\omega=\omega_{k}$, where $\omega_{k}$ is the corresponding natural frequency. A superposition of such waves belonging to the same mode of oscillation is described by

$$
\begin{equation*}
\sum_{k}\left(\omega-\omega_{k}\right) \psi_{k}(\boldsymbol{r}, t)=0 \tag{50}
\end{equation*}
$$

Here, $\psi_{k}$ is an elementary plane wave corresponding to the dependence $\psi_{k} \propto \exp (-\mathrm{i} \omega t+\mathrm{i} \boldsymbol{k} \cdot \boldsymbol{r})$. We can easily see that

$$
\omega=\mathrm{i} \frac{\partial}{\partial t}
$$

so that for $\psi(\boldsymbol{k}, t)=\sum_{k} \psi_{k}$ we obtain

$$
\begin{equation*}
\mathrm{i} \frac{\partial \psi}{\partial t}=\sum_{k} \omega_{k} \psi_{k}(\boldsymbol{r}, t) \equiv H \psi \tag{51}
\end{equation*}
$$

The expression on the right-hand side of this equation can be regarded as the result of the action of some operator $H$ on spatial coordinates of a function $\psi$. For example, in the case of an acoustic wave travelling along $x$ the frequency is $\omega_{k}=k c_{\mathrm{s}}$, so that $H=-\mathrm{i} c_{\mathrm{s}} \mathrm{\partial} / \partial x$. For other waves the expression for $H$ may be more complex. We shall consider here only one special case, namely the evolution of the envelope of a wave packet.

We shall temporarily introduce the symbol $\psi^{\prime}$ for the wave function (this is the simplest term for $\psi^{\prime}$ ): $\psi^{\prime}=\sum_{k} \psi_{k}$. Let us now assume that $\psi^{\prime}=\psi(r, t) \exp \left(-i \omega_{0} t+\mathrm{i} k_{0} x\right)$. Here, $\omega_{0}=\omega_{k}\left(k=k_{0}\right)$ is the frequency of a wave with the wave number $k=k_{0}$ and the function $\psi(r, t)$ plays the role of the amplitude, i.e. of the envelope of the wave with given values of $k=k_{0}$ and $\omega=\omega_{0}$. We shall assume that
$\psi(r, t)$ is a smooth function of the variables $r$ and $t$. Then, the expansion of $\psi(\boldsymbol{r}, t)$ into a sum of harmonics of the type $\exp (-\mathrm{i} \nu t+\mathrm{i} \boldsymbol{\kappa} \cdot \boldsymbol{r})$ contains frequencies $v \ll \omega_{0}$ and the wave vectors $\kappa \ll k_{0}$. We can use this to expand $\omega_{k}$ as a Taylor series in $\boldsymbol{\kappa}=\boldsymbol{k}-\boldsymbol{k}_{0}$ :

$$
\begin{equation*}
\omega_{k} \cong \omega_{0}+v_{\mathrm{g}} \kappa_{x}+\frac{1}{2} v_{\mathrm{g}}^{\prime} \kappa_{x}^{2}+\frac{1}{2 k_{0}} v_{\mathrm{g}} \kappa_{\perp}^{2} . \tag{52}
\end{equation*}
$$

Here,

$$
v_{\mathrm{g}}=\frac{\partial \omega_{k}}{\partial k}, \quad v_{\mathrm{g}}^{\prime}=\frac{\partial v_{\mathrm{g}}}{\partial k}=\frac{\partial^{2} \omega_{k}}{\partial k^{2}}
$$

and the term with $\kappa_{\perp}^{2}$ appears because

$$
k=\left(k_{x}^{2}+\kappa_{\perp}^{2}\right)^{1 / 2} \cong k_{x}+\frac{1}{2 k_{0}} \kappa_{\perp}^{2}
$$

If we substitute the above relationships in Eqn (51), then after cancellation of the common phase factor $\exp \left(-\mathrm{i} \omega_{0}+\mathrm{i} k_{0} x\right)$ the result is

$$
\begin{equation*}
\mathrm{i} \frac{\partial \psi}{\partial t}=-\mathrm{i} v_{\mathrm{g}} \frac{\partial \psi}{\partial x}-\frac{1}{2} v_{\mathrm{g}}^{\prime} \frac{\partial^{2} \psi}{\partial x^{2}}+\frac{1}{2 k_{0}} v_{\mathrm{g}} \Delta_{\perp} \psi \tag{53}
\end{equation*}
$$

This equation describes the evolution of the envelope of a wave packet. It is known as the Leontovich parabolic equation. If the spatial dimensions of a wave packet are very large, the last two terms on the right-and side of Eqn (53) can be ignored and then the wave packet simply propagates at a group velocity $v_{g}=\partial \omega_{x} / \partial k$. We note that since Eqn (53) applies to an envelope, the absolute phase of the complex amplitude $\psi$ ceases to have meaning: the replacement of $\psi$ with $\psi \exp (\mathrm{i} \alpha)$, where $\alpha=$ const, leaves Eqn (53) unchanged.

## 11. Correlation function

Under real conditions one frequently encounters the situation in which many waves are excited simultaneously and they have different frequencies and amplitudes. For example, in the case of acoustic waves this is the noise in a large city or of a large number of people at a railway station. A large set of waves can be regarded as wave chaos and can be described by the methods of statistical physics. The elements of such a description can be understood by considering the simplest case of one mode of oscillations in a onedimensional medium. The analysis can be carried out again using a wave function $\psi(x, t)$ because in linear waves all the other physical quantities experiencing small oscillation can be expressed in terms of $\psi$. For example, let us assume that

$$
\begin{equation*}
\psi(x, t)=\sum_{k} a_{k} \exp \left(-\mathrm{i} \omega_{k} t+\mathrm{i} k x\right) \tag{54}
\end{equation*}
$$

where $a_{k}$ is the amplitude of a wave whose wave number is $k$ and $\omega_{k}$ is the corresponding oscillation frequency. The time and space averages of $\psi(x, t)$ obviously vanish. If the number of waves is large and their phases vary with time, their average values $\left\langle a_{k}\right\rangle$ can also be regarded as zero. This is known as the random phase approximation. However, if we take the product $a_{k} a_{k}^{*}$, its average value is independent of the phase, so that we can assume that $\left\langle a_{k} a_{k}^{*}\right\rangle \equiv a_{k}^{2}$, where $a_{k}^{2}$ is a certain average value of the square of the amplitude. The quantity $a_{k}^{2}$, considered as a function of $k$, is known as the spectral function or simply the spectrum.

We shall now find $\left\langle\psi(x, t) \psi^{*}\left(x^{\prime}, t^{\prime}\right)\right\rangle$, where the angular brackets denote the averaging over random phases and amplitudes of the elementary waves. This quantity is known as the correlation function. If we substitute here the sums of Eqn (54) for $\psi$ and $\psi^{*}$ and bear in mind that after averaging the phases only the products $a_{k} a_{k}^{*}$ remain, we find that

$$
\begin{align*}
& \left\langle\psi(x, t) \psi^{*}\left(x^{\prime}, t^{\prime}\right)\right\rangle \\
& \quad=\sum a_{k}^{2} \exp \left[-\mathrm{i} \omega_{k}\left(t-t^{\prime}\right)+\mathrm{i} k\left(x-x^{\prime}\right)\right] . \tag{55}
\end{align*}
$$

This is the familiar expression which makes it possible to calculate the correlation function from the known spectral density.

It should be noted that the average of the square of the modulus of the wave function at a given point $x$ is simply

$$
\begin{equation*}
|\psi|^{2}=\sum_{k} a_{k}^{2} \tag{56}
\end{equation*}
$$

The time dependence of the correlation function on $t-t^{\prime}$ is determined, in accordance with Eqn (55), both by the spectral composition of the waves and by the dispersion law $\omega=\omega_{k}$.

## 12. The Schrodinger equation

Quantum mechanics shows that all microparticles have wave properties. Each free particle corresponds to a wave frequency $\omega$ and a wave vector $\boldsymbol{k}$, so that the energy of this particle is $\omega$ and the momentum is $\boldsymbol{p}=\hbar \boldsymbol{k}$, where $\hbar$ is the Planck constant. For a free particle of mass $m$ the energy is $p^{2} / 2 m=\hbar^{2} k^{2} / 2 m$. In other words, $\varepsilon=\hbar \omega_{k}=\hbar^{2} k^{2} / 2 m$. Consequently, Eqn (51) for a free particle assumes the form of the Schrödinger equation

$$
\begin{equation*}
\mathrm{i} \hbar \frac{\partial \psi}{\partial t}=-\frac{\hbar^{2}}{2 m} \nabla^{2} \psi \tag{57}
\end{equation*}
$$

It should be noted that, like the Leontovich equation (53), the Schrödinger equation is invariant under replacement of $\psi$ with $\psi \exp (\mathrm{i} \alpha)$, where $\alpha$ is a constant phase shift. If a particle is not free but moves in a potential $U(\boldsymbol{r})$, the total energy $\varepsilon$ is equal to the sum of the kinetic and potential energies $\quad \varepsilon=\left(p^{2} / 2 m\right)+U$ and, consequently, the Schrödinger equa-tion can be written in the form

$$
\begin{equation*}
\mathrm{i} \hbar \frac{\partial \psi}{\partial t}=-\frac{\hbar^{2}}{2 m} \nabla^{2} \psi+U(\boldsymbol{r}) \psi \equiv H \psi \tag{58}
\end{equation*}
$$

where $H$ is the Hamiltonian operator.
By analogy with Eqn (51), Eqn (57) may be regarded as applicable to a classical wave field. Then, $|\psi|^{2}$ can be interpreted as the density of identical Bose particles at a given point in space. The application of the Schrodinger equation to one particle requires interpretation of $|\psi|^{2}$ as the probability density. We shall discuss later this approach in more detail.

At this stage we shall consider the equation of free motion (57) for one particle in one dimension only, when $\psi=\psi(x, t)$.

Eqn (57) also resembles the diffusion equation, and the only difference is an additional factor of $i$ on the left-hand side. Therefore, the approach to its solution can be similar to that used in the solution of the parabolic diffusion equation. For example, let us assume that at $t=0$ the wave function is

$$
\begin{equation*}
\psi_{0}(x)=\left(\frac{1}{\pi b^{2}}\right)^{1 / 4} \exp \left(-\frac{x^{2}}{2 b^{2}}\right) \tag{59}
\end{equation*}
$$

This function is normalised to unity, i.e.

$$
\int_{-\infty}^{+\infty} \psi_{0}^{2} \mathrm{~d} x=1 .
$$

It describes a wave packet localised in an interval $\Delta x$ of a scale $b$. Direct substitution can demonstrate that at subsequent moments the solution of the Schrodinger equation becomes

$$
\begin{equation*}
\psi(x, t)=\left(\frac{b^{2}}{\pi}\right)^{1 / 4} \frac{1}{a} \exp \left(-\frac{x^{2}}{2 a^{2}}\right) \tag{60}
\end{equation*}
$$

where $a^{2}=b^{2}-\mathrm{i}(\hbar t / m)$. With the aid of this function we can readily calculate the mean-square value $\left\langle x^{2}\right\rangle$. It is given by the expression

$$
\begin{equation*}
\left\langle x^{2}\right\rangle=b^{2}+\frac{\hbar^{2} t^{2}}{m^{2} b^{2}} \tag{61}
\end{equation*}
$$

This expression corresponds exactly to the familiar uncertainty relationship

$$
\begin{equation*}
\Delta x \Delta p>\hbar . \tag{62}
\end{equation*}
$$

In other words, if the initial localisation $\Delta x$ amounts to $b$, then the velocity uncertainty $\Delta v \propto \hbar / m b$ results in the expansion of the wave packet at a rate $\Delta v$, and at high values of time the localisation $\Delta x$ increases as $\Delta x \propto t(\hbar / m b)$.

The localised perturbation (59) can be confined by a harmonic potential $U=\frac{1}{2} \alpha x^{2}$. We can easily show that the 'elastic' constant is then $\alpha=\hbar^{2} / m b^{4}$ and the energy of an oscillator is $\varepsilon=\frac{1}{2} \hbar \omega_{0}=\frac{1}{2} \hbar(\alpha / m)^{1 / 2}$. In the limit $b \rightarrow \infty$ the elastic constant becomes $\alpha \rightarrow 0$.

## 13. Particle in a thermostat

We shall consider a quantum particle with one degree of freedom in a thermostat which is a box of length $L$ along the $x$ axis. At this stage we shall not be interested in the motion of the particle along the $y$ and $z$ axes. If the walls of this box reflect the particle perfectly, the following eigenfunctions should be selected for this particle:

$$
\begin{equation*}
\psi_{n}=\exp \left(-\mathrm{i} \omega_{n} t\right)\left(\frac{2}{L}\right)^{1 / 2} \sin \frac{\pi n x}{L} . \tag{63}
\end{equation*}
$$

Here, $\omega_{n}=\varepsilon_{n} / \hbar$ is the eigenfrequency, $\varepsilon_{n}=p^{2}{ }_{n} / 2 m=$ $(\pi n \hbar / L)^{2} / 2 m$ is the eigenenergy, and $n=1,2,3, \ldots$ is any natural number. Normalisation of the functions $\psi_{n}$ is selected so that $L\langle | \psi^{2}| \rangle=1$.

If the walls of the box are at a temperature $T=$ const, the particle should reach thermodynamic equilibrium with the walls at the same temperature. This means that, in accordance with the Boltzmann formula or, equivalently, in accordance with the canonical distribution in statistical mechanics, the probability $p_{n}$ of finding a particle in a state $n$ is

$$
\begin{equation*}
p_{n}=Z^{-1} \exp \left(-\beta \varepsilon_{n}\right) \tag{64}
\end{equation*}
$$

Here, $Z$ is a normalisation factor and $\beta=1 / T$. Since $\sum_{n} p_{n}=1$, it follows that

$$
\begin{equation*}
Z=\sum_{n} \exp \left(-\beta \varepsilon_{n}\right) \tag{65}
\end{equation*}
$$

and for this reason the function $Z$ is called the partition function.

Sometimes the free energy $F=-T \ln Z$ is used instead of $Z$ and then Eqn (64) becomes

$$
\begin{equation*}
p_{n}=\exp \left(-\beta \varepsilon_{n}+\beta F\right) . \tag{66}
\end{equation*}
$$

This relationship can be used to find readily the average energy of the particle:

$$
\varepsilon=\langle\varepsilon\rangle=\sum_{n} p_{n} \varepsilon_{n}
$$

and other thermodynamic quantities. The relevant relationships can be simplified by assuming that the dimensions of the system are very large so that the energy of the lower level is $\varepsilon_{i} \ll \mathrm{~T}$. Then the summation over $n$ in Eqn (65) can be replaced by an integral on the assumption that $n$ is a continuous variable. It is now quite easy to find the expression for $Z$ :

$$
\begin{equation*}
Z=\left(\frac{2}{\pi} \frac{L}{b_{0}}\right)^{1 / 2} \tag{67}
\end{equation*}
$$

where

$$
b_{0}=\frac{\hbar}{(m T)^{1 / 2}} .
$$

Similarly, we can obtain the expressions for other physical quantities. In particular, the average energy is

$$
\begin{equation*}
\varepsilon=\langle\varepsilon\rangle=\frac{1}{2} T . \tag{68}
\end{equation*}
$$

The first law of thermodynamics for a one-dimensional particle can be written in the form

$$
\begin{equation*}
\mathrm{d} \varepsilon=-p \mathrm{~d} L+T \mathrm{~d} s \tag{69}
\end{equation*}
$$

Here, $\varepsilon, p$, and $s$ are the energy, pressure, and entropy per particle; $L$ is the length of the box (vessel). Let us assume that, for example, the walls of the box are impermeable to heat, so that $\mathrm{d} s=0$ and the process is adiabatic. All the probabilities $p_{n}$ should then retain their previous values, which is possible only if

$$
\begin{equation*}
T L^{2}=\text { const } \tag{70}
\end{equation*}
$$

This is the adiabatic law for the one-dimensional case. If we assume that $p=T / L$, where $1 / L$ plays the role of the 'particle density', it follows from Eqn (69) that the law of conservation of energy for $\mathrm{d} s=0$ and the relationship (70) are satisfied. The entropy is best defined in accordance with statistical mechanics:

$$
\begin{equation*}
S=-\sum_{n} p_{n} \ln p_{n}=\beta(\varepsilon-F) \tag{71}
\end{equation*}
$$

We have used here the relationship (66). On the other hand, Eqn (71) becomes

$$
\begin{equation*}
F=\varepsilon-T s \tag{72}
\end{equation*}
$$

For a free particle in a very large box we find that

$$
\begin{equation*}
s=\ln \frac{L}{b_{0}}+s_{0}, \quad b_{0}=\frac{\hbar}{(m T)^{1 / 2}}, \tag{73}
\end{equation*}
$$

where

$$
s_{0}=\frac{1}{2} \ln \frac{2}{\pi}
$$

are small numbers which can be ignored compared with $\ln \left(L / b_{0}\right)$. Eqns (68) and (73) obtained above and the
relationship $p=T / L$ are valid only if $L \gg b_{0}$. If the length $L$ is reduced so that it approaches $b_{0}$, it is no longer permissible to change from the summation over $n$ to integra-tion. We can see that on approach to $L \rightarrow b_{0}$ the lower energy level $\varepsilon_{1}$ becomes comparable to $T$. Consequently, at low values of $L$ a particle drops to the lowest energy level and the energy of this particle increases as $L^{-2}$ as $L$ is reduced. The entropy $s$ is zero and the pressure $p$ varies as $L^{-3}$. We seem to reach the 'size' of the particle $b_{0}$ and further deformation of the particle requires a greater energy.

We can therefore broadly say that $b_{0}$ is the characteristic 'size' of the particle at a temperature $T$ and then the ratio $L / b_{0}$ represents the number of cells in which a particle with temperature $T$ can be placed.

The distribution of the probabilities $p_{n}$ can be used to find one further quantity $\rho\left(x, x^{\prime}\right)$, which is called the density matrix in quantum mechanics. The density matrix resembles the correlation function for a random set of classical waves. It is defined by the following expression:

$$
\begin{equation*}
\rho\left(x, x^{\prime}\right)=\left\langle\psi(x) \psi^{*}\left(x^{\prime}\right)\right\rangle=\sum_{n} p_{n} \psi_{n}(x) \psi^{*}\left(x^{\prime}\right) \tag{74}
\end{equation*}
$$

If $L \gg b_{0}$, the function $\rho\left(x, x^{\prime}\right)$ depends only on the difference $x-x^{\prime}$. Therefore, in calculating Eqn (74) we can average over $x$ for a given difference $x-x^{\prime}$. Next, since in accordance with Eqns (66), (72), and (73) the probability is

$$
p_{n}=\left(\frac{\pi b^{2}}{2 L^{2}}\right)^{1 / 2} \exp \left(-\beta \varepsilon_{n}\right)
$$

the density matrix $\rho$ can be calculated quite rapidly. Once again the summation over $n$ can be replaced by integration and then $\rho\left(x, x^{\prime}\right)$ is given by

$$
\begin{equation*}
\rho\left(x, x^{\prime}\right)=\frac{1}{L} \exp \left[-\frac{\left(x-x^{\prime}\right)^{2}}{2 b_{0}^{2}}\right] \tag{75}
\end{equation*}
$$

where $b_{0}^{2}=\hbar^{2} / m T$.
The above expression can be represented in the form

$$
\rho\left(x, x^{\prime}\right)=\frac{\sqrt{2 \pi} b_{0}}{L} \rho_{0}\left(x, x^{\prime}\right)
$$

where

$$
\begin{equation*}
\rho_{0}\left(x, x^{\prime}\right)=\left(\frac{1}{2 \pi b_{0}^{2}}\right)^{1 / 2} \exp \left[-\frac{\left(x-x^{\prime}\right)^{2}}{2 b_{0}^{2}}\right] \tag{76}
\end{equation*}
$$

The function $\rho_{0}\left(x, x^{\prime}\right)$ is normalised to unity, i.e. the integral of $\rho_{0}$ with respect to $x^{\prime}$ is unity. The function $\rho_{0}\left(x, x^{\prime}\right)$ seems to represent the density distribution over $x^{\prime}$ for a fixed value of the point $x$, which in its turn can be placed in one of the cells with the probability $2 \pi b_{0} / L$.

## 14. Wave-particle dualism

There is some similarity between the formal expressions for the density matrix in quantum mechanics and for the correlation function of a random classical wave field. However, actually these physical objects differ strikingly from one another. The wave function in quantum mechanics applies in the simplest case to just one particle. Broadly speaking it is real only where the particle exists and has little meaning outside the particle. We can state this differently. In quantum mechanics all the physical quantities are obtained as a result of the action of some operators on the wave
function. Accordingly the average values of these quantities can be found after assigning to them the weight $|\psi|^{2}$. Hence it is clear that the absolute phase and the absolute amplitude of a wave function have no physical meaning and can be selected for convenience in calculations. Therefore, major relative changes in the amplitude at distant points do not significantly alter the physical quantities if the gradient $\psi$ changes only negligibly. For this reason the $|\psi|^{2}$ function assumes the meaning of a distribution of probabilities and not a distribution of the real density or of the wave motion, as is true of classical fields.

We shall apply this approach to a particle in a thermostat. As demonstrated earlier, the energy of a free particle in such a thermostat is independent of the length $L$ provided $L \gg b_{0}$.

We shall repeat now the same thought experiment on a quantum particle as was done earlier on a classical particle. Specifically, we shall insert an impermeable barrier into our thermostat and thus divide it into two parts. The perturbation of the thermostat by this barrier is slight and the particle still has the energy $\frac{1}{2} T$, provided $L \gg b_{0}$. The particle is then only in one of the halves, but as far as the external world is concerned this does not imply anything: repetition of this experiment still leaves a situation in which $|\psi|^{2}$ is equal to $1 / L$ in either of the two halves.

Let us assume now that we know of a method of identifying which of the halves contains the particle. In quantum mechanics such detection of a particle in one of the states is called 'measurement'. As soon as this measurement is carried out, the value of $|\psi|^{2}$ in the empty half disappears and it is doubled in the half with the particle. The wave function collapses. The entropy of the particle decreases by $\ln 2$, i.e. it decreases by one bit.

If this had not resulted in any change in the external world, we could have constructed a perpetual motion machine of the second kind. Since this is impossible, it follows that our measurement should be accompanied by an irreversible increase in the entropy in the external world by an amount which is at least one bit. We can say that the collapse of the wave function is a real physical event, but it is induced not so much by a device as by an irreversible process in the external world.

One might try to 'capture' the particle in a smaller interval $L_{0} \ll L$. If again we have $L_{0} \gg b_{0}$, the barrier does not alter the energy of the particle and reduces its entropy to $\ln \left(L_{0} / b_{0}\right)$, i.e. by an amount $\ln \left(L / L_{0}\right)$. The entropy of the external world should increase by at least the same amount. We can see that this collapse may occur quite spontaneously provided only $L_{0}>b_{0}$. We can say that the quantity $b_{0}$ represents the effective 'size' of the particle and at distances much greater than $b_{0}$ a quantum particle differs little from a classical one. In other words, in spite of its wave properties, a quantum particle in a large vessel may behave like a classical particle.

We shall now consider this problem from a somewhat different standpoint. The wave function of Eqn (63) can be represented by the sum

$$
a\left[\exp \left(-\mathrm{i} \omega_{n} t+\mathrm{i} k_{n} x\right)-\exp \left(-\mathrm{i} \omega_{n}-\mathrm{i} k_{n} x\right)\right]
$$

where $a$ is a certain amplitude and $k_{n}=\pi n / L$ is the wave number. We can see that the above sum is the superposition of two waves, one of which travels to the right and the other to the left. The wave number $k_{n}$ is proportional to the momentum of the particle: $p_{n}=\hbar k_{n}$. This means that $\psi_{n}$ describes a particle which travels between perfectly reflecting walls and has a constant momentum $p_{n}$. Such motion can
continue indefinitely if the walls are immobile. This state is known as pure in quantum mechanics.

However, in the presence of heat exchange with the walls this pure state cannot be conserved: according to Eqn (64) each such state corresponds to just a certain probability $p_{n}$. If we are dealing not with one particle, but with a very large number (for example, $N \gg 1$ ) of identical bosons, then in each state there may be many particles. In this case the probability $p_{n}$ multiplied by the total number of particles $N$ would correspond simply to the Maxwellian distribution and then the density matrix $\rho\left(x, x^{\prime}\right)$ multiplied by the total number of particles could be considered as a classical correlation function of a random wave field $N^{1 / 2} \psi(x, t)$. In this case we could begin with the natural assumption that the phases of the eigenfunctions $\psi_{n}$ are mutually random, so that averaging of them gives the average correlation function.

When we consider one particle, the behaviour becomes purely probabilistic. At each moment in time the particle can be in only one of its mutually incoherent states: one particle cannot have many momenta at the same time. If we suddenly open one of the covers and allow such a particle to escape far from the containing vessel, we could detect this particle with just one of the possible values of the momentum. In exactly the same way we cannot have two values of the momentum in the case of a classical particle. Therefore, $p_{n}$ represents only the probability of finding a particle in a state $n$. Loss of the mutual coherence of the $\psi_{n}$ functions because of heat exchange with the wall leaves one possibility: the particle remains in one of the mutually incoherent states. There seems to be a hidden 'collapse' of the wave function, but this collapse is not yet real: the external world may still have no information on the level at which the particle is located.

If the particle is acted upon very slowly, for example if the end barrier is moved or new barriers are introduced, the particle may have sufficient time to go over from one level to another and assume on average the Maxwellian distribution. From the point of view of the external world such a particle behaves as a small thermodynamic-equilibrium system. In this case again we are dealing with a density matrix $\rho\left(x, x^{\prime}\right)$, but the changes or actions corresponding to it must be averaged over time intervals much longer than the time needed for the establishment of thermodynamic equilibrium.

If we attempt to localise the particle sufficiently rapidly in space or to find its momentum, the automatic result is the collapse of the probabilties accompanied by the collapse of the wave function. It is this event that cannot take place without an increase in the entropy in the external world. In other words, the collapse is a miniature irreversible process which resembles microscopic birth or death.

This example shows that a real measurement event in quantum mechanics can be represented as a combination of two actions: the preparation of the $\psi$ function for expansion into mutually incoherent components and the collapse into one of these components. The loss of coherence may occur simply because of external noise or thermal motion, but the collapse during measurement requires a real nonequilibrium process which creates information in the measuring instrument and generates at least an equal amount of entropy in the external world. Under real conditions these two components of the measurement may be combined or they may be difficult to distinguish. They are more important for logical clarity than for practical realisation.

We shall continue our discussion of the behaviour of a particle in a box or vessel of length $L$ but we shall now assume
that only one of the ends of the box is maintained at a temperature $T$ and the other end is very cold, i.e. we shall assume that it is practically at absolute zero. Then the particle may carry a heat flux $q$. We can readily estimate its maximum value $q_{\max }$. This value is reached if in the first collision with the warm wall the particle receives a thermal energy $\frac{1}{2} T$. When it arrives at the cold wall the particle transfers to it an energy of the same order and then returns to the warm wall for the next portion of heat. Let us assume that $v_{\mathrm{t}}=(T / m)^{1 / 2}$ is the average thermal velocity. The time of flight to the second wall and return to the first wall cannot be less than $2 L /$ $v_{\mathrm{t}}$. Therefore, the heat flux is limited to $q_{\max }=T v_{\mathrm{t}} / 4 L$. The real heat flux may be much smaller: its value depends on how effectively the particle can exchange heat with the walls in each collision.

We shall now assume that initially the particle is introduced in some localised state of Eqn (59) with an initial localisation width $b$. Moreover, if this particle moves at a velocity $v_{0}=p_{0} / m$, the corresponding wave function can easily be shown to be simply $\psi(x, t)=\psi\left(x-v_{0} t, t\right) \exp (-\mathrm{i} \omega t+\mathrm{i} k x)$, where $\hbar k=p_{0}=m v_{0}$ and $\hbar \omega=p^{2}{ }_{0} / 2 m$. If $v_{0}$ is of the order of the thermal velocity $v_{t}$, the time of flight from one wall to the other is $t \propto L / v_{\mathrm{t}}$. According to Eqn (61) in this time the particle spreads out additionally, but if $b^{2}$ is selected so that the width of this spread conserves the scale of the initial width, i.e. if $b^{2} \sim \hbar t / m \sim \hbar / v_{\mathrm{t}} m$, we can assume approximately that the wave packet retains its width. We can see that $b^{2}=L b_{0}$, where $b^{2}{ }_{0}=\hbar / m T$ determines the minimum width of localisation of the density matrix. If $L \gg b_{0}$, the quantity $b$ considerably exceeds $b_{0}$, but it is $\left(b_{0} /\right.$ $L)^{1 / 2}$ times less than $L$.

Since the system under consideration is far from equilibrium, we cannot exclude the situation in which each collision with a cold wall may be recorded, i.e. may be 'measured' by the wall itself.

This process implies that phonons created by the particle reflected inelastically by the cold wall can induce irreversible strains in the wall or give rise to other excitations of soft modes which are then 'remembered' in one form or another. Irreversible processes inside the wall may automatically 'measure' the particle, i.e. they may induce an additional collapse of the wave function. The quantity $\left(L b_{0}\right)^{1 / 2}$ then determines the minimum size of the wave packet: this size is attained when the 'measurement' occurs in each impact. It is evident that the wave packet can then resemble a classical particle. However, the width of this packet $b \geqslant\left(L b_{0}\right)^{1 / 2}$ is considerably greater than the minimum scale $b_{0}=\hbar(m T)^{1 / 2}$. It follows from the uncertainty relation that $b$ corresponds to the momentum uncertainty $\Delta p \sim \hbar / b$. We can see that each wave packet is broadened over several energy levels, so that $\Delta v / v_{\mathrm{t}} \sim b_{0} / b$. If the localisation width is $b \ll L$, the particle may travel as a nonequilibrium wave packet and, in order to fill the whole permitted range of velocities $\sim v_{\mathrm{t}}$, it has to make many collisions with the walls.

If we begin to increase the temperature of the cold wall so that it approaches the temperature of the warm wall, the degree of nonequilibrium decreases. There is a corresponding broadening of the width of the nonequilibrium packet $b$. In the limit we again have a situation in which the wave function of a particle spreads out over the whole length $L$. If then the interaction with the walls is negligible, the particle will be for a long time at each energy level $\varepsilon_{n}$, undergoing a slow Brownian motion between the levels because of the interaction with the thermal vibrations of the walls.

It thus follows from the above discussion that the waveparticle dualism in quantum mechanics can depend to a great extent on the external conditions under which a particle finds itself.

The apparatus of quantum mechanics makes it possible to describe a wide range of specific physical situations. In particular, the description of the thermal motion of the particle in the form of moving packets may be provided on the basis of the density matrix $\rho\left(x, x^{\prime}\right)$, but this matrix will differ from the thermodynamic density matrix of Eqn (74), because for each wave packet the phases of the neighbouring harmonics are correlated with one another, i.e. each such packet looks like a superposition $\sum_{n} c_{n} \psi_{n}$. If the averaging of the density matrix in Eqn (74) is carried out, the result is the expression

$$
\begin{equation*}
\rho\left(x, x^{\prime}\right)=\sum_{n, m} W_{n m} \psi_{n} \psi_{m}^{*}, \tag{77}
\end{equation*}
$$

in which the off-diagonal matrix elements $W_{n m}$ need not vanish. Moreover, if $W_{n m}=c_{n} c_{m}^{*}$, we obtain simply the unique pure state which corresponds to the selected wave packet.

If again we draw an analogy with the classical field, we can say that in the pure state the phases of the individual waves are correlated with one another, i.e. they are not fully chaotic. The transition to thermal equilibrium is accompanied by a change to chaotic phases and by the loss of coherence. A particle can then be only in one of two mutually incoherent states. Therefore, when the phases are completely chaotic the off-diagonal terms in Eqn (77) disappear and we have the usual definition of the equilibrium density matrix with the Boltzmann energy distribution of the probabilities $W_{n n}$.

## 15. Radioactive decay

Our idealised model of a one-dimensional thermostat will be used to consider a process which resembles the radioactive decay of a nucleus. In a thermostat of length $L$ we place a minature trap for a particle by introducing an additional barrier at a distance $b$ from one of the ends, for example from the end on the left. Let us assume that $b \ll b_{0}$, where $b_{0}=\hbar(m T)^{1 / 2}$ is the width the minimum localisation of the particle whose energy is $\sim T$. The first (lowest) energy level in this small trap, $\varepsilon_{\text {min }}^{*}=\pi^{2} \hbar^{2} / 2 m b^{2}$, is much higher than the temperature. Therefore, the temperature of the left-hand end of the thermostat plays no role: the particle is already at a high level $\varepsilon_{\text {min }}^{*}$ and the next level may be so high that the probability of the particle reaching it is negligible.

Let us assume that our particle is in a pure state of energy $\varepsilon_{\text {min }}^{*}$ in a small trap of size $b \ll L$. Let us now make a small aperture in the barrier so that the wave function of the particle can leak through it slowly into the main part of the thermostat. Naturally, instead of such an aperture we can employ an energy barrier, which is analogous to a barrier through which a wave function can leak in a process known as tunnelling, which is subbarrier leakage of the particle to the free region.

This process is exactly analogous to the $\alpha$ decay of a heavy nucleus. Since for $L \gg b$ the wave function in the free region is always small compared with the $\psi$ function in a small section, the wave function of this small section decreases exponentially with time: $\psi_{b} \propto \exp (-t / 2 \tau),\left|\psi_{b}\right|^{2} \propto \exp (-t / \tau)$. If there
are very many particles in a small section, for example if there are initially $N_{0}$ particles, their number decreases with time as $N_{0} \exp (-t / \tau)$. However, we have only one particle so that we have to speak of probabilities.

It is quite obvious that the probability of finding a particle in a small section is $p_{1}=\exp (-t / \tau)$, and the probability of finding the particle in the main thermostat is $p_{2}=$ $1-\exp (-t / \tau)$. For a classical particle this is exactly how an irreversible process of penetration of this particle across a 'slimmed down' barrier would take place. The spatial part of the entropy of Eqn (28) for a particle in a thermostat would have been considerably greater than unity and a monotonic rise of the probability $p_{2}$ would have been automatically accompanied by a monotonic rise of the entropy (the entropy of the small section can be assumed to be zero). However, a quantum particle behaves in a somewhat more complex manner.

Let us assume initially that the right-hand end of the thermostat represents a pure barrier at absolute zero. Then it would seem that the probabilities $p_{1}$ and $p_{2}$ should apply to the small section and to the thermostat, and that their values should be described by the relationships given above. However, we now have a contradiction because the entropy $S=-\sum_{i} p_{i} \ln p_{i}$ should increase first, reaching its maximum at $p_{1}=p_{2}=1 / 2$, and then should fall to zero at $p_{1} \rightarrow 0$ and $p_{2} \rightarrow 1$. In fact, the state under discussion, composed of a superposition of the wave functions in the small section and in the main part, is pure if there is no external disturbance. Before measurement this is the only state. Only after repeated measurements can we obtain the values of the probabilities of finding the particle in the small section $p_{1}$ or in the large part $p_{2}$.

The state is pure before measurements even if it does not have a fixed energy. For example, if the decay is very slow, then at some moment in the middle of the process one can increase adiabatically slowly the length $L$ by shifting the more distant end barrier. The energy of that part of the wave function which is outside the small section (trap) then decreases strongly compared with $\varepsilon_{\min }^{*}$. The $\psi$ function leaking out from this small section has still the energy $\varepsilon_{\text {min }}^{*}$. Therefore, the $\psi$ function would form as a superposition of two states with different energies.

Before the measurements the superposition is again a pure state. An external force of frequency corresponding to the difference between the energy levels can freely change the ratio of the amplitudes of the two sublevels. Therefore, before the measurements it is not possible to detect the radiation decay.

However, if it is assumed that the right-hand end area is not perfectly reflecting, the first collision of the particle with this barrier shows that the particle has escaped somewhat earlier from a small initial volume. Such a collision, accompanied by the excitation of phonons in the end barrier, serves as a measurement after which the $\psi$ function of the small volume practically disappears and the $\psi$ function in the main part of the thermostat is normalised to unity. In other words, an inelastic collision with the wall causes the collapse of probabilities (which up to that moment exist as possibilities and not as real numbers), namely, the result is that $p_{1} \rightarrow 0$ and $p_{2} \rightarrow 1$. Only multiple repetition of the process of decay can give the dependence $N=N_{0} \exp (-t /$ $\tau$ ), on the basis of which we can gain an idea about the evolution of the averages in an ensemble of the a priori probabilities $p_{1}$ and $p_{2}$.

After the first collision the particle is fixed inside the thermostat because it is 'measured' together with the approximate path from the 'decay' point and then the process of establishing equilibrium with a thermostat begins. The particle exchanges phonons with the wall, its average energy decreases, and the distribution between the levels approaches the Maxwellian form. Finally, the particle energy reaches its average thermal value $\varepsilon=\frac{1}{2} T$ and the energy residue $\varepsilon_{\min }^{*}-\frac{1}{2} T$ is transferred to the wall in the form of a portion of heat $\Delta Q=\varepsilon_{\min }^{*}-\frac{1}{2} T$. The internal entropy of the particle then reaches the value given by Eqn (73), i.e. it becomes $s \cong \ln \left(L / b_{0}\right)$, and the entropy of the thermostat rises by an amount $\Delta S=\Delta Q / T=\left(\varepsilon_{\min }^{*} / T\right)-\frac{1}{2}$. All this second stage of the process is essentially irreversible and is accompanied by an entropy increase.

However, we shall now return to the first stage, ie. to the first collision. Let us assume that $\psi_{b}$ is the normalised wave function in the small section and $\psi_{L}$ is the normalised wave function in the free volume of the cold thermostat. These two functions correspond to the same energy when $\varepsilon \rightarrow \infty$. For finite values of $\varepsilon$ the wave function can be represented by the superposition

$$
\begin{equation*}
\psi=c_{b} \psi_{b}+c_{L} \psi_{L} \tag{78}
\end{equation*}
$$

where $c_{b}$ and $c_{L}$ are the corresponding amplitudes. It follows from the rules of quantum mechanics that the quantities $\left|c_{b}\right|^{2}$ and $\left|c_{L}\right|^{2}$ correspond to the probabilities of the relevant states when repreated measurements are carried out. Each of these measurements induces the collapse of the function into either $\psi_{b}$ or $\psi_{L}$. In the case under discussion the role of the measuring instrument is played by the second end of the large volume and the measurement itself induces the $\psi \rightarrow \psi_{L}$ collapse. According to quantum mechanics, the measuring instrument cannot be used to induce deliberately this collapse. It simply awaits the moment when the collapse takes place and after many measurements makes it possible to find $p_{2}(t)=1-\exp (-t / \tau)$. The collapse itself is essentially random. Einstein said that the process looks as if 'God played dice". He assumed that a deeper meaning lies behind this randomness. However, since this meaning has not been found, we are left with the orthodox quantummechanical point of view.

We can thus conclude that our process of relaxation of an equilibrium state begins with the first inelastic collision of a particle with the thermostat wall. In quantum mechanics this collision represents a random collapse of the wave function: $\psi \rightarrow \psi_{L}$. One pure state of the particle is transformed to another pure state, but an irreversible process of emission of a phonon takes place in the wall and the phonon escapes to the external world. A memory of this phonon may remain in the form of one bit of information concealed somewhere deeply in the wall. In principle, if some automaton is placed alongside the wall, this bit of information can be transformed into the subsequent action of the automaton in accordance with an algorithm provided in advance. For example, the automaton could use part of the 'decay' energy to do work.

We shall carry out one more thought experiment. We shall assume that a narrow trap of width $b \ll b_{0}$ is located not at the end, but exactly in the middle of a thermostat of length $L$. We shall also postulate that the trap has two small apertures so that the particle may equally well escape to the left- or right-hand half-volumes.

The process of relaxation again begins from the first collision. This may now be the collision with the left-hand or right-hand wall. This first collision leaves one bit of informa-tion in the relevant wall. This is followed by a process of relaxation which again creates an entropy $\Delta S=\Delta Q / T=\left(\varepsilon_{\min }^{*} / T\right)-\frac{1}{2}$ in the appropriate wall. The entropy of the particle is then equal to $s \cong \ln \left(L / 2 b_{0}\right)=\ln \left(L / b_{0}\right)-\ln 2$, because the region occupied by the particle is equal to only half the length $L$. The entropy of the particle is now one bit less. Then the displacement of the central barrier in the direction of the empty volume could carry out work. However, we then have to know in which half the particle is actually located and this requires the use of the same bit of information that has escaped into the wall in the first collision. Therefore, once again we cannot utilise the thermal energy of the particle without 'payment'.

The 'radioactive' decay considered by us is thus an irreversible process of relaxation to thermodynamic equilibrium. It begins with an irreversible collapse of the wave function which creates entropy in the external world. This is followed by thermal relaxation which increases the entropy of the particle and the entropy of the external world. In our example the energy is stored in the particle itself and, therefore, dissipation does not require an additional energy or the additional introduction of a negative entropy from outside the system.

## 16. Schrodinger's cat

The wave function $\psi$ thus has the following meaning. In the case of many identical particles the quantity $|\psi|^{2}$ is proportional to the density of particles at a given point in space. In the limit of one particle the quantity $|\psi|^{2}$ becomes the probability of finding this particle at a given point. The wave function evolves in accordance with the linear Schrödinger equation, which admits the superposition of linear solutions, i.e. the superposition of different states. In this superposition the functions $\psi$ themselves and not their squares are used. Therefore, the usual law for the additional probability applies only if the wave functions of different states do not overlap in space or if they have mutually uncorrelated phases. Such decorrelation may appear as a result of the interaction of a particle with the external environment. If this environment fixes one of the states of the particle, the wave function seems to collapse.

The external world is assumed to be classical, i.e. macroscopic bodies are postulated to be not wave but point (or extended) objects with fixed coordinates that vary with time in accordance with the laws of classical mechanics. However, this approach means that there is a yawning gap between the wave microworld and the classical macroworld: in the usual approach it is not possible to go over from one world to the other. In fact, when the approach is made from the microworld side, it is necessary to include, in the Schrödinger equation description of many particles, objects of increasing size including in the limit the whole Universe. On the other hand, when the approach is from the classical macroworld side, it is natural to extend the classical description right down to the smallest particles and the smallest dimensions. The boundary between the microworld and the macroworld is indeterminate and its discussion has raised and continues to raise many problems.

This contradiction has been stated most clearly by Schrödinger [26] in his famous thought experiment with a cat. Let us assume that there is a Geiger counter which records $\alpha$ particles originating from $\alpha$ decay of nuclei and passing through this counter. This counter is in contact with a device which can break an ampoule with potassium cyanide when the particle crosses the counter. If the ampoule and a live cat are placed together under a glass jar, it follows from standard quantum mechanics that there is the possibility of a superposition of live and dead cats if the counter has not made its measurement and is in a state of superposition of recorded and unrecorded particle flights.

Obviously there is no real paradox. The actual recording of a flight of a particle is an irreversible process accopanied by the collapse of the wave function and only this can be followed by the operation of an automaton which breaks the ampoule. Superposition of this state with the process of detection of a particle which has not taken place is simply impossible. However, orthodox quantum mechanics does not cover the process of irreversible measurements and the feasibility of unreal superposition shows that there is an intermediate region between the quantum microworld and the classical macroworld where standard quantum mechanics does not apply and which requires a special discussion. We shall return to this topic later.

## 17. Irreversibility of our environment

Not only life on Earth but other nonequilibrium processes on our planet depend to a large extent on solar radiation. The solar constant, which is the flux of energy at the average distance of the Earth from the Sun, amounts to $1.4 \times$ $10^{6} \mathrm{erg} \mathrm{cm}^{-2} \mathrm{~s}^{-2}$. If we ignore albedo, we find that this energy is eventually converted to heat at a temperature $T$ of about 300 K . We can easily estimate the corresponding rate of entropy per unit area per unit time $\dot{S}_{0} \cong 3 \times$ $10^{19} \mathrm{~cm}^{-2} \mathrm{~s}^{-1}$. Since the temperature of the Sun is about $6 \times 10^{3} \mathrm{~K}$-i.e. about 20 times higher than the temperature of the Earth - and since solar radiation is contained within a very small solid angle, we can readily estimate that the proportion of the solar entropy in $\dot{S}_{0}$ represents no more than $1 \%$. In other words, the solar energy has a very high degree of order and before it is converted into heat it carries an information flux $\dot{I}_{0}$ of the same order of magnitude as $\dot{S}_{0}$. The information flux $\dot{I}_{0}$ has a giant value of $\sim 4 \times$ $10^{19} \mathrm{bit} \mathrm{cm}^{-2} \mathrm{~s}^{-1}$. It is incomparably greater than any one of the information fluxes deliberately created by man.

Naturally, this information flux should be divided by the number of molecules to which it is transferred and which gives a much more modest value per molecule. For example, if the flux $\dot{I}$ is divided by the number of molecules in the terrestrial atmosphere, in $1 \mathrm{~cm}^{2}$ of the Earth's surface, then the information flux per molecule does not exceed one bit per week. However, even this number is not so small. We must also bear in mind that not all the energy is converted into heat. Some of the solar energy is captured by the atmosphere and serves as the source of air currents, winds, clouds, precipitation, etc. Another part of this energy is used in the photosynthesis of proteins by the plant kingdom and the proteins in turn serve as the food and life support of the animal world. Therefore, in addition to degradation of the ordered solar energy and its conversion into heat, there is a simultaneous process of self-organisation and complexification of the structures in the world surrounding us. All the
growth processes occur only because they are accompanied by a powerful increase in the entropy due to a reduction in the degree of ordering of the energy arriving from the Sun.

Taken as a whole the atmosphere and biosphere of the Earth represent a complex open system. A steady-state balance of the energy fluxes is established by the emission of thermal radiation from the Earth into space: the energy arrives on the Earth with a low entropy and escapes with a much higher entropy. However, we cannot say that the irreversible process of the entropy increase is perfectly monotonic in all the components of our complex system. On the contrary, the overall entropy increase is accompanied by the process of creation of ordered structures and a reduction in the local values of the entropy. It is this global entropy rise that makes possible the opposite process of local organisation and development of order. The situation is similar to an irrigation system in which mechanisms are used to raise water from a lower to a higher level: the water falling back sets in action a water wheel which transports part of the water upwards. The large flux downwards creates a small flux upwards.

Stephen Hawking, a major physicist of our time, uses the following lecture demonstration of irreversibility. He shows a small episode on cine film where a cup with coffee slides down from a table, falls on the floor, and breaks; the coffee spreads over the floor. He comments on this episode as follows: '"all of us are familiar with this scenario'. He then adds that "the reverse process is simply impossible" and shows the same event in the reverse order: a pool of coffee returns back to the cup, which is restored from the fragments, and then the cup with coffee jumps and comes to rest on the surface of the table. Naturally, nobody believes in the possibility of such a miracle although perhaps not all of us think what happens to the entropy.

After all, the whole episode begins from the point at which a cup with coffee stands on a table. A million years ago nothing has indicated that such a cup would appear as a result of a purely irreversible process accompanied by continuous degradation of the global order and a monotonic increase in the total entropy in the process of thermalisation of the energy flux arriving from the Sun. It is this general degradation of energy that has made possible the opposite process, again irreversible, involving a reduction in the local entropy and the creation of highly organised matter. It is this process that has finally resulted in the appearance of the cup with coffee on the table.

We can say that not only in a general open system, the passage of ordered energy through which is accompanied by entropy creation, but also in separate open parts of this system there are two simultaneous processes: degradation leading to chaotic thermal motion of molecules and selforganisation accompanied by an increase in the complexity of structures and the growth of the associated portion of information (i.e. the entropy decreases).

A full understanding of the irreversibility must simultaneously allow for both processes: self-organisation with a slight reduction in the entropy and degradation of the order with energy thermalisation and entropy creation at a much higher rate. Both destruction and ordering are essential for the understanding of irreversibility.

## 18. Who throws the dice?

The probabilistic description of events in quantum theory raises the natural question of whether a more complex deterministic pattern of physical processes lies behind the random events and we are simply unable to describe this pattern because of its complexity. Attempts to introduce into quantum theory so-called hidden variables have been made on many occasions. However, it has been shown convincingly - beginning with the work of von Neumann, and then later - that the logical structure of quantum theory is closed and excludes the possibility of introducing hidden variables. (The reasons given by von Neumann himself are not quite correct, as pointed out for example by Mermin [34], but this does not change the situation.)

What is the origin of randomness? Who in the final analysis throws the dice? In order to understand what it is all about, we must consider first the game of dice itself. Two players in turn throw a cube with six faces bearing numbers from one to six. This is indicated usually by the number of dots on the faces. After throwing the dice cube the players determine - in accordance with rules established in advance - who won. Usually each of the faces appears perfectly randomly with a probability of $\frac{1}{6}$. However, why does this happen at random?

It happens because before the cube comes to rest it turns over many times. Let us assume that, for example, the cube makes 100 such revolutions. The same face appears for the second time if the cube is thrown with a precision of at least $1 / 600$. Here the number 6 allows for the number of faces. It is practically impossible to repeat the throw of the cube with this precision. Moreover, if the throw of the cube produces repeatedly the same result, the second player may suspect the first of cheating. It follows that in this case the random behaviour is due to the hands of the players. The fact that the numbers appear at random is more due to the inability of the players to coordinate precisely their movements than to their conscious behaviour.

Instead of throwing dice one could consider a device ensuring a perfectly chaotic motion of a cube in a box closed to the players. The players would open the top of the box from time to time and read the result. In this case we are dealing with an objectively occurring random process. The motion of the cube in the box can be regarded as an analogue of chaotic thermal motion. Consequently, the cube in the closed box has the entropy $S=\ln 6$. As soon as the box is opened and both players look at the cube the entropy of the cube collapses to zero, because the cube face is now fully determined with the probability of unity. Each of the players acquires the information $I=\ln 6$, which may then be followed by the following events: one of the players may become pale and the other may smile with pleasure. The information is received when each of the players destroys without fail some of his ordered structure and transfers to chaos, i.e. to the entropy, at least $I=\ln 6$ of his own information.

We shall now assume that our players have decided to replace the cube with a Brownian particle so as to exclude completely any possibility of regular motion. It is unimportant which position the players decide to regard as the winning position and we shall simply assume that they periodically measure the position of the particle and then return it to the origin of the coordinate system. Let us assume that a dust particle is moving in air in a container of $1 \mathrm{~cm}^{3}$ in volume. The
motion of this particle is completely independent of the observers and is absolutely chaotic. It is here that we might say that "God plays dice". However, let us not be in haste.

We shall now consider to what extent this motion is random. One cubic centimetre of air contains something of the order of $N \approx 10^{19}$ molecules. Consequently, the volume per molecule is $V_{0} \approx 10^{-19} \mathrm{~cm}^{3}$. If we want to localise each of the molecules in volume not smaller than $V_{0}$, it follows from Eqn (29) that the configuration part of the entropy of the gas is a quantity which is at least $N \approx 10^{19}$. We shall now assume that we want to localise, i.e. 'freeze', this state. This state then has the information $I \approx S \approx 10^{19}$. We also desire to control this complex system by correcting it at time intervals $\tau \approx l / c_{\mathrm{s}} \approx 10^{-11} \mathrm{~s}$. Here, $l \approx 10^{-6} \mathrm{~m}$ is the average distance between molecules and $c_{\mathrm{s}}=300 \mathrm{~m} \mathrm{~s}^{-1}$ is the velocity of sound. We can see that the control of the motion of the gas requires an information flux, converted into the entropy, of the order of $\sim 10^{30} \mathrm{~s}^{-1}$. This is $10^{10}$ times greater than can be supplied by the solar energy flux reaching $1 \mathrm{~cm}^{2}$. In other words, if somebody would want to help one of the players, he would have to posses a spiritual potential capable of maintaining an ordered motion of molecules by changing into chaotic the flux of information on the scale arriving from the Sun on a surface of $1 \mathrm{~cm}^{2}$.

It is therefore more correct to say that God does not play dice but the reverse is true: because of complete neutrality of God to this game, it occurs on the basis of pure randomness created by the spontaneous nature.

The randomness may be regarded as an integral property of matter which is in thermal motion. Observations of a random process reveal chaotic successive numerical values of a random quantity. Therefore, the most rigorous approach to the description of this process is based on the concepts of probability, or distribution functions of probabilities, if the random quantity can be regarded as continuous. If random events are considered, then in the presence of information links they should be analysed allowing simultaneously for the complex events which concern the observer or, more generally, occur in the external world.

## 19. Brownian motion

We shall now consider the motion of a Brownian particle in greater detail. We shall begin with the one-dimensional motion of a classical particle. The most convenient approach to the description of the motion of such a particle is based on the Langevin equation

$$
\begin{equation*}
\frac{\mathrm{d} v}{\mathrm{~d} t}+\gamma v=\frac{F}{m} \tag{79}
\end{equation*}
$$

Here, $v$ is the velocity of the particle, $\gamma$ is the coefficient of friction, $m$ is the mass of the particle, and $F$ is a random force created by random collisions of molecules with the particle.

Let us assume that at $t=0$ the velocity of the particle is zero. It then follows from Eqn (79) that

$$
\begin{equation*}
v=\frac{1}{m} \int_{0}^{t} \exp \left[-\gamma\left(t-t^{\prime}\right)\right] F\left(t^{\prime}\right) \mathrm{d} t^{\prime} \tag{80}
\end{equation*}
$$

If the average value of the force is $\langle F\rangle=0$, the average velocity is also zero. However, the mean-square value differs from zero and is
$\left\langle v^{2}\right\rangle=\frac{1}{m^{2}} \int_{0}^{t} \int_{0}^{t} \exp \left[-2 \gamma t+\gamma\left(t^{\prime}+t^{\prime \prime}\right)\right]\left\langle F\left(t^{\prime}\right) F\left(t^{\prime \prime}\right)\right\rangle \mathrm{d} t^{\prime} \mathrm{d} t^{\prime \prime}$.

Since the collisions of molecules occur very frequently, the force $F$ has a very small time-correlation scale, so that we can assume approximately that $\left\langle F\left(t^{\prime}\right) F\left(t^{\prime \prime}\right)\right\rangle=$ $F^{2} \tau \delta\left(t^{\prime}-t^{\prime \prime}\right)$. Here, $F^{2}$ is a characteristic scale of the square of the force and $\tau$ is the characteristic correlation time. The expression (81) becomes

$$
\begin{equation*}
\left\langle v^{2}\right\rangle=\frac{F^{2} \tau}{2 m^{2} \gamma}[1-\exp (-2 \gamma t)] \tag{82}
\end{equation*}
$$

We can see that at low values of $t$ the mean-square value $\left\langle v^{2}\right\rangle$ increases with time as $\left\langle v^{2}\right\rangle=F^{2} \tau m^{-2} t$ and, in the limit $t \rightarrow \infty$, tends to the constant value $F^{2} \tau / 2 m^{2} \gamma$. Under thermal equilibrium conditions this limiting value is $T / m$, so that we obtain

$$
\begin{equation*}
D_{v} \equiv \frac{F^{2} \tau}{2 m^{2}}=\frac{\gamma T}{m} . \tag{83}
\end{equation*}
$$

Here, $D_{v}$ denotes the velocity diffusion coefficient. According to Eqn (82) at low values of $t$ we have $\left\langle v^{2}\right\rangle=2 D_{v} t$. We shall also introduce the velocity distribution function $f(v, t)$. If there is no random function, then all the velocities decrease as $v_{0} \exp (-\gamma t)$. In this case the distribution function would vary in accordance with the law $f(\nu, t)=\exp (\gamma t) f_{0}[\nu \exp (\gamma t)]$ where $f_{0}\left(v_{0}\right)$ is the velocity distribution at $t=0$.

We can easily see that then

$$
\frac{\partial f}{\partial t}=\gamma \frac{\partial}{\partial v}(v f)
$$

On the other hand, if $\gamma$ vanishes and there is a random force, velocity diffusion occurs. In fact, we then have
$f(v, t+\Delta t)=f(v-\Delta v, t) \cong f(v, t)-\frac{\partial f}{\partial v} \Delta v+\frac{1}{2} \frac{\partial^{2} f}{\partial v^{2}}(\Delta v)^{2}$,
where $\Delta v$ can be found from Eqn (80). The averaging procedure gives $\langle(\mathrm{D} v)\rangle^{2}=2 D_{v} \mathrm{D} t$. Consequently, expanding the left-hand side in $\Delta t$ and allowing $\Delta t$ to approach zero, we obtain the diffusion equation. Together with the term containing $\gamma$ it becomes

$$
\begin{equation*}
\frac{\partial f}{\partial t}=\gamma \frac{\partial}{\partial v} v f+D_{v} \frac{\partial^{2} f}{\partial v^{2}} . \tag{84}
\end{equation*}
$$

This is known as the Fokker-Planck equation. After a long time a steady-state distribution is attained and it follows from Eqn (83) that this is the Maxwellian distribution

$$
\begin{equation*}
f_{m}=\left(\frac{m}{2 \pi T}\right)^{1 / 2} \exp \left(-\frac{m v^{2}}{2 T}\right) \tag{85}
\end{equation*}
$$

The time needed for the establishment of the Maxwellian distribution is usually very short.

If the time intervals are regarded as much greater than 1/ $\gamma$, the motion of a particle along the $x$ coordinate can be regarded as diffusion. In fact, we can find the quantity

$$
\begin{equation*}
\left\langle x^{2}\right\rangle=\int_{0}^{t} \int_{0}^{t}\left\langle v\left(t^{\prime}\right) v\left(t^{\prime \prime}\right)\right\rangle \mathrm{d} t^{\prime} \mathrm{d} t^{\prime \prime} . \tag{86}
\end{equation*}
$$

Since $v$ is described in Eqn (80) in terms of an integral of a random force, and since the force is regarded as $\delta$-correlated,
it follows that

$$
\left\langle v\left(t^{\prime}\right) v\left(t^{\prime \prime}\right)\right\rangle=\frac{T}{m} \exp \left[-\gamma\left(t^{\prime}-t^{\prime \prime}\right)\right]
$$

and Eqn (86) leads to the relationship $\left\langle x^{2}\right\rangle=2 D t$, where the diffusion coefficient is $D=2 T / a \gamma m$. Consequently, by analogy with the discussion leading to Eqn (84), we can readily obtain the equation for the probability of the distribution $p(x, t)$ of a particle along the $x$ coordinate:

$$
\begin{equation*}
\frac{\partial p}{\partial t}=D \frac{\partial^{2} p}{\partial x^{2}} \tag{87}
\end{equation*}
$$

The description of the evolution of the particle velocity thus leads to the Fokker-Planck equation and the diffusion of a particle with a steady-state velocity distribution is described by Eqn (87) if $\gamma t \gg 1$.

If initially the particle is localised with respect to its velocity and coordinates, and its entropy is very small, the entropy increases monotonically with time. In the velocity distributions this occurs because of the approach to the Maxwellian distribution, whereas the configuration component of the entropy increases because the particle occupies an increasing interval along $x$ by a diffusion process.

We shall now consider the Brownian motion of a quantum particle. The numerical aspect of the problem will not be stressed here and we shall therefore consider only an extremely simplified variant of such motion. We shall discuss the example, considered qualitatively in Section 14, of a particle in a finite one-dimensional box with walls which are at different temperatures. In this way we shall directly allow for the occurrence of a nonequilibrium process.

Let us assume that a particle of mass $m$ is inside a box of length $L$ and one end of this box is at a temperature $T_{1}$ and the other is at a lower temperature $T_{2}<T_{1}$. Temperature $T_{2}$ of the cooler is always regarded as lower than the temperature $T_{1}$ of the heater. Only one-dimensional motion will be considered.

Let the particle be initially localised in the interval $b \ll L$. We can use the classical approach during a certain time interval before the wave packet spreads out. Let $v$ be the velocity of the particle. This velocity changes randomly owing to collisions with the walls, so that after a certain time the changes in the velocity can be described by the Langevin equation. The motion along the $x$ coordinate is of little interest to us: it simply represents free flight from one wall to the other and back again. The time of flight between the walls is $L / v$, so that the frequency of collisions with each of the walls is $v / 2 L$.

Let us assume that each collision with the wall results in the loss of some of the momentum by the particle, so that its velocity decreases by $\alpha \nu_{\mathrm{t}}$, where $v_{\mathrm{t}}=\left(T_{1} / m\right)^{1 / 2}$ and the coefficient $\alpha$ is simply a number smaller than unity. Then the loss of the momentum in collision with one of the walls can be described as uniform deceleration at the rate

$$
-\frac{\gamma}{2} v \cong \frac{\alpha v_{T}}{2 L} v .
$$

If we assume approximately that the loss of momentum on the hot and cold walls does not differ greatly from the above value, then the losses on both walls are described by $\gamma=$ $\alpha \nu_{\mathrm{t}} / L$. There is a well-known fluctuation-dissipation theorem which shows that the mechanism responsible for the dissipation also creates fluctuations. In our case this is the heating of the particle by diffusion. Eqn (83) allows us to
write down the Fokker-Planck equation in the previous form (84), but for $T_{2} \neq T_{1}$ we have different diffusion coefficients for the left-hand and right-hand walls, so that on average

$$
\begin{equation*}
D_{v}=\frac{\gamma}{2 m}\left(T_{1}+T_{2}\right) . \tag{88}
\end{equation*}
$$

We can easily see that the velocity distribution of the particle approaches the Maxwellian form with the average temperature $T_{0}=\frac{1}{2}\left(T_{1}+T_{2}\right)$.

We shall now discuss the question of how long we can use the classical approach. We can resolve this question if we know how fast a wave packet spreads with time. We shall consider first the fate of the temperature of the particle before it has reached its equilibrium value. We shall multiply Eqn (84) by $m v^{2}$ and integrate it with respect to velocity. If we assume that the distribution over the velocity is close to Maxwellian, we obtain

$$
\begin{equation*}
\frac{\partial T}{\partial t}=-\gamma\left(T-T_{1}\right)-\gamma\left(T-T_{2}\right) \tag{89}
\end{equation*}
$$

We have used here the relationship (88) for $D_{v}$.
According to Eqn (89) the particle tends to equilibrium with each of the walls at a rate $\sim \gamma$. Under equilibrium conditions, i.e. when $T=\frac{1}{2}\left(T_{2}+T_{1}\right)$, the particle receives heat from the hot wall at a rate of $q=\frac{1}{2} \gamma\left(T_{1}-T_{2}\right)$ and transfers it at the same rate to the cold wall. In other words, $q$ is the heat flux carried by the particle. This heat transfer delivers to the cold wall an entropy flux

$$
\begin{equation*}
\dot{S}_{\mathrm{e}}=\frac{q}{T_{2}}=\frac{\gamma}{2}\left(\frac{T_{1}}{T_{2}}-1\right) \tag{90}
\end{equation*}
$$

This entropy flux may induce an irreversible process in the wall, such as the storage of information that the particle has collided with the wall. We can say that this entropy increase is evidence of a kind of 'measurement' of the particle. It can be regarded as a process of destruction of mutually coherent parts of a wave packet. Broadly speaking, if $\Delta S_{\mathrm{e}} \approx 1$, a wave packet may split into two mutually incoherent halves. This results in the collapse of the wave function: one of its halves vanishes and this increases the stored information, i.e. it increases the 'knowledge' at the cold wall with respect to the packet, but such knowledge is at the expense of an increase in the wall entropy by $\Delta S_{\mathrm{e}} \approx 1$. If this process is repeated many times, the wave packet retains on average a localised state. The width of localisation of the packet can be estimated from Eqn (61), which describes the spreading of the packet with time if its initial localisation is $\left\langle x^{2}\right\rangle=b^{2}+\left(\hbar^{2} t^{2} / m^{2} b^{2}\right)$. If the packet is 'pressed' against the cold wall in successive collisions, the spreading width should be of the same order as the initial width of the packet. In other words, if $t \approx 1 / \dot{S}_{\mathrm{e}}$, then $b^{2} \approx^{2} / m^{2} b^{2} \dot{S}_{\mathrm{e}}^{2}$. Hence, we find the localisation width

$$
\begin{equation*}
b \cong\left(\frac{\hbar}{m \dot{S_{\mathrm{e}}}}\right)^{1 / 2} \tag{91}
\end{equation*}
$$

It follows from Eqn (90) that at the temperature $T_{1}$ much higher than $T_{2}$, the value of $\dot{S}_{\mathrm{e}}$ is of the order of $\gamma=\alpha \nu_{\mathrm{t}} / L$, where $\alpha$ is a numerical factor which can be regarded as of the order of unity. In this way Eqn (91) yields the estimate $b \approx\left(b_{0} L\right)^{1 / 2}, \quad$ where the quantity $b_{0}=\hbar(m T)^{-1 / 2}$ introduced earlier represents a characteristic localisation width of the density matrix. Even for an electron when $L=1 \mathrm{~cm}$ and $T=10 \mathrm{eV}$ an estimate of $b$ gives a fairly
small value: $b \approx 10^{-4} \mathrm{~cm}$. In other words, nonequilibrium results in a strong localisation of the particle and imparts those properties that are usually attributed to a microparticle.

If $b \ll L$, then the behaviour of such a particle can be described by classical mechanics. Each of the possible wave packets has a 'velocity' width which is of the order of $v_{\mathrm{t}} b_{0} / b \ll v_{\mathrm{t}}$. Consequently, the velocity distribution of the particle can be regarded as continuous and characterised by a distribution function $f(v)$.

## 20. Microworld and macroworld

The discovery of quantum mechanics has immediately raised a number of problems and the solutions to some of these problems cannot yet be regarded as fully satisfactory. This does not prevent the use of the powerful theoretical apparatus of quantum theory when applied to an extensive range of physical phenomena. Quantum mechanics has not yet failed to account for practically any experimental observation (this applies primarily to the nonrelativistic theory). Nevertheless, we are left with the question of why the theory predicts only probabilities of random processes in the microworld, and also with the question why these processes are random. The most important is the difficulty of establishing where the boundary is between quantum and classical physics.

It is obvious to everyone that classical macroscopic bodies have no visible wave properties. On the other hand, all microparticles behave exactly as predicted by quantum mechanics, demonstrating the universal wave nature of the microworld. Therefore, where is the intermediate region and if it does exist, what is the theoretical apparatus needed to describe it?

In this section we shall discuss this problem in a purely qualitative way in order to make clear the direction of thought which will be followed later. The main point is the constant information coupling which is characteristic of objects in our environment. All of nature is exposed to solar radiation, which shines with all colours; this is sufficient for all living organisms continuously to observe their environment. A similar coupling or link via light may also exist between inanimate objects. In quantum mechanics this means that the positions of macroscopic bodies are being constantly 'measured' by the surrounding living organisms and inanimate bodies. Therefore, the $\psi$ waves of macroscopic bodies are subject to constant destruction of coherence. We live in a world of destroyed coherence and further continuous destruction of coherence. We shall now try to estimate where the natural boundary is between the microworld and the macroworld. We shall do this on the basis of Eqn (91), describing the width of a nonequilibrium wave packet.

Let us consider a macroscopic body of density $\rho=1 \mathrm{~g} \mathrm{~cm}^{-3}$ and of transverse size $L$. The mass of this body is $m=\rho L^{3}$. An information flux reaches the body directly or indirectly and this flux is then converted into entropy, which can be estimated from $\dot{S}_{\mathrm{c}}=L^{2} \dot{S}_{0}$, where $\dot{S_{0}} \approx 10^{19} \mathrm{~cm}^{-2} \mathrm{~s}^{-1}$ is the flux reaching the Earth from the Sun.

We shall now consider a body for which the localisation width of Eqn (91) is of the order of its own dimensions. Such a body is obviously somewhere on the boundary between the macroworld and the microworld. Substituting the relevant quantities, we obtain the following estimate:

$$
\begin{equation*}
L \approx\left(\frac{\hbar}{\rho \dot{S_{0}}}\right)^{1 / 7} \approx 10^{-5} \mathrm{~cm} \tag{92}
\end{equation*}
$$

Naturally, this is a very rough estimate, just sufficient to show that all visible bodies belong to the macroworld and this applies to the objects seen under a microscope. We recall that a wavelength of visible light is $\lambda \approx 10^{-4} \mathrm{~cm}$, so that the minimum dimensions of a macroscopic body of $\sim 0.1 \mu \mathrm{~m}$ obviously do not exceed $\lambda$.

It follows that all the objects in our macroscopic environment have wave packets which are 'collapsed' to dimensions much smaller than their transverse size and which appear to us as sharply defined and capable of description in the classical terms of solid, liquid, or gaseous bodies. It is sufficient to introduce the corresponding classical variables and to apply classical theoretical mechanics. In other words, we are dealing with conventional dynamics.

On the other hand, in the case of microparticles the wave behaviour predominates. If we want to know something about the behaviour of a microparticle, we must ensure that it interacts with a macroscopic instrument which intervenes between the particle and the observer. This instrument, including its components, is all the time in a state of information exchange with the environment. Therefore, the wave function of the instrument 'lives' under conditions of unavoidable destruction of its coherence. It is the destruction of coherence of the instrument that occurs in a purely random manner because of the many links with the classical objects in the environment.

The instrument itself is constructed in such a way that different states of a microparticle correspond to different readings. The selection by the external world of one of the readings automatically destroys the coherence of the wave function of the microparticle. All this looks like a random process but, if it is repeated many times, then certain features of statistical relationships are revealed and they can be described on the assumption of transformation of a pure ensemble into a mixed one. The quantity $|\psi|^{2}$ then plays the role of the probability density. The instrument simply indicates in which part of the complete set of states acquired by the instrument is the particle located during a given measurement.

## 21. Behaviour of a microparticle

Any object or living organism interacting with the surrounding world manifests only a small proportion of its properties or structural possibilities. The usual atomic approach postulates that all these internal properties can be revealed step by step, i.e. they can be explained completely if we know all the properties of the small components of the object and their interactions with one another. Following the same approach we shall consider the simplest object, namely a small particle which manifests only its dynamics, i.e. its mechanical properties.

In classical mechanics such an object is called a material point, i.e. a body of very small dimensions without any internal structure. All that such a point has are its mass, position in space, and response to external forces in accordance with Newton's second law. In the one-dimensional case this response is described by

$$
\begin{equation*}
m \ddot{x}_{0}=\dot{p}_{0}=F \tag{93}
\end{equation*}
$$

Here, $x_{0}$ is the coordinate of the point, $m$ is its mass, $p_{0}=m x_{0}$ is the momentum, $\dot{x}_{0}$ is the velocity, $\ddot{x}_{0}$ is the acceleration, and $F$ is the force acting on the point. In the case of a potential field the force is $F\left(x_{0}\right)=-\partial U / \partial x_{0}$, where $U\left(x_{0}\right)$ is the potential energy. This approach to a small particle seems to be absolutely irreproachable and most rigorous. However, it does not always satisfactorily describe the interaction of such a particle with the external world. In fact, if the particle is inside a thermostat and is subjected to very slow actions, such as a change in the volume occupied by it or a change in its average kinetic energy, it is more correct to describe it in terms of thermodynamic quantities, which include the temperature, volume, entropy, internal and free energies, etc. The descrip-tion of the object should correspond exactly to the interaction of this object with the external world.

In the thermodynamic description there is no need to consider the instantaneous position of a particle in space and it is sufficient to know only its average characteristics. Naturally, in this case the description is based on incomplete information about the particle. A more general approach to an incomplete description of the particle is based on the introduction of the distribution function $f(x, v, t)$ for the probability of finding the position and velocity of the particle near $x$ and $v$, respectively, so that the quantity $f(x, v, t) \Delta x \Delta v$ corresponds to the probability that the coordinate lies within the interval $\Delta x$ and the velocity within $\Delta v$. Evolution of the function $f(x, v, t)$ is governed by the kinetic equation

$$
\begin{equation*}
\frac{\partial f}{\partial t}+v \frac{\partial f}{\partial x}-\frac{1}{m} \frac{\partial U}{\partial x} \frac{\partial f}{\partial v}=0 \tag{94}
\end{equation*}
$$

The description given by Eqn (94) is most appropriate to a small particle. It naturally covers both limiting cases: in one case when the velocity and coordinate distributions have time to assume the Maxwell-Boltzmann form and the thermodynamic limit is reached, and also in the case when

$$
\begin{equation*}
f_{\mathrm{e}}=\delta\left(x-x_{0}\right) \delta\left(v-v_{0}\right) \tag{95}
\end{equation*}
$$

and we are dealing with a material point whose coordinates are $x_{0}(t)$ and $v_{0}(t)$. Eqn (95) contains formal $\delta$ functions which always vanish except at the points where their argument vanishes. In physical discussions it is convenient to assume that $x$ and $v$ are not continuous but discrete variables, so that the space of $x$ and $v$ is split into very small cells of dimensions $\Delta x$ and $\Delta v$. Then the function (95) should be equal to $(\Delta x \Delta v)^{-1}$ in just one cell and in all the other cells it vanishes.

If the distribution function of Eqn (95) is substituted into Eqn (94) and integrated with a weight $x$ and then with a weight $v$, the result is Eqn (93). Therefore, the description of the motion of a particle in terms of the dynamic variables $x$ and $p=m v$ corresponds to the maximum localisation of Eqn (95). We shall assume that the coordinate $x$ varies only in a finite segment of length $L$ and the velocity is limited from above to the value $c$. Then the total number of cells in the phase space is $N=L c(\Delta x \Delta v)^{-1}$. Consequently, the function described by Eqn (95) corresponds to a state with just one occupied cell, i.e. with the maximum information $I=\ln N$ and zero entropy. If we select smooth distributions, then the corresponding values of the entropy increase with the number $\Gamma$ of occupied cells: $S=\ln \Gamma$. Consequently, the amount of information decreases: $I=\ln N-\ln \Gamma$. In the limit when the particle uniformly fills the whole segment $L$ and the velocity distribution becomes Maxwellian, the
entropy reaches its maximum value for a given average energy. Consequently, the amount of information for this state should be regarded as zero because the velocity limit at a given temperature $T$ is simply $c \approx(T / m)^{1 / 2}$.

A description based on the distribution function readily deals with the problem of going to the thermodynamic limit. However, even this description has a characteristic, but not very obvious defect: it applies only to a small particle in the macroworld. Essentially this particle is in constant information 'contact' with the external world and the information interaction is sufficiently small not to affect in any way the dynamics of this world.

The kinetic description does in fact admit a solution of the type given by Eqn (95). It readily follows from the kinetic equation (94) that $x_{0}$ and $p_{0}$ satisfy Eqn (93). This consequently means that if the $x$ coordinate has the value $x_{0}(t)$ at time $t$, and the value $x_{0}(t+\Delta t)$ at time $t+\Delta t$, its velocity is defined as $\left[x_{0}(t+\Delta t)-x_{0}(t)\right] / \Delta t$. In other words, the velocity measurement can be made by measuring the coordinate twice: at time $t+\Delta t$ and at time $t$. Only if we are sure that the second measurement does not disturb the state of the particle during the first measurement, can we speak of the existence of the velocity $v_{0}$ and, consequently, of the momentum $p_{0}$ which occurs in the dynamic equation (93). It is understood that the measurement and the interaction of the particle with the instrument are objective processes. Therefore, it is more correct to say that the dynamic equations are based on the assumption that the particle has a constant information link with the external world and this link does not disturb the dynamic properties of the particle. These are the characteristics appropriate for relating them to the objects in the macroworld. However, in the case of the microworld particles it follows from quantum mechanics that the old assumptions about the simultaneous existence of the position and momentum of a particle are incorrect.

Quantum mechanics is based on a completely new appproach to the process of measurement or, more accurately, to the information interaction of a microparticle with the macroworld objects. The fundamental principle of quantum theory says that the process of measurement, which would seem to permit an infinitesimally small exchange of energy between a particle and an instrument, has nevertheless a significant influence on the microparticle dynamics. Each measurement considerably alters the state of the microparticle and therefore a second measurement applies only to a new state and the previous state is disturbed by the measurement itself. The question is: how can one describe a microparticle in such a case?

The first conclusion is obvious. If any measurement destroys something, then the required relationship can be identified only by many similar experiments. This means that each measurement may give results somewhat different from the other measurements: the events are random and only the average statistical results can reveal the investigated relationship. However, this is also true when the description is based on the distribution function $f(x, v, t)$. Where is then the difference? The difference is that the macroscopic description admits the solution of Eqn (95) with the minimum entropy and this solution is based on the feasibility of second measure-ments without disturbing the state of a particle, whereas in the microworld any measurement disturbs this state.

We shall call the solution of Eqn (95) with the minimum entropy a pure classical ensemble: repeated measurements carried out on such an ensemble always give the same result. It is this fact that corresponds to $S=0$. However, in addition to this pure state, Eqn (95) also predicts states with $S \neq 0$. We shall call them mixed states. Any mixed state can be regarded as a composite of pure states:

$$
\begin{equation*}
f(x, f, t)=\int c\left(x_{0}, v_{0}, t\right) f_{0} \mathrm{~d} x_{0} v_{0} \tag{96}
\end{equation*}
$$

where $c\left(x_{0}, v_{0}, t\right)$ is a function of the variables $x_{0}$ and $v_{0}$, whose form is identical with $f(x, v, t)$.

However, a quantum particle can also be in a mixed state: this is simply a randomly selected member of a statistical ensemble with a certain distribution of probabilities between individual states which can be called pure. A particle in a mixed state interacts with the external world as if only some of its information potential is participating in this interaction. In the limit of maximum entropy and minimum information about a quantum particle we can once again apply a thermo-dynamic description in terms of temperature and entropy.

Quantum mechanics states that even in a pure quantum state a particle interacting with a macroinstrument behaves as a random object which requires a statistical description. We shall try to show why the logic of quantum mechanics naturally leads to the wave equation. We shall assume that we have an instrument which can measure the position of a particle. After each measurement the state of the particle is destroyed, i.e. it is converted either into a state which cannot be pure or to a different pure state but different from the initial one. The second measurement of the coordinate of the pure state cannot give the result which would be related directly to the first measurement of position. Therefore, the most natural assumption is that the position can be measured by establishing some statistical relationship. Let us assume that $p_{x}(x)$ is the probability density for obtaining the results of the measurements of the position of a particle in the interval $(x, x+\mathrm{d} x)$.

In addition to the coordinate $y$ a particle has a second dynamic characteristic, which is its momentum $p$. The momentum $p$ cannot be measured with an instrument but is used in the determination of its position: if repeated measurements that do not destroy the state are forbidden, the velocity of the particle cannot be found by measuring its coordinates. Consequently, the momentum has to be measured with a different instrument, for example on the basis of the recoil momentum after reflection from a perfectly reflecting barrier in the instrument, which makes it possible then to measure the momentum acquired by this barrier. Let us assume that $p_{p}(p)$ is the probability of finding the particle momentum in the interval $(p, p+\mathrm{d} p)$.

We can therefore carry out two incompatible types of measurements: either we determine the position of a particle or its momentum. These measurements may be carried out on the same state of the particle, but by completely different instruments. In each case we speak of a complete set of measured quantities and, consequently, a complete measurement.

Let the state of the particle evolve with time. Then, the probabilities found by the instruments are functions of time: $p_{x}=p_{x}(x, t), p_{p}=p_{p}(p, t)$. We shall now try to see what we can deduce about these probabilities by clear and logical physical reasoning. We shall simplify such reasoning by
assuming that our particle moves freely, i.e. that $U=0$. The energy is then $\varepsilon=p^{2} / 2 m$ and, therefore, determination of the momentum distribution function gives automatically the energy distribution function.

Let us next assume that a state is created with a precisely specified momentum $p=p_{0}$ and a precisely specified energy $\varepsilon_{2}=p^{2}{ }_{0} / 2 m$. From the point of view of mechanics such a state is absolutely stationary (steady) in time and, therefore, the probability $p_{x}$ should no way depend on time. However, this automatically means that $p_{x}$ should be independent of $x$ because uniform motion along $x$ should again yield a stationary state.

It therefore follows that a stationary state corresponds to the probabilities $p_{x}=$ const and $p_{p}=\delta\left(p-p_{0}\right)$. We shall now consider a slightly nonstationary (unsteady) state when $p_{x}$ is a slowly varying function of the coordinate $x$ and of the time $t$, and $p_{p}(p)$ is a strongly localised function near the value $p=p_{0}$.

If the width of the localisation region of the particle along $x$ is very large, then the instrument used to measure the position of the particle need not be known very precisely: a slow change in $p_{x}$ along $x$ means that the scale of $\Delta x$ in the measurement of the particle can be large. However, we then have an almost classical particle and it is natural to assume that $p_{x}(x, t)$ travels at the velocity of the particle, which is close to $p_{0} / m: p_{x}=p_{x}\left(x-p_{0} t m^{-1}\right)$. We thus obtain an analogue of Eqn (95) except that the scale of localisation along $x$ is fairly large and the probabilities $p_{x}$ and $p_{p}$ apply to the readings of different instruments, so that they cannot be combined into one product of Eqn (95).

By analogy with Eqn (95) we shall assume that there is a function $W(x, p)$ such that, when integrated with respect to $p$ or $x$, yields the following expressions for $p_{x}(x)$ and $p_{p}(p)$, respectively:

$$
\begin{equation*}
p_{x}(x)=\int W(x, p) \mathrm{d} p, \quad p_{p}(p)=\int W(x, p) \mathrm{d} x \tag{97}
\end{equation*}
$$

The function $W(x, p)$ is called the Wigner function. In the classical case $W(x, p)$ should be identical to the distribution function of $x$ and $p$, but in the quantum case this is not true because measurements of the values of $x$ and $p$ are carried out by different instruments. Consequently, $W(x, p)$ need not be a positive or even real function. Moreover, the Wigner function may not be expandable into the product of the function of just $x$ and another function of just $p$. Finally, in the case of a smooth distribution of $p_{x}(x)$ with respect to $x$ the function $W(x, p)$ may be regarded as close to $W\left(x-p t m^{-1}, p\right)$ with the dependence on the second argument localised strongly near $p=p_{0}$. So far none of this is in conflict with the classical distribution of the probabilities. The transition to the quantum description requires the appearance of a quantity with the dimensions of length which would indicate at what length scale the new physics applies. It has been found, however, that there is no such universal quantity with the dimensions of length. Instead there is a universal quantity $\hbar$, which is the Planck constant with the dimensions of action.

We can try to find the minimum length with the aid of $\hbar$ and $p$. For example, we can relate this length to the quantity $\hbar / p$ and, consequently, relate the momentum to the reciprocal of length. It would be more natural from the mathematical point of view to assume that the momentum is proportional to the derivative $\hbar \partial / \partial x$. However, the average value of the derivative $\partial / \partial x$ vanishes for any distribution $p_{x}(x)$. There-
fore, this derivative should not act on the probability but on some other quantity, for example, a new quantity such as the wave phase in a wave packet. By analogy with classical fields it is necessary to introduce a wave function $\psi$. More exactly, if $\psi_{k}(x)=\exp (\mathrm{i} k x)$ then $p$ is defined as follows:

$$
p \psi_{k}=-\mathrm{i} \hbar \frac{\partial \psi_{k}}{\partial x}=\hbar k \psi_{k}
$$

Therefore, for a plane wave the momentum is $p=\hbar k$, where $k$ is the wave number and the Planck constant $\hbar$ indicates that for one and the same particle the momentum $p$ corresponding to a given value of the wave number $k$ cannot be as small as we please, but is limited from below by the quantum of action.

The function $\psi$ should be linked somehow to the probabilities $W(x, p), p_{x}(x)$, and $p_{p}(p)$. We shall assume that $\psi$ depends also on time, for example, in accordance with the law

$$
\psi_{k}(x, t)=\exp (-\mathrm{i} \omega t+\mathrm{i} k x)
$$

The frequency $\omega$ should depend first of all on the wave number, i.e. $\omega=\omega_{k}=\omega_{k}(k)$. If this is true, the superposition of the function $\psi_{k}$ can be used to plot wave packets propagating at a group velocity. These packets with a set of wave numbers near a specific value $k=k_{0}=p_{0} / \hbar$ correspond to the probability $p_{k} \approx p_{x}\left(x-p_{0} t m^{-1}\right)$ for a very wide distribution of the probability density $p_{x}$ along the $x$ axis. Since in the case of a wave packet we have $p_{0} m=v_{\mathrm{g}}=\partial \omega_{k} / \partial k$ and since $p_{0}=\hbar k$, we find directly the dispersion law: $\omega_{k}=\hbar k^{2} / 2 m$. Analogy with the construction of the classical wave packets leads in a natural manner to the relationship

$$
\begin{equation*}
p_{x}(x, t)=\psi^{*}(x, t) \psi(x, t)=|\psi(x, t)|^{2} \tag{98}
\end{equation*}
$$

Next, we can close the relationships in Eqn (97) on the assumption that the Wigner function is

$$
\begin{equation*}
W(x, p)=\frac{1}{\pi \hbar} \int \psi^{*}(x) \psi\left(x^{\prime}\right) \exp \left[\frac{\mathrm{i} p}{\hbar}\left(x-x^{\prime}\right)\right] \mathrm{d} x^{\prime} \tag{99}
\end{equation*}
$$

In other words, $W(x, p)$ corresponds to the Fourier transformation of the density matrix $\psi^{*}(x) \psi\left(x^{\prime}\right)$ along one of the coordinates and the probability $p_{p}(p)$ is equal to the modulus of the square of the amplitude in the Fourier expansion $\psi(x)$ in terms of harmonics of the type $\exp (i k x)$, where $x=p / \hbar$. The logic of the information interaction of a microparticle with a macroinstrument, considered on the assumption that the state of the particle is destroyed by measurement and that there is a quantum of $\hbar$ action, thus unavoidably leads to wave mechanics, and then to the Schrodinger equation and the probabilistic interpretation of $\psi$.

## 22. Perception

Measurements in quantum mechanics represent an entirely irreversible information process and this process can be described if we introduce some auxiliary reasoning and constructs. We recall once again that in the information processes it is not the energy aspect which is important (although it is essential), but the meaning of the transmitted and received signals. The dynamics is the exchange of momentum and energy, and the informatics is the exchange of symbols between partners participating in the information interaction process.

We shall consider here the information aspect of the process of measurement which can be called the detection or perception of signals processed first by an analyser. A


Figure 1. Perception occurs as a result of a transition from the state a to the state $b$. In response to an event $u_{i}$ in an observed object $U$ an analyser $A$ detects an event $a_{i}$ due to the initial storing of information in a feeder system $F$.
measured object $U$ (unity) interacts directly with an analyser A, which is in one of the states $a_{i}$. F or the sake of simplicity, we shall assume that these states are equivalent and related to the object of measurement in such a way that all $\mathrm{a}_{i}$ are equiprobable. Let us assume that the total number of $a_{i}$ states is $N$. Then the probability of any one of these states is $1 / N$. A measurement is regarded as performed if a state $a_{i}$ is recorded by a detector acting as an information 'perceptor'. The information can then be recognised. In this section we shall deal specifically with recognition itself.

We shall assume that our detector consists of two parts: an information perceptor $P$ and a feeder $F$ which feeds informa-tion to the system $P$. Let us assume that $P$ and $F$ each have $N$ components. Let us also assume that before receiving information the system $P$ is 'clean', i.e. it is free of any initial information.

A system free of information can be represented in two variants: it is either simply a set of $N$ clean cells or these cells have states which vary rapidly and at random, for example from zero to unity and back again. In the former case we are dealing simply with a memory $M$, which we shall also consider later. In the latter case we have what is apparently a thermo-stat with pure thermal motion or its chaotic analogue. Therefore, we shall denote it by the letter $C$ (capacity). In the case of $M$ we have a system with zero information and zero entropy, whereas in its initial state the information in C is zero and the entropy has its maximum value $S=\ln N$. In this section we shall consider purely physical processes by which information can be exchanged with the external world and therefore in our discussion it will be more convenient to consider an initially stochastic system with zero information and maximum entropy, i.e. a system of the type C .

We shall now discuss the system $F$ which can serve as a source which can feed information to the perceptor $P$ in the state C . We shall also assume that F consists of $N$ cells, but only one cell is filled or excited. The entropy of $F$ is zero and its information content is $I=\ln N$.

We shall now consider the model of perception in the purely classical case when the occurrence of a random value of the measured quantity is not related to the measurement process. Let us assume that the system U has $N$ states $\mathrm{u}_{i}$ which can be realised with the same probability $1 / N$. We shall assume that between the states of the measured objects $u_{i}$ and the states of the analyser $a_{i}$ there is a one-to-one correspondence, so that the occurrence of $\mathrm{u}_{i}$ automatically results in the event $\mathrm{a}_{i}$.

Therefore, in the classical case we seem to have just one event $\left(\mathrm{u}_{i}, \mathrm{a}_{i}\right)$. For example, a change in the direction of wind $\mathrm{u}_{i}$ alters the direction of a weathercock $\mathrm{a}_{i}$, and if these directions are observed every hour, a chain of random events
$\left(\mathrm{u}_{i}, \mathrm{a}_{i}\right)$ is recorded. However, these events have not yet been recognised and 'understood'. Recognition requires an irreversible recording $a_{i}$ somewhere in the instrument. This can be done as follows (Fig. 1).

After an event $u_{i}$ occurs in the system $U$, the analyser consecutively scans the cells of $U$ and when it detects $u_{i}$, it responds to $u_{i}$ by an event $a_{i}$ at the expense of the information stored in the feeder system F . This system is constructed in almost the same way as the analyser A, but it is in the state of maximum order, when only one of the cells is filled and the other cells are empty. The initial information about the event $\mathrm{u}_{i}$, for example, that one of the faces of a cube comes up, is $\ln N$, where $N$ is the number of cells in U . The analyser A responds by the appearance of a signal in the cell $\mathrm{a}_{i}$ corresponding to $\mathrm{u}_{i}$. The information in the system A then increases abruptly from zero to $\ln N$ and the entropy falls from the initial value $\ln N$ to zero. In accordance with the second law of thermodynamics such a fall of the entropy is possible only because of a change to chaotic behaviour in the feeder system F , the entropy of which rises from zero to $\ln N$.

It should be pointed out that the total entropy is conserved in the interaction of the systems A and F. This means that the process of information transfer from $F$ to $A$ is fully reversible: from the state of the systems $A$ and $F$ in Fig. 1 b can return to the state of A and F in Fig. 1a without violation of the second law of thermodynamics. Therefore at this stage the appearance of the signal $a_{i}$ in the analyser does not represent the irreversible storage of information.

However, we must bear in mind that in practice such an ideal instrument is difficult to construct because of natural external disturbances. Therefore, a more realistic device for signal detection may look as shown in Fig. 2.

If because of irreversible processes some of the information is lost by the feeder $F$, it is necessary to 'fatten up' the feeder itself and expand it, i.e. it is necessary to increase the information stored in it. This can be done by increasing the number $N_{*}$ of its cells so that $N_{*}>N$ and $I_{*}>I$. Now in the course of storing the information which is in A some of the stored information $I_{*}$ may be lost, i.e. it may be destroyed by dissipative interactions with the surrounding medium. The information $W=I_{*}-I=\ln \left(N_{*} / N\right)$ is then given up to the external medium in the form of losses (waste). We can say that the interaction with a medium in equilibrium increases the entropy of the $\mathrm{A}+\mathrm{F}$ system by $W$. We shall introduce a quantity $\eta_{I}=I / I_{*} \leqslant 1$ which can be naturally called the information efficiency or effectiveness. Clearly, all the real processes have the efficiency $\eta_{I}<1$.

If $\eta_{I}<1$, the process of recording the signal becomes irreversible: it is no longer possible to return simply from Fig. 2b to Fig. 2a. We can therefore say that the signal $\mathrm{a}_{i}$ is now


Figure 2. Record of an event $\mathrm{u}_{\mathrm{i}}$ in an analyser A when the information $I_{*}$ in the feeder F is not fully utilised: the rest of the information $W=I_{*}-I$ is lost by dissipation in the environment.
'perceived' or 'stored'. We can easily see that apart from the analyser A, shown in Fig. 2, we can imagine one or more similar additional analysers. All of them 'see' the same signal $u_{i}$. Let us assume that $U$ is a cube with numbered faces and the event $\mathbf{u}_{i}$ is a random coming up of one of its faces. Analysers of the A type are the players checking which face of the cube appears on the top. They record the events $a_{i}$ by seeing them and storing them in their memory.

The players may want to throw the cube again. However, they should prepare for the recognition of the results of the new throw. This is done by 'cleaning' the analyser A, i.e. by removing the recorded signal (Fig. 2b) and transferring it to a different memory system by exactly the same process as that by which the signal $\mathrm{a}_{i}$ has been obtained from the signal $\mathrm{u}_{i}$. Moreover, it is necessary to 'fatten up' the feeder F with information. This can be done by exactly the same process which F has used to 'interrogate' A, namely we must have a feeder $\mathrm{F}^{\prime}$ with the information $I_{* *}>I_{*}$ and then with the efficiency $\eta_{I}=I_{*} / I_{* *}<1$ we can switch F to the state in Fig. 2a. In other words, in order to throw the cube and recognise the results of such events it is necessary to 'feed' information from somewhere outside.

The information process in Fig. 2 includes two irreversible processes: the actual event $\mathrm{u}_{i}$ and the recording of the event $a_{i}$ with the aid of $F$ and with the loss of information $I_{*}-I$, i.e. with an increase in the entropy of the combined system $\mathrm{A}+\mathrm{F}$. The event $\mathrm{u}_{i}$ arrives from outside and can be either random or may represent a more regular sequence. The process in the system $A+F$ is a typical irreversible process accompanied by an increase in the entropy and the loss of the information $W$. The random or nonrandom nature of a chain of consecutive recognition events is determined entirely by the input, i.e. by the presence or absence of regularity in the sequence of events $u_{i}$.

We shall now assume that the object $U$ is not of classical but of quantum nature. Then the cells in Fig. 2a correspond to eigenfunctions of the complete basis and the object U itself is a superposition of these eigenfunctions. The analyser A must also be a quantum object, so that its cells corresponding to arbitrary readings of the instrument are also eigenvectors correlated with U. The actual measurement occurs when the irreversible system F associated with the external world comes into play. It is this system that performs the irreversible 'measurement' process corresponding to the transition from Fig. 2a to Fig. 2b. This increases the entropy of the system $F$ and the systems $U$ and $A$ collapse simultaneously into $u_{i}$ and $a_{i}$. The quantum systems $U$ and $A$ come into contact with the external classical world via the system F. This contact destroys the coherence of the initial
states of the systems $U$ and $A$ and leads to an immediate collapse of these systems into $\mathrm{u}_{i}$ and $\mathrm{a}_{i}$. When measurements of the same type are repeated, the collapse occurs at random in one and then in another cell, so that the pattern averaged over many measurements assumes the nature of a mixed state described statistically by, for example, the density matrix.

The system $F$ is thus the boundary between the quantum and classical worlds. In the time interval between the measurements the quantum description can be expressed in terms of the evolution of the wave function $\psi$ of a pure state or of the density matrix $\rho$ of a mixed state.

These discussions are fairly general and do not yet give a direct recipe for the description of the real link and interaction between the quantum microworld and the classical macroworld. Before we can provide this description we have to become acquainted with one more very important phe-nomenon, which is the presence of fluctuations. However, a further general comment should be made first.

Complex physical objects can be described with different degrees of detail and this is not only because an exact description requires much effort or a large volume of computer calculations. After all, all our experience shows that when interacting with the external world a physical object never reveals all its potential inner complexity. Consequently, an incomplete or even a phenomenological description of physical phenomena or of physical objects sometimes better describes the crux of the matter and provides an under-standing of what occurs. This applies in particular to collec-tive phenomena when an enormous number of particles is involved in general collective motion. For example, gas dynamics describes better and more satisfactorily the wind flows than does simply molecular dynamics. Exactly in the same way the processes associated with the thermal motion of atoms are described better and can be understood more easily if use is made of thermodynamics and statistical physics. We can state this differently: there are physical phenomena for which an incomplete description with neglect of excess detail is more satisfactory and even reflects better the essence of the processes under consideration.

In particular, a statistical description involving the introduction of probabilities and averaging over the velocity distributions is a better description of objects composed of a very large number of atoms. If the number of atoms is reduced, then against the background of a probabilistic description, which does not lose its meaning averaged over many processes of the same type, individual processes begin to appear and play an increasingly important role. These can be called fluctuations and we can quite arbitrarily select the
degree of detail in their description. For example, the motion of a Brownian particle can be described as diffusion. Frequent repetition of the measurements makes it possible to charac-terise this motion as a random Markov chain. In the limit, if the particle is followed over very short time intervals, we can see a very complex path of such a particle. In any case, when applied to classical physics, there is no doubt that we can provide the description as accurately as required. However, this is not true of a quantum particle: the observation is accompanied by the particle interaction with the macroworld and this interaction cannot be as small as we please. It is desirable to become acquainted first with fluctuations in order to find ways of achieving a better understanding of the effects in question.

## 23. Fluctuations and irreversibility

The thermodynamic relationships used above in Sections 36 apply solely to the average values. Under real physical conditions such averaging may occur, in a sense, spontaneously because of the slowness of the processes that take place. The formal averaging should therefore also be carried out only with respect to time. It has been shown in statistical physics that in the case of a large number of particles the averaging in question can be carried out not only with respect to time but also over the phase space, which finally leads to a canonical distribution. However, the discrete-i.e. atomic-structure of matter does not appear completely and it is manifested by fluctuations, which are small deviations from statistical equilibrium. In this section we shall consider the simplest examples of fluctuations and discuss their relationship to irreversibility.

Once again let us consider an ideal gas. Let $N$ noninteracting classical particles be located in a volume $V$. Then the average density of the particles is given by $n=N / V$. The thermal motion of the particles has the effect that the number of particles in a certain small volume $V_{0}$ is not exactly equal to $N_{0}=N V_{0}$, but fluctuates near this value. These fluctuations can be found quite simply. Let us assume that a sum over all the particles

$$
\begin{equation*}
n_{\mu}(\boldsymbol{r})=\sum_{i} \delta\left(\boldsymbol{r}-\boldsymbol{r}_{i}\right) \tag{100}
\end{equation*}
$$

represents the real microscopic density of the particles at the point $\boldsymbol{r}$. Here, the $\delta$ function of the type $\delta\left(\boldsymbol{r}-\boldsymbol{r}_{i}\right)$ is localised at the point of location of the $i$ th particle. We can now assume that $n_{\mu}=n+\delta n$, where $n$ is the average density and

$$
\begin{equation*}
\delta n(\boldsymbol{r})=n_{\mu}-n=\sum_{i} \delta\left(\boldsymbol{r}-\boldsymbol{r}_{i}\right)-n \tag{101}
\end{equation*}
$$

The function $\delta n(r)$ represents a set of spikes against the background of a uniform negative value $(-n)$. The average value is $\langle\delta n\rangle=0$. The magnitude of fluctuations is usually found by introducing a correlation function $\left\langle\delta n(\boldsymbol{r}) \delta n\left(\boldsymbol{r}^{\prime}\right)\right\rangle$. It follows formally from Eqn (101) that

$$
\begin{align*}
\left\langle\delta n(\boldsymbol{r}) \delta n\left(\boldsymbol{r}^{\prime}\right)\right\rangle= & \delta\left(\boldsymbol{r}-\boldsymbol{r}^{\prime}\right)\left\langle\sum_{i} \delta\left(\boldsymbol{r}^{\prime}-\boldsymbol{r}_{i}\right)\right\rangle \\
& +\sum_{i \neq j}\left\langle\delta\left(\boldsymbol{r}-\boldsymbol{r}_{i}\right) \delta\left(\boldsymbol{r}-\boldsymbol{r}_{j}\right)\right\rangle-n^{2} \tag{102}
\end{align*}
$$

Here, the first term is simplified to the components with the same indices $i$ and $j$ and it is assumed that $\delta\left(\boldsymbol{r}-\boldsymbol{r}_{i}\right) \delta\left(\boldsymbol{r}^{\prime}-\boldsymbol{r}_{i}\right)=\delta\left(\boldsymbol{r}-\boldsymbol{r}^{\prime}\right) \delta\left(\boldsymbol{r}^{\prime}-\boldsymbol{r}_{i}\right)$, and in the second
term only the com-ponents with $i \neq j$ are retained. We can easily see that the resultant relationship can be represented in the form

$$
\begin{equation*}
\left\langle\delta n(\boldsymbol{r}) \delta n\left(\boldsymbol{r}^{\prime}\right)\right\rangle=n\left[\delta\left(\boldsymbol{r}-\boldsymbol{r}^{\prime}\right)-\frac{1}{V}\right] \tag{103}
\end{equation*}
$$

This relationship can be used to find readily the fluctuation of the number of particles in a small volume $V_{0}$. This can be done by integrating twice with respect to $\boldsymbol{r}$ and $\boldsymbol{r}^{\prime}$ in the volume $V_{0}$. The result is the deviation $\delta N_{0}$ from the average value $N_{0}=V_{0} n$

$$
\begin{equation*}
\left\langle\delta N_{0}^{2}\right\rangle=N_{0}\left(1-\frac{V_{0}}{V}\right) \tag{104}
\end{equation*}
$$

If $V_{0} \ll V$, we obtain the familiar relationship $\delta N_{0}=N_{0}^{1 / 2}$ and for $V_{0}=V$ we find the natural result that the total number of particles in the whole volume $V$ does not fluctuate. The relative fluctuations of the density become particularly large when only one particle is in the volume $V$ : $N=1, n=1 / V$.

We can easily show that Eqn (104) retains its form also for one particle so that the fluctuations of the particle number in the volume $V_{0}$ are described by

$$
\begin{equation*}
\left\langle\delta N_{0}^{2}\right\rangle^{0}=\frac{V_{0}\left(V-V_{0}\right)}{V^{2}} \tag{105}
\end{equation*}
$$

This relationship is absolutely symmetric relative to the internal $V_{0}$ and external $\left(V-V_{0}\right)$ volumes. In the limits $V_{0} \rightarrow 0$ and $V_{0} \rightarrow V$ the fluctuations disappear. For a small value of $V_{0}$, compared with $V$, we have $\left\langle\delta N_{0}^{2}\right\rangle=V_{0} / V$. This relationship has an obvious physical meaning: $V_{0} / V$ is simply the fraction of time when within the volume $V_{0}$ there is only one particle, so that $\left\langle\delta N_{0}^{2}\right\rangle$ is unity multiplied by the probability that the particle is found inside the volume $V_{0}$. We can see that for one particle the square of the fluctuation in the particle number $\left\langle\delta N_{0}^{2}\right\rangle$ is equal to the average number of particles $N_{0}=V_{0} / V$ in the volume $V_{0}$. In other words, the fluctuations are very large and, therefore, they may significantly influence the logic of some of our earlier conclusions.

We shall now return to the process of work done at the expense of the thermal energy by a single particle, with the aid of Maxwell's daemon, i.e. to the measurement of the position or velocity of a particle considered in Section 4. For simplicity, we shall begin with the one-dimensional case on the assumption that the particle is in a thermostat with two ends separated by a distance $L$ from one another along the $x$ axis. The particle collides with the ends and on average maintains the Maxwellian velocity distribution with a temperature $T$. If the effective mass $M$ of an acoustic wave created by the impact of the particle on the end considerably exceeds the mass $m$ of the particle under investigation, then each such collision with the end changes the absolute velocity of the particle only by a small fraction $\sim m / M$ of its value. The smallness of $m / M$ is due to the fact that phonons in a substance consisting of heavy atoms are also 'heavy' and slow. If $m / M \ll 1$, an atom experiences many collisions before there is any deviation from the Maxwellian distribution. The process of relaxation is then similar to a random walk described by the Langevin equation. After many collisions the Maxwellian distribution will definitely be restored and the process can easily be described in terms of the Brownian motion of the momenta.

It would be more convenient to consider a simpler case when $m=M$. It is then necessary to assume approximately
that only one collision is sufficient to reach the Maxwellian distribution of the particles flying away from the wall irrespective of the velocity that these particles have arriving at the wall. For example, let us assume that one particle with a Maxwellian velocity distribution fills the whole thermostat of length $L$. We shall now begin our thought experiment on the work done. Let a detector record the collisions of particles by the end wall and then after a time shorter than $b / v_{\mathrm{t}}$ a barrier is inserted at a distance $b$ from the end and imprisons the particle in a small section. A slow shift of the barrier in the direction of the other end makes it possible to perform negative work $W$ on this barrier at the expense of the thermal energy of the particle, which is $\frac{1}{2} T$ in the thermal equilibrium case.

It would seem that this experiment is a threat to the second law of thermodynamics. In fact, in recording the impact of a particle at the wall we obtain only one bit of information, because against the background of many time intervals without any collisions and without any new information, only one interval provides the signal: 'collision'. This is equal to one bit of information. Consequently, 'assimilation' of this information and the subsequent insertion of a barrier requires an increase in the external entropy $S_{\mathrm{e}}$ by an amount $\ln 2$. It seems that the gain in the work can be much greater: after all, the initial volume $b$ can be expanded to $L$, which may be much larger than $b$. Consequently, the entropy increases by an amount $\ln (L /$ $b) \gg 1$. However, we must not be too hasty with our conclusions. It is found that the validity of the discussion requires allowance for the presence of fluctuations.

We shall begin with a certain general comment. The famous fluctuation-dissipation theorem has been proved in statistical physics: its meaning is that the mechanism of any dissipation is simultaneously the mechanism of creation of fluctuations. This is the reason why fluctuations never die out but are maintained at the level dictated by the discrete, i.e. atomic, nature of matter.

In our case the dissipation, i.e. the conversion of the probability of the distribution of the particle velocities to the Maxwellian form, is created by collisions with the warm ends. It is the collisions with the ends that maintain the fluctuations. We shall now consider this process in somewhat greater detail.

We shall begin by assuming that there is a cloud of many, for example $N \gg 1$, particles. Let $f(x, v, t)$ be the distribution function along the coordinate $x$ and the velocity $v$ of these particles. If this cloud is incident on a perfectly reflecting wall, it is reflected without distortion and consequently no irrevers-ible processes occur. However, if this wall is 'warm', then the reflection of each atom becomes inelastic and after many repeated reflections the distribution function approaches the Maxwellian form. It would seem that dissipation can 'disperse' and smear out all the perturbations, including those associated with the atomic structure of the cloud. However, this is not true: the microscopic distribution function $f_{\mu}$ always has the form

$$
\begin{equation*}
f_{\mu}=\sum_{i} \delta\left(x-x_{i}\right) \delta\left(v-v_{i}\right) \tag{106}
\end{equation*}
$$

where the coordinates of the particles $x_{i}$ and $v_{i}$ obey the microscopic equations of dynamics and therefore provide the full description of the motion of the particles. The fluctuations of the type described by Eqn (102) should therefore be conserved. Creation of fluctuations can be
regarded as a random process in which one particle is apparently taken out from the incident flux which has a smooth distribution function. A particle, with certain definite values of $x_{i}$ and $v_{i}$ obeying on average the statistics of inelastic scattering, then appears in the reverse flux. Therefore, an inelastically reflecting wall is simultaneously responsible for the mechanisms of the smoothing out of fluctuations and creation of new fluctuations by a random 'injection' of particles accompanied by simultaneous removal of particles from their average distribution function.

If the number of particles is large, the fluctuations themselves are small and the mechanism of their formation has a weak perturbing effect on the average distribution.

We shall now consider what happens when $N=1$. Let us assume that a cloud with the probability density $p(x, v, t)$ is moving towards a wall. Somewhere inside this cloud the one and only particle is 'hidden'. A layer $\Delta x=v \Delta t$ of this cloud reaches the wall in a time $\Delta t$. The thickness of this layer is proportional to the running value of the velocity $v$. Since we are interested only in the probability $p(x, v, t)$, then in our model we should take the fraction $\int p(x, v, t) v \Delta t \mathrm{~d} v$, which applies to the probability of finding a particle in this layer, and convert it to the Maxwellian distribution with reflection from the wall. If the Maxwellian distribution applies to the cloud incident on the wall, then the cloud reflected by the wall has the same Maxwellian distribution.

A completely different description language should be used if we wish to follow in detail the fluctuations. In the whole cloud with $p(x, v, t)$ there is only one particle at an unknown location. We can easily see that an inelastic collision of the particle with a wall can be regarded as a random event, which instantaneously destroys the a priori probability $p(x, v, t)$ and transforms it into the $\delta$ function of the type $\delta\left(v-v_{0}\right) \delta\left(x-v_{0} t\right)$, where $v_{0}$ is the velocity after the collision at the moment $t=0$ and at the point $x=0$. A collapse takes place, i.e. the extended probability cloud $p(x, v, t)$ collapses into a narrow localised $\delta$ function. Repeated collisions of the particle with the end wall repeat the process: the a priori probability for the incident cloud disappears at some moment and the particle with the associated d-function probability density flies away from the wall.

If the cloud contains, say, $N$ particles with the same distribution $p(x, v, t)$, then $N p(x, v, t)=f(x, v, t)$ and each collision 'picks out' just one particle so that the distribution function is transformed 'instantaneously' from $N p(x, v, t)$ into $(N-1) p(x, v, t)+\delta\left(v-v_{0}\right) \delta\left(x-v_{0} t\right)$. If $N \gg 1$, the random process of the appearance of fluctuations does not disturb very strongly the initial Maxwellian distribution. In this case the fluctuations occur in accordance with the $N^{1 / 2}$ law and at high values of $N$ the fluctuations are relatively weak.

In the case of a single particle however, the fluctuations are very large: they transform the motion of the particle into classical flight from one end to the other with a random change in the velocity after each collision. Nevertheless, it seems obvious that if the motion of a piston-shaped barrier is very slow, so that the collisions can be regarded as very frequent, the time-averaged distribution of the particle velocities can be regarded as Maxwellian. However, this is not quite true because there are fluctuations.

Let us assume that a barrier, located initially at a distance $b$ from the left-hand end and separating the particle from the rest of the volume of size $L \gg b$, begins to move slowly at a
constant velocity $\varepsilon v_{\mathrm{t}}$, where $\varepsilon \ll 1$ is a very small number and $v_{\mathrm{t}}=(T / m)^{1 / 2}$ is the average thermal velocity. In the smaller section the particle will mainly exhibit the Maxwellian behaviour so that the work will be done on the barrier and the power corresponding to this work is

$$
P=-\dot{W}=\frac{T}{b} \dot{x}=\varepsilon v_{\mathrm{t}} \frac{T}{b}
$$

since the average pressure of the particle on the wall is $T / b$. However, there is a low probability that after reflection by the warm wall the velocity of the particle will be less than $\varepsilon v_{t}$. Such a particle can never catch up with the moving barrier and, consequently, with a probability of the order of $\varepsilon$ the pressure on the barrier can unexpectedly fall to zero and never recover. This can occur after $1 / \varepsilon$ collisions with the wall (on average). Because the average time between the collisions is of the order of $b / v_{t}$, it follows that before the disappearance of the pressure the barrier becomes displaced through a distance $\Delta x=\varepsilon v_{\mathrm{t}}\left(b / v_{\mathrm{t}}\right) / \varepsilon \approx b$. In other words, when the work is done we can only double the section containing the particle and then the pressure on the barrier (piston) disappears abruptly and no further work can be done. The entropy of the particle can then be increased by just one bit, i.e. exactly equal to the amount by which the entropy of the environment increases in the initial detection of the collision of the particle with the wall. Further work can be done by stopping the barrier and waiting until the particle catches up with the barrier and then recovers its Maxwellian distribution. Having detected that at least one 'full-weight' collision with the barrier, we again expand the section with the particle. However, such a measurement requires an expense of information, i.e. it requires an increase in the entropy of the environment by one bit. The subsequent expansion of the volume can again increase the entropy of the particle by just one bit, doubling the volume until the pressure on the barrier disappears again. Once again, Maxwell's daemon has to take account of the second law of thermodynamics.

We note that if instead of one particle we have $N$ particles which do not interact with one another, the disappearance of the pressure due to the motion of the barrier applies to each particle separately so that on average the pressure disappears and the volume is doubled even if $N \gg 1$.

At first sight it seems that our example is too artificial and that we can find the conditions under which one detection of a collision with the end can 'fence off' the particle at a distance $b$ from the end. Then by expanding the volume to the total length $L$ it should be possible to increase the entropy of the particle by $\ln (L / b)$ and do the relevant work at the expense of the thermal energy. However, this is not true. Let us consider, by way of example, a more realistic case when the particle is in a cylindrical thermostat of radius $a$. The Maxwellian distribution is then established by collisions with the side walls, so that the appearance of slow longitudinal motion of the particle at a velocity less than that of the barrier (piston) plays no major role: the particle can rapidly recover its longitudinal velocity because the side walls convert the distribution to the Maxwellian form.

Let us assume that $v_{*}=v_{t} / a$ is the average frequency of collisions of the particle with the side walls. It is convenient to consider the limit in the case when $b \gg a$; it is then that $v_{*}$ governs the rate of recovery of the Maxwellian distribution. It would seem that at a low expansion velocity $\dot{x}=\varepsilon v_{\mathrm{t}}$, where $\varepsilon \ll 1$, there should be no problem in recovering the

Maxwellian distribution and the average pressure on the wall should always be $T / b$. However, this is not true: the diffusion effects begin to play a role. At each collision with a side wall there is a random change in the transverse and longitudinal components of the velocity. Therefore, the motion of a particle in the longitudinal direction becomes a random walk with the diffusion coefficient $D=v_{*} a^{2}=v_{\mathrm{t}} a$. A part of the vessel of length $b$ is filled by such a particle in the time $t=b^{2} / D$, so that the velocity of the barrier should not exceed $\dot{x}=b / t=D / b=v_{\mathrm{t}} a / b$ to guarantee that the distribution of the probability over $x$ does not become 'detached' from the moving barrier. However, this is not sufficient. Each collision with the side wall results in localisation of the distribution function of the particle near the point of collision and then this localised cloud wanders for a long time along the $x$ axis. The time $t \approx b^{2} / D$ is the time of wandering inside the segment of width $b$ from one end to the other, so that the average collision frequency in the presence of a barrier is of the order of

$$
v=\frac{1}{t}=\frac{D}{b^{2}}=\frac{v_{\mathrm{t}}}{b} \frac{a}{b}
$$

For $a / b \ll 1$ this quantity is much less than the average collision frequency $v_{\mathrm{t}} / b$ of a freely moving particle. It follows that in a narrow tube the work done by the particle on a piston is considerably less and the high rate of conversion to the Maxwellian distribution does not help at all, but even interferes with the work.

It follows that fluctuations play a major role and particularly when one particle is involved. It should be stressed that fluctuations can be considered as an intrinsic part of an irreversible process: dissipation disperses the consequences of fluctuating behaviour, but it also creates new fluctuations. In particular, the process of conversion to the Maxwellian distribution of a particle at a warm wall can be regarded as a randomly repeated process of annihilation of the probability density flowing to the wall and of creation of a state localised narrowly along $x$ and $v$ : this is a typical collapse of the distribution of the probabilities for a classical particle.

We shall now consider a quantum particle which is subject to the uncertainty limits $\Delta x \Delta p>\hbar$, where $\Delta x$ is the uncertainty along the coordinate and $\Delta p$ is the momentum uncertainty. Let us assume that this particle is in a mixed state, for example, that it represents a set of wave packets with the probability $\left|C_{i}\right|^{2}$ of finding it in the $i$ th packet. Such a particle differs little from the classical one if each of the wave packets does not spread out during the observation time. The collision of the particle with a warm wall has the effect that only one of the wave packets exists in reality after an inelastic reflection: the other packets do not contain the particle and they become automatically annihilated. And if in this process the important aspect for the external environment is the state in which the particle flies away, we can say that a collapse of the mixed state takes place.

The situation is somewhat more complex in the collapse of a pure state. We shall assume that a very wide almost monochromatic packet with $\Delta x \Delta p \ll \hbar$ is incident on a wall. The quantity $|\psi|^{2}$ for such a packet plays the role of the probability distribution and, therefore, it may in principle collapse in the same way as the density of the probability distribution for a classical particle. If $|\psi|^{2}$ is the classical probability distribution, then an inelastic collision accompanied by storage and information about the collision in the
body itself can randomly capture a particle from the cloud $|\psi|^{2}$ and completely annihilate the incident part, emitting a strongly localised reflected part of the probability density. Something similar occurs in the case of a quantum particle. If the incident wave packet is split into wide strips of thickness $\Delta x$, then for a sufficiently large value of $\Delta x$ the collapse occurs in only one of the layers. The very fact of localisation with respect to $x$ automatically broadens the distribution of the momenta by an amount $\hbar / \Delta x$. This broadening cannot be greater than the measure of an inelastic collision. If, for example, the velocity of a particle is altered by a collision by a value of the order of $v_{t}$, then the minimum size of an inelastically reflected packet can be $\hbar / m v_{\mathrm{t}}=b_{0}$. Since our particle is reflected inelastically from many atoms of the wall, it follows that for $\Delta v \approx v_{\mathrm{t}}$ the quantity $b_{0}$ corresponds to the coherence length of the packet. It is natural to assume that the particle is captured by just one of the coherent packets. If for any reason the probabilistic localisation of a particle is considerably greater than the coherence width, it means that we again have a mixed state with a certain distribution of the probabilities $\left|C_{i}\right|^{2}$ of finding a particle in the $i$ th pure state.

We shall return again and again to the phenomenon of collapse of a wave function. Here, we shall stress once more the irreversible nature of the fluctuation effects. We shall consider once more just one particle in a long tube, which is a thermostat of length $L$. If the particle is localised at some point in the middle part of the tube, it begins to diffuse along the $x$ axis and this is characterised by a diffusion coefficient $D$. This is a typical irreversible process accompanied by an increase in the entropy in accordance with law $S \propto \ln (D t)$. On the other hand, each collision with a side wall can be regarded as a random event, which localises a particle in a distance of the order of $a$. Each such collision is also an irreversible process, but it is accompanied not by an increase in the entropy but by its reduction. What is the reason for this?

This is a very fundamental problem and should be discussed in greater detail.

Let a particle be in a tube of length $L$ separated into $N=L / b$ identical cells each of length $b$ (Fig. 3a). In the initial state the position of the particle is unknown and its entropy is $S=\ln (L / b)$. Let the particle now be in the shaded cell in Fig. 3b. The process of appearance of the particle in the cell can be regarded as, for example, a pure fluctuation collapse: the particle 'informs' of its position by one collision with a side wall, and all the other probabilities vanish. A random event occurs and in this event the entropy of the particle falls to zero, but this must be accompanied by appearance of information in the wall of the shaded cell in Fig. 3b. This information may be subjected to the following subsequent processing: the receiving system may 'store' this information as one of the $N=L / b$ possibilities, so that the corresponding value $I$ is $I=\ln (L / b)$. In this case the quantity $I$ would represent the information imparted to the external observation system about the position of the particle and we would have $I+S=$ const $=\ln (L / b)$.

If the collision of a particle with a wall triggers an external information system, then the information obtained can be used to produce the work due to the thermal motion. If this is not done, then the received information is simply forgotten in the course of the irreversible increase in the entropy. The work done can be maximised if maximum use is made of the opportunities for the recognition of information and subsequent increase in the entropy of the particle during expansion of the free volume.


Figure 3. Fluctuations localise a particle from an initially homogeneous state (a) in the only cell (b).

In the other limiting case we may do nothing with the obtained information and not even 'recognise' it, so that we do not know where the particle is really located. Then an equilibrium path of the particle along the cube in Fig. 3a represents a simple thermal random walk with a constant (on average) entropy $S=\ln (L / b)$.

It therefore follows that the information about a particle can be dealt with by very different methods. Let us assume that, for example, the wall of each cell in Fig. 3a contains a sensor which records the impact of a particle with the wall. If the transverse size of the tube is greater than the cell length $b$, repeated collisions in one cell will be rare, so that between the collisions the particle can 'jump' from one cell to another. Repeated measurements can give the average frequency of collisions with the walls $\tau^{-1}$ and the mean-square displacement $(\Delta x)^{2}$ between collisions. These quantities can be used to calculate the diffusion coefficient $D=\frac{1}{2}(\Delta x)^{2} \tau^{-1}$ and then one can use a statistical description of the Brownian motion of a particle. This description of an individual particle is effectively replaced by an ensemble of identical systems and then the average evolution of this ensemble corresponds to the probabilistic, i.e. incomplete, description of the particle dynamics.

The corresponding diffusion process is irreversible and is accompanied by a monotonic increase in the entropy if in the initial state the entropy is not maximal. For example, if initially an ensemble of identical systems corresponds to the position of a particle in the shaded cell with zero entropy, then during the subsequent time intervals the probability 'cloud' expands in accordance with the law $(\overline{\Delta x})^{2}=2 D t$ and the entropy increases logarithmically with time. However, if just one individual particle is selected from this ensemble, then the very first collision with a wall, recorded by a particular cell, leads to the collapse of the a priori probability $p(x, v, t)$ in one cell and this is accompanied by an abrupt fall of the entropy of the particle to zero.

It seems that this collapse is independent of the presence or absence of sensors on the walls. In fact this is not correct: the process of collapse is closely related to the system of measurement. For example, all electrical signals from sensors can be brought together at one input and then only the occurrence of the collision is recorded, i.e. it seems that the collapse is recorded without indication of the cell where it occurs. If irreversible sensors such as minature photographic plates are used, the occurrence of the collapse is recorded for one of the cells, but as yet without an immediate recognition by external recording instruments.

It is very clear from the last examples that one bit of information is sufficient to record the actual collapse. The collapse is an irreversible process or, more exactly, a random event corresponding to a collision and the amount of the corresponding new information $I$ and the reduction in the previous entropy $S$ depend primarily on the ensemble which models the average statistical characteristics of the particle
motion. Consequently, the values $S$ and $I$ apply not so much to the particle as to the combined system formed by the particle and its immediate environment, including the observation and perception of information about the particle.

It is convenient to separate these two physical events. The first event is a random appearance of a particle in one of the possible positions with corresponding 'marking' of the cell number, which we will call the hint. The hint is the appearance of a number on a thrown dice. The hint and perception or storage of information represents what we shall call 'observation' or 'measurement'. The complete measurement process can involve a major change, which is the 'collapse' of the a priori probability and, consequently, may be accompanied by a large change in the entropy which applies to the ensemble representing the collective aspect of the dynamics of a given particle.

A measurement in classical mechanics can thus be represented as a combination (linkage) of a hint, which is the appearance of a given numerical value of the measured quantity, and subsequent recording and storage of this value. Overall, this is an irreversible process accompanied by an increase in the entropy of the external world.

We shall consider one more aspect of the interaction of a classical particle with a macroscopic object. We have assumed so far that the walls of a vessel or a thermostat have rigorously defined geometric dimensions and the probabilistic descrip-tion is required only for a particle. This is not generally true: the walls themselves are in random motion and, moreover, external instruments may be insufficient for accurate and constant tracking of the positions of macroscopic bodies. We shall assume that the positions of the walls are determined with an error and that there is a probability density which applies to the distribution of the wall positions (Fig. 4).

The continuous curve in Fig. 4 is the main (on average) boundary of a solid and the dashed lines represent its uncertainty. A 'cloud' of the probability distribution $p(x, v, t)$ is incident on this wall. After inelastic reflection by this body there may be one bit of information about the collision and the particle flies away from the body. If far from the body this particle is detected either by an instrument or as a result of a spontaneous irreversible process in which the actual appearance of a particle plays a definite role in the subsequent evolution of the system, the reflected particle should be regarded


Figure 4. Conversion of a 'cloud' of the probability distribution incident on a wall into one reflected particle. The solid wall has an uncertainty represented formally by dashed curves. The wavy line is the information on collision 'stored' in the macroscopic body and the continuous arrow is the information which escapes to the external world.
as present in a localised state: $p=\delta\left(\boldsymbol{r}-v_{0} t-\boldsymbol{r}_{0}\right) \delta\left(v-v_{0}\right)$, where $\boldsymbol{r}_{0}$ and $v_{0}$ are the coordinates immediately after collision at $t=0$.

As we can see, under these conditions after the particle bounces off from the boundary of the body the probability distribution of the particle collapses and nothing as yet occurs in relation to the uncertainty about the body itself: the particle may bounce off from the continuous line or from any of the dashed lines (Fig. 4). However, if not one but two incident particles are bounced off from the same point on the wall, the point of their intersection may be localised and information may be transferred to the external world that only the point on the continuous line is real. Two more particles will be sufficient to determine the angle of direction of the continu-ous line in the plane of the figure. Not only the probabilities of the motion of the particles but also the probabilities of the position of the solid collapse. We can see that only one information link with the external world is sufficient for the distribution function of the probabilities of the position of the solid body to collapse into a state corresponding to a quite definite position of the classical object (naturally, accurate to within thermal fluctuations of its boundary).

We shall now consider what happens when a quantum particle is incident and then reflected by a macroscopic body with a fixed boundary. If the incident 'cloud' $|\psi|^{2}$ of the density matrix of the particle is sufficiently extended, its behaviour may differ little from that of a classical particle. Irrespective of whether the incident state is pure or mixed, only a strongly localised 'cloud' is reflected from the boundary of the body as a result of an inelastic interaction (with a corresponding 'measurement' inside the body). The only restraint is the uncertainty of the position and momentum, corresponding to the Heisenberg relationship. However, if the boundary of the macroscopic body itself is subject to an uncertainty corresponding to an additionally extended wave function of this body, the situation changes. The subsequent reflections of particles one after another result in narrowing of the localisation of the macroscopic body. And since the spreading of the quantum packet of the macroscopic body is very slow, the final result is that the packet acquires features of a classical object. It is the interaction with the external world that converts macroscopic bodies into classical objects with localised wave functions. In general, when microparticles interact with other objects, we encounter two extremes: the microworld of light particles and the macroworld of very massive bodies, with some intermediate range between them representing bodies with small but macroscopic scales.

## 24. Measurement in quantum theory

We shall now consider in greater detail the problem of measurement in quantum mechanics. To do this, it is convenient to turn back to Fig. 1 which shows schematically the process of perception of the information resulting from an event $u_{i}$ in a physical system U. The rows A and F in Figs 1 and 2 demonstrate how this information may be perceived. Only the first row $U$ is important to us and it demonstrates the actual event $u_{i}$. Measurement in classical or quantum systems begins with such an event.

We shall adopt the logic and notation of the book by J Schwinger Quantum Kinematics and Dynamics (Reading, MA: W A Benjamin, 1970) and assume that an event $u_{i}$


Figure 5. Selective measurement of $M\left(u_{i}\right)$ corresponds, in quantum theory, to a random event when a physical quantity $U$, which can assume one of N values, is in the only cell $u_{i}$. The dashed outline represents the possibility of adding another physical quantity $v_{i}$ to the system.
corresponds to a measurement which results in the selection of $u_{i}$ out of a set of other possible realisations of the physical quantity $U$. We shall denote this event by the symbol $M\left(u_{i}\right)$. It follows from Fig. 5 that the symbol $M\left(u_{i}\right)$ simply means that a quantum particle, which can be in any one of $N$ possible states, enters the cell $u_{i}$ in this event. In other words, $M\left(u_{i}\right)$ corresponds to the collapse $U \rightarrow u_{i}$.

The process itself can be explained somewhat differently. Let us assume that a quantum particle with a wave function $\psi$ is incident on an instrument, i.e. on a macroscopic body which has a complex internal structure. At the entry to this instrument there is an analyser which specifically shows how the information on the incident particle can be perceived. In quantum mechanics this means the method of determination of the eigenvectors $\psi_{i}$ into which the $\psi$ function can be expanded: $\sum_{i} c_{i} \psi_{i}$. The eigenfunctions $\psi_{i}$ are assumed to be orthogonal and normalised, and each of them corresponds to an eigenvalue $u_{i}$ of a physical quantity $U$. Let there be $N$ such functions, so that the corresponding number of cells (Fig. 5) is $N$. The instrument for selective measurement in Fig. 5 retains only the function $\psi_{i}$ in each measurement event. All the other functions are annihilated, so that the collapse $\psi \rightarrow \psi_{i}$ takes place and this corresponds simply to the circumstance that the particle being measured has 'fallen' into the state $i$. It should be stressed that in such selective measurement followed by perception of the relevant information the absence of the particle from one of the remaining empty cells also implies nothing for the subsequent 'history': only the information with the value of the physical quantity $U$ equal to $u_{i}$ is real and meaningful.

A single measurement event can be regarded as purely random and it is obviously insufficient to establish the structure of $\psi$ function. Only repeated measurements can give the average values of the probabilities of realisation of the values $u_{i}$ of the investigated physical quantity $U$. Repeated measurements create a mixed ensemble which can be described by a density matrix, but this is not essential. Our attention will be concentrated on the measurement itself.

The symbol $M\left(u_{i}\right)$ thus corresponds to a selective measurement event which retains (or 'transmits') a particle with the value $u_{i}$ of $U$ and rejects (i.e. annihilates) all the other states.

The class of selective measurements can be extended in a natural manner. Namely, we can assume that a screen D is placed inside the instrument and it selects some part of the $\psi$ function incident on the instrument (Fig. 6). Then the cells outside the screen aperture will never be recorded. In particular, if the screen obstructs all the cells apart from the


Figure 6. Screen D can reduce the number of states that can be recorded by an analyser A (the inaccessible region is shaded).
$i$ th, it selects a subensemble with the value $u_{i}$ of the quantity $U$. If all the cells are obstructed, the result is a measurement which would be natural to denote by the symbol 0 . On the other hand, if the screen is opened completely and the actual value of $u_{i}$ is not recorded, then such an extremely nonselective measurement can be designated by the symbol 1 . It is then quite easy to introduce the operations of addition and multiplication of the measurement symbols. The addition operation creates a measurement which is not so selective and the result of measurements is a subensemble corresponding to all the values of $U$ occurring in the sum, i.e. one of these values cannot be distinguished from the others by a measurement of this kind. Measurement which corresponds to the sum over all $u_{i}$, i.e. which transmits the whole ensemble without dividing it into subensembles, should obviously be regarded as unity. The operation of multiplication of the measurement symbols implies consecutive measurements (from right to left). It follows from the physical meaning of such operations that the addition is commutative and associative, and the multiplication is associative.

The properties of elementary operations denoted by selective measurement symbols can be written in the form

$$
\begin{align*}
& M\left(u_{i}\right) M\left(u_{i}\right)=M\left(u_{i}\right), \quad M\left(u_{i}\right) M\left(u_{j}\right)=0 \text { for } i \neq j, \\
& \sum_{i} M\left(u_{i}\right)=1, \tag{107}
\end{align*}
$$

where the symbols 0 and 1 represent measurements which either reject or transmit all the systems. The meaning of such measurements implies that they have the following algebraic properties:

$$
\begin{align*}
& 1 \cdot 1=1, \quad 0 \cdot 0=0, \quad 1 \cdot 0=0 \cdot 1=0 \\
& 1+0=1  \tag{108}\\
& 1 \cdot M\left(u_{i}\right)=M\left(u_{i}\right) \cdot 1=M\left(u_{i}\right),  \tag{109}\\
& 0 \cdot M\left(u_{i}\right)=M\left(u_{i}\right) \cdot 0=0, M\left(u_{i}\right)+0=M\left(u_{i}\right) .
\end{align*}
$$

We have considered so far the one physical quantity $U$ whose action on the wave function $\psi$ is regarded in quantum mechanics as the operator

$$
U \psi=U \sum_{i} c_{i} \psi_{i}=\sum_{i} c_{i} u_{i} \psi_{i}
$$

Naturally, the number of physical quantities can be large. Therefore, there can be many measuring instruments. Two physical quantities $U^{(1)}$ and $U^{(2)}$ are regarded as compatible if measurement of one of them does not destroy the knowledge obtained by an earlier measurement of the other. They
are denoted by the symbols $U$ and $V$ in Fig. 5. The corresponding selective measurements $M\left(u_{i}^{(1)}\right)$ and $M\left(u_{j}^{(2)}\right)$, carried out in one sequence or another, create an ensemble of systems in which the quantities $U^{(1)}$ and $U^{(2)}$ have simultaneously quite definite values $u_{i}^{(1)}$ and $u_{j}^{(2)}$. We can therefore introduce a symbol for a composite selective measurement:

$$
\begin{equation*}
M\left(u_{i}^{(1)}, u_{j}^{(2)}\right)=M\left(u_{i}^{(1)}\right) M\left(u_{j}^{(2)}\right)=M\left(u_{j}^{(2)}\right) M\left(u_{i}^{(1)}\right) \tag{110}
\end{equation*}
$$

A complete set of compatible physical quantities $U^{(1)}, U^{(2)}, \ldots, U^{(k)}$ is understood to be a set of physical quantities in which each pair is compatible and at the same time there are no other physical quantities which are compatible with each of $U^{(s)}$. Then

$$
\begin{equation*}
M(u)=\prod_{s=1}^{k} M\left[u_{i}(s)\right] \tag{111}
\end{equation*}
$$

describes a complete measurement. It corresponds to the maximum number of tags which can be obtained simultaneously without uncontrolled measurement of the value of each of them in an earlier measurement. The symbolic properties of complete measurements correspond also to Eqns (108) and (109).

We shall make one further step on the way of expanding the class of measurements. We shall consider measurements which may alter the state. Let the symbol $M\left(u_{i}, u_{j}\right)$ denote a selective measurement in which only the systems in the state ' $j$ ' are accepted and the systems appear in the state ' $i$ '. The value of $u_{j}$ of the physical quantity $U$ is thus converted to $u_{i}$. It is naturally understood that $M\left(u_{i}, u_{i}\right)=M\left(u_{i}\right)$. We can easily show that measurements of the $M\left(u_{i}, u_{j}\right)$ type have the property

$$
\begin{equation*}
M\left(u_{i}, u_{j}\right) M\left(u_{k}, u_{l}\right)=\delta_{j k} M\left(u_{i}, u_{l}\right), \tag{112}
\end{equation*}
$$

where $\delta_{j k}=0$ when $j \neq k$ and $\delta_{i i}=1$. In fact, the factor $M\left(u_{i}, u_{j}\right)$ on the left transmits only the states for which the value of $U$ is $u_{j}$, i.e. $u_{k}$ must be equal to $u_{j}$. It should be noted that if the factors in Eqn (112) are transposed, the result is

$$
\begin{equation*}
M\left(u_{k}, u_{l}\right) M\left(u_{i}, u_{j}\right)=\delta_{l i} M\left(u_{k}, u_{j}\right), \tag{113}
\end{equation*}
$$

which shows that multiplication of measurement symbols (112) and (113) is noncommutative.

In addition to the complete set of $U$, there may be also other complete set $V, W, \ldots$, which are mutually incompatible. Measurement symbols of the type $M\left(u_{i}, u_{j}\right)$ may be derived for each of these sets. Naturally, the question arises of a selective measurement linking two such sets. The symbol $M\left(u_{i}, v_{j}\right)$ denotes a measurement process which rejects all the values, except $v_{j}$, of a physical quantity $V$ and which emits a system into the state $u_{i}$ of a physical quantity $U$.

We shall now construct a combined measurement $M\left(u_{i}, v_{j}\right) M\left(w_{k}, z_{l}\right)$. The final result of such a measurement is selection of the state $z_{l}$ and its transfer to $u_{i}$, i.e. this should be a selective measurement of the $M\left(u_{i}, z_{l}\right)$ type. However, during the first stage a system appears in the state $w_{k}$ and only the second stage represents selection of the state $v_{j}$. If $v_{j}$ and $w_{k}$ have the same value, i.e. $V=W$, we should obtain an intermediate factor $\delta_{j k}$ which is either zero or unity. However, in the general case of different values $V \neq W$, we can expect the transfer factor to fluctuate from one measurement to another. On the average, only a certain fraction of the states originating from $W$ is recognised in the second stage of measurement by the composite instrument. It is therefore natural to assume that on average

$$
\begin{equation*}
M\left(u_{i}, v_{j}\right) M\left(w_{k}, z_{l}\right)=\left\langle v_{j} \mid w_{k}\right\rangle M\left(u_{i}, z_{l}\right) \tag{114}
\end{equation*}
$$

Here, the quantity $\left\langle v_{j} \mid w_{k}\right\rangle$ is a number such that the matrix $\left\langle v_{j} \mid w_{k}\right\rangle$ represents the statistical relationship between the states of $V$ and $W$. In the special case when $V \equiv W$, we obviously have

$$
\begin{equation*}
\left\langle v_{j} \mid v_{k}\right\rangle=\delta_{j k} \tag{115}
\end{equation*}
$$

Since $M\left(u_{i}, u_{i}\right)=M\left(u_{i}\right)$ it follows from the general relationship (114) that

$$
\begin{align*}
& M\left(v_{j}\right) M\left(w_{k}, z_{l}\right)=\left\langle v_{j} \mid w_{k}\right\rangle M\left(v_{j}, z_{l}\right),  \tag{116}\\
& M\left(u_{i}, v_{j}\right) M\left(z_{l}\right)=\left\langle v_{j} \mid z_{l}\right\rangle M\left(u_{i}, z_{l}\right) \tag{117}
\end{align*}
$$

The appearance of numerical factors of the $\left\langle v_{j} \mid w_{k}\right\rangle$ type in Eqns (114)-(117) is of fundamental importance in quantum theory. Formally, these factors appear as the property of measurements, but we shall see later that they in fact lead to the principle of superposition of states, which is the main positive principle of quantum mechanics.

However, we shall begin first with some of the simplest consequences of the relationships obtained above. If we assume that $w_{k}=z_{l}$ in Eqn (116) and use the relationship $M\left(w_{k}, w_{k}\right)=M\left(w_{k}\right)$, we obtain

$$
\begin{equation*}
M\left(v_{j}\right) M\left(w_{k}\right)=\left\langle v_{j} \mid w_{k}\right\rangle M\left(v_{j}, w_{k}\right) . \tag{118}
\end{equation*}
$$

We shall now consider the triple products $M\left(v_{j}\right) M\left(u_{i}\right) M\left(w_{k}\right)$, which we shall sum over all intermediate values of $u_{i}$. Then, on the one hand, we obtain simply the quantity (118) and, on the other, in the case of the first pair of factors we can use the relationship (118) and then apply (117) so that the result is

$$
\begin{align*}
& \sum_{u_{i}} M\left(v_{j}\right) M\left(u_{i}\right) M\left(w_{k}\right)=\left\langle v_{j} \mid w_{k}\right\rangle M\left(v_{j}, w_{k}\right) \\
& \quad=\sum_{u_{i}}\left\langle v_{j} \mid u_{i}\right\rangle\left\langle u_{i} \mid w_{k}\right\rangle M\left(v_{j}, w_{k}\right) . \tag{119}
\end{align*}
$$

A comparison of the last two expressions yields

$$
\begin{equation*}
\sum_{u_{i}}\left\langle v_{j} \mid u_{i}\right\rangle\left\langle u_{i} \mid w_{k}\right\rangle=\left\langle v_{j} \mid w_{k}\right\rangle . \tag{120}
\end{equation*}
$$

In the special case when $V \equiv W$, we find that

$$
\begin{equation*}
\sum_{u_{i}}\left\langle v_{j} \mid u_{i}\right\rangle\left\langle u_{i} \mid w_{k}\right\rangle=\delta_{j k} \tag{121}
\end{equation*}
$$

So far these are purely algebraic relationships which are a simple consequence of the definitions of the measurement symbols introduced above. However, let us consider a simpler product $M\left(v_{j}\right) M\left(u_{i}\right) M\left(v_{j}\right)$. By analogy with the last expression in Eqn (119) (without summation over $u_{i}$ ), we obtain

$$
\begin{equation*}
M\left(v_{j}\right) M\left(u_{i}\right) M\left(v_{j}\right)=p\left(u_{i}, v_{j}\right) M\left(v_{j}\right), \tag{122}
\end{equation*}
$$

where

$$
\begin{equation*}
p\left(u_{i}, v_{j}\right)=\left\langle v_{j} \mid u_{i}\right\rangle\left\langle u_{i} \mid v_{j}\right\rangle=p\left(v_{j}, u_{i}\right) \tag{123}
\end{equation*}
$$

The meaning of the left-hand side of Eqn (122) is that the instrument first selects the state $v_{j}$ and then selects from it that part which can be transmitted as $u_{i}$; the instrument then emits again the state $v_{j}$. Therefore, $p\left(u_{i}, v_{j}\right)$ plays the role of the probability that the state $v_{j}$ passes through the whole instrument, i.e. that $p\left(u_{i}, v_{j}\right)$ can be interpreted as the
probability of detection of the state $u_{i}$ in the initial state $v_{j}$. The probability $p$ should be regarded as a real nonnegative quantity. It is therefore natural to assume that the factors in Eqn (123) are mutually conjugate:

$$
\begin{equation*}
\left\langle v_{j} \mid u_{i}\right\rangle=\left\langle u_{i} \mid v_{j}\right\rangle^{*} \tag{124}
\end{equation*}
$$

It follows from the relationships (121) and (124) that the matrix $S$ with the matrix elements $S_{i j}=\left\langle u_{i} \mid v_{j}\right\rangle$ should be regarded as unitary.

The above relationships can be written in a more usual form if we use the language of wave mechanics. Let $\psi$ be the measured wave function. It can be represented as a superposition of the eigenfunctions of the operators $U$ and $V$ :

$$
\begin{equation*}
\psi=\sum_{i} c_{i} \psi_{i}=\sum_{j} b_{j} \varphi_{j} \tag{125}
\end{equation*}
$$

where $U \psi_{i}=u_{i} \psi_{i}$ and $V \varphi_{j}=v_{j} \varphi_{j}$. Naturally, the measurement symbols should also be regarded as operators, so that

$$
\begin{equation*}
M\left(u_{i}\right) \psi=c_{i} \psi_{i}, \quad M\left(v_{j}\right) \psi=b_{j} \varphi_{j} \tag{126}
\end{equation*}
$$

Each of the functions $\varphi_{l}$ can be expressed as a linear superposition of the functions $\psi_{k}$ simply because of the properties of the eigenfunctions (it is assumed that the Hermitian operators $V$ and $W$ have all the necessary properties). Let us assume that the relevant relationship is

$$
\begin{equation*}
\psi_{i}=\sum_{j} S_{i j} \varphi_{j} \tag{127}
\end{equation*}
$$

We shall apply the operator $M\left(v_{k}\right)$ to this relationship. On the right-hand side of this relationship it leaves only one term with $j=k$ and on the left-hand side instead of $M\left(v_{k}\right)$ we can use an operator $M\left(v_{k}\right) M\left(u_{i}\right)$ since $M\left(u_{i}\right) \psi_{i}=\psi_{i}$. However, we can now apply a relationship of the type given by Eqn (118), from which it follows that

$$
\begin{equation*}
M\left(v_{l}\right) M\left(u_{i}\right) \psi_{i}=\left\langle v_{l} \mid u_{i}\right\rangle \varphi_{l} \tag{128}
\end{equation*}
$$

since $M\left(v_{l}, u_{i}\right) \psi_{i}=\varphi_{l}$. Consequently, we find that

$$
\begin{equation*}
S_{i l}=\left\langle v_{l}, u_{i}\right\rangle, \quad \text { i.e. } \quad \psi_{i}=\sum_{j}\left\langle v_{j} \mid u_{i}\right\rangle \varphi_{j} \tag{129}
\end{equation*}
$$

This is the relationship for going over from one representation to another. If, as usual, it is assumed that $\psi_{i}$ and $\varphi_{l}$ each represents an orthonormal basis, we readily find that

$$
\begin{equation*}
b_{j}=\sum_{i}\left\langle v_{j} \mid u_{i}\right\rangle c_{i} \tag{130}
\end{equation*}
$$

Derivation of the expression for $\left|b_{j}\right|^{2}$ shows that the superposition of terms of the $c_{l}^{*} c_{i}$ type appears on the righthand side of Eqn (130). Let us assume that the difference between the phases of mutually orthogonal amplitudes has no direct physical meaning and can be random. Then, averaging over the phases gives

$$
\begin{equation*}
\left|b_{j}\right|^{2}=\sum_{i} p\left(u_{i}, v_{j}\right)\left|c_{i}\right|^{2} \tag{131}
\end{equation*}
$$

where the transition probability is given by Eqn (123). Now, Eqn (131) looks like the relationship between the probabilities if the quantities $\left|b_{j}\right|^{2}$ and $\left|c_{i}\right|^{2}$ are interpreted as probabilities. An assumption is made (sometimes implicitly) that the phases of the individual modes in the expansion of the $\psi$ function in terms of the eigenfunctions are mutually random. Therefore, a measurement in quantum mechanics
intrinsically depends on the appearance of incoherence in the case of mutually orthogonal modes.

Let us return to Fig. 5. Selective measurement of $M\left(u_{i}\right)$ identifies only the cell in which there is a physical quantity $U$ if the corresponding information is received by the instrument. The wave function then collapses into $\psi_{i}$ and all the coeffi-cients $c_{k}$ (with the exception of $c_{i}$ ) disappear. The coefficient $c_{i}$ is transformed into unity. Therefore, within the framework of this logic a measurement seems to be composed of two closely related elements: the measurement symbol $M\left(u_{i}\right)$ 'points a finger' to the quantity $u_{i}$ and the perception of $u_{i}$ causes the collapse of the function $\psi \rightarrow \psi_{i}$ and of the probability distribution $\left|c_{k}\right|^{2} \rightarrow \delta_{i k}$.

The measurement symbol is thus constructed in such a way that it does not yet define fully the results of measurements. In particular, if in Fig. 5 a 'line' of cells $u_{i}$ of the quantity $U$ is followed by a similar 'line' of the quantity $V$, a process is possible in which the first line transmits only $\psi_{i}$, i.e. $\psi$ is transformed into $c_{i} \psi_{i}$ before it is measured. Therefore, the measurement symbol plays the role of a hint, i.e. the role of an indication which quantity may be measured. Then $\psi_{i}=\sum_{i}\left\langle v_{j} \mid u_{i}\right\rangle \varphi_{j}$ drops into the second 'line' and only there the final collapse $\psi \rightarrow \varphi_{j}$ with a corresponding composite measurement takes place. If this is repeated many times by changing the cells $i$, the result is Eqn (131) relating the probabilities $\left|b_{j}\right|^{2}$ and $\left|c_{i}\right|^{2}$ by the transition probability $p\left(u_{i}, v_{j}\right)$.

Since in a composite measurement $M\left(v_{j}, u_{i}\right)$ the order of operations is definite, an opportunity arises for arranging a sequence of measurements in time. In particular, if the level $U$ corresponds to the function $\psi(0)$, then the level $V$ in Fig. 5 may be attributed to the function $\psi(t)$ related to $\psi(0)$ by

$$
\begin{equation*}
\psi(t)=S \psi(0)=\exp \left(-\frac{i}{\hbar} H t\right) \psi(0) \tag{132}
\end{equation*}
$$

where $H$ is the Hamiltonian operator. Then also the $S$ matrix assumes the meaning of the operator of a transition from the initial time $t=0$ to some subsequent moment $t$. The wave function $\psi(t)$ satisfies the Schrodinger equation.

We can now understand more precisely the principle of superposition. The relationship given by Eqn (131) means that for any linear superposition of the wave functions, i.e. for any expansion of the $\psi$ function in terms of any one of its bases, there is a ready probabilistic interpretation of what occurs in the course of measurements: the squares of the amplitudes give the probabilities and the squares of the matrix elements give the transition probabilities. The whole evolution in time of these future possible elements of the measurement process is governed by the linear Schrodinger equation.

Why have we arrived at quantum mechanics? If we return once more to the beginning of the present section and follow again the logic of introduction of the measurement symbols $M\left(u_{i}\right)$, it is far from self-evident that they will lead to quantum mechanics rather than to classical mechanics. A hint of the appearance of elements of the quantummechanical approach first appears in Eqn (114) when a matrix element $\left\langle v_{j} \mid w_{k}\right\rangle$, i.e. a number different from the trivial zero and unity in Eqns (107) - (109), is introduced for the first time. In constructing the symmetric expression (122) we encounter the square of the matrix element (123) which can be interpreted as the transition probability.

However, the main and decisive step towards wave mechanics is made in going over from Eqn (130) to

Eqn (131), when the hypothesis of random phases of mutually orthogonal amplitudes makes it possible to draw the conclusion that only the squares of the amplitudes of the $\left|c_{i}\right|^{2}$ type have a real physical meaning and are interpreted as corresponding probabilities.

We shall now draw attention to the following circumstance. All the operations on the transition symbols are conventionally carried out from right to left. However, they can be read also from left to right naturally in a somewhat different sense. For example, the relationship (114) can be read from left to right on the assumption that the first operator on the left perceives $u_{i}$ and emits $v_{j}$, the second recognises $w_{k}$ and emits $z_{l}$, and so on. However, the same process can be constructed in the usual order, i.e. we can obtain

$$
\begin{equation*}
M\left(z_{l}, w_{k}\right) M\left(v_{j}, u_{i}\right)=\left\langle w_{k} \mid v_{j}\right\rangle M\left(z_{l}, u_{i}\right) \tag{133}
\end{equation*}
$$

According to Eqn (124), the matrix element that appears here is $\left\langle w_{k} \mid v_{j}\right\rangle=\left\langle v_{j} \mid w_{k}\right\rangle^{*}$. The meaning of the relationship (133) read element by element from right to left has been interpreted above as consecutive operations in time, i.e. in the direction from the past to the future. Consequently, if Eqn (114) is read from left to right this means that the operations are carried out in the direction from the future to the past. We can see that the entire difference between Eqns (114) and (133) lies simply in the complex conjugate of the operators. Therefore, beginning from an introduction of the composite-measurement symbols and later we have already been following the tracks of quantum mechanics by establish-ing a symbolic basis for fully reversible operations. The whole subsequent logic automatically leads to the conclusion of full reversibility of quantum-mechanical processes: time can be reversed simply by going over to complex conjugates. The relationships (122) and (123) can therefore be interpreted as the condition of 'encounter' of the past with the future. At this point an instrument can break the quantum-mechanical causal relationship between the past and the future, but the probabilistic relationship (123) is super-imposed on the results of measurements.

It is clear from the discussion that the logic of quantum mechanics establishes a rigid boundary between two classes of events.The fundamental object of quantum mechanics is the study of fully reversible processes beginning from a certain externally imposed state and right up to the entry into an instrument where a strongly irreversible process of the collapse of a wave function takes place. Wave mechanics describes the evolution of a wave function and predicts only the probabilities of some specific results of measurements. Therefore, wave mechanics represents powerful apparatus for the study of the possibilities rather than an 'earth-bound' theory of real processes. This can be seen particularly clearly in the what is known as the many-world interpretation of quantum mechanics [27], but we shall not discuss this topic.

Instead, we shall consider a specific example of a typical quantum process which is radioactive decay. This process is shown schematically in Fig. 7.

Let us assume that at a point R there is a radioactive nucleus R which emits $\alpha$ particles as a result of its decay. According to quantum mechanics, the wave function of an $\alpha$ particle consists of a part localised inside a nucleus and of a wave flux leaking outside the nucleus, represented formally by the wavy arrows in Fig. 7. Let us assume that an instrument detecting $\alpha$ particles consists of a set of thin


Figure 7. Schematic representation of an instrument for detecting an $\alpha$ particle emitted as a result of radioactive decay of a nucleus R. The dashed lines represent a variant of a measurement in which the exact position of the nucleus $R$ is not known.
plates each of which contains an enormous number of cells connected to counters that can record the passage of an $\alpha$ particle through one of the cells. All the cells operate in passive mode and transmit no information to the memory of an instrument if there are no signals in these cells.

Let us assume that at some moment $t$ a cell $u_{i}$ in the very first plate records the passage of an $\alpha$ particle and transmits this information for the processing and perception inside the instrument. Enormous changes should then occur in the $\psi$ function. Relative to the future there is a total collapse of the wave function into a compact wave packet, which then intersects all the other plates along the dashed line shown in Fig. 7. Consequently, only such cells correspond to the $\alpha$ particle. However, an equally surprising effect occurs with respect to the past. The whole wave function for the past also collapses into a wave packet moving in the direction towards the cell $u_{i}$ so that the particle velocity $v$ and the distance $L$ from R to $u_{i}$ can be used to find approximately the time $t-(L / v)$ when this collapse occurs. At this calculated moment not only does the whole external wave function collapse into a wave packet (apart from the wave packet travelling towards the instrument), but the wave function of the $\alpha$ particle inside the nucleus is annihilated.

If the initial wave function of the nucleus itself has not been localised but covers the dashed circle in Fig. 7, the first measurement in the first plate is insufficient to determine the $\alpha$-particle ray inside the dashed tangents to this circle. However, the second measurement in a subsequent plate fixes the ray itself along which the $\alpha$ particle is travelling. In fact, measurements on one $\alpha$ particle are insufficient to determine the position of the nucleus along the ray path and to find the moment of $\alpha$-particle decay, i.e. the moment of collapse of the wave function in the nucleus itself. There is some uncertainty in this respect. However, if a small grain of the radioactive element is available, then repeated measurements on $\alpha$ particles can fix the position of this grain (if this information has not been obtained by another method).

Let us now assume that instead of a radioactive nucleus R there is its classical imitator, ie. a small classical 'warm' trap which can emit classical particles with the same lifetime. We shall assume that the measuring instrument is constructed similarly. We can easily see that the corresponding classical 'decay' has much in common with that shown in Fig. 7. Instead of the flux of $\psi$ outside the 'nucleus', we now have a probability density flux. Measurement of $u_{i}$ again collapses its flux into a compact bunch, which in the future crosses all the plates of the instrument $P$ and in the past there is a collapse of the probability associated with the wave function collapse. Apart from the uncertainty relationships, the two processes are very similar. However, in the classical case we know exactly that there are particles with classical paths and
the probabilistic description is used simply because we do not fully know the 'lifetime' of a classical particle inside the trap. In the quantum case there is no exact path and the probability becomes an integral property of the evolution of the quantum object. It would seem that the quantum and classical descriptions should be identical when applied to objects and processes which are similar to one another, apart from the uncertainty relationship. However, this is not true.

Quantum mechanics has been developed as a theory of fully reversible processes, but only 'between measurements'. It does not include the concept of a classical body and there is no description of the processes of the interaction between microscopic and macroscopic bodies. The strength of quantum mechanics is that it describes universally all phenomena, but only in the framework of reversible processes. This is also its weakness, because the process of measurement is an external effect which is not catered for by this theory. We can see from Fig. 7 in which direction this theory should be extended.

Let us move the instrument $P$ increasingly further away from the nucleus R , so that $L \rightarrow \infty$. The process of collapse of the $\psi$ function near the nucleus is obviously not affected. In other words, the decay of the nucleus and the corresponding collapse of the $\psi$ function can be regarded as the result of a measurement carried out in the far future. If we retain the causality, the influence of the future can be only a causeless random process. Thus, a more general view of the evolution of quantum systems should include the causal development of events in accordance with the Schrodinger equation and 'causeless' collapse events representing the result of 'measurements in the future'. We shall consider a little later how these collapse events occur.

## 25. The Einstein - Podolsky - Rosen paradox

At the very early stages of the development of quantum theory, in 1935, Einstein, Podolsky, and Rosen published a paper [2] entitled 'Can quantum-mechanical description of physical reality be considered complete?" The authors formulated their famous paradox which has stimulated lively discussions that are still continuing.

Einstein, Podolsky, and Rosen (EPR) considered two quantum-mechanical systems which interact for a time with one another and then cease to interact. For example, these may be two particles which, having interacted at a short distance, fly far apart. If measurements are carried out on the first system, different results of these measurements correspond to different states of the second system described by different wave functions, although in fact there is no physical action on the second system. A pair of particles with a wave function which cannot be factorised into a product of wave functions of each of the particles is usually called an EPR pair. States with the wave function which cannot be separated into products of individual functions have been called by Schrodinger 'entangled states'. Such states are characterised by a fairly strong internal correlation. It is precisely because of this correlation that a measurement on one particle alters the wave function of the second particle even if the latter is very far from the first particle. At first sight this seems to be an absolutely paradoxical situation indicating the existence of some nonlocal interaction, known as the 'absence of local reality'. The formal resolution of this paradox was provided by Bohr almost immediately after publication of the EPR paper. Bohr
pointed out that in quantum mechanics it is not permissible to speak of a state independently of its environ-ment and, in particular, independently of the measuring instruments. If in the measurement of the momentum of one particle it is possible to predict uniquely the momentum of a second particle, then this occurs precisely because of a certain configuration of the instruments. The exact knowledge of the momentum of the second particle is obtained for a specific distribution of the instruments and it is this macroscopic configuration that makes it possible to reveal internal correlations in a quantum system.

However, this approach to the resolution of the EPR paradox has not satisfied all and other ideas have been put forward. We shall discuss this paradox by considering a more specific example shown in Fig. 8.


Figure 8. Decay of an unstable molecule $R$ into two particles and recording of a particle by an instrument P ; C is a gas cloud

An unstable particle $R$ dissociates into two particles $M$ and $m$. This may be the dissociation of an excited molecule in two atoms or the $\alpha$ decay of a radioactive nucleus or the decay of the Auger effect type when an excited atom emits an electron and drops into a stable state. If in the initial state the particle $R$ is at rest, the particles $M$ and $m$ emerging in opposite directions have identical but oppositely directed momenta. Let us assume that an instrument $P$ measures the momentum $\boldsymbol{p}$ of the particle m . Obviously, this procedure also measures the momentum of the particle $M$ although the distance between the particles may be very large and there is practically no direct effect of the instrument on the second particle. Measurement of the momentum of the particle M converts its wave function into a wide wave packet in the form of a plane wave with the momentum $\boldsymbol{- p}$ and with an amplitude which decreases slowly towards the edges of the packet. Clearly, the particle M can no longer move along the directions represented by the dashed and chain lines in Fig. 8. A real collapse of the wave function of the particle M occurs in the direction identified by the continuous line. If the momentum of the particle $M$ is now measured, it is found to be equal to $-\boldsymbol{p}$. However, if we try to measure the coordinate of the particle M , the appropriate instrument destroys the wide wave packet because the momentum of the particle M is perturbed. We can say that the instrument $P$ prepares the initial state of the wave function of the particle $M$ and then this wave function evolves in accordance with the Schrodinger equation. The most interesting aspect is that the instrument alters the wave function for the particle $M$ by its influence only on the second partner of the EPR pair, i.e. on the wave function of the particle $m$.

Let us look at Fig. 8 once again. If instead of a quantum system we have had a classical imitator, then the collapse would have been quite obvious and natural. In fact, the continuous, dashed, and chain lines would have corresponded to different random events. Their probabilities
would have been perfectly identical and then the recording of the particle $m$ by the instrument $P$ would have meant that this measurement would have involved precisely the event which can be recorded with the instrument $P$. This instrument would have obtained the information $I=-\sum_{i} p_{i} \ln p_{i}$, and the probabilities $p_{i}$ would have collapsed to unity for the continuous line and would have vanished for all the other states. This irreversible measurement process should be accompanied by irreversible reception processes, as pointed out above, so that the combined entropy of the particle and the instrument cannot decrease.

However, as demonstrated earlier, wide quantum packets behave in practice as localised particles. Therefore, the behaviour shown in Fig. 8 should not differ very greatly from that of the classical imitator. We shall consider the case when the mass of the light particle $m$ is considerably less than the mass of the heavy particle $M$. The velocity of the light particle is then considerably higher than the velocity of the heavy particle, so that the former is the first to escape to the external world. We shall now remove the instrument $P$ and replace it with a gas cloud $C$. The light particle reaches this cloud and is 'self-measured' becoming a participant of a nonequilibrium process. We can state this as follows: the separate wave packets of the light particle lose their mutual coherence because of the interaction with the cloud $C$ and the initially pure state of the light particle becomes mixed. The entropy of this particle increases from zero to $I=-\sum_{i} p_{i} \ln p_{i}$, where $p_{i}$ are the probabilities of the incoherent packets and $i$ is the packet number. The correlation between $M$ and $m$ ensures that the same occurs to the heavy particle: it loses the 'purity' of its state and acquires the same entropy $S$. If an irreversible collapse process now occurs in the cloud (for example, at the expense of the energy of the particle $m$ itself), the probabilities $p_{i}$ collapse, so that only one state with the probability of unity remains. This is accompanied by the collapse of the wave function of the particle $M$. We can say that this collapse is a direct consequence of the forbiddenness of the Schrodinger's cat state: there cannot be a superposition of states belonging to very different scenarios of history, i.e. to evolution of the nonequilibrium world. It should be stressed once again that the collapse of a wave function is related directly to the contact (direct or indirect) of a quantum object with the external world.

Macroscopically isolated systems with a low level of the thermal noise can also exhibit quantum properties. Such objects sometimes are called 'quantum cats' in scientific jargon [28, 29]. Only their contact with the classical world, for example, in the form of dissipation, can transform their statistical properties to the realm of classical logic $[30,31]$.

The collapse of a wave function is thus more likely to be the property of the environment of a quantum object and not of the object itself: it is the external world that first converts $|\psi|^{2}$ into a set of probabilities $p_{i}$ and then, by nonequilibrium evolution, transforms the probabilities into a set of zeroes and one unity for the state into which the collapse takes place. The collapse is a random process of the dice-throwing type. It is this process that is outside the framework of the traditional apparatus of quantum theory, which is a theory of reversible processes. The collapse events can be allowed for by explicitly supplementing the equations of evolution with appropriate operators which would allow for the real irreversible evolution of quantum systems with time. We shall see later how this can be done.

## 26. Bell's inequalities

Quantum theory has become the normal tool of modern physics. Nevertheless, there is a certain suspicion of doubt which sends tremors through the souls of theoreticians and experimentallists, and a direct check of the fundamental principles of wave mechanics is still a topical subject. In particular, it would be desirable to exclude by direct experiments an alternative but inconsistent theory of hidden variables. In 1964, J S Bell published a paper [3] in which the problem of hidden variables was considered from a new standpoint. He demonstrated that the hypothesis of the existance of hidden variables, i.e. the hypothesis of the statistics of events close to the classical ideas, makes it possible to derive a number of inequalities which do not agree with quantum theory and can be checked by direct experiments. We shall consider here only the simplest example of Bell's inequalities.

This example is based on an analysis of a correlation experiment of the EPR type, but in a simpler form proposed by D Bohm [32]. Let us assume that a molecule in a singlet state dissociates into two atoms, each of $\operatorname{spin} \frac{1}{2}$ (in units of $\hbar$ ). After dissociation the combined momentum of the two particles is still zero because of the law of conservation of momentum. Let us first check that the quantum approach does indeed lead to the EPR paradox. If the projection $\sigma_{x}$ of the momentum of the first particle along the $x$ axis is measured, the second particle is automatically found to be in the state $-\sigma_{x}$. In the classical case the projection of the spin of the second particle onto other axes, for example $y$ and $z$, would have vanished automatically. However, this is not true in the quantum case. The operator $\sigma_{y}$ does not commute with $\sigma_{x}$, and its eigenvalues are $\pm \frac{1}{2}$. Therefore, measurements of $\sigma_{y}$ for the second particle give values $\pm \frac{1}{2}$ with equal probabilities.

We shall now derive one of Bell's inequalities. We shall assume that the second particle has simultaneously all three components of the spin which in the measurements can give the values $\pm \frac{1}{2}$. We now make many measurements and obtain an ensemble of states. Let us use $X^{+}$for the case when $\sigma_{x}=\frac{1}{2}$ and $X^{-}$for the case when $\sigma_{x}=-\frac{1}{2}$, and so on. We then find that
$N\left(X^{+}, Y^{-}\right)=N\left(X^{+}, Y^{-}, Z^{+}\right)+N\left(X^{+}, Y^{-}, Z^{-}\right)$,
where $N$ is the number of measurements in an ensemble with properties identified by the parentheses. We similarly obtain
$N\left(Y^{-}, Z^{+}\right)=N\left(X^{+}, Y^{-}, Z^{+}\right)+N\left(X^{-}, Y^{-}, Z^{+}\right)$,
$N\left(X^{+}, Z^{-}\right)=N\left(X^{+}, Y^{+}, Z^{-}\right)+N\left(X^{+}, Y^{-}, Z^{-}\right)$.
It follows clearly from the above three relationships that

$$
\begin{equation*}
N\left(X^{+}, Y^{-}\right) \leqslant N\left(Y^{-}, Z^{+}\right)+N\left(X^{+}, Z^{-}\right) \tag{137}
\end{equation*}
$$

We shall now return to two particles which fly apart in a singlet state, so that the spins of the two particles are exactly opposite. For example, if the first particle is in the state $X^{-}$, then the second is in the state $X^{+}$. Similarly, if we reverse the signs of the second (i.e. the first when reading from right to left) terms inside the parentheses in Eqn (137), we obtain the relationship

$$
\begin{equation*}
n\left(X^{+}, Y^{+}\right) \leqslant n\left(Y^{-}, Z^{+}\right)+n\left(X^{+}, Z^{+}\right) \tag{138}
\end{equation*}
$$

where $n\left(X^{+}, Y^{+}\right)$denotes the number of particle pairs in the ensemble and in each pair the first particle has the spin
projection $\sigma_{x}=+\frac{1}{2}$, the second has $\sigma_{y}=+\frac{1}{2}$, etc. The relationship (138) is in fact one of Bell's inequalities. Naturally, not one but several such relationships can be written down. They correspond to the assumption of the existence of hidden variables and, consequently, they may be violated in the case of real quantum objects because the spin components are noncommutative.

Experiments carried out on different quantum objects have shown that Bell's inequalities are in fact violated, as expected on the basis of quantum theory. We shall mention here only two of this series of experiments. A Aspect and his colleagues [39] experimentally studied the correlations of the polarisation of photon pairs with the aid of analysers varying with time. They were switched at a rate faster than $c / L$, where $c$ is the velocity of light and $L$ is the distance between the detectors (analysers). The results were in full agreement with quantum mechanics: there are correlations at any moment in time and they are not transferred by any signal. Bell's inequalities are violated with an error amounting to five standard deviations. In the latest experiment reported by T E Kiess et al. [40] it was found that Bell's inequalities are violated with an error of 22 standard deviations. Many experiments on violation of Bell's inequality thus confirm reliably the principle of quantum mechanics and exclude 'local realism', i.e. they exclude the existence of hidden variables.

## 27. Quantum cryptography

The EPR paradox and the associated Bell's inequalities appear as certain kinds of nonlocal interactions, i.e. as force-free transfer of information over a large distance, possibly even at superluminal velocity. Therefore, the feasibility of constructing a superluminal telegraph has been discussed frequently in the literature. We can easily see that in the direct variant of a single EPR pair, the EPR paradox is not appropriate for this purpose. In fact, a measurement carried out on the first particle of the pair is purely random and cannot be predicted and controlled in advance. The situation does differ very much from the classical case, when black and white balls are concealed separately in different boxes and the boxes are far apart. Opening one of the boxes immediately reveals the colour of the ball inside it and that in the second box. There is no transfer of information in this case: this is an a priori known correlation of the probabilities. The quantum case differs only in that before the opening of the box the balls have no colour. However, as soon as the first box is opened and the ball is 'illuminated' (in the same way as happens to a photographic plate), it immediately reveals a colour, and the second ball then acquires an opposite colour inside the second box. In quantum mechanics other types of measurements with noncommutative operators can be carried out on the second box, but in the determination of the 'colour' it is not possible to detect anything else except the colour opposite to that of the first ball. Since the first event is purely random, a correlation between isolated EPR pairs cannot be used directly for instantaneous information transfer [5].

However, EPR correlations are interesting because of their possible use in encoding the transmitted messages. The fundamental idea is based on the circumstance that any interference with a quantum system, such as 'eavesdropping', destroys a pure state and therefore cannot remain undetected if the pure states are used correctly. This has been called
quantum cryptography [6-9]. We shall now consider two very simple examples of quantum cryptography. The simplest variant is probably that proposed by A K Ekert [6]. It is based on Bell's inequalities. Two participants of the process of information transmission and reception prepare many EPR pairs, atoms with spin $\frac{1}{2}$, which have a net spin of zero. These pairs are divided into two halves between the participants, who then measure the spins in accordance with a matched programme, so as to violate Bell's inequalities. It is found that a suitable selection of the measurement programme can maximize these violations [10]. In this correlated experimental scheme the information appears in the course of measurements and cannot be obtained by eavesdropping. Deliberate interference by a third person ignorant of the measurement programme is readily detected.

There are also more direct ways of quantum encoding information without recourse to Bell's inequalities [7, 8]. One of the most interesting variants of quantum information transfer has been proposed by C H Bennett et al. [9]. They have called it 'teleporting' of a quantum state with the aid of double, classical, and EPR channels. The term teleporting comes from science fiction where it describes a process by which a person or an object disintegrates at one place and complete information on its structure is transferred elsewhere, where an exact copy is formed.

In the present case the following process is considered. A singlet EPR pair of identical particles, each with spin $\frac{1}{2}$, is first prepared. One particle remains with the information sender and the second is sent to the receiver. Let us assume now that the sender wishes to teleport to the receiver a third particle in an unknown state $|\phi\rangle$ with spin $\frac{1}{2}$. It is found that the sender needs to carry out only a simultaneous measurement of the spin in the double system comprising the first partner of the EPR pair and the third particle. This can be done by means of Bell's basis [10], which contains four orthogonal states: one singlet and three triplet (with spins 0 and 1, respectively). However, the three particles, i.e. the EPR pair and the third particle in the unknown state $|\phi\rangle=a|\uparrow\rangle+b|\downarrow\rangle$, have a total of four orthogonal states and they can be represented by the superposition of certain eigenfunctions containing the amplitudes $a$ and $b$. After measurement, the sender can send the classical information on these amplitudes to the receiver, who can construct the state $|\phi\rangle$ by simple rotation of the instrument about the $x, y, z$ axes relative to the second partner of the ESR pair.

## 28. Random wave function

We shall now consider a somewhat more complex thought experiment by means of which we shall study, in a very simplified form, the evolution of a quantum system which is in a constant information coupling with the external world. We shall specifically assume that a quantum particle is 'measured' not by instruments, but by the interaction with the nonequilibrium environment. An appropriate experimental set up is shown in Fig. 9.

Let us assume that a microparticle of mass $m$ is enclosed between two horizontal walls, which are kept at somewhat different temperatures: $T_{1} \neq T_{2}$. Then, this particle collides in turn with the upper and lower wall and, if the interaction is inelastic, it can transport portions of heat from the hot to the cold wall. At the cold wall each such portion spreads into the interior of the wall material and can be 'measured' there. More accurately, such a portion of heat resembles, in a sense,


Figure 9. Microparticle moving along they axis collides consecutively and inelastically with the walls of a vessel. The temperature of the upper wallT $_{1}$ is not equal to the temperature of the lower wallT $T_{2}$ and the particle transfers (on average) heat from one wall to the other. A plug in the form of a macroscopic body $M$ is introduced into the left-hand end of the vessel so that motion along $x$ may result in a collision of the microparticle with the macroscopic body.
the process of radioactive decay, because it may collapse in a natural manner: we can say that it is 'measured in the future'. The collapse of each portion of heat causes collapse of the $\psi$ function of the microparticle and, consequently, the motion of the particle along the $x$ axis can be regarded as a sequence of random collapse events.

Since the interaction of the particle with the walls is random, the $\psi$ function is random. We can describe it by developing suitable rigorous methods of statistical physics, but at the moment it would be more useful to concentrate on the qualitative aspects of the phenomenon. We shall therefore adopt major simplifications and describe the $\psi$ function only by its moments, i.e. by the average values of the position of the wave packet $\langle x\rangle$, its width $\left\langle(x-\langle x\rangle)^{2}\right\rangle$, and velocity $u=\langle p\rangle / m$, where $p$ is the momentum and the angular brackets denote averaging with the weight $|\psi|^{2}$.

We shall assume that immediately after an inelastic collision with a wall the wave function assumes the form of a localised packet, Eqn (59), with the momentum $p=u / m$ :

$$
\begin{equation*}
\psi_{0}=\left(\pi b^{2}\right)^{1 / 4} \exp \left(i k x^{\prime}-\frac{x^{\prime 2}}{2 b^{2}}\right) \tag{139}
\end{equation*}
$$

where $k=u / \hbar m, x^{\prime}=x-\xi, \xi$ is the coordinate of the 'centre of mass' of the packet, and $b^{2}=\left\langle\left(x^{\prime}\right)^{2}\right\rangle$ is its width; the initial time is taken as zero. At the subsequent times the wave function evolves so that the packet spreads in accordance with Eqns (60) and (61):

$$
\begin{equation*}
\psi_{t}(x, t)=\exp \left(-\mathrm{i} \omega t+\mathrm{i} k x^{\prime}\right) \psi_{0 t}\left(x^{\prime}-u t, t\right) \tag{140}
\end{equation*}
$$

where $\omega=\hbar k^{2} / 2 m$ and the function $\psi_{0 t}\left(x^{\prime}, t\right)$ is given by Eqn (60), i.e. by

$$
\begin{equation*}
\psi_{0 t}=\left(\frac{b^{2}}{\pi a^{4}}\right)^{1 / 4} \exp \left(-\frac{x^{\prime 2}}{2 a^{2}}\right) \tag{141}
\end{equation*}
$$

where $a^{2}=b^{2}-\mathrm{i}(\hbar / m) t$. According to Eqn (61) the width of the packet increases as

$$
\begin{equation*}
\left\langle x^{\prime 2}\right\rangle=b^{2}+\frac{\hbar^{2} t^{2}}{m^{2} b^{2}} \tag{142}
\end{equation*}
$$

Let us assume that $\tau$ is the average time between consecutive 'measurements'. Then, on average, the width of the packet before a collision with a wall is given by Eqn (142) at $t=\tau$. The square of the width of localisation $b^{2}$ of the
wave packet immediately after an inelastic collision with the cold wall is governed by the processes of 'measurement' of a portion of heat, which occur in the wall itself. The average time $\tau$ is proportional to the distance between the walls and is a free (fitting) parameter. It can be selected in such a way that the second term in Eqn (142) is much greater than the first. In other words, after 'measurement' a wide wave packet contracts to dimensions much less than its previous width. We shall use $a_{0}^{2}$ to denote the quantity described by Eqn (142) at $t=\tau$, i.e. immediately before an inelastic collision. Averaging Eqn (142) during the time interval between collisions allows us to determine the average value $\overline{a^{2}}=\frac{1}{2}\left(a_{0}^{2}+2 b^{2}\right)$. Therefore, the wave packet oscillates in width between $b^{2}$ and $a_{0}^{2}$ and the mean-square width is $\overline{a^{2}}$.

Between two consecutive collisions the wave packet is displaced on average by $\Delta \xi=u \tau$. If we assume that the velocities $u$ have the Maxwellian distribution with a temperature $T=\frac{1}{2}\left(T_{1}+T_{2}\right)$, the average value of the square of the displacement along $\xi$ before the second collision will be $T \tau^{2} / m$. The displacements due to the widening of the packet and its free motion are statistically independent of one another, so that the probability $p(\xi, t)$ of the position $\xi$ of the wave packet is described by the diffusion equation:

$$
\begin{equation*}
\frac{\partial p}{\partial t}=D \frac{\partial^{2} p}{\partial \xi^{2}}, \quad D=\frac{1}{2 \tau}\left(a_{0}^{2}+\frac{T \tau^{2}}{m}\right) \tag{143}
\end{equation*}
$$

It therefore follows that a random walk of a quantum particle can be regarded as the diffusion of a wave packet with the probability distribution $p(t, \xi)$ of its centre of mass whose evolution is described by Eqn (143). The distribution of the velocities $u$ can be regarded as Maxwellian and the square of the width $\left\langle x^{2}\right\rangle$ is periodically either compressed to $b^{2}$ after a collision or expands to $a_{0}^{2}$ before the next collision.

We shall now consider how such a randomly moving quantum particle interacts with a macroscopic body whose mass is $M$. Let the wave function of the macroscopic body be $\Psi(X)$, where $X$ is the coordinate of the right-hand boundary of the body with which the particle may collide. Before the interaction with the particle the wave function $\Psi(X)$ can be regarded as stationary, because for $M \gg m$ the quantum spreading of $\Psi(X)$ can be ignored. In the course of its motion along the container the microparticle sooner or later collides with the macroscopic body and after reflection from it flies to the right. Since the coordinate of the macroscopic body is undetermined, superposition of the wave packets occurs.

Let $\Psi(X)$ be localised near $X \approx 0$, as shown in Fig. 9. We can easily see that if $X>0$ the reflected packet is found to be slightly further to the right, but for $X<0$ it is slightly further to the left, compared with the reflection by the boundary $X=0$. Therefore, after reflection the point $x-X<0$ assumes the position $x^{\prime}=-x+2 X>0$ and the wave function becomes

$$
\begin{equation*}
\Psi(X) \psi(-x+2 X, t) \tag{144}
\end{equation*}
$$

where the $\psi$ function of the microparticle is given by Eqn (140). In other words, in this approximation the parts of the wave packet which would have been outside the surface of the macroscopic body in the case of free motion, i.e. would have been at $x-X<0$, is simply reflected elastically from the $X$ plane and is found at the point $X-(x-X)=2 X-x$.

Before the next collision with the thermostat wall the wave function of the microparticle resembles the wave packet of Eqn (141). However, immediately after the collision this
wave function contracts into a narrow wave packet with the square of the width amounting to $b^{2}$. There cannot be any super-position of different wave packets: the 'measurement' leaves only one packet and annihilates all the other components of the $\psi$ function. A real collapse of the wave function takes place. The coordinate $X$ of the macroscopic body is also measured to within $a_{0}$ : after the collision accompanied by 'measurement' the coordinate of the microparticle $x$ is known to within $b$, so that the factor $\Psi(X)$ in the wave function (144) also collapses and we have $\left\langle(X-\text { const })^{2}\right\rangle=a_{0}^{2}$, i.e. the coordinate is $X \approx X_{0}=$ const to within $\frac{1}{2} a_{0}$.

The setup in Fig. 9 represents essentially an indirect measurement of the coordinate of the macroscopic body. After each such indirect measurement the initial function $\Psi(X)$ collapses into a packet of width less than $\frac{1}{2} a_{0}$. Before the collapse, the quantity $|\Psi(X)|^{2}$ plays the role of the classical probability density of the distribution of the average values of the coordinate $X$ for each of the localised wave packets which are generated randomly by collisions of the microparticle with a horizontal wall.

Our thought experiment suggests a way of developing a theory describing the phenomena involving natural processes of the collapse of the wave function. First, a random intrusion of collapses into the evolution of the system with time shows that the main role begins to be played not by quasistationary states, but by wave packets which spread with time. Apart from quantum-mechanical properties, these packets acquire features of extended classical objects, so that it is natural to apply to them a mixture of classical and quantum descrip-tions. The naturally appearing noncoherence of states imparts to macroscopic bodies the usual nature of determinate behaviour and 'throws bridges' across between the classical macroworld and the quantum microworld.

The reader naturally asks the question: why is it necessary to assume that $T_{1} \neq T_{2}$ ? It would seem that an ordinary thermostat can destroy the quantum coherence and create mixed states of the wave-packet type from any pure state. In fact, this is not so much the question of the real behaviour, but of logical grounding. Even if $T_{1}=T_{2}$, there are grounds for speaking of the absence of coherence, but we cannot state with certainty that the noncoherence is accompanied by 'measurement' and collapse of wave functions. Random collapse processes are natural in systems which are not in thermodynamic equilibrium.

It will probably be necessary to come back to this topic in future discussions. At this stage we shall consider in greater detail how to describe the evolution of a random wave function. Let $\psi(x, t)$ be a random wave function representing one-dimensional motion of a particle when random measure-ments take place. The evolution of $\psi$ with time can be described by the equation

$$
\begin{equation*}
\mathrm{i} \hbar \frac{\partial \psi}{\partial t}=H \psi+M \psi=-\frac{\hbar^{2}}{2 m} \frac{\partial^{2} \psi}{\partial x^{2}}+M(\psi) \psi . \tag{145}
\end{equation*}
$$

Here, $M$ is a random 'measurement' operator. We shall assume that $M$ intrudes randomly in the evolution of the system, i.e. in accordance with the Poisson distribution with an average time interval $\Delta t=\tau$ between 'measurements'. Immediately after a 'measurement', which will be assigned the time $t=0$, the function $\psi$ collapses into a packet of Eqn (139). Since this packet contains two parameters $u$ and $\xi$, the definition of the operator $M$ should include an instruc-
tion how these parameters should be dealt with. Before the next collision the solution $\psi$ is given by Eqn (140), i.e. $\psi_{t}=\exp (-\mathrm{i} H t / \hbar) \psi_{o}$, where $t$ is measured from the preceding collision. We shall assume that $b \ll a$, so that we can write approximately

$$
\begin{equation*}
p\left(u, u^{\prime}, \xi, t\right)=p\left(u, u^{\prime}\right)\left|\psi_{t}(x=\xi)\right|^{2} \tag{146}
\end{equation*}
$$

Here, $p\left(u, u^{\prime}, \xi, t\right)$ is the distribution of the probabilities of the quantities $u$ and $\xi$ in the case when this collision occurs after a time interval $t$ from the first; $p\left(u, u^{\prime}\right)$ is the probability of a change in the velocity of the packet from the value $u^{\prime}$ before 'measurement' to the value $u$ after 'measurement'. In the special case of complete 'forgetting' of the velocity before 'measurement', the probability $p\left(u, u^{\prime}\right)$ is independent of $u^{\prime}$, so that the quantity $p\left(u, u^{\prime}\right)=p(u)$ has the Maxwellian distribution with the temperature $T$. If there are many such 'measurements', the transition probability can be averaged over time:

$$
\begin{equation*}
p\left(u, u^{\prime}, \xi\right)=\int_{0}^{\infty} \exp \left(-\frac{t}{\tau}\right) p\left(u, u^{\prime}, \xi, t\right) \frac{\mathrm{d} t}{\tau} \tag{147}
\end{equation*}
$$

Here the first factor under the integral represents the probability of the absence of a collision up to the moment $t$ and the quantity $\mathrm{d} t / \tau$ is equal to the probability of 'measurement' in the interval $\mathrm{d} t$. If we go over to the time-average probability, the number of collisions in the interval $\Delta t$ should be taken as $\Delta t / \tau$. In this way the proposed logic automatically leads to a classical Markov chain and the quantum approach is needed only to find the probabilities of a transition from one 'measurement' to the next. As a result, in the case of many repeated measurements we obtain the diffusion equation (143) for $p(\xi, t)$ with the Maxwellian particle velocity distribution. It is possible to go over from these probabilities to the density matrix $\rho\left(x, x^{\prime}\right)=\left\langle\psi^{*}(x) \psi\left(x^{\prime}\right)\right\rangle$. However, we can see that there is no urgent need for this. Average wave packets, which occur in Eqn (147), play the role of the basis in which the density matrix is diagonal: $\rho\left(x, x^{\prime}\right)$ represents a random sampling of one of such packets with a probability which is prescribed from outside by the measurement operator $M(\psi)$. Consequently, the description of the statistical properties of a random wave function is influenced greatly by the properties of the 'measurement' itself and the free flight of the particle between one 'measurement' and the next determines only the diffusion coefficient $D$.

Let us see once again how the operator $M(\psi)$ acts on the right-hand side of Eqn (145). As established above, before the action of this operator the wave function looks like a relatively wide wave packet of Eqn (141) and its width is $\sim a_{0}$. Immediately after the measurement a packet of width $\sim b \ll a_{0}$ is formed. Thus, $|\psi|^{2}$ has the value $\sim 1 / a_{0}$ before the measurement and $\sim 1 / b$ after it. The pattern looks like a wave packet of width $\sim a_{0}$ cut into segments of width $b$. The number of such segments is $a_{0} / b$. At each measurement the particle falls into one of these segments and, therefore, the square of the wave function increases by the factor $a_{0} / b$. Therefore, the operator $M(\psi)$ is strongly nonlinear: it destroys the wave function in all the cells except one and in that cell it increases the amplitude of the function by $\left(a_{0} / b\right)^{1 / 2}$. This is a natural consequence of the collapse with retention of the normalisation. However, it is found that changes in the density matrix, and particularly in $\rho(x, x)=|\psi(x)|^{2}$, are much less remarkable (if they occur at all). Naturally, we need at least $a_{0} / b$ measurements to ensure that the particle
drops once into each of the segments formed by 'cutting' the original packet. Therefore, the average (over $a_{0} / b$ measurements) probability that the particle falls into a given segment is approximately the same as for the original packet. If we bear in mind that the probability of measurement is assumed by us to be proportional to $\left|\psi_{t}(x)\right|^{2}$, we find that this probability is exactly the same as after the measurement.

The operator $M(\psi)$ differs little from a macroscopic instrument in respect of its action on $\psi_{t}$ : it causes the collapse of the wave function in accordance with the rules of measure-ment theory in quantum mechanics, i.e. the collapse is into one of the mutually orthogonal states. If these measurements are treated in terms of the conversion of a pure ensemble into a mixed one, we can easily see that the density matrix $\rho\left(x, x^{\prime}\right)$ changes very little as a result of such measurements. In fact, the oscillatory dependence of the density matrix on $x-x^{\prime}$ is governed primarily not by the dimensions of the wave packets, but by the Maxwellian distribution of the momenta. Therefore, the description of a mixed state in terms of the density matrix is not sufficiently sensitive to determine whether the collapse does take place: averaging over an ensemble easily destroys the relevant very fragile information.

In the description of the measurements resulting in the destruction of the coherence and in the collapse of a wave function one needs a more rigorous approach in which the processes in question are described explicitly.

It is known from quantum theory that there are no universal measurements: the construction of an instrument or, more exactly, the macroscopic conditions during measure-ment, entirely determine which specific physical quantities can be measured and to which eigenvectors these quantities correspond.

We shall consider here just one more specific example of a natural measurement process, which is radioactive decay. In the $\alpha$-decay of a radioactive nucleus the wave function of the $\alpha$ particle consists of two parts: a function localised near the nucleus and decreasing with time as $\exp (-t / 2 \tau)$ and a stationary function outside the nucleus, describing the flux of the function to infinitely distant regions. The decay can be described by a random measurement operator $M(\psi)$. However, in this case the operator $M(\psi)$ should select at random a certain ray emerging from a nucleus and should form on this ray a wave packet which travels away along the ray at the velocity of the emitted $\alpha$ particle. The action of the operator $M(\psi)$ can be regarded arbitrarily as the result of a measurement carried out at infinity and, consequently, in the infinite future. In the Schrodinger equation it appears as a random collapse of a wave function. We can readily establish again that the collapse itself represents a strongly nonlinear process, which occurs at random, but if we consider just the quantity $|\psi|^{2}$, as representing the probability density of an ensemble of identical systems, then the relevant processes in question appear as very 'soft' that do not alter the average value $\left.\left.\langle | \psi\right|^{2}\right\rangle$ over the ensemble.

## 29. Mesoworld

We can now consider how the mesoworld is constructed, i.e. the range of physical phenomena between the classical macroworld and the quantum microworld. It is natural to start with the microworld which can be regarded as the base for describing all the other phenomena.

Quantum theory itself is logical and closed. The dynamical behaviour of microscopic objects with any number of particles, including quantised fields, is described by the generalised Schrodinger equation

$$
\begin{equation*}
\mathrm{i} \hbar \frac{\partial \psi}{\partial t}=H \psi \tag{148}
\end{equation*}
$$

where $\psi$ is the wave function of the whole system, dependent on a large number of coordinates, and the operator $H$ is the Hamiltonian of the system. The wave function can be regarded as a vector in Hilbert space, so it is sometimes called the state vector.

The Schrodinger equation describes the evolution of a quantum system beginning from a certain initial state and ending with a measurement, i.e. with the interaction between this system and a macroscopic instrument. The instrument is assumed to be a purely classical object and is described in terms of classical variables. Between these two limits there is a mesoworld, the understanding of which is absolutely essential if we want to link the microworld and the macroworld. We shall begin discussing this topic again with a thought experiment.

Let us assume that a radioactive tracer is deposited on the surface of a solid macroscopic body (Fig. 10a). If an $\alpha$ particle emitted by the tracer is 'measured' by an instrument, the measurement gives also the coordinate $y$ on the surface of the macroscopic body: after the measurement we can reliably say that the radioactive nucleus is known to be located at the point $A_{*}$ and not at some other point $B_{*}$. Before the measurement the points $B_{*}$ and $A_{*}$ may be equivalent either because the wave function of the macroscopic body has the same amplitude at these points (pure state) or simply because an external observer does not know at which of the two points, $A_{*}$ or $B_{*}$, the tracer is really located. In the first case the measurement on an $\alpha$ particle automatically leads to the collapse of the wave function of the macroscopic body, whereas in the latter case the a priori uncertainty of the probability of finding the tracer is lifted and the probability 'collapses', i.e. the external world receives information on the position of the tracer. It should be pointed out that the $\alpha$ particle itself could escape along any one of the rays emerg-ing from the point $\mathrm{A}_{*}$, but the ray identified in Fig. 10a by an arrow identified by $\alpha$. This means that the wave function of the $\alpha$ particle collapses on one of the rays.


Figure 10. If on the surface of a macroscopic body there is a radioactive tracer $\mathrm{A}_{*}$, detection of an $\alpha$ particle which is emitted can be used to measure the y coordinate of the macroscopic body (the dashed lines in Fig. 10a show the unrealised paths). In the presence of external perturbations the $\psi$ function of the a particle splits into noncoherent wave packets b and of these only one packet is real: the a particle 'falls' accidentally into this packet.

We can imagine that the selection of one of the rays occurs because weak external perturbations shift somewhat the relative phases of the wave functions for different rays, converting them into ray packets. If this is true then the phases may be shifted along the ray so that the wave function of the $\alpha$ particle may change from a spherical wave diverging from the point $A_{*}$ into a set of wave packets shown in Fig. 10b. The particle itself can be in just one of these packets identified by a solid circle (in Fig. 10b) and the other packets are just missed opportunities for the particle to occupy them. We can now easily show that this representation is correct. In fact, if we turn back to Eqns (130) and (131), we can show once again that the squares of the amplitudes are transformed into the probabilities only if the difference between the phases of the various orthogonal states become chaotic. Nonintersecting wave packets in Fig. 10b seem to represent such mutually orthogonal states. The phase chaos seems to split the coherent wave function into separate blocks and the value of $|\psi|^{2}$ for each of these blocks is converted into the probability. In terms of the transition from a pure to a mixed state such conversion of the phases to chaos seems to be a slightly changing transition: $|\psi|^{2}$ changes nowhere in magnitude and it is only converted into the probability, i.e. into the a priori possibility. A particle may drop into only one of the cells, i.e., into one of the wave packets, but an external observer does not know it yet. Only a real observation results in the collapse of the probability into one of the cells. In accordance with our terminology the dropping of a particle into a cell is simply the measurement symbol or hint that a future measurement will find the particle precisely in this cell. This is more an informational than a dynamical effect. Consequently, one bit of information is required to label one of the cells as the cell with the particle.

The distribution of the cells in Fig. 10b and the dimensions of the cells depend largely on the intensity of the external perturbations that destroy the coherence of the pure state. However, since the $\psi$ function of a particle emerging from a radioactive nucleus spreads progressively from the nucleus $\mathrm{A}_{*}$, in the real world the destruction of the coherence sooner or later does occur: if external perturbations are reduced, the cells in Fig. 10b simply move to the right and become larger. The particle itself sooner or later falls into one of the cells and, therefore, objectively (i.e. independently of the observer) the collapse of the $\psi$ function seems to take place. The subsequent measurement collapses only the probability and thus confirms that the radioactive decay has taken place and that the $\alpha$ particle flies along one of the possible rays.

The collapse of the wave function occurs into just one of the wave packets within which the phase is not greatly distorted. Therefore, initially the collapse of the $\psi$ function and then the measurement of $|\psi|^{2}$ relatively far from $\mathrm{A}_{*}$ determines (approximately) the whole pattern of the initial emission of the particle from the nucleus and its subsequent motion along a ray. We can say that the collapse of the wave function is spontaneous, but it is more logical to assume that it is induced by a 'measurement' at a very large distance from the nucleus $A_{*}$. We can now imagine a situation when this 'measurement' occurs very far, i.e. practically at infinity, from $\mathrm{A}_{*}$ and then the collapse of the $\psi$ function can be regarded as the result of a 'measurement' carried out at an infinite distance in the infinite future.

We can thus formulate the principle of collapse. It is similar to the principle of emission of radiation in
electrodynamics. We recall that in discussing the emission of an electromagnetic wave by a periodic dipole a singlevalued solution can be obtained by applying a boundary condition in the form of the absence of an advance (i.e. arriving from infinity) wave. This principle is not universal: if the radiation is emitted in a closed resonator cavity, then undoubtedly we have to allow for waves both leaving the source and arriving there. However, in open space an advance wave is assumed to be absent, which is equivalent to a very weak attenuation of the wave, i.e. a slight initial dissipation. The suppression of an advance wave makes the solution single-valued.

In the case of quantum particle we must also impose the condition of absence of a wave arriving from infinity. Moreover, we may assume that a 'measurement' occurs in infinity, i.e. that the collapse of a wave function takes place there. This distant collapse leads to the collapse of the $\psi$ function into a wave packet over the whole path of motion of a particle along a ray on which this particle is found. The net result is that the measurement itself seems to be unnecessary.

We can extend the principle of collapse to all the quantum particles escaping from macroscopic bodies to infinity. We thus seem to say that a wave function diverging to infinity cannot be a real object if this function corresponds to a single particle. Such a particle should manifest itself by the interaction with the external world only in a finite volume, because it corresponds to the only reality.

We shall now consider a situation in which the boundary of a solid is illuminated with a flux of light and the photons reflected by the boundary fly away to infinity (Fig. 11).

Let us assume that the whole surface of this body is blackened and only a small region near a point $A_{*}$ is specularly reflecting. This region is fully analogous to the radioactive tracer discussed above: the photons reflected from this region can be measured at infinity and the external environment then obtains information on the position of the point $A_{*}$. The position of this point can therefore be found irrespective of whether the initial position has been indeterminate in the classical sense (i.e. because of a certain distribution of the position probability) or because the macroscopic body has an extended wave function. In particular, the eye can be the measuring instrument and then the behaviour looks as if the observer has caused the collapse of the wave function of both a photon and of the macroscopic body. It is understood that in this case the collapse is in no way related to the observational and intellectual capabilities of the human observer: the detection of a reflected photon by any instrument has exactly


Figure 11. Boundary of a solid $M$ with a blackened surface carries a small mirror, $\mathrm{A}_{*}$, which reflects photons. 'Measurement' of these photons at infinity determines the coordinates of the point $A_{*}$ with an error up to one wavelength.
the same effect. Moreover, the participation of this photon in any irreversible process occurring as far as we please from $\mathrm{A}_{*}$ can be regarded as a 'measurement' accompanied by the collapse of the wave function.

Let us now assume that the mirror $\mathrm{A}_{*}$ is in fact the pointer of an instrument so that various coordinates $y_{i}$ of the position of $A_{*}$ on the surface of the body correspond to different values of $u_{i}$ of a certain physical object $U$. If the object $U$ is of quantum nature, then the quantities $u_{i}$ can also assume certain discrete values. Therefore, the photons reflected from $\mathrm{A}_{*}$, which can be 'measured' somewhere far from $\mathrm{A}_{*}$, cause the collapse of the $\psi$ function of the mirror, i.e. of the instrument pointer. This results in measurement of a physical quantity $U$ when it has a certain value $u_{i}$ and the result of this measurement may be recognised by the external world and, if desired, the reflection of the photon from $\mathrm{A}_{*}$ can be stored in a memory. We can easily see that two irreversible processes occur in such a measurement. Initially because the phases become chaotic the single coherent state splits into layers which represent a set of wave packets. The only $\psi$ function is split into segments with a slight phase distortion, but the particle (photon) can be in only one of the coherence regions. The wave function seems to split into a set of probabilities and only inside one of the packets is there a pure state of the particle. We can say that its wave function represents a somewhat 'softer' entity than the distribution of the probabilities or of information: different parts of the wave function of the pure state still retain a certain 'affinity via the phases'.

We can then easily see what the classical world represents. In the case of large macroscopic bodies with a large mass even few 'measurements' are sufficient for the localisation of the wave packet to be negligible compared with the dimensions of a body. Such bodies acquire 'forever' the characteristics of classical objects with precisely specified coordinates. The classical world is the world of large dimensions and masses. Therefore, any measurements which are then applied to classical objects are accompanied by the collapse of the wave functions in accordance with the recipes of quantum mechanics. The objectively occurring 'measurements' performed by the nonequilibrium solar radiation are quite sufficient for the reduction of the wave packets of classical bodies in our environment.

The mesoworld intermediate between the classical and quantum worlds is much more complex and much richer with regard to the physical phenomena which occur in it. The bodies in the mesoworld exhibit both classical and quantum properties. The mesoworld is the world of a set of alternative collapses of wave functions and probabilities, and it is the world in which the coherences are born and die.

It is therefore evident that there is no universal recipe for the description of the mesoworld: the variety of specific physical conditions in which a combination of classical and quantum processes is possible is now far too great. One of the possible approaches to this description is given in the preceding section, but there may be many other approaches.

It seems to be now somewhat clearer what processes cannot exist in the mesoworld. For example, superpositions of the Schrodinger's cat type should definitely be forbidden. In fact, the actual recording of radioactive decay excludes the possibility of its superposition with a nucleus which has not decayed. Even more so there cannot be a superposition of a live and a dead cat, because a living organism represents a macroscopic body which is closely linked to the external world and is therefore subject to continuous 'measurement'.

We can see that the logic of the mesoworld excludes many imaginable but physically impossible states of the macroworld when wide coherent wave functions, incompatible with real conditions, may be attributed to macroscopic bodies. The microworld may include mixed, i.e. classical-quantum, states and processes when the coherence of the packets may be annihilated and created again. This is why the mesoworld is a very mobile and live world of collective phenomena organised in a complex manner.

We shall now turn back to Fig. 11 and consider in greater detail how 'measurements' in the external world can localise the wave function of a massive particle. We shall consider a quantum system of two particles: a very heavy particle with a mass $M$ and a light particle with a mass $m \ll M$. We shall try to see how the measurements carried out on the light particle result in an indirect 'measurement', i.e. in compression of the wave function of the heavy particle. Let the wave function of the system be $\Psi(\boldsymbol{r}, \boldsymbol{R}, t)$, where $\boldsymbol{r}$ and $\boldsymbol{R}$ are the position vectors of the light and heavy particles, respectively. Let us assume that the tracer $\mathrm{A}_{*}$ in Fig. 11 corresponds to some local potential of the interaction $U(|\boldsymbol{r}-\boldsymbol{R}|)$ between the particles under consideration. Let this potential be not only strongly localised, but also not very deep, so that the Born approxima-tion can be used to describe the interaction, i.e. the scattering of the light by the heavy particle. More specifically, let us consider the following situation. Let a plane wave of the light particle with coordinate dependence of $\exp \left(i \boldsymbol{K}_{\mathrm{i}} \cdot \boldsymbol{r}\right)$ be incident on the body with the mass $M$. Here, $\boldsymbol{k}_{\mathrm{i}}$ is the wave vector of the incident wave (Fig. 12). We shall assume that the marker $\mathrm{A}_{*}$ is near the origin of the coordinate systems so that $R \ll r$.


Figure 12. Plane wave of a light particle is incident on the scattering centre $\mathrm{A}_{*}$ representing a 'marker' of a heavy particle. The scattered wave with the wave vector $\boldsymbol{k}_{\mathrm{f}}$ falls on an instrument P and it is measured there.

If, for example, $R=0$, the incident wave can be described as $\exp (i \boldsymbol{k} \cdot \boldsymbol{r})$ and the scattered wave is $r^{-1} \exp \left(\mathrm{i}_{\mathrm{f}} \cdot \boldsymbol{r}\right)$ [the dependence of the type $\exp (-\mathrm{i} \omega t)$ is ignored]. If the scattering centre $A_{*}$ is now displaced to the point $\boldsymbol{R}$, the incident wave acquires at the scattering point the factor $\exp \left(\mathrm{i} \boldsymbol{k}_{\mathrm{i}} \cdot \boldsymbol{R}\right)$, so that after the scattering a wave $|\boldsymbol{r}-\boldsymbol{R}|^{-1} \exp \left(\mathrm{i} \boldsymbol{k}_{\mathrm{f}}|\boldsymbol{r}-\boldsymbol{R}|+\mathrm{i} \boldsymbol{k}_{\mathrm{i}} \cdot \boldsymbol{R}\right)$ arrives from this point at the point $\boldsymbol{r}$. If an instrument $P$ is located very far, then the quantity $|\boldsymbol{r}-\boldsymbol{R}|$ in the pre-exponential function can be regarded simply as $L=$ const. This quantity is unimportant to our analysis, but the phase factor will be considered in greater detail. It will be assumed that the heavy particle is at rest and that the scattering of the light particle is elastic, so that $\left|\boldsymbol{k}_{\mathrm{f}}\right|=\boldsymbol{k}_{\mathrm{i}} \mid=k$. Moreover, for the sake of simplicity, it will be assumed that the incident wave propagates only along the $x$ axis, i.e. it has the form $\exp (i k x)$ and that the instrument is also located on the same axis. Moreover, we shall assume that $\boldsymbol{R}$ has only two coordinates, which we shall denote by the capitals $X$ and $Y$. If we assume that $y, Y \ll L$, we obtain the
following expression for the phase factor of the scattered wave, accurate to within $Y^{2}$ :

$$
\exp \left[\mathrm{i} k x+\frac{\mathrm{i} k}{2} \frac{(y-Y)^{2}}{L}\right]
$$

We shall now assume that this wave is measured by the instrument P , i.e. that a packet localised along the $y$ axis and of the envelope $\psi_{\mathrm{P}}=\exp \left(\mathrm{i} k x-y^{2} / 2 b^{2}\right)$ separates from this wave; here, $b$ is the localisation width of the packet. The wave function can be 'projected' onto this state simply if it is multiplied by $\psi_{\mathrm{P}}^{*}$ and the result then can be integrated with respect to $x$ and $y$. The final expression is proportional to an exponential function:

$$
\begin{equation*}
\Phi=\exp \left[-\frac{k^{2} b^{2} Y^{2}}{2\left(L^{2}+k^{2} b^{4}\right)}\left(1-\frac{\mathrm{i} L}{k b^{2}}\right)\right] \tag{149}
\end{equation*}
$$

This exponential function should be multiplied by the wave function of the heavy particle $\psi(Y)$. We can see that measurement on the light particle automatically localises the heavy particle with the form factor $\Phi$. If we ignore the imaginary part in the exponential function of $\Phi$ and retain only the real part of the form factor $\Phi_{\text {real }}$, it can be represented in the form

$$
\begin{equation*}
\Phi_{\text {real }}=\exp \left(-\frac{Y^{2}}{2 a^{2}}\right) \tag{150}
\end{equation*}
$$

where

$$
a^{2}=\frac{L^{2}+k^{2} b^{4}}{k^{2} b^{2}}
$$

Two limiting cases, $L^{2} \ll x^{2} b^{4}$ and $L^{2} \gg k^{2} b^{4}$, can easily be dealt with by means of the above expressions. In the former case (of a nearby instrument) we have $a^{2}=b^{2}$, i.e. the localisation of a heavy particle is identical to the localisation of a light particle. In the other limiting case, $L^{2} \gg k^{2} b^{4}$, we have $a=1 / \theta k$, where $\theta=b / L$ is the vertex angle of a cone supported by the instrument at the position of the particle. Since $k=2 \pi / \lambda$, where $\lambda$ is the wavelength, then even for small angles $\theta \approx 1 / 10$ the size of the localisation of the particle is of the order of the wavelength.

The complete expression (149) for $\varphi$ contains also an exponential function with an imaginary phase which may lead to additional localisation for $L^{2} \gg k^{2} b^{4}$ for small departures from coherence, but at this stage we shall not go into such detail.

In the above discussion the concept of an instrument is used frequently. However, it is quite unnecessary: if the localisation of the light particle occurs somewhere in the external world, i.e. because of the interaction with an unknown 'instrument', this process still leads to the localisation of the wave function of the heavy particle. If then $L^{2} \gg k^{2} b^{4}$ and the angle $\theta$ is small (but not negligible) the localisation of the heavy particle will be on the scale of the wavelength of the light-particle wave function.

Fig. 12 has just been used to consider the localisation along the $y$ axis, but the same instrument of finite size along the $z$ axis performs similar localisation of the wave function along $z$. If the instrument and the incident wave are both rotated by $90^{\circ}$, then localisation along the $x$ axis is possible. Naturally, in spontaneous 'measurements' on many incident particles a complete localisation along all directions can be achieved.

We shall assume that such 'measurements' on different incident particles are repeated one after another at a certain average frequency $\dot{N}=.1 / \tau$, so that after a long time interval $t$ a large number $N=\dot{N} t=t / \tau$ of such measurements takes place. Each measurement gives a new factor of the type given by Eqn (150), so that after $N$ measurements the factor becomes $\Phi^{N}$. The localisation region then becomes $a / N^{1 /}$ ${ }^{2}=a(\tau / t)^{1 / 2}$. For a very large mass the localisation continues until such a particle becomes a classical object with a fully determinate classical coordinate $Y$ and, consequently, $X$ and $Z$. If the mass is not very large, the process of localisation is compensated for by the opposite effect of the broadening of the wave packet between consecutive measurements. If we recall Eqn (61) for the broadening of a wave packet with time, we can estimate the minimum localisation of the particle with mass $M$ :

$$
\begin{equation*}
a_{\min } \approx\left(\frac{\hbar \tau}{M}\right)^{1 / 2}=\left(\frac{\hbar}{M \dot{N}}\right)^{1 / 2} \tag{151}
\end{equation*}
$$

This expression defines an approximate boundary between the macroworld and the mesoworld: if from the practical point of view the value of $a_{\text {min }}$ can be regarded as zero, then we are dealing with a macroscopic body which has classical coordinates. In the opposite case, a body has both the properties of a classical object, and may also exhibit wave properties, i.e. we are dealing with a mesoworld object.

It therefore follows from the standpoint of wave functions that the macroworld is the world of 'fossils': all the past 'measurements' carried out on the macroscopic bodies in the preceding epochs had the result that these bodies have 'lost' their former wave properties and have become classical objects with sharply defined coordinates. In 'contact' with the microworld and mesoworld they can only 'recall' their wave properties by a simultaneous 'collapse' of the wave function of the microworld and a random choice of the coordinate of the classical object, for example, the position of a pointer of a measuring instrument. The last process can be described also in terms of the model of the scattering of a light particle by a marker $A_{*}$, shown schematically in Fig. 12. All that is required is that the instrument P should distinguish various positions of the marker $\mathrm{A}_{*}$ in one coherent state. This can be done most simply by supplementing the instrument $P$ with a set of collimators and detectors, so that each collimator-detector pair is directed to its own point with the coordinate $Y_{i}$. These discrete coordinates $Y_{i}$ should be selected in such a way that the form factors of the different positions of the marker $\mathrm{A}_{*}$ do not overlap along the coordinate $Y$. Then the very first operation of one of the detectors measures one of the coordinates $Y_{i}$ (the relevant collimator is directed to this coordinate).

If instead of an instrument $P$ there is an external world, the scattering of just one light particle immediately causes the collapse of the wave function of the macroscopic object along the coordinate $Y$. If the microparticle is 'emitted' into the external world through a system of collimators, such that each of them is directed to just one of the discrete positions $Y_{i}$, then each microparticle collapse is accompanied by the collapse $Y \rightarrow Y_{i}$. Repetition of such 'measurements' can establish the statistical distribution of the coordinate $Y_{i}$. In this way we can find the density matrix of a mixed state of a macroscopic object after 'measurement', i.e. after destruction of the coherence by the scattering and subsequent 'escape' of the scattered microparticle to the external world.

## 30. Collapse of wave functions

Some examples of the collapse of wave functions are discussed above: in radioactive decay and in indirect 'measurements' as a result of the collapse of the wave function of a particle that escapes to the external world and loses its coherence there. As a rule, the collapse applies more properly to the probabilities of 'expectation' by the external world, i.e. it in fact corresponds to a random appearance of a specific value of one or another physical quantity. However, it is convenient to refer the collapse to the wave function itself. When this approach is adopted, the collapse can be described by a 'measurement operator' in the generalised Schrodinger equation (145). The term 'measurement operator' indicates that this operator should cover also the quantum measure-ment processes. In the latter case the operator is identical to the projection operator: it projects the initial state $\psi$ onto an eigenvector $\psi_{i}$, which corresponds to an eigenvalue $u_{i}$ of a physical quantity $U$ which is being measured. Therefore, in discussing real measurement processes we can use instead the von Neumann projection operators [33]. However, our more general problem requires not only the description of the deliberately made measurements, but of the natural processes of the collapse of wave functions which represent, together with the reversible dynamics in accordance with the Schrodinger equation, the real 'quantum history'.

In this section we shall consider what general properties should the measurement operator have in the simplest case of one or two particles. First of all, let us point out that in Eqn (145) the operator $M(\psi)$ is in the form of a term alongside the kinetic energy and the total energy $\hbar \omega$. Therefore, the operator should have the dimensions of energy, i.e. of the ratio $\hbar / t_{0}$, where $t_{0}$ is a characteristic measurement time. Therefore, interference of the operator in the evolution of a quantum particle should in general perturb not only the wave function, but also the energy of the particle. In other words, measurement may be accompanied by the exchange of energy with the external environment. However, this energy may be negligible if either the measurement takes a very long time or the collapse occurs into such wide wave packets that the corresponding energy change can be ignored. For example, when the physical quantity $U$ is such that its operator commutes with the Hamiltonian of the particle is being measured, there is no energy perturbation and this measurement need not destroy the stationary state.

The simplest example of measurement of a physical quantity is the Stern-Gerlach experiment in which the passage of a beam of particles across an inhomogeneous magnetic field splits this beam into components corresponding to different values of the projection of the spin onto the $z$ axis. Let us assume that, for example, the spin projection $\frac{1}{2} \sigma_{z}$ can have only two values: $\pm \frac{1}{2}$. Then, a detector will register only two lines. Let a wave packet of rectangular shape along the longitudinal coordinate enter the detector. If the length of the packet is $L$, then the spatial part of the square of the wave function is $|\psi|^{2}=1 / L$. Before reaching detectors (in this case there are two of them) the packet splits into two components with magnitudes $\sigma_{z}= \pm 1$. We shall denote by $v$ the velocity of the particle and by $\tau$ the time taken to record the particle inside the detector. If $\tau \ll L / v$, the wave packet can be 'cut' mentally into layers of width $l=v \tau$. Obviously, each of the detectors will analyse a packet layer by layer and, with the probability $l / 2 L$, one of the detectors will record a particle
and at the same time measure the value of $\sigma_{z}$. In this example we can assume that the probability of the distribution along $x$, equal to $|\psi|^{2}$, reaches the detector input, so that in fact the collapse of probabilities takes place. However, it is not very wrong to say that each of the detectors induces the collapse of a wave function into a layer of width $l$ in one of the detectors and destroys the rest of the wave function. The result does not change if we assume that even before reaching a detector the wave function collapses into one of its components, for example that with $\sigma_{z}=+1$ or with $\sigma_{z}=-1$, and then this component collapses into a layer of width $l$ in one of the detectors. Introduction of a suitable operator into the generalised Schrodinger equation makes it possible to describe the collapse of a wave function as a consequence of the collapse of probabilities (throwing of dice) when a quantum particle is information-coupled to the external world.

We shall now consider how an operator acts in the case of radioactive decay. Once again it is convenient to consider the probabilities and not the wave functions. Moreover, it is preferable to replace $|\psi|^{2}$ by a very large number of identical radioactive nuclei and consider the number of nuclei. Let the dependence $N=N_{0} \exp (-t / \tau)$ indicate how the number of radioactive nuclei decreases with time ( $N_{0}$ is the initial number of nuclei and $1 / \tau$ is the decay rate constant).

We shall select a time interval $\Delta t \ll \tau$ and consider what happens to the wave functions of $\alpha$ particles emitted from the nuclei. We shall select some value $R=R_{0}$ for the distance from the radioactive source and assume that the $\alpha$ particles are 'measured' outside this radius, i.e. they fall within the environment of atoms and molecules in which they follow nonequilibrium tracks. Let $v$ be the velocity of the $\alpha$ particles. Then in a time $\Delta t$ they travel a distance $\Delta R=v \Delta t$. Since the source power is $N / \tau$, then in a layer of volume $4 \pi R^{2} v \Delta t$ there are $\Delta N=N \Delta t / \tau$ particles.

Let the interval $\Delta R$ be selected in such a way that it corresponds to the localisation width of a wave packet after 'measurement'. If in a transverse direction the localisation width is also of the order of $\Delta R$, then in the layer in question of volume $4 \pi R^{2}{ }_{0} \Delta R$ there are $N_{S}=4 \pi\left(R_{0} / \Delta R\right)^{2}$ cells and each of the collapsing particles may fall into one such cell. The probability of reaching the cell per particle is

$$
\Delta p=\frac{\Delta N}{N} \frac{1}{N_{\mathrm{S}}}=\frac{1}{\tau} \frac{(\Delta R)^{3}}{4 \pi v R_{0}^{2}}
$$

We can easily see that the value of $1 / \tau$ per particle is equal to the flux across a surface of radius $R_{0}: 1 / \tau=|\psi(R)|^{2} \times 4 \pi R^{2} v_{0}$. Therefore, the probability $\Delta p$ is simply $\Delta p=|\psi(R)|^{2}(\Delta R)^{3}$. Consequently, the operator can be regarded as equal to the form factor which collapses a particle into one of the cells of volume $(\Delta R)^{3}$ and annihilates the $\psi$ function in all the other cells. This process can be regarded as repeated randomly near a sphere of radius $R_{0}$ at an average repetition frequency $1 / \Delta t=v / R_{0}$. If the decay process is slow, then the sum of the probabilities of falling into all the cells of volume $4 \pi R_{0}^{2} \Delta R$, i.e. $|\psi|^{2} \times 4 \pi R_{0}^{2} \Delta R$, may prove to be much less than unity. This means that the process of collapse near the sphere of radius $R_{0}$ should be repeated many times until the particle falls into at least one of the cells. However, if it falls there, then $|\psi|^{2}$ in the cell should be increased to $(\Delta R)^{-3}$ and elsewhere the wave function should become annihilated.

A packet which collapses into a volume $(\Delta R)^{3}$ spreads out further away from the source, so that after a long time its transverse localisation varies, in accordance with Eqn (61), as $\Delta x \propto \hbar t / m \Delta R$. The radial coordinate of a packet also
increases with time as $v t$ so that at high values of $t$ the packet remains within a cone of size

$$
\frac{\Delta x}{v t}=\frac{\hbar}{m v \Delta R}
$$

Hence, it is clear that the selection of $R_{0}$ is fairly arbitrary: provided only $R_{0} \gg \Delta R$, further propagation of wave packets depends little on the value of $R_{0}$. Moreover, the formation of packets can be represented by an evolution, reversed in time, down to the scale $R_{0} \geqslant \Delta R$. The measurement operator can therefore be regarded as performing random sampling on a wave packet with the aid of a form factor localised both along the radius and with respect to the angle. The actual moment of collapse is not so important: it may only be indicated as a 'trend' near a certain point of radius $R=v t$ and a wave packet may form with a shift along the time axis, i.e. much later. The 'trend' itself corresponds to one bit of information about the subsequent formation of a wave packet near a given point with definite angular coordinates and with the radial coordinate $R=v t$. Correspondingly, the measurement operator also permits freedom of change in the localisation width of a wave packet around randomly selected coordinates of its centre.

Real 'measurement' of an $\alpha$ particle my occur at a further large distance from a nucleus. However, the wave packet obtained then may be 'projected' into the future. Consequently, the collapse of a wave function can be regarded as a spontaneous contraction of a wave function into a wave packet as a result of 'premonition' of a future measurement.

The collapse is an irreversible process, but its evolution in time does not resemble the usual causal evolution from the past to the future, passing successively through all the intermediate phases: such a detailed description of the wave function would have required intermediate measurements and these measurements would have destroyed the coherence of the state. Therefore, the operator in the generalised Schrodinger equation is much more complex than the operators acting on physical quantities.

If allowance is made for the collapse, the relationship between the past and future may be more complex than that viewed from the standpoint of the usual classical determinism. Its analysis can conveniently be made by simplified measurements.

Let us consider, for example, an EPR pair in the Bohm variant when two particles with spin $\frac{1}{2}$ are prearranged initially, at $t=0$, in a singlet state and then they fly apart in different directions to large distances from one another. We shall assume that at the moment $t_{1}$, the $x$ component of the spin $\sigma_{x 1}$ is determined for the first particle and the result is $\sigma_{x 1}=+1$. This means that in view of the correlation the second particle collapses to a state with $\sigma_{x 2}=-1$. Since an observer cannot control the selection of $\sigma_{x 1}$ and the values $\sigma_{x 1}= \pm 1$ appear completely at random, the collapse of the second spin $\sigma_{x 2} \rightarrow-1$ cannot be used for instantaneous transfer of information from the first observer to the second. Let us now assume that the second observer measures the component $\sigma_{y 2}$ and with a probability 0.5 obtains the value $\sigma_{y 2}=+1$. Such a measurement on a correlated EPR pair would have led to a corresponding reduction of the state of the first particle: $\sigma_{y 1} \rightarrow-1$. If both observers are not too far apart, the collapse of the wave functions will be induced by the first (on the time scale) measurement. However, when the particles are far apart, the concept of simultaneity loses its absolute meaning: in the

Lorentz-invariant theory the sequence of events on a spacelike plane depends on the velocity of the coordinate system.

The question arises how the processes of the wavefunction reduction can be made consistent with the Lorentz invariance. We shall consider this in more detail. We shall begin with the measurement $\sigma_{x 1}=1$ on the first particle at a moment $t$. This measurement automatically leads to $\sigma_{x 2}=-1$ for the second particle. Since no action has been taken in relation to the second particle, it seems natural that the spin $\sigma_{x 2}=-1$ of the second particle is quite definite not only at times longer than $t_{1}$ but also when $t<t_{1}$. In other words, the measurement yielding $\sigma_{x 1}=1$ not only makes it possible to predict $\sigma_{x 2}=-1$ for a future measurement of $\sigma_{x 2}$, but it also matches the value $\sigma_{x 2}=-1$ in the past. However, similar reasoning can also be extended to the measurement of the spin of the second particle. The measured value of this spin $\sigma_{y 2}=+1$ means, because of the correlation, that $\sigma_{y 1}$ assumes the reliable value $\sigma_{y 1}=-1$, both after the measurement which gives $\sigma_{y 2}=1$ and before this measurement. We now seem to be faced with a paradox. Namely, if $\sigma_{x 1}=1$ is measured first, then the spin of the second particle assumes the value $\sigma_{x 2}=-1$ and the subsequent measurement yielding $\sigma_{y 2}=1$ destroys the coherence and correlation of the EPR pair. If the first measurement gives $\sigma_{y 2}=1$, then the value $\sigma_{y 1}=-1$ is established automatically for the first particle and the subsequent measurement $\sigma_{x 1}=1$ destroys the correlation of the EPR pair. The sequence of measurements can be different in different Lorentz-invariant coordinate systems. Therefore, it is not clear which scenario is the true one.

In fact, there is no major conflict. The intermediate correlated states of the $\sigma_{x 1}=1, \sigma_{x 2}=-1$ and $\sigma_{y 2}=1$, $\sigma_{y 1}=-1$ types are only the possible measured states and not the real ones. In reality the measured $\sigma_{x 1}=1, \sigma_{y 2}=1$ measured state does not conserve the correlation in the EPR pair, although at first sight the operators $\sigma_{x 1}$ and $\sigma_{y 2}$ seem to be commuting, i.e. simultaneously measurable. In fact, the measurement of one of the operators perturbs the measurement of the second one because of a correlation (for details see Ref. [11]). Therefore, the measurements of $\sigma_{x 1}$ and $\sigma_{y 2}$ give uncorrelated pairs $\sigma_{x 1}= \pm 1$ and $\sigma_{y 2}= \pm 1$ without any causal relationship between the measurements.

We shall now turn back to the correlated measurements of the $\sigma_{x 1}=1, \sigma_{x 2}=-1$ type. Here we have a strong correlation. Therefore, it is quite unimportant which of the measurements is carried out first: the second measurement gives a definite result both forward and backward on the time axis. The second measurement simply reveals the result which had been predetermined either by the past or future measurement on the first particle. Naturally, here we are speaking of the first and second particles in a fairly arbitrary manner: the two particles are equivalent. Therefore, in the case of measurement of $\sigma_{x 1}$ and $\sigma_{x 2}$, it is more logical to assume that the selection of $\sigma_{x 1}=1$ and $\sigma_{x 2}=-1$ is perfectly spontaneous, i.e. that it occurs before a real measurement. It is even more logical to assume that we are dealing here only with a hint symbol of a measurement which has not yet been converted into a real measured result. Only the measurement event itself selects the values of $\sigma_{x 1}$ and $\sigma_{x 2}$. The situation is similar to that of radioactive decay: an EPR pair seems spontaneously to select the polarisation matched to the instruments. This polarisation is then measured by the instruments.

It should be stressed that in all three cases under discussion - the measurement of $\sigma_{x 1}$ and $\sigma_{x 2}$ in an EPR pair, the radioactive decay, and the collapse of the wave function of a heavy particle in the case of an indirect 'measurement' when a correlated light particle escapes to infinity - no energy perturbation is introduced from outside. It is sufficient to assume one bit of information appears somewhere in the external world, which is fully compatible with the nonequilibrium state of this world. However, in the case of a purely local analysis, the collapse can be regarded as spontaneous. It can be described by a 'measurement' operator $M$ in the generalised Schrodinger equation. This operator can be regarded as random with a not very accurate indication of moment of time when this operator causes the collapse.

We shall now consider a more complex example of a system of many particles in the form of a low-density gas of 'classical' particles. We shall postulate that the gas of atoms (or molecules) with a density $n$ is at a temperature $T$. Let us assume that $\sigma$ is the cross section for collisions between particles, so that $r_{0}=\sigma^{1 / 2}$ represents the average size of an atom and $\lambda=1 / n \sigma$ is the average mean free path. The density parameter $\varepsilon=n r_{0}^{3}$ of a low-density gas is very small: $\varepsilon \ll 1$. Let us assume that initially the gas is classical, i.e. that the average localisation size of the wave functions (which will be denoted by $b$ ) is much less than the average distance between the atoms $n^{-1 / 3}$. That gas can be regarded as consisting of a set of separate wave packets. We shall now try to understand what happens to these wave packets and whether the initial pattern of separate wave packets is also retained subse-quently.

Let us assume that $v_{\mathrm{t}}=(T / m)^{1 / 2}$ represents the average thermal velocity. The quantity

$$
\tau=\frac{\lambda}{v_{\mathrm{t}}}=\frac{1}{n \sigma v_{\mathrm{t}}}
$$

is equal to the average time between pair collisions of atoms with one another. We shall select an arbitrary wave packet and consider what happens to it. If initially the value $b$ is less than $r_{0}$ and this inequality is retained for a time $\tau$, the wave packet evolves in the same way as a classical particle, colliding and being scattered by other atoms at a frequency $v=1 / \tau$. Therefore, after a time $\tau$ the initial wave packet disappears and becomes transformed into a scattered wave. However, if $b>r_{0}$, then the initial packet also disappears in the time $\tau$, but it is not converted into one but $\left(b / r_{0}\right)^{2}$ scattered waves, because in travelling a distance $\lambda$ the initial packet meets $\left(b / r_{0}\right)^{2}$ different atoms. In the second case, when $(b /$ $\left.r_{0}\right)^{2}>1$, the separate scattered waves lose their mutual coherence and the original atom drops into one of the scattered waves with a probability $\left(r_{0} / b\right)^{2}$. In the first case, when $b<r_{0}$, at the first scattering event the packet may still retain its individuality, but a few subsequent collisions are sufficient to make the scattered waves so complex that the coherence between its individual parts will be lost completely. The particle may again drop into just one of the scattered packets.

We can therefore draw the conclusion that in the course of its subsequent evolution the gas may retain its 'packet structure'. We may easily estimate also the dimensions of the (on average) steady-state wave packets. We can do this by applying the uncertainty relation. Let $b$ be the spatial localisation of the packet. Then, the uncertainty in the velocities $\Delta v$ amount to $\Delta v \approx \hbar / m b$ and the average size of a
packet does not change greatly in the time $\tau$ between the collisions if $\Delta v \tau \approx b$. Hence, we find that

$$
\begin{equation*}
b=\left(\frac{\hbar}{m v_{\mathrm{t}} \sigma n}\right)^{1 / 2}=\left(b_{0} \lambda\right)^{1 / 2} \tag{152}
\end{equation*}
$$

where $b_{0}$ denotes $b_{0}=\hbar / m v_{\mathrm{t}}$ of the order of the de Broglie wavelength of an atom moving at thermal velocity.

Let us now estimate the value of the quantity described by Eqn (152) for air at atmospheric pressure. If we substitute (in the cgs system of units) the values $\hbar=10^{-27}, m=3 \times 10^{-23}$, $v_{\mathrm{t}}=3 \times 10^{4}, \quad \sigma=10^{-15}$, and $n=3 \times 10^{19}$, we find approximately that $b \approx 2 \times 10^{-7}$. This is slightly less than the average distance between the atoms: $n^{-1 / 3} \approx 3 \times 10^{-7}$. We can see that air can be regarded arbitrarily as a gas of classical particles in the sense that the effective dimensions of the wave packets of the nitrogen and oxygen molecules do not exceed the average distance between these molecules.

Our discussion ignores the circumstance that in the case of identical bosons the wave function should be symmetrised over all the molecules. However, this circumstance does not greatly affect our conclusions, provided we indeed have $b<r_{0}$. In view of the identity of the particles the wave packet of one particle is indistinguishable from the wave packet of a second colliding particle, but the packet structure of the gas is still conserved.

In our discussion the key feature is the initial assumption that a particle may drop into one of the component and mutually uncorrelated parts of a complex wave packet. It is this assumption that underlies the statistical interpretation of the square of the wave function $|\psi|^{2}$. Conversion of one packet into a set of uncorrelated packets is equivalent to conversion of a pure state into a mixed one, i.e. we find that the wave function is apparently converted into probabilities. Since in this case there is no external measurement, formally there is no collapse of the probabilities either. The transfer of a particle into one of the component parts of a packet discussed here is not observable in the external world. Consequently, although the evolution of the wave packets with the corresponding collapse events does take place inside the gas, in the external world this appears only in the form of probabilities which should be described statistically with the aid of the density matrix. Since the events in which the wave functions of an equilibrium gas collapse are not observable, they cannot by themselves lead to the collapse of the distribution of probabilities for macroscopic bodies which are interacting with the gas. However, the gas molecules may destroy the coherence of the wave functions of macroscopic bodies (this is true, for example, of a particle in Brownian motion). This requires the destruction of the coherence between the individual components in a superposition of the wave functions of a macroscopic body. The phases of the components change with time if their frequencies are altered somewhat and then after a time $t$ it is possible to change the phase by $\sim 1$ if $\Delta \omega \approx 1 / t$. Since $\Delta \omega$ corresponds to an energy perturbation $\Delta E=\hbar \omega$, the loss of coherence requires an exchange of energy $\Delta E \approx \hbar / t$.

We can summarise the above discussion by drawing the following conclusions. In generalised quantum mechanics, which includes a description of irreversible 'measurement' processes, there are two types of collapse of wave functions. The first corresponds to a nonequilibrium 'measurement' process in the external world, it relates purely to information, can occur without any perturbation, and appears as
spontaneous contraction of a wave function. These are the processes that occur in measuring instruments. The processes of the second kind are associated with uncontrolled perturbations. They appear as the loss of coherence with energy exchange and are described in terms of conversion of pure ensembles into mixed ones. They include above all the processes associated with the thermal motion of particles. They can be described naturally in terms of the density matrix, which in the case of a weak perturbation of the diagonal terms is automatically converted into the probability distribution function.

In the case of the collapse associated with the appearance of information in the external world the most natural description is that involving the 'measurement' operator $M$. The action of this operator leads to random transitions of the wave function to new more localised states. This represents a kind of consecutive application of the projection operators. Each such projection operator can be described in terms of the Dirac notation in the form $P=\left|\psi_{\mathrm{c}}\right\rangle\left\langle\psi_{\mathrm{c}}\right|$, where $\psi_{\mathrm{c}}$ is the wave function after the collapse. The action of the operator $P$ is given by

$$
\begin{equation*}
P_{\mathrm{c}} \psi=\psi_{\mathrm{c}} \int \psi_{\mathrm{c}}^{*} \psi \mathrm{~d} x=\left|\psi_{\mathrm{c}}\right\rangle\left\langle\psi_{\mathrm{c}} \mid \psi\right\rangle \tag{153}
\end{equation*}
$$

if $\psi(x)$ depends only on one variable $x$. In general, the action of $P_{\mathrm{c}}$ on $\psi$-i.e. the projection of $\psi$ on $\psi_{\mathrm{c}}$ - is equal to the scalar product of the wave functions $\psi_{\mathrm{c}}$ and $\psi$, multiplied by the wave function $\psi_{\mathrm{c}}$. The operator $M$ differs from $P$ only in this respect: after projection the wave function should be normalised to unity. Moreover, as a rule the operator $M$ represents a set of many random projections. Since both functions $\psi$ and $\psi_{c}$, evolve with the same Hamiltonian operator $H$, in many cases it is not important at which precise moment the operator $M$ is applied: in particular, it can be assumed that it is applied from the future if in the interval from a given time $t$ to infinity there are no other collapse events.

## 31. Second quantisation

In systems of many identical particles it is more convenient to use the apparatus of second quantisation. We shall discuss this only to the extent necessary for subsequent considerations. Let us assume that $\psi_{N}\left(x_{1}, \ldots, x_{N}, t\right)$ is the wave function of identical particles dependent only on one of the spatial coordinates $x_{i}$ of the $i$ th particle out of the total number $N$. For simplicity, we shall assume that these particles satisfy the Bose-Einstein statistics, i.e. that the wave function is symmetric in the variables $x_{i}$. In the second quantisation language there is no need to fix the number of particles $N$ and it is possible to admit the possibility of creation and annihilation of the particles, as well as changes in the occupancy numbers of different quantum states. Therefore, instead of one function $\psi_{N}$ we can consider a set of functions $\psi_{0}, \psi_{1}, \psi_{2}, \ldots, \psi_{N}, \ldots$ with different number of particles. These functions can be arranged in column form and we can thus obtain what is known as the state vector in Fock space. Instead of considering the functions $\psi_{N}$, it is convenient to introduce operator functions acting on the state vector. These operator functions represent a secondquantisation field or simply quantum field. In the case of the Fock space under discussion, two operator fields are specified: the creation operator $\Psi^{+}(x)$ and the annihila-tion operator $\Psi^{-}(x)$ of one particle. Sometimes in the case of the
operator $\Psi^{-(x)}$ the minus index is omitted and then $\Psi^{+}$ denotes simply the Hermitian conjugate operator relative to $\Psi$. The operator $\Psi$ acting on the function $\psi_{N}$ transforms it into $\psi_{N-1}$, and the operator $\Psi^{+}$transforms $\psi_{N-1}$ into $\psi_{N}$. The explicit expression for these operators is

$$
\begin{align*}
& \Psi(x) \psi_{N}\left(x_{1}, \ldots, x_{N}\right) \\
& \quad=\frac{1}{N^{1 / 2}} \sum_{l=1}^{N} \delta\left(x-x_{l}\right) \psi_{N-1}\left(x_{1}, \ldots, x_{l-1}, x_{l+1}, \ldots, x_{N}\right), \tag{154}
\end{align*}
$$

$\Psi^{+}(x) \psi_{N-1}\left(x_{1}, \ldots, x_{N-1}\right)=N^{1 / 2} \psi_{N}\left(x, x_{1}, \ldots, x_{N-1}\right)$.
We shall add here one more useful relationship describing the commutator of the operators $\Psi(x)$ and $\Psi^{+}(y)$ :

$$
\begin{align*}
{\left[\Psi(x), \Psi^{+}(y)\right] } & \equiv \Psi(x) \Psi^{+}(y)-\Psi^{+}(y) \Psi(x) \\
& =\delta(x-y) \tag{156}
\end{align*}
$$

The action of these operators can be understood most simply by considering the example of noninteracting particles when the symmetrised function $\psi_{N}$ may be assumed to be the product of one-particle functions $\psi_{l}$ :

$$
\begin{equation*}
\psi_{N}\left(x_{1}, \ldots, x_{N}\right)=\prod_{i=1}^{N} \psi_{1}\left(x_{i}\right) \tag{157}
\end{equation*}
$$

This state is known as Bose-Einstein condensation. We can now see that the operator $\Psi(x)$ transforms the function $\psi_{1}\left(x_{l}\right)$ into the $\delta$ function $\delta\left(x-x_{l}\right)$ and then summation is carried out over all the functions $\delta\left(x-x_{l}\right)$. The operator $\Psi_{N}^{+}$ adds one more factor $\psi_{1}(x)$ to the product (157). We can say (and this is true, apart from a factor $1 / N^{1 / 2}$ or $N^{1 / 2}$ ) that the operator $\Psi(x)$ replaces one of the functions $\psi_{1}\left(x_{l}\right)$ with the $\delta$ function at a point $x_{l}$, so that one of the factors of the $\psi_{1}\left(x_{l}\right)$ type disappears. The operator $\Psi^{+}(x)$ simply adds one more particle with the wave function $\psi_{1}(x)$, i.e. it increases by unity the number of factors of the $\psi_{1}\left(x_{i}\right)$ type. The relationships given by Eqns (154) and (155) allow us readily to find the expression for the operator $\Psi^{+} \Psi$, which is called the particlenumber operator $N$ :

$$
\begin{align*}
N \psi_{N}\left(x_{1}, \ldots, x_{N}\right) & \equiv \Psi^{+}(x) \Psi(x) \psi_{N}\left(x_{1}, \ldots, x_{N}\right) \\
& =\sum_{l=1}^{N} \delta\left(x-x_{l}\right) \psi_{N}\left(x_{1}, \ldots, x_{N}\right) \tag{158}
\end{align*}
$$

The particle-number operator is diagonal and the integral is $\int \Psi^{+}(x) \Psi(x) \mathrm{d} x \psi_{N}=N \psi_{N}$, i.e. it is simply equal to the number of particles $N$ for each of the functions $\psi_{N}$. The meaning of the operators $\Psi(x), \Psi^{+}$, and $N$ can be understood better if we average them, i.e. if we integrate with respect to $x_{1}, \ldots, x_{N}$ with the weight $\left|\psi_{N}\right|^{2}$. In the simplest case we find that integration of Eqn (157) gives

$$
\begin{align*}
& \langle\Psi(x)\rangle_{N}=N^{1 / 2} \psi_{1}^{*}(x) \\
& \left\langle\Psi^{+}(x)\right\rangle_{N}=(N+1)^{1 / 2} \psi_{1}(x)  \tag{159}\\
& \langle N\rangle_{N}=N\left|\psi_{1}(x)\right|^{2}
\end{align*}
$$

In the more general case, when $\psi_{N}$ is not equal to the product (157), instead of $\psi_{1}(x)$ there is a one-particle function averaged over all $N$ variables with the exception of one. The operators $\Psi(x)$ and $\Psi^{+}(x)$ can be used to find readily the projection operators. Let us assume, for example, that as a result of measurement the initial function $\psi_{1}$ collapses into $\psi_{\mathrm{c}}$. If there is only one particle, then this collapse is due to the
collapse operator (153). If we are dealing with $N$ particles, we have to ensure the collapse of each particle in turn and even symmetrise the resultant wave function so as to allow for the Bose-Einstein symmetry. If Eqn (154) is integrated with respect to $x$ and the weight used is $\psi_{\mathrm{c}}(x)$, then the right-hand side is exactly the wave function in which one of the factors $\psi_{1}\left(x_{l}\right)$ is replaced with $\psi_{\mathrm{c}}\left(x_{l}\right)$, i.e. the particle for which the collapse $\psi_{1}\left(x_{l}\right) \rightarrow \psi_{c}\left(x_{l}\right)$ takes place. We can now obtain an equation of the type given by Eqn (153) if we just add the matrix element $\left\langle\psi_{\mathrm{c}} \mid \psi_{1}\right\rangle$. We can readily see that this can be done by integration of Eqn (155) with the weight $\psi_{\mathrm{c}}^{*}(x)$. In this way we obtain by a natural procedure the following expression for the projection operator:

$$
\begin{equation*}
P_{\mathrm{c}}=\int \psi_{\mathrm{c}}^{*}(x) \Psi^{+}(x) \mathrm{d} x \int \psi_{\mathrm{c}}(y) \Psi(y) \mathrm{d} y \tag{160}
\end{equation*}
$$

It is understood that this expression is accurate apart from a normalisation factor (if the final wave function is normal-ised). If we use the relationships (154) and (155), the action of the projection operator (160) on the wave function of $N$ particles $\psi_{N}\left(x_{1}, \ldots, x_{N}\right)$ can be written in the form

$$
\begin{equation*}
P_{\mathrm{c}} \psi_{N}\left(x_{1}, \ldots, x_{N}\right)=\sum_{l=1}^{N} \psi_{\mathrm{c}}\left(x_{l}\right) \int \psi_{\mathrm{c}}^{*}\left(x_{l}\right) \psi_{N}\left(x_{1}, \ldots, x_{N}\right) \mathrm{d} x_{l} \tag{161}
\end{equation*}
$$

In other words, the matrix element is calculated for the wave function $\psi_{N}$ relative to the wave function $\psi_{\mathrm{c}}^{*}\left(x_{l}\right)$ and it is then multiplied by $\Psi_{\mathrm{c}}\left(x_{l}\right)$; next, the resultant expression is symmetrised for all the particles. After the collapse the symmetry of the type given by Eqn (157) is lost but the new expression applied to the function (157) is still quite simple.

Introduction of the operators $\Psi$ and $\Psi^{+}$is convenient because their use makes the description of the dynamics of quantum particles very similar to the description of the dynamics of fields. Therefore, the operator $\Psi(x)$ is usually called the quantum field. Let us consider what this means.

First of all, some additive physical quantities can be expressed very simply in terms of the operators $\Psi(x)$ and $\Psi^{+}(x)$. For example, the momentum of one particle $p_{x}$ is expressed in the form of the operator

$$
p_{x}=-\mathrm{i} \hbar \frac{\partial}{\partial x}
$$

which is applied to the wave function. If we define the operator

$$
\begin{equation*}
P_{x}=\int \Psi^{+}(x) p_{x} \Psi(x) \mathrm{d} x \tag{162}
\end{equation*}
$$

it then follows from Eqns (154) and (155) that its action on $\psi_{N}\left(x_{1}, \ldots, x_{N}\right)$ is equivalent to the sum of the momenta of all $N$ particles. Similarly, the one-particle Hamiltonian $\int \Psi^{+}(x) H(x) \Psi(x) \mathrm{d} x$ averaged over $x$ and applied to $\psi_{N}$ is equal to the sum of the Hamiltonians of $N$ particles. Moreover, even in the presence of pair forces with the potential $U^{(2)}(|x-y|)$ the total Hamiltonian may be written in the following compact form

$$
\begin{align*}
H= & \int \Psi_{+}(x) H^{(1)} \Psi(x) \mathrm{d} x \\
& +\int \Psi^{+}(x) \Psi^{+}(y) U^{(2)}(|x-y|) \Psi(x) \Psi(y) \mathrm{d} x \mathrm{~d} y \tag{163}
\end{align*}
$$

Here, $H^{(1)}$ is a one-particle Hamiltonian in the field of forces with the potential $U^{(1)}(x)$, and $U^{(2)}(|x-y|)$ is the potential of the interaction between the particles. In the description of the dynamics of many particles it is frequently convenient to go over to the occupancy number representation. This is done as follows. Let $\psi_{n}(x)$ be the complete orthonormal basis. As this basis we shall consider, for example, standing waves of the type

$$
\psi_{n}(x)=\sin \left(\frac{\pi n x}{L}\right)
$$

However, in general, $\psi_{n}$ are complex variables. The operators $\Psi(x)$ and $\Psi^{+}(x)$, considered as functions of the variable $x$ can be expanded into the series:

$$
\begin{equation*}
\Psi(x)=\sum_{i} a_{i}^{-} \psi_{i}(x), \quad \Psi^{+}(x)=\sum_{i} a_{i}^{+} \psi_{i}^{*}(x) \tag{164}
\end{equation*}
$$

The functions $\psi_{n}$ represent an orthonormal basis, so that

$$
\begin{equation*}
a_{i}^{-}=\int \psi_{i}^{*}(x) \Psi(x) \mathrm{d} x, \quad a_{i}^{+}=\int \psi_{i}(x) \Psi^{+}(x) \mathrm{d} x \tag{165}
\end{equation*}
$$

The amplitudes $a_{n}^{-}$and $a_{n}^{+}$are also operators. The wave functions $\psi_{N}\left(x_{1}, \ldots, x_{N}\right)$, considered as functions of the variables $x_{1}, \ldots, x_{N}$, can also be expanded in the basis $\psi_{n}(x)$. It is then convenient to go over to the occupancy number representation. The wave function with the serial number $n_{i}$ can be denoted by $\left|n_{i}\right\rangle$ and it is possible to show that the action of the operators $a_{i}^{-}$and $a^{+}{ }_{i}$ is quite simple:

$$
\begin{align*}
a_{i}^{+}\left|n_{i}\right\rangle & =\left(n_{i}+1\right)^{1 / 2}\left|n_{i}+1\right\rangle \\
a_{i}^{-}\left|n_{i}\right\rangle & =n_{i}^{1 / 2}\left|n_{i}-1\right\rangle \tag{166}
\end{align*}
$$

The operator $a_{i}^{+} a_{i}^{-}=n_{i}$ is diagonal and the operator of the total number of particles is

$$
\begin{equation*}
N=\sum_{i} a_{i}^{+} a_{i}^{-}=\sum_{i} n_{i} \tag{167}
\end{equation*}
$$

We have ignored so far the time dependence of the wave functions $\psi_{N}$. In fact, these functions satisfy the Schrodinger equation, which can be written in the following compact form:

$$
\begin{equation*}
\mathrm{i} \hbar \frac{\partial \psi}{\partial t}=H \psi \tag{168}
\end{equation*}
$$

where $\psi$ is a column of the functions $\psi_{N}\left(x_{1}, \ldots, x_{N}\right)$ and the Hamiltonian operator is given by Eqn (163). The symbolic solution of Eqn (168) is

$$
\begin{equation*}
\Psi=\exp \left(-\frac{\mathrm{i} H}{\hbar} t\right) \psi(0) \tag{169}
\end{equation*}
$$

where $\psi(0)$ is the state vector at the initial time.
The equations for quantised fields become even more striking if we try to carry out all the averaging of the physical quantities with the weighting function $\psi(0)$. This implies going over from the Schrodinger to the Heisenberg representation: the wave function is regarded as independent of time, but all the operators acquire a time dependence. In the case of the operator $\Psi$ this means

$$
\begin{equation*}
\Psi(t)=\exp \left(\frac{\mathrm{i} H}{\hbar} t\right) \Psi_{0} \exp \left(-\frac{\mathrm{i} H}{\hbar} t\right) \tag{170}
\end{equation*}
$$

where $\Psi_{0}$ is the time-independent operator introduced earlier by means of Eqns (154) and (155). It follows from the relationship (170) that

$$
\begin{equation*}
\mathrm{i} \hbar \frac{\mathrm{~d} \Psi}{\mathrm{~d} t}=-[H, \Psi] \tag{171}
\end{equation*}
$$

where the square brackets denote commutation. Eqns (170) and (171) describe the evolution of a pure state. They are fully reversible and ignore all the interactions with the external world. If an interaction of this type can be reduced to just heat exchange with the external medium and the loss of coherence of the states, the corresponding processes can be described simply by introducing the density matrix of mixed states. However, if these processes include nonequilibrium collapse events, the situation becomes more complex. Each of these events can be described by the projection operator of the type given by Eqn (159). It can be represented in the form $P_{\mathrm{c}}=A_{\mathrm{c}}^{+} A_{\mathrm{c}}$ where the amplitudes $A_{\mathrm{c}}$ and $A_{\mathrm{c}}^{+}$are, respectively,

$$
\begin{equation*}
A_{\mathrm{c}}=\int \psi_{\mathrm{c}}(x) \Psi(x) \mathrm{d} x, \quad A_{\mathrm{c}}^{+}=\int \psi_{\mathrm{c}}(x) \Psi_{\mathrm{c}}^{+} \mathrm{d} x \tag{172}
\end{equation*}
$$

The appearance of random projection operators can be allowed for in Eqn (168) by an additional term of $M \psi$. The generalised Schrodinger equation then ceases to be reversible and the corresponding evolution of the quantum field, described by Eqns (170) and (171), takes place only between the collapse events. In general, the evolution of a quantum system becomes much more complex and, most importantly, ceases to be reversible. The irreversibility is the result of an information link or coupling between a given quantum system and the nonequilibrium external world.

## 32. Molecular chaos

Our main task is to describe nonequilibrium processes, which may be of quantum or classical nature. In these processes we shall distinguish two types of behaviour: the approach to an equilibrium because of dissipation, accompanied by an increase in the corresponding entropy; and the reverse process of development of self-organisation with a reduction in the entropy at the expense of an increase in the environment entropy. It is natural to begin with the first of these irreversible processes, i.e. with the approach to equilibrium. This approach represents monotonic destruction of order or unavoidable dispersal of the initial fluctuations only if the deviation from equilibrium is small. In systems far from equilibrium there are usually more complex nonlinear processes in which in the case of some of the degrees of freedom there is no destruction but complication of the structures. Naturally, it is necessary to begin with the simplest case of a small deviation from equilibrium. Moreover, it is natural to start from the simplest physical system.

This simplest system is a low-density gas of classical particles. Let us assume that the density parameter of the gas is $\varepsilon=n a^{3}$ (where $n$ is the density and $a$ is the average size of the molecule) is much less than unity. This means that the average distance $l=n^{-1 / 3}$ between the molecules is considerably greater than the size of the molecules. Therefore, the gas molecules travel freely a distance $\lambda=1 / n a^{2} \gg a$ and they collide with one another relatively rarely.

Under these conditions the behaviour of the gas is described by the Boltzmann equation

$$
\begin{equation*}
\frac{\partial f}{\partial t}+v \cdot \nabla f=\operatorname{St}(f) \tag{173}
\end{equation*}
$$

Here, $f(\boldsymbol{r}, v, t)$ is the distribution function of the particle velocities at a point $\boldsymbol{r}$ and at time $t$. The collision term $\operatorname{St}(f)$, quadratic in $f$, describes the evolution of the distribution function when allowance is made for pair collisions of molecules with one another. The collision term is derived on the assumption that both particles flying towards one another have the same distribution function $f$ and that the collisions are random. This approximation postulates that 'molecular chaos' is established in the gas, i.e. that molecules are in no way correlated before collisions. This assumption is perfectly natural from the physical point of view, but it needs a definite justification which will be discussed below.

The collision term in Eqn (173) explicitly introduces an irreversibility, as demonstrated by the famous $H$ theorem of Boltzmann. The question is frequently asked how does this irreversibility arise: after all, in finding the change in the distribution function owing to pair collisions it would seem that there have been no explicit assumptions about irreversibility. Moreover, the collision term itself is derived on the assumption that the dynamics of the colliding particles is reversible. Consequently, it is the assumption of 'molecular chaos' that leads to irreversibility. It is necessary to consider the origin of 'molecular chaos' and how such chaos is then involved in the irreversible evolution of the distribution function.

We shall first consider the process of approach to thermodynamic equilibrium. The kinetic equation (173) can be written very roughly in the form

$$
\begin{equation*}
\frac{\partial f}{\partial t}+v \cdot \nabla f=-v\left(f-f_{0}\right) \tag{174}
\end{equation*}
$$

where $f_{0}$ is the Maxwellian distribution with density $n$, average velocity $\boldsymbol{u}$, and average energy $\left\langle\frac{1}{2} m v^{2}\right\rangle$, which are exactly the same as analogous quantities in the function $f$. On the right-hand side of Eqn (174) the coefficient $v$ represents the average frequency of pair collisions: $v=v_{t} / \lambda$ and $v_{\mathrm{t}}=(T / m)^{1 / 2}$. If the distribution function $f$ is homogeneous in $\boldsymbol{r}$, it then follows from Eqn (174) that any deviation from $f_{0}$ decays with time as $\exp (-v t)$. However, if the initial distribution function depends both on $v$ and on $r$, its evolution is much more complex. In particular, for spatial scales considerably greater than the mean free path $\lambda$, the collisions establish quite rapidly, in a time $\tau=1 / v$, a local equilibrium distribution function with the density $n$, tempera-ture $T$, and velocity $\boldsymbol{u}$ dependent on the spatial coordinates.

In the subsequent time the evolution of the quantities $n, \boldsymbol{u}$, and $T$ is described by the equations of gas dynamics. If allowance is made for the terms proportional to $1 / v$, these equations contain dissipative terms with the viscosity and thermal conductivity. They lead to a much slower decay of fluctuations with time. Since the kinematic viscosity and thermal diffusivity are of the order of $\chi \approx v_{\mathrm{t}}^{2} / v$, perturbations of the order of $\lambda$ decrease only as $\exp \left(-\chi t / \Lambda^{2}\right)=$ $\exp \left(-v_{t}^{2} t / v \Lambda^{2}\right)$. We can see that the higher the collision frequency $v$, the slower the decay of the corresponding perturbations. The slowest decay is exhibited by perturbations of the maximum scale $\Lambda=L$, where $L$ is the size of the container enclosing the gas. In the final analysis, it follows from the kinetic equation (173) that all the fluctuations should disappear and the function $f$ should eventually
become the equilibrium Maxwellian distribution independent of the spatial coordinates.

However, the derivation of the Boltzmann equation (173) is subject to a certain inaccuracy. Although the collision term describes, by definition, random collisions of molecules, only the time-average part of this term is retained. It will be more correct to consider the collision term a random quantity, so that Eqn (173) should be supplemented by a fluctuation term equal to the difference between the true random collision term and its average part $\operatorname{St}(f)$. It is found that this correction, which plays a role of an external random force, prevents complete relaxation of $f$ and continuously reactivates thermal fluctuations. It is convenient to divide these fluctuations into two classes: individual and collective. The individual fluctuations occur on a scale less than the mean free path when the motion of the molecules can be regarded as free. However, on a scale greater than $\lambda$ we have to speak of collective fluctuations.

These collective fluctuations can be described by the equa-tions of gas dynamics. The corresponding motion of the gas splits into two classes: acoustic vibrations and incompressible motion of the type represented by small vortices. Each thermal mode of the acoustic vibrations has an energy $T$, equal to the sum of the kinetic ( $\frac{1}{2} T$ ) and potential $\left(\frac{1}{2} T\right)$ energies. For each vortex there is an energy $\frac{1}{2} T$. The number of such modes is a quantity of the order of $(L / \lambda)^{3}$. Hence, it is clear that the energy of thermal motion of collective modes, $\sim(3 T / 2) L^{3} / \lambda^{3}$, is much less than the total of thermal energy $(3 T / 2) L^{3} n$. Their ratio is $1 /$ $\lambda^{3} n=a^{6} n^{2}=\varepsilon^{2} \ll 1$.

Therefore, if in Eqn (173) we replace the average collision term with its true random value, we find that the distribution function evolves not to the Maxwellian form $f_{0}$, but to the quantity $f_{0}+\widetilde{f}$, where $\tilde{f}$ corresponds to thermal fluctuations. In the presence of these thermal fluctuations a small fraction, $\varepsilon^{2} \ll 1$, of the energy is related to the collective degrees of motion and the rest, $(3 n T / 2)\left(1-\varepsilon^{2}\right)$, of the thermal energy is represented by the individual degrees of motion. The total energy of thermal motion is $\frac{3}{2} n T$, in accordance with the laws of thermodynamics.

Two questions remain: why nonthermal perturbations of the distribution function relax and whether thermal fluctuations can be regarded as reversible. In order to understand the basis for these questions it is desirable to consider one further possible approach to the description of the dynamics of a gas with the aid of the Liouville equation. This equation describes the evolution of the distribution function $W\left(\boldsymbol{r}_{1}, v_{1} ; \boldsymbol{r}_{2}, v_{2} ; \ldots ; \boldsymbol{r}_{N}, v_{N}, t\right)$ for all $N$ particles in the gas in a 6 N -dimensional phase space:

$$
\begin{equation*}
\frac{\partial W}{\partial t}+\sum_{i}^{N} v_{i} \cdot \nabla_{i} W+\sum_{i}^{N} g_{i} \cdot \nabla_{v_{i}} W=0 . \tag{175}
\end{equation*}
$$

Here, $\boldsymbol{g}_{i}$ is the acceleration of the $i$ th particle. In classical mechanics the motion of particles is described by the Hamilton equations

$$
\begin{equation*}
\frac{\partial r_{i}}{\partial t}=\nabla_{p_{i}} H, \quad m g_{i}=\frac{\mathrm{d} \boldsymbol{p}_{i}}{\mathrm{~d} t}=-\nabla_{r_{i}} H \tag{176}
\end{equation*}
$$

where

$$
\begin{equation*}
H=\sum_{i} \frac{p_{i}^{2}}{2 m}+\sum_{i j} U\left(\left|\boldsymbol{r}_{i}-\boldsymbol{r}_{j}\right|\right) \tag{177}
\end{equation*}
$$

represents the Hamiltonian operator, i.e. the sum of the kinetic and potential energies of the particles, $\boldsymbol{p}_{\boldsymbol{i}}=m v_{i}$ is
the momentum of the $i$ th particle; and $g_{i}$ is the acceleration of the $i$ th particle. It follows from the Hamilton equation (176) that the probability $W$ is conserved along a path in the phase space so that Eqn (175) can be written in the form $\mathrm{d} W / \mathrm{d} t=$ 0 , where $\mathrm{d} / \mathrm{d} t$ is the total time derivative along the path. In other words, the flux in the phases space is incompressible.

The Liouville equation (175) is fully reversible: the statement about the constancy of $W$ along a path is true both as far as the future and the past are concerned. Therefore, it is not at all obvious that the statistical description should lead to irreversibility: after all this is not evident in the Liouville equation. More likely the opposite is true: the probability $W$ is constant along a path and for each of the configurations of the initial coordinates and velocities of the particles the path is uniquely defined both in the future and in the past.

Let $\boldsymbol{r}_{\boldsymbol{i}}(t)$ and $v_{i}(t)$ describe the motion of a point in the phase space for a certain number of initial values. We can easily see that the function

$$
W_{\mu}=\prod_{i} \delta\left[\boldsymbol{r}_{i}-\boldsymbol{r}_{i}(t)\right] \delta\left[v_{i}-v_{i}(t)\right]
$$

satisfies the Liouville equation. At each moment in time this function corresponds to one unique point in the phase space. Any function $W$ can be regarded as the average value $W_{\mu}$ of the initial data taken with a certain weight $F_{0}\left(\boldsymbol{r}_{i 0}, v_{i 0}\right)$. We shall now consider the microscopic density in a sixdimensional phase space

$$
\begin{equation*}
F_{\mu}^{\prime}(\boldsymbol{r}, v)=\sum_{i} \delta\left(\boldsymbol{r}-\boldsymbol{r}_{i}\right) \delta\left(v-v_{i}\right) \tag{178}
\end{equation*}
$$

This function plays the same role in relation to $W$ as the second-quantisation operator N in relation to the wave function $\psi_{N}\left(x_{1}, \ldots, x_{N}\right)$. The only difference is the addition of the variables $v_{i}$ to the configurational variables $\boldsymbol{r}_{\boldsymbol{i}}$. Averaging the function (178) with the weight $W$ makes it possible to find a single-particle distribution function $f(r, v, t)$. If $W_{\mu}$ is used instead of $W$, the result is the following time-dependent microscopic density:

$$
\begin{equation*}
F_{\mu}(\boldsymbol{r}, v, t)=\sum_{i} \delta\left[\boldsymbol{r}-\boldsymbol{r}_{i}(t)\right] \delta\left[v-v_{i}(t)\right] \tag{179}
\end{equation*}
$$

This function can be averaged over the initial data and we then again obtain $f(\boldsymbol{r}, v, t)$. We can easily see that the function (179) satisfies the Klimontovich microscopic equation:

$$
\begin{equation*}
\frac{\partial F_{\mu}}{\partial t}+v \cdot \nabla F_{\mu}+g \cdot \nabla_{v} F_{\mu}=0 \tag{180}
\end{equation*}
$$

where

$$
\begin{equation*}
\boldsymbol{g}=-\frac{1}{m} \nabla \int U\left(\left|\boldsymbol{r}-\boldsymbol{r}^{\prime}\right|\right) F_{\mu}\left(\boldsymbol{r}^{\prime}, v^{\prime}, t\right) \mathrm{d} \boldsymbol{r}^{\prime} \mathrm{d} v^{\prime} \tag{181}
\end{equation*}
$$

Eqn (180) is also fully reversible. However, if averaging is carried out over $r$ and over a time interval of the order of the collision time, then under the conditions of molecular chaos the last term may be transformed into a collision term. It is then quite obvious that the assumption of molecular chaos, i.e. the absence of correlation in the motion before collisions is precisely that key element which introduces irreversibility into the Boltzmann kinetic equation.

In order to understand the meaning of the assumption about molecular chaos, we must consider in greater detail the scenario of motion of a single molecule.


Figure 13. Molecule $M$ collides consecutively with molecules $1,2,3$. If the velocity of the initial motion is perturbed slightly (dashed line), the deviations from the initial path increase from one collision to the next.

Fig. 13 The continuous line in Fig. 13 represents the path of one of the molecules, and the spheres labelled $1,2,3$ describe the potential of the scattering of a molecule M by the molecules $1,2,3$. The radius of each ball is regarded as approximately equal to $a$; the average path between the collisions is taken to be $\lambda=1 / n a^{2}$. The initial velocity of the molecule M is assume to receive a certain perturbation $\tilde{v}$, so that its path follows the dashed line that deviates by an angle $\gamma_{0}=v / v_{\mathrm{t}}$ relative to the initial path. We can readily see that in the next collision this angle increases by a factor $\lambda / a$ and after the $k$ th collision it is of the order of $\gamma_{0}(\lambda / a)^{k}$. If the average distance along the path is denoted by $s \approx k \lambda$ and if $s=v t$, the result is

$$
\begin{equation*}
\gamma=\gamma_{0} \exp (\kappa t) \tag{182}
\end{equation*}
$$

where

$$
\kappa=\frac{v}{\lambda} \ln \frac{\lambda}{a}
$$

is known as the Kolmogorov - Sinai entropy.
Hence, we can see that even for a very small initial perturbation $\gamma_{0} \ll 1$ the value of $\gamma$ may become of the order of unity for

$$
k=\frac{s}{\lambda} \cong \ln \left(\frac{1}{\gamma_{0}}\right)\left(\ln \frac{\lambda}{a}\right)^{-1}
$$

The dashed path subsequently represents collisions with very different molecules. Therefore, the motion of a molecule in a gas is very sensitive to the initial perturbations.

We shall now show that in addition to the initial perturbations there are those which are contributed by the external environment. We shall do this by allowing for those thermal fluctuations which correspond to the collective degrees of freedom. If there are no inelastic interactions of molecules with the walls, these fluctuations become a component part of the motion of a system of $N$ particles in a 6 N -dimensional phase space. In other words, it should be regarded as reversible. In the hydrodynamic approximation such collective fluctuations are created by a random component of the collision term and the corresponding energy of the fluctuations is a fraction of the order of $\varepsilon^{2}$ of the thermal energy. In the presence of inelastic collisions with the walls a new source of thermal noise appears and its power seems to be equivalent to the noise due to the boundary layer of a gas of thickness $\lambda / L$. Consequently, the fraction of the noise due to the external source is

$$
\xi=\varepsilon^{2} \frac{\lambda}{L} \ll 1
$$

In other words, perturbations on a scale

$$
\gamma_{0} \approx \frac{\tilde{v}}{v_{\mathrm{t}}} \approx \varepsilon\left(\frac{\lambda}{L}\right)^{1 / 2}
$$

are introduced from outside, i.e. they are absolutely random and uncontrollable. Hence we can see that after

$$
k=\ln \left(\frac{1}{\gamma_{0}}\right)\left(\ln \frac{\lambda}{a}\right)^{-1}=\ln \left[\frac{1}{\varepsilon}\left(\frac{L}{\lambda}\right)^{1 / 2}\right]\left(\ln \frac{\lambda}{a}\right)^{-1}
$$

collisions the initial conditions are completely forgotten since there cannot be any reversibility.

We shall estimate $k$ for air. Then (in the cgs system of units) we have $n=3 \times 10^{19}, a^{2}=10^{-15}, \lambda=3 \times 10^{-5}$, and $\varepsilon=n a^{3}=10^{-3}$, so that if $L=1$, we obtain approximately $k \approx 2$. After the first two or three collisions the correlation of the motion of molecules is lost completely! Therefore, the simplest mechanism for the excitation of the acoustic noise by the walls is sufficient for the instability of the particle paths to lead to molecular chaos under very usual conditions in air at normal temperature and pressure. A weak interaction of the gas with the walls is sufficient for molecular chaos to be established in the gas and the consequence of this chaos is an irreversible approach to an equilibrium.

The influence of the external environment can be allowed for quantitatively (but still approximately) by adding small additional terms to the kinetic equations. Let us begin with Eqn (180) for the microscopic function $F_{\mu}$. The presence of external perturbations has the effect that in each collision of the molecules an additional inaccuracy appears in the collision parameters and the scale of this inaccuracy is $\xi$. In other words, a random distortion of the path of a departing particle of the order of $\xi$ is introduced by each collision. We can say that each of the molecules experiences an additional random force which produces additional impulses at an average frequency $v$ equal to the frequency of pair collision. This situation corresponds exactly to the Langevin equation (79) and therefore the average result of such impulses can be allowed for exactly in the same way as has been done in the derivation of Eqn (84) allowing for the Einstein relationship Eqn (83). Consequently, the right-hand side of Eqn (180) should be supplemented by an additional term

$$
\mathrm{K} f=D_{v} \nabla_{v}\left(\frac{v}{v_{\mathrm{t}}^{2}} f+\nabla_{v} f\right)
$$

where $D_{v}=\xi^{2} v_{\mathrm{t}}^{2} v$ is the diffusion coefficient in the velocity space and $v_{\mathrm{t}}^{2}=T / \mathrm{m}$.

Here $K$ is the diffusion operator in the velocity space. It acts on the function $F_{\mu}$ and leads to broadening of the $\delta$ functions, so that with time the uncertainty $(\Delta v)^{2}$ near each of the $\delta$ functions increases linearly with time, i.e. $\Delta v \propto v_{\mathrm{t}}\left(\xi^{2} v t\right)^{1 / 2}$. After a time $t \approx \tau=1 / v$ this uncertainty represents a small quantity $\Delta v \approx \xi v_{\mathrm{t}}$, i.e. the estimate obtained is exactly the same as for $\tilde{v}$. The subsequent collisions characterised by diverging paths lead to a much faster broadening of localised 'lumps' in Eqn (179) which appear as a result of averaging of $F_{\mu}$ over the impulses received from the external environment. If the localisation of the individual components $\left\langle F_{\mu}\right\rangle$ becomes of the scale $a$, then Eqn (181) for the acceleration of particles can no longer be used because in the course of collisions the average value $\left\langle F_{\mu}(\boldsymbol{r}, v, t) F_{\mu}\left(\boldsymbol{r}^{\prime}, v^{\prime}, t\right)\right\rangle$ cannot be regarded as equal to the product of the average values of the function $F_{\mu}$. A way of bypassing this difficulty is well known: it is necessary to assume that the functions $F_{\mu}(r, v, t)$ and
$F_{\mu}\left(\boldsymbol{r}^{\prime}, v^{\prime}, t\right)$ are uncorrelated before the collision. Then the average value of the third term in Eqn (180) automatically leads to the Boltzmann equation. Therefore, the 'molecular chaos' is in fact created by the external world and is enhanced by a large factor due to the motion of the molecules.

The random interference of the external world can also easily be allowed in the Liouville equation (175) by adding again the corresponding additional terms. We can readily determine how these terms look. In fact, if the solution of Eqn (175) is taken to be the microscopic function

$$
W_{\mu}=\prod_{i} \delta\left[\boldsymbol{r}_{i}-\boldsymbol{r}_{i}(t)\right] \delta\left[v_{i}-v_{i}(t)\right]
$$

it follows from the Langevin approximation that we should allow for the additional velocity diffusion of each of the particles. Therefore, the generalised Liouville equation is of the form

$$
\begin{align*}
\frac{\partial W}{\partial t} & +\sum_{i}^{N} v_{i} \cdot \nabla_{i} W+\sum_{i}^{N} g_{i} \nabla_{v_{i}} W=K W \\
& \equiv D_{v} \sum_{i}^{N} \nabla_{v_{i}}\left(\frac{m v_{i}}{T} W+\nabla_{v_{i}} W\right) \tag{183}
\end{align*}
$$

We can see that the diffusion operator $K$ leads to broadening in time of a localised function $W_{\mu}$. We can say that $K$ represents the operator of entropy creation by the external world. The need to allow for the microscopic irreversibility has been pointed out by Prigogine [12]. His approach also includes the entropy creation operator $M$ which is close in meaning to the operator $K$ (we are using $K$ only because the symbol $M$ is employed as the measurement operator). In contrast to Prigogine's treatment, Eqn (183) gives the explicit form of the operator $K$.

The existence of the operator $K$ qualitatively alters the structure of Eqn (183). This equation is no longer reversible: the velocity diffusion, no matter how slow, makes Eqn (183) parabolic and an increase in the entropy occurs in just one direction, from the past to the future. According to Eqn (183) the probability $W$ is not simply transferred along the paths, but also diffuses weakly from one path to another. As a result of this we may speak of the 'molecular chaos'.

However, we have assumed so far that the external environment itself is close to equilibrium. It is found that the situation changes radically when the external environment is not in equilibrium. A physical system which can exchange energy and entropy with the external world is known as open. Many open physical systems have the property of formation of complex nonlinear structures and processes. This is why they are called complex physical systems or systems with self-organisation.

## 33. Convection

Let us consider convection of a liquid in a gravitational field as the simplest example of an irreversible process in an open system. Such convection readily appears in any layer of a liquid when it is heated from below. The warmer parts of the liquid become lighter as a result of thermal expansion and they are pushed upwards by the Archimedean force, and become replaced by cold masses. There is a large number of convective flows. Let us consider just one of the simplest examples of convection in a closed torroidal vessel (Fig. 14) of radius $R$.


Figure 14. Liquid in a ring-shaped vessel is heated from below by a heater kept at a temperature, $T_{\mathrm{m}}$, considerably higher than the temperature $T_{0}$ of a cooler in the upper part of the vessel. When the temperature difference is sufficiently large, convection appears in the liquid.

We shall assume that the liquid in this vessel is heated from below and cooled from above. More exactly, we shall assume that the wall temperature is

$$
T_{\mathrm{w}}=T_{0}+\frac{1}{2}\left(T_{\mathrm{m}}-T_{0}\right)(1+\cos \theta),
$$

where the angle $\theta$ is identified in Fig. 14. Let $T=T(\theta, t)$ be the temperature of the liquid averaged over the transverse cross section of the toroidal vessel. If $\gamma$ denotes the coefficient of heat exchange with the vessel walls, then the temperature $T$ can be described by the equation

$$
\begin{equation*}
\frac{\partial T}{\partial t}+\frac{v}{R} \frac{\partial T}{\partial \theta}=-\gamma\left(T-T_{\mathrm{w}}\right) \tag{184}
\end{equation*}
$$

The liquid is regarded as weakly compressible, so that the velocity $v$ is independent of $\theta$ and can vary only with time. Let $\alpha$ be the volume thermal expansion of the liquid, so that its density $\rho$ acquires a correction $\delta \rho=\rho \alpha T$. In a gravitational field this change in the density gives rise to a force $g \delta \rho$. Its projection along the azimuthal direction is $-g \operatorname{dd} \rho \sin \theta=$ $-g \rho \alpha T \sin \theta$. If this force is averaged over the angle $\theta$, the result is a certain average value. It is this average value that acts on a ring of the liquid, so that

$$
\begin{equation*}
\rho \frac{\partial v}{\partial t}=-g \rho \alpha\langle T \sin \theta\rangle-\eta v \tag{185}
\end{equation*}
$$

Here the last term with the coefficient $\eta$, proportional to the viscosity of the liquid, allows for the deceleration of the liquid by the walls.

Averaging of Eqn (184) over the angle $\theta$ shows that the angle-independent part of the temperature reaches a steadystate value $\langle T\rangle=\frac{1}{2}\left(T_{0}+T_{\mathrm{m}}\right)$. Let $T=\langle T\rangle-T_{1} \sin \theta+T_{2} \cos \theta$. Then $T_{1}$ and $T_{2}$ are described by the equations

$$
\begin{align*}
& \frac{\partial T_{1}}{\partial t}+\frac{v}{R} T_{2}=-T_{1}  \tag{186}\\
& \frac{\partial T_{2}}{\partial t}-\frac{v}{R} T_{1}=-\gamma\left(T_{2}-\frac{T_{\mathrm{m}}-T_{0}}{2}\right)
\end{align*}
$$

After averaging $T \sin \theta$ over the angle $\theta$ the equation for the velocity becomes

$$
\begin{equation*}
\rho \frac{\partial v}{\partial t}=+\frac{g \rho \alpha}{2} T_{1}-\eta v \tag{188}
\end{equation*}
$$

As we can see, in the system of three equations for the variables $v, T_{1}$, and $T_{2}$, two of them - Eqns (186) and (187) —are nonlinear and only Eqn (188) is linear. The system of
equations (186) and (187) has been derived by Lorenz [13], who has drawn attention to its very interesting features. Usually the Lorenz equations are written in the dimensionless form

$$
\begin{align*}
\frac{\mathrm{d} X}{\mathrm{~d} t} & =\sigma(Y-X) \\
\frac{\mathrm{d} Y}{\mathrm{~d} t} & =-Y+r X-X Z  \tag{189}\\
\frac{\mathrm{~d} Z}{\mathrm{~d} t} & =-b Z+X Y
\end{align*}
$$

where $\sigma, r$, and $b$ are dimensionless parameters and the quantities $X, Y, Z$ are selected in such a way that they are proportional to $v, T_{1}$, and $T_{2}-\frac{1}{2}\left(T_{\mathrm{m}}-T_{0}\right)$, respectively.

The system of equations (189) is usually called the Lorenz model. Its simplest solution is $X=Y=Z=0$. It describes the temperature distribution in a liquid at rest. However, the system (189) has one more steady-state solution:

$$
\begin{equation*}
X_{0}=Y_{0}= \pm[b(r-1)]^{1 / 2}, \quad Z_{0}=r-1 \tag{190}
\end{equation*}
$$

Naturally, this solution exists only for $r>1$. A comparison of the second equation of this system (189) with Eqn (187) readily shows that the parameter $r$ is proportional to the temperature difference $T_{\mathrm{m}}-T_{0}$. Therefore, the solution with a nonzero velocity $v \propto X$ appears only after the difference $T_{\mathrm{m}}-T_{0}$ exceeds a certain critical value. Steady-state convection then can occur either anticlockwise ( $v \propto X>0$ ) or clockwise ( $v \propto X<0$ ). According to Eqn (190), the convection velocity is proportional to the square root of the supercriticality $(r-1)$.

Let us consider the linear approximation of Eqn (189) when the variables $X, Y$, and $Z$ are small. In this case the quadratic terms should be dropped and the third equation of the system (189) is unrelated to the first two and describes the decay of $Z$ with time. The first two linear equations have a solution in the form of an exponential function such that both $X$ and $Y$ are proportional to $\exp (\gamma t)$. The growth increment $\gamma$ of small perturbations is described by the dispersion equation:

$$
\begin{equation*}
\gamma^{2}+(\sigma+1) \gamma+\sigma(1-r)=0 . \tag{191}
\end{equation*}
$$

If $r<1$, two roots of this equation are negative, i.e. small perturbations decay with time. However, if $r>r_{\mathrm{c}}=1$, one eigenvalue becomes positive, i.e. an instability appears. If $r-1 \ll 1$, i.e. if the supercriticality is small, then one of the modes varies slowly with time: in many branches of physics these are called soft modes. The quantity $X-Y$ for a soft mode is small, so that $Y \approx X$. Therefore, in the second and third equations of the system (189) we can ignore the derivatives with respect to time, so that $Y$ and $Z$ are described by the quasiequilibrium values: $Y=X, Z=b^{-1} X Y=b^{-1} X^{2}$. These approximate relationships can be used to write down the second equation of the system (189) in the form

$$
\begin{equation*}
F \equiv-\frac{\mathrm{d} X}{\mathrm{~d} t}+(r-1) X-\frac{1}{b} X^{3}=-\frac{\mathrm{d} X}{\mathrm{~d} t}-\frac{\mathrm{d} U}{\mathrm{~d} X}=0 . \tag{192}
\end{equation*}
$$

This equation should be regarded as the equation of motion of a material point with zero mass and with a coordinate $X$, acted upon by a force $F$ equal to the viscous friction $-\mathrm{d} X / \mathrm{d} t$, and by a force exerted by the potential

$$
\begin{equation*}
U=-\frac{r-1}{2} X^{2}+\frac{1}{4 b} X^{4}+\text { const } . \tag{193}
\end{equation*}
$$

According to Eqn (192), a 'particle' slides slowly down to the bottom of a potential well $U$. The potential $U$ is very different for $r<1$ and $r>1$ (Fig. 15).


Figure 15. Dependence of the potential energy $U$ on $X$ in the case when (a) $r<1$ and (b) $r>1$. At $r>1$ the particle in a stable steady state can be only in one of the minima of $U: 1$ or 2 .

If $r<1$, the potential $U$ rises in both directions from $X=0$, so that $X=0$ is a stable steady state. When the parameter $r$ passes through unity in the direction $r-1>1$, a bifurcation appears: one unstable and two stable positions are acquired by a material point with the coordinate $X$ and these positions correspond to steady states.

If we recall that the coordinate $X$ is proportional to the velocity of the liquid $v$, we can say that steady-state convection in the clockwise or anticlockwise direction appears when the supercriticality is small.

The supercriticality parameter $r$ is frequently called the control parameter: it is imposed from outside. If the parameter $r$ rises and passes through unity, the liquid responds by an instability at the point $r=1$ and by a subsequent spontaneous untwisting of the flow so that $v \propto(r-1)^{1 / 2}$. We can say that if $r<1$, the liquid carries no information associated with its motion: the liquid is simply at rest. However, if $r>1$, the liquid has two equilibrium states differing in respect of the sign of the rotational velocity. This means that one macroscopic degree of freedom can 'remember' one bit of information which for $r>1$ is maintained stably by convection flow. It is readily seen that within the framework of the adopted macroscopic equations one bit of information appears directly at $r=1$ when the flow velocity of the liquid is still zero: if $r$ is reduced from $r>1$ in the direction of $r=1$, the bit of information about the liquid is conserved up to $r=1$. We can say that at $r=1$ the liquid exhibits a 'trend', i.e. a hint of the appearance of the macroscopic order parameter $v$. Then a negligibly small excess of $r$ over unity transforms this trend into one bit of information. A further increase in $r$ does not change in any way the information capacity of our convection ring regarded as a memory cell: only the velocity $v$ changes, i.e. the cell becomes more resistant to disturbances.

Real liquids or gases are subject to thermal fluctuations. In particular, the degree of freedom associated with circular flow corresponds to an energy $\frac{1}{2} T$. Consequently, if $r<1$, there is a grass-line background of the thermal noise of different modes. In the limit $r \rightarrow 1$, a circular-motion mode seems to separate from other modes: its amplitude becomes greater since the decay decreases and the pumping of the thermal noise remains the same as before. It is this phenomenon that can be called a trend: the mode amplitude stands out by its height among 'coevals'. The passing of the parameter $r$ through unity converts this epithermal mode into macroscopic flow.

Let us return to the complete Lorenz model of Eqn (189). It has three steady-state solutions for $r>1$, but only two of them described by Eqn (190) are stable when the supercriticality is small. Let us consider what happens if we increase the parameter $r$ without limiting it to small values. The first question, whether the equilibrium of Eqn (190) is stable, can again be considered by a linear approximation near an equilibrium. An appropriate analysis shows that there is a second critical value $r_{\mathrm{c} 2}$ above which a second bifurcation takes place. However, this is not all. The system of equations (189) is found to have many solutions representing different modes of motion. The most surprising among them, discovered by Lorenz [14], corresponds to the values of the parameters $r=28, \sigma=10$, and $b=8 / 3$. This solution is known as the 'strange attractor'. Lorenz has shown that the system ( $X, Y, Z$ ) undergoes complex chaotic motion, resembling a dance, around two unstable foci. Beginning from any point with small values of $X, Y, Z$ the system goes over to an unstable focus around which it begins to describe spirals of amplitude increasing with time, i.e. to follow paths along an untwisting spiral. After a certain number of such turns the system abruptly moves to the second focus around which it again forms paths along an untwisting spiral. After several turns of the second spiral the system jumps back to the first spiral in order to repeat approximately the same motion. However, there is no periodicity in such motion: the times for which the system is close to one of the foci and the number of turns in each of the spirals seem to be perfectly random. Chaotic motion appears in a perfectly determinate dynamical system with three coordinates $X, Y$, and $Z$.

Although the form of the equations of motion in the system (189) is very simple, it does not give a clear idea how the chaos appears. With this in mind, one of the equations should be written in a more complex form. The variable

$$
Y=X+\frac{1}{\sigma} \frac{\mathrm{~d} X}{\mathrm{~d} t}
$$

from the first equation in the system (189) should be substituted into the second and third equation, and then the variable

$$
Z=\frac{1}{b}\left(X Y-\frac{\mathrm{d} Z}{\mathrm{~d} t}\right)
$$

should be substituted into the second equation. We thus obtain an equation for $X$ which again is not closed, but it now has the following form which is very convenient for qualitative analysis:

$$
\begin{equation*}
\frac{1}{\sigma} \frac{\mathrm{~d}^{2} X}{\mathrm{~d} t^{2}}=F=-\frac{1+\sigma-X^{2}}{\sigma} \frac{\mathrm{~d} X}{\mathrm{~d} t}-\frac{\mathrm{d} U}{\mathrm{~d} X}+\frac{X}{b} \frac{\mathrm{~d} Z}{\mathrm{~d} t} \tag{194}
\end{equation*}
$$

Here $U$ is the same potential (193) with two humps (for $r>1$ ). Without the last term Eqn (194) has the form of the equation of motion of a material point in a potential well $U$ with a friction force proportional to $\mathrm{d} X / \mathrm{d} t$ and a friction coefficient whose sign is reversed from positive to negative for $X^{2} \geqslant 1+\sigma$. The last term has the form of an elastic force with a time-dependent elasticity coefficient

$$
-\frac{1}{b} \frac{\mathrm{~d} Z}{\mathrm{~d} t}
$$

If the derivative $\mathrm{d} Z / \mathrm{d} t$ is not small, this term depends on two other variables $Y$ and $Z$ which have complex time dependences, and acts as a variable driving force. If the correlation
between $\mathrm{d} Z / \mathrm{d} t$ and $X$ is ignored, it seems to be a random force. In other words, a material point in a two-hump potential well moves under the action of a random force and the friction coefficient can be both positive or negative. This accounts for qualitatively for the nature of motion of a strange attractor, but its quantitative analysis is preferably carried out by going back to the initial system of equations (189).

At high values of the parameter $r$ the Lorenz system has not only a strange attractor, but a whole series of different dynamical regimes. These regimes will not be discussed in detail here (the reader is directed to, for example, Ref. [34]). We shall simply point out that these solutions are associated with convection and that convection itself is related to an enhancement of the heat flux from a heater to a cooler. In fact, a vertical heat flux $q$, calculated unit area of the transverse section of the vessel, is $q=c v T \sin \theta$, where $c$ is the specific heat. Averaging of $q$ over the angle $\theta$ gives $q=c v T / 2$, i.e. $\langle q\rangle \propto X Y$. If the system (189) is averaged over time, it is found that the time-average heat flux is proportional to $b \bar{Z}$, where the bar over $Z$ denotes time averaging. In a stable state it is found that $Z=0$ and in a regime with convection the relationship $\langle\bar{q}\rangle \propto b(r-1)$ is obeyed approximately. In other words, convective motion is related directly to enhancement of heat transfer, i.e. to an increase in the rate of rise $\dot{S}_{\mathrm{e}}$ of the entropy of the environment.

The convection in the Lorenz model is found to have the main common features of dissipative nonlinear processes considered in the approximation of a small number of the order parameters. An increasing deviation from equilibrium (i.e. an increase in the difference $T_{\mathrm{m}}-T_{0}$ ) gives rise initially, beginning from a certain critical value of this control parameter, to spontaneous new nonzero order parameters ( $T_{1}$ and $v$ in this case). A further increase in the supercriticality causes these parameters to grow, i.e. a steady-state bifurcation of Eqn (190) appears and there is a corresponding increase in the rate of dissipation, i.e. in $\dot{S}_{\mathrm{c}}$. Then, a further increase in the supercriticality $T_{\mathrm{m}}-T_{0}$ gives rise to a second bifurcation, so that the coordinate parameters $X, Y, Z$ become dynamical variables of a complex nonlinear system (189). A further increase in $T_{\mathrm{m}}-T_{0} \approx r-1$ considered within the framework of the system (189) may result in interchange of the various modes. New order parameters may appear in a real physical system and they describe higher harmonics of the motion of the liquid. As the number of harmonics increases, the motion becomes more and more complex: for simplicity, we shall call it simply turbulent. Such turbulent motion together with heat transfer from the heater to the cooler represents a complex scenario of approach to an equilibrium in a system far from it.

## 34. Self-organisation

Our convection system is a typical example of selforganisation: as the nonequilibrium parameter $r$ increases, a liquid becomes unstable at $r=1$ and then steady-state convection is established; at high values of $r$ there are various modes of stochastic motion of the strange-attractor type. There is an enormous number of other types of nonlinear self-organisation which are described in books [14-19] and in the literature cited there.

Many of the nonlinear systems are so beautiful that they are quite suitable for brightening up the interiors of modern
apartments or houses. For example, the structure shown in Fig. 16, which is known as the 'plasma sphere', can be seen quite frequently in the shops of Western cities.


Figure 16. 'Plasma sphere' in which glow-discharge 'plasma snakes' appear, rise upwards, and pairwise merge there.

This is indeed a sphere made of a dark glass, $\frac{1}{2} \mathrm{~m}$ in diameter, filled with luminous moving 'snakes'. Each snake is a plasma formation representing a weakly luminous filamentary discharge. Such a discharge is very similar to the luminous discharge in conventional neon tubes used for illumination of major cities at night. It is known as the glow discharge: it appears when the electric current in a lowpressure gas is not too high. In the plasma sphere the glow filamentary discharge develops between a spherical metal electrode located at the centre of the system and a weakly conducting metallised surface of the outer glass sphere.

Each discharge snake, and up to twenty of them may appear simultaneously, is directed on average in the radial direction. However, like a live snake, it bends slightly and vibrates all the time and there are several bending periods along its length. At each of its ends the snake has a characteristic trident, which like a tiny cat's paw, shimmers continuously and collects charges from the appropriate electrode. These snake discharges are in continuous motion. In addition to the twisting, each of the snakes slowly rises upward and this is obviously the result of convection. The snakes collect in the upper position and merge in pairs, so that some of them disappear permanently. On the other hand, in the lower part of this system there is a continuous process of creation of new snakes, they multiply, split into two, and rise upwards, where they disappear.

In spite of its complexity, the whole pattern can be readily understood from the physical point of view. Naturally, it is theoretically much simpler to consider an absolutely symmetric glow discharge between the inner and outer electrodes. However, such a discharge is unstable: heating of the gas and a reduction in its local density with a corresponding reduction in the electric resistance mean that an electric current flows more readily along relatively narrow tubular channels. The discharge splits into plasma filaments. Since these filaments are lighter, they float up under the Archimedean force. The interaction of the filaments with the gas currents and with one another is the reason for the formation of a pattern of snakes, which is organised in a
complex manner and which resembles the mythological head of the Medusa (one of the Gorgons). We can see why cat's paws form at the ends of each snake. If the conductivity of the electrodes is low, the density of the surface charge directly opposite the discharge becomes less and the end of a snake with the opposite charge is likely to split and travel from point to point collecting the surface charge. The qualitative behaviour is thus absolutely clear and even somewhat boring. Nevertheless, the plasma sphere fascinates and attracts by its apparent mystery: it resembles a living thing with conscious motion.

The example of the plasma sphere makes it possible to follow once again all the main characteristics and components of self-organisation. It begins in a system which is close to its stability margin. An instability, which in this case is the splitting of the discharge into filaments, begins from a hint of the appearance of a future filament. Only one bit of information is sufficient for each hint. As the external nonequilibrium parameter (in this case current) increases, real formation of filaments takes place. The initial spherical symmetry is disturbed: we can say that spontaneous symmetry breaking occurs. As the gas in the filaments becomes heated, convection begins to play a role, i.e. the next bifurcation occurs and a new order parameter (gas-dynamic velocity) appears. The cat's paw at the end of each snake represents one further bifurcation with its own instability mechanism. On the whole, a complex nonlinear physical system with chaotic motion is formed. The motion is maintained for a long time if the system is open: an electric current from an external source must be passed continuously through the plasma sphere. Moreover, this source should supply energy in a sufficiently ordered form: in accordance with the Brillouin terminology, a negative entropy, i.e. an entropy with the opposite sign, should be 'injected' into the system.

Let us now consider what this means in the case of the plasma sphere. The snakes exist only because of local heating of the gas inside a filamentary discharge. In other words, the gas inside the filament should be heated but the system as a whole remains at room temperature, i.e. the excess heat is transferred to air across the glass sphere. If there is a flow of heat, it means that entropy is created. The plasma sphere converts some of the highly organised electric energy into heat, which is then disipated into surrounding space. Entropy is created continuously inside the sphere and it flows together with heat into the surrounding space. If the entropy flux is blocked, a plasma filament 'dies'. In the plasma sphere we need to remove continuously the 'slag' of the newly created entropy: figuratively speaking, we have to introduce an entropy with the opposite sign. If entropy is a measure of disorder, then entropy with a negative sign (negentropy) is a measure of order.

It follows that self-organisation requires two elements to 'feed' it: energy and negentropy. Only their sum can ensure a steady maintenance of a structure of a nonlinear dissipative system. Systems of this kind are usually called complex open physical systems.

A clearer idea of the entropy (or information) fluxes can be gained by considering the simplified scheme in Fig. 17. In this figure the abscissa represents the absolute maximum entropy of a system M when it is isolated from the external world and the ordinate gives the internal entropy $S_{i}$ with the opposite sign, i.e. the negentropy. If the system is initially perturbed from a stable equilibrium, its negentropy increases to $S_{0}$. In monotonic relaxation processes (Fig. 17A) the


Figure 17. Schematic representation of the entropy fluxes in a nonequilibrium system M: (A) monotonic increase of the internal entropy $S_{\mathrm{i}} ;\left(\right.$ B) in the presence of $N$ pivots a part of the entropy $S_{\mathrm{i}}^{\prime}$ decreases against the general background of an increase in the entropy $S_{i}$; (C) if entropy flows through the external system, the reduction in the part of the entropy of the system $S_{\mathrm{i}}^{\prime}$ may be due to a negative entropy (information) of the external medium.
negentropy of the system smoothly slides down, apparently 'falling' in the direction of the 'force of gravity'. However, a relaxation process of this type is relatively rare. F or example, if in a perfectly homogeneous gas the particle velocity distribution function is disturbed identically at all the points, a fast slide of the type shown in Fig. 17A occurs. However, if such a perturbation is not homogeneous in space, the picture changes drastically. Collisions convert the distribution function to the Maxwellian form only locally: the local equilibrium order parameters (density, velocity, and temperature) evolve then in accordance with the laws of gas dynamics. Collisions then even prevent fast relaxation to an equilibrium: the more frequent the collisions, the smaller are the kinetic dissipation coefficients. In Fig. 17B this variant is identified as dissipation with a connectivity (number of links) $N$, which prevents a uniform slide of the entropy level. We apparently have a lever with its large arm dropping and a small arm which can rise. In other words, dropping down to the 'centre of mass' is accompanied by a rise of a small part of the system: an increase in the order appears in the number of degrees of freedom, i.e. self-organisation takes place.

The variant in Fig. 17C shows a more complex selforganisation scenario when the second arm with a connectivity $N^{\prime}$ prevents dropping of the system negentropy because of the flow of entropy to the external world. The result is an open system in which ordering is due to an increase in the entropy $S_{\mathrm{e}}$ of the external world.

The entropy $\Delta S_{i}$ flowing out of the system M represents heat either partly or mainly. This means that the entropy flux is accompanied by an energy flux. Consequently, a steady state in a self-organised system $M$ cannot be maintained simply by removal of the excess entropy: energy has to be applied to the input of this system. The energy delivered to the system should be more organised compared with heat: the entropy per unit of this energy should be less than $T_{\mathrm{e}}^{-1}$, where $T_{\mathrm{e}}$ is the temperature of the external medium. In other words, it is necessary to feed the system simultaneously with both energy and negentropy.

On Earth the most powerful source of energy and negentropy is solar radiation. It is this radiation that sets in motion the mechanisms of self-organisation in the atmosphere, oceans, and biosphere of the Earth.

## 35. Approach to equilibrium and collapse

Many of the processes discussed above represent relaxation, i.e. the approach to thermodynamic equilibrium either in systems initially far from such an equilibrium or in systems disturbed from equilibrium by other relaxation processes. Let us try to discuss them from a general standpoint. Let us begin with the simplest example of one particle in a potential well (Fig. 18).


Figure 18. Classical material point $M$ in a potential well (a) undergoes weakly damped vibrations and transfers its energy to the external medium. A quantum particle (b) can transfer its energy by emitting radiation quanta.

If a classical particle M is placed in the potential well shown in Fig. 18a, then this particle vibrates periodically with an amplitude which decreases monotonically because of the friction force. The energy $\varepsilon$ of the vibrations is gradually transferred to the medium until this energy reaches the thermal vibration level, which is equal to the temperature of the medium $T$.

It is quite obvious why this occurs. After all, a displacement of the particle $M$ from its equilibrium position, imparts to it a greater energy, i.e. it 'drives' an energy $\varepsilon$ much greater than temperature into one degree of freedom. If this energy is simply converted into the thermal energy of the medium, the entropy of the medium would increase by $\Delta S=\varepsilon / T$. Consequently, our initial state is far from equilibrium and has a negentropy $\varepsilon / T$. The simplest scenario of disappearance of this negentropy is the dissipation of the energy of ordered oscillations of the point particle M simply into heat, which does indeed occur because of the friction forces.

Naturally, there are also more complex scenarios of such a transition. For example, the point M can be used as a weight to set in motion a pendulum clock and then the process of dissipation of energy into heat, i.e. the disappearance of the negentropy, is different.

Near the bottom of the potential well in Fig. 18a the point particle $M$ oscillates thermally. These oscillations can be described following Langevin, i.e. the frictional force may be supplemented by random impulses from the external medium. On average, the total energy of the particle is then $T$.

Let us now consider a quantum particle (Fig. 18b). The energy of such a particle is quantised. For example, in a harmonic potential the energy of a level $n$ is $\varepsilon_{n}=\hbar \omega_{0}\left(\frac{1}{2}+n\right)$, where $\omega_{0}$ is the vibration frequency. The initial state of the particle need not correspond to just one level. F or example, in the case of a harmonic oscillator it is possible to construct what are called coherent states from a superposition of wave functions of different levels. However, even in the more complex case of an anharmonic oscillator we can select as the initial wave function any superposition of eigenfunctions.

However, the specific nature of selection then is manifested quite rapidly during subsequent evolution.

Each level $n$ has its own eigenfrequency $\omega_{n}=\varepsilon_{n} / \hbar$, where $\varepsilon_{n}$ is the corresponding eigenenergy. Therefore, a particle M can transfer energy to the medium if it 'itself searches for' resonances at frequencies $\omega_{m n}=\omega_{m}-\omega_{n}$. If such resonances are found, then even in the case of a very weak interaction with a medium the particle M may transfer energy to the medium. However, the process of relaxation does not begin from this event. If the oscillator is anharmonic, the frequen-cies $\omega_{12}, \omega_{23}, \ldots$, do not coincide and they therefore find different resonances in the external medium. The first effect is the loss of the phase difference between the various levels. The wave functions of different levels lose their mutual coherence and, therefore, the wave function of the particle collapses into one of the levels. The probability of this collapse is equal to the square of the amplitude. In view of energy conservation, a similar collapse should also occur in the system in which the particle is 'prepared' in a state of superposition of several levels. In other words, the collapse of the function of the observed particle seems to be transferred to the system that prepares the particle for the subsequent observation.

When the particle 'settles' at a particular level, the processes of emission of quanta $\hbar \omega_{12}, \hbar \omega_{23}$, etc. begin, i.e. the particle 'drops' down the levels. Finally, at the lower levels, thermal equilibrium is established: the particle either emits or absorbs quanta and reaches equilibrium by emitting thermal radiation. The last process is also accompanied by quantum transitions, i.e. by the collapse to one level, then to another, and so on. At any given moment one quantum particle can be only at one level if coherent links of some quantum states of radiation with other states are not established deliberately in the external medium. The transitions accompanied by the emission of radiation or emission of field quanta represent the collapse of the wave function of the particle. We can say that the concept of collapse is implicit in the concept of quanta: instead of the term quantum transitions, we can equally well consider the collapse of wave functions.

We shall now consider a somewhat more complex thought experiment. We shall assume that the potential of Fig. 19a is deformed adiabatically slowly in such a manner that instead of one minimum, two potential energy minima are created. We shall also assume that a particle is at the lowest energy level.

More rigorously, in a potential with two wells and a large hump between them we have to allow for the presence of two lower levels. One of them (the lower) corresponds to a symmetric wave function and the other to an antisymmetric function. If the initial state is set very accurately, i.e. if it is set by adiabatic deformation of the lower state in the initial single well, then the particle is in a symmetric state.

Let us now consider various thought experiments on this particle. First of all, let us assume that two wells (Fig. 19a) are sufficiently far apart to seem to 'lock' the wave functions in two 'boxes'. The energies of the symmetric and antisymmetric functions are then found to be practically identical. Con-sequently, the energy levels in the left-hand and right-hand wells are also identical and the wave functions of the particle in each of the wells are correspondingly equal to the half-sum and half-difference between the symmetric and antisymmetric functions.


$\sigma$

Figure 19. Lower energy levels in a two-hump well (a) represent a weakly split doublet. If a weak external interaction destroys the coherence of the right-hand and left-hand parts, the wave function collapses into one of the potential wells. The corresponding symmetry breaking can be detected from outside, which is equivalent to measurement: out of two possible states (b) only one is real.

Let us now begin to 'heat' the particle by bringing it into contact with an external thermostat. Then, in addition to the lower levels, higher levels begin to play a part and the particle may be transferred to these levels with the probability given by the Boltzmann distribution. However, a different effect is more important to us. The thermal noise disturbs the coherent coupling between the right-hand and left-hand wells. The particle may then exist only in one of the boxes. Consequently, the wave function does exist in one of the boxes but not in the other. The collapse of the wave function takes place, but at this stage without the collapse of the probabilities: the probability of finding the particle in one of the boxes is still exactly equal to one-half. The situation is exactly similar to that considered in the first sections of this paper. We have a particle in a thermostat separated by an internal barrier. We can try to find in which of the boxes inside the thermostat the particle is located. This requires an appropriate measurement, which is accompanied by an irreversible process in the external world. After such a measurement the distribution of the probabilities for the particle collapses into the state $(0,1)$. The entropy of the particle then decreases by one bit and an irreversible process accompanied by an increase in the entropy by at least one bit should take place in the external world. In other words, we are dealing with a typical information process.

Let us now assume that no such measurement takes place and we again begin to lower temperature to zero. In the final state the particle is again at the lowest the level in one of the wells, but now we can say precisely that one of the boxes is empty and the particle is inside just one of these boxes. In other words, only a very weak contact between the system and the external world and the corresponding destruction of the coherence seems to have resulted in a 'spontaneous symmetry breaking', so that the wave function of the particle collapses into one of the boxes. We can again try to detect in which of the boxes (wells) the particle is located, but this time the process of measurement does not affect the wave function and simply causes the collapse of the probability distribution.

Let us consider the situation when this measurement is not carried out. We can then imagine a mixed ensemble of many states, half of which have the half-sum and the other half the half-difference of the symmetric and antisymmetric wave functions. In other words, in half of the states the particle is in the left-hand well and in the other half it is in the right-hand well. Then each of the states begins to oscillate since the corresponding wave function passes periodically from one well to the other. The frequency of such oscillations is $\Delta \varepsilon / \hbar$, where $\Delta \varepsilon$ is the difference between the energies of the symmetric and antisymmetric states. If the frequency of these
oscillations is known, we can predict the moment when the wave function again collects in one of the wells from the previously isolated boxes. If we now detect the presence of the particle in one of the wells, we automatically induce the collapse of the previous probabilities, i.e. we shall determine to which subensemble a given particle belongs.

Instead of this very long description of a scenario, let us now consider a simpler example. Let us assume that we wish to carry out a measurement, i.e. to detect in which of the wells (Fig. 19a) the particle is more likely to be located after the second well has been formed by slow deformation of a single well (Fig. 18a). Let the instrument be constructed in such a way that after measurement the wave function collapses into just one well, i.e. that only one of the possible states is real (Fig. 19b). Such an instrument should destroy the coherence of the initial state and create a mixture of symmetric and antisymmetric functions. This means that it should transfer an energy of at least $\frac{1}{2} \Delta \varepsilon$ to the particle. However, together with the collapse of the wave function (in one of the wells) the instrument also induces the collapse of the probabilities in a specific well. This means that the instrument of the external world should experience an increase in the entropy by at least one bit, i.e. $\ln 2$.

Our very simple example has thus made it possible to deal with a number of problems. First of all, the collapse of the wave functions can be separated from the collapse of the probabilities. As established above, the thermal motion itself is sufficient to destroy the coherence and cause the collapse of the wave function into one of the possible states. As long as this collapse is not observed outside, it is better to speak of the conversion of a pure ensemble into a mixed one: we have here an irreversible process with a set of probabilities in the final state and our particle is a member of this ensemble. We can say that the collapse is a fluctuation and if we are not particularly interested in fluctuations, we can use this average statistical description with the corresponding probabilities, i.e. the density matrix of the mixed state.

However, the collapse itself is an interesting physical process and it could be studied in the case of quantum measurements. A quantum measurement is a process organised in such a way that the wave function and the probabilitles collapse simultaneously. We seem to have a unique information process. According to Fig. 2, a measurement does indeed cause the collapse of the wave function and of the prob-abilities. Information about a quantum object is perceived, i.e. stored in a perceptor $P$, and this is accompanied by a simultaneous 'release' of a value $U_{i}$ of a physical system $U$. All this is possible because of a feeder F , which transfers some of the negentropy (information) to the perceptor and some of the entropy $W=I_{*}-I$ is ejected in the form of 'slag' to the surrounding medium. In the physical object itself the coherence is lost and the wave function collapses into just one of the states. Therefore, a quantum measurement repre-sents a process far from equilibrium and it can be considered as one of the scenarios of approach to equilibrium.

We shall consider one further aspect of the model in Fig. 19a. If the potential wells in Fig. 19a are moved sufficiently far apart, the system obtained can be regarded as a memory cell. A particle placed in one of the boxes can now remain there indefinitely, retaining one bit of information. This bit can be imparted initially by deformation of the well of Fig. 18a so that it forms the potential of Fig. 19a and this is accompanied by an
asymmetric correction so that a particle drops to the lowest level in just one of the wells. This asymmetric correction is the control parameter which initially conserves one bit of information and then deposits it in a cell. Another variant of creating an asymmetry is more cumbersome: the second well in Fig. 19a could be formed far from the first well (where the particle is located) and then bring it closer in the empty form and thus create the symmetric cell of Fig. 19a. One further variant involves the formation of many cells of the type shown in Fig. 19a and then collapse into them of the wave functions by 'heating' the cells; this is followed by separation of the cells into 'right-handed' and 'left-handed' by measurement. The cells can then be used to construct a text. However, even more attractive is the possibility (if it can be realised) to record the text directly by the collapse of wave functions inside the cells.

A set of many cells of the type shown in Fig. 19a with the wave functions collapsed inside them can be regarded as a form of a 'quantum memory'. The text can be retrieved from such a memory simply by identifying which well contains the wave function. There is not any need to destroy the existing quantum state: all that is necessary is an external feeder with a sufficient store of the classical negentropy and a detector which recognises which well is filled.

We shall now consider the approach to equilibrium of a system of many particles, namely a gas, which is one of the simplest systems of this kind. As demonstrated earlier, classical analysis of the motion of atoms or molecules in a gas leads naturally to molecular chaos and to the Boltzmann equation. The approach of a dense gas to equilibrium, considered within the framework of the Boltzmann equation, goes over naturally to the description based on the equations of gas dynamics with dissipation.

However, gas atoms are not classical particles but microparticles. How can one construct a more logical picture of the process? We can do this by considering a separate quantum packet of a randomly selected particle. Since the particles are indistinguishable, it is best not to speak of a specific particle, but of a wave packet corresponding to one particle. The motion of such wave packet causes its scattering by other packets and its shape may resemble that of an expanding cloud with a complex jagged shape. Some parts of this cloud rapidly lose their mutual coherence, so that the particle must unavoidably fall into one of the parts. We can say that the wave packet of such a particle collapses into a more compact wave packet. However, this collapse is not observed. The collapse of wave functions inside a gas are indistinguishable from thermal fluctuations: they cannot be measured from outside and they are not accompanied by the collapse of the observed probabilities. Therefore, inside a small macroscopic volume the process of relaxation occurs in exactly the same way as for classical particles. Consequently, the distribution function becomes locally Maxwellian and the gas acquires macroscopic order parameters (temperature, density, velocity). Macroscopic gas particles have a very short de Broglie wavelength, so that their wave functions can be regarded as collapsed into quasiclassical functions. Therefore, the equations of classical gas dynamics, with all their consequences, can be applied to the gas as a whole.

## 36. Sokolov effect

The collapse of wave functions of gas atoms cannot usually be observed. However, it does not mean that the process is
always hidden and we shall now describe an effect in which the role of such a collapse is the dominant one. We shall discuss the effect discovered experimentally by Yu L Sokolov $[36,37]$ and we shall call it the Sokolov effect. This effect was detected in atomic interferometry experiments [36] carried out with the use of apparatus shown schematically in Fig. 20.


Figure 20. Experimental setup for atomic interferometry. Here, S is the source of hydrogen atoms in a metastable state $2 S ; 1$ and 2 are double slits with a longitudinal electric field in the gaps; $D$ is a detector of radiation due to $2 P \rightarrow 1 S$ transitions.

A beam of excited hydrogen atoms in a metastable state $2 S$ is generated in a special source S : hydrogen ions are extracted from a small plasma source and accelerated to an energy of about 20 keV ; they are then subjected to a charge exchange process in a gaseous target. Beyond this target many excited atoms are formed, but at a distance of a few metres only the metastable $2 S$ atoms remain. These atoms are first transmitted by a pair of miniature slits (1) to which a longitudinal electric field is applied, i.e. a small capacitor is formed. Each atom becomes polarised in this longitudinal electric field, i.e. a state representing a superposition of the $2 S$ and $2 P$ states is formed. The state $2 P$ is no longer metastable and a dipole transition causes its decay to the ground state accompanied by the emission of a Lyman-alpha photon which can be recorded with a detector, D. In addition to the first capacitor (1) there is also another capacitor (2) connected rigidly to the detector D .

After crossing the first capacitor (1) the state $2 P$ decays monotonically, so that the detector D of the Lyman-alpha radiation yields a monotonically falling dependence as the detector is moved away from the slits (1) to a distance $L$ (Fig. 21a). However, if the second capacitor (2) with an electric field $E_{2}$ is placed in front of the detector, the coordinate dependence of the intensity of the Lyman-alpha radiation $I_{2 P}(L)$ shows a very clear interference pattern (Fig. 21b). It appears because the energy levels of the $2 S$ and $2 P$ states are separated from one another by the Lamb shift with a frequency of about $10^{9} \mathrm{~Hz}$. The velocity of a hydrogen atom with an energy 20 keV is about $2 \times 10^{8} \mathrm{~cm} \mathrm{~s}^{-}$ ${ }^{1}$. Con-sequently, the phase of the $2 P$ amplitude oscillates relative to the phase of the $2 S$ amplitude, changing by an amount $2 \pi$ in a distance of about 0.2 cm . The second capacitor (2) creates a new admixture of the $2 P$ amplitude to the $2 S$ state and adds it to the component created earlier in the first capacitor (1). Since the phases of these two components are different, atomic interference between the two $2 P$ amplitudes becomes possible (Fig. 21b).

It must be stressed once again that we are not speaking of interference between two 'rays', but of interference between two $2 P$ amplitudes inside the same atom. Since the $2 S+2 P$ superposition appears at an electron cloud shifted relative to the proton, interference apparently takes place between two dipole shifts of an electron.

The Lamb shift is by itself small, but the real difference between the energy levels of the $2 S$ and $2 P$ states in this
experiment is even less and is practically zero. This is because the shift of the levels is induced by static electric fields which do not alter the total eigenfrequency of an energy level. Therefore, a small change in the energy of an electron level is compensated for exactly by a corresponding change in the kinetic energy of an atom. We can say that if allowance is made for the kinetic energy of the proton, the levels $2 S$ and $2 P$ have the same energy, so that the transitions $2 S \rightarrow 2 P$ and $2 P \rightarrow 2 S$ are apparently equivalent to a change in the polarisation of a particle without a change in its energy.

These atomic interference experiments are in good agreement with standard quantum theory. They have made it possible to measure the Lamb shift with a very high precision. However, this is not of prime interest to us. The most important feature is that in the Sokolov experiments a small admixture of the $2 P$ amplitude appears after passage through the second gap 2 even in the absence of any electric field at this gap. Moreover, this admixture appears also when a $2 S$ atom simply flies near a metal plate (Fig. 21c). Precisely this phenomenon will be called the Sokolov effect.

The effect is fairly strong. Its results appear as if a longitudinal electric field of the order of several volts per centimetre were present near a slit or a metal plate, but under the experimental conditions the appearance of a field of this order of magnitude is quite impossible. Since the effect is strong, interference can be observed even when both capacitors are replaced with metal plates (Fig. 21d). Deliberately designed experiments have shown that this 'daemon's field' (which is the name used by Sokolov) polarises a $2 S$ atom in such a way that its electron shell stands a way somewhat from the proton.


Figure 21. Experimental setup for atomic interferometry. (a) After crossing a capacitor with a field $\mathrm{E}_{1}$ a mixture of $2 S$ and $2 P$ amplitudes is formed, atoms in the state 2 P decay, and the corresponding intensity $I_{2 P}$ falls monotonically when the distance L from the slit is increased. (b) Second slit with the field $E_{2}$ gives rise to interference patterns (dashed curve) in a weak field $E_{2}$ and (solid curve) in a strong field $E_{2}$. (c) Weak interference appears even when the second capacitor is replaced by metal plate M. (d) Even weaker interference is detectable when both capacitors are replaced with metal plates M.

a


Figure 22. (a) Atom A with dipole moment $\boldsymbol{d}$ flies at a velocity $v_{0}$ above a metal at a distance $l_{0}$ from its surface. The thickness of the layer of the interaction with the conduction electrons is $l$. (b) In a coordinate system
in which an atom is at rest there are two electrons, $\mathrm{e}^{\prime}$ and e , which travel at an angle to the metal surface and are then reflected, escaping into the bulk of the metal.

All attempts to account for the Sokolov effect by the force interaction of an atom with fluctuating electric fields or with an image field in a metal have not been successful: they give contributions which are several orders of magnitude less than the required value. However, the effect can be explained on the basis of the ideas on the collapse of wave functions of electrons in a metal [38].

In this explanation it is necessary to consider in greater detail the interaction of an excited $2 S$ atom with electrons in a metal. For simplicity, we shall consider the case when an atom A flies along the surface of a metal at a constant distance $l_{0}$ from this surface (Fig. 22). The velocity of the atom is $v_{0}$. In Sokolov's experiments the hydrogen atoms had an energy of about 20 keV , so that $v_{0} \approx 2 \times 10^{8} \mathrm{~cm} \mathrm{~s}^{-1}$. This value is somewhat greater than the velocity $\nu_{\mathrm{F}}$ of electrons in a metal at the upper levels of the Fermi distribution, which amounts to about $10^{8} \mathrm{~cm} \mathrm{~s}^{-1}$.

Let the $z$ axis be parallel to the motion of the atom. The $2 S \rightarrow 2 P$ dipole transition along this axis corresponds to a dipole moment $d \approx 3 a_{0} e$, where $a_{0}$ is the Bohr radius and $e$ is the electron charge. We can say that such an atom interacts with a conduction electron whose number is $i$ at a point $M_{i}$ (Fig. 22a) in accordance with the Coulomb law with the potential $U_{i}=e d z / r^{3}$. This interaction is very weak. In no way can it disturb the coherence of a wave function of an atomic electron, but it can shift somewhat the phases of the $S$ and $P$ amplitudes. This process corresponds to the creation of a small admixture of the $P$ state from the original $S$ state by an off-diagonal matrix element of the interaction. Our task is to estimate this admixture created by all the conduction electrons.

Let $n$ be the density of the conduction electrons. The order of magnitude of this density is given by $n \approx a_{0}^{-3} T / E_{\mathrm{F}}$, where $T$ is the absolute temperature (in energy units) and $E_{\mathrm{F}}$ is the Fermi energy. We shall introduce $l=n^{-1 / 3}=a_{0}\left(E_{F} /\right.$ $T)^{1 / 3}$ for the average distance between the conduction electrons at the upper Fermi levels.

The interaction of an atom with each conduction electron is formally described by a quantity $U_{i}$, but the total interaction of all the electrons with an atom is not the sum of all the $U_{i}$ contributions: it is strongly suppressed by the presence of an ion core and by the correlations of electrons with one another. The main correlation, which can be called a local quasineutrality, appears in the bulk of a metal. Each charge, including the charge of every electron, is surrounded by a shell with the opposite charge. In a plasma, i.e. in a system of charged classical particles, the characteristic chargescreening length is equal to the Debye radius. In metals this
length is approximately equal to the average distance $l$ between the conduction electrons. We can therefore assume that only those electrons which are in a surface layer of thickness $l$ do actually interact with an atom.

Each electron travelling from the interior of a metal in the direction of its boundary and then reflected back into the interior remains for a time $\Delta t=l / \nu_{\mathrm{F}}$ in the surface layer. Therefore, it seems that during the time interval $\Delta t$ all the surface electrons are replaced with new ones that arrive from the interior of the metal.

We shall now consider one of the electrons in the surface layer of thickness $l$. If there is only one electron, then in the time $\Delta t$ it should induce a change $\Delta a_{P i}$ of the amplitude of the $2 P$ state (on the assumption that the amplitude of the $2 S$ state is close to unity), which is of the order of $\Delta a_{P i} \approx U_{i} \Delta t / \hbar$. However, the total change in the amplitude $\Delta a_{P}$ is not always equal to the sum of the random quantities $\Delta a_{P i}$. This is because the contributions $\Delta a_{P i}$ are correlated in such a way that the sum of the amplitudes practically vanishes. This condition can be called external quasineutrality: the motion of electrons in the surface layer is perturbed weakly in such a way that the total electric field outside the metal is weak, i.e. it does not exceed the thermal noise value $E \propto T^{-1 / 2} l_{0}^{-3 / 2}$. Therefore, the effect of the total change in the amplitude $\Delta a_{P}$ seems to be always excluded. However, this is not quite true.

The interaction of the conduction electrons with an atom should in fact be considered as a single quantum process so that the phase shift $\Delta a_{i}$ applies not only to an atom but also to a conduction electron whose number is $i$. After the interaction this electron escapes into the interior of the metal and there, because of the loss of coherence, its wave function collapses so that out of a wide packet of a wave reflected on the surface only a small fraction remains after the collapse of the $\psi$ function. We can say that each pure state of the wave packet is converted into a mixed state, but then a nonforce correction may also appear in the phase $\Delta a_{P i}$. This effect is similar to a correlation pair of particles in the EPR paradox: the collapse of the wave function of one of the particles which has already ceased to interact alters the wave function of the particle correlated with it. The EPR paradox is not of the force but of the correlation type, such as, for example, the Pauli principle. Therefore, the correlation phase shifts do not obey the quasineutrality rule and the sum of the phase shifts does not vanish: these phase shifts are not due only to the average electric field acting on an atom, but also to the processes in the interior of a metal.

We shall now try to estimate the order of magnitude of this effect. We shall do this by adopting a coordinate system moving together with an atom (Fig. 22b). In this coordinate system the electrons in a metal move relative to an atom so that they approach the boundary and are reflected from it at an acute angle relative to the surface (Fig. 22b).

We shall consider two electrons, $\mathrm{e}^{\prime}$ and e, reflected from the boundary of a metal at exactly symmetric points, $\mathrm{M}^{\prime}$ and M , at the same time. A macroscopic electric field of such electrons is exactly zero simply because of the symmetry. We can readily see that, on average, all the electrons interacting with an atom can be divided into such symmetric pairs so that, if thermal fluctuations are ignored, the macroscopic interac-tion of the conduction electrons with the atom is, on average, zero. However, let us consider in greater detail what happens to each electron after reflection from the boundary of the metal (Fig. 23).


Figure 23. Wave-function packet of an electron e is reflected from the boundary of a metal and spreads into its interior. Somewhere in the region C the wave function collapses because of the scattering by impurities, inhomogeneities, or other electrons. Correlation is transferred from this collapse to a point M and to an atom A .

Let the wave function of an electron e appear as a small wave packet which is incident on the boundary of the metal, is reflected by it, and then escapes into the interior. Let the size of this packet be of the order of $b$. Then the wave function of an electron immediately after reflection can be approximated by Eqn (139) on the assumption that the $x$ axis is directed along the motion of the packet and the wave number is $k=u / \hbar m$, where $u$ is the component of the electron velocity along the $x$ axis. The order of magnitude of this component is $u \approx v_{\mathrm{F}}$. Further away from the point $M$ the wave function of a free electron evolves in accordance with the relationships given by Eqns (140) -(142), i.e. the wave packet spreads out. However, in the distance equal to the mean free path $\lambda$, i.e. at the point $C$ in Fig. 23, the wave packet is 'torn into tatters' to such an extent by the scattering on impurities, inhomogeneities, and other electrons that separate parts of the packet are no longer coherent. Under these conditions the original packet splits into several parts and the electron remains in only one of them. In other words, the wave function collapses to dimensions of the order of $b$ and the remaining parts of the packets are annihilated.

The collapse of the wave functions should not change the average electron energy. However, if a wide wave packet collapses into a compact packet of size $b$, it follows directly from the uncertainty relationship that its energy should increase by an amount

$$
\frac{3}{2 m}\left(\frac{\hbar}{b}\right)^{2}
$$

The average energy of the particle does not change if its velocity decreases by an amount $\Delta u$, so that

$$
m u \Delta u=\frac{3}{2 m}\left(\frac{\hbar}{b}\right)^{2} .
$$

The quantity

$$
\alpha=\frac{\Delta u}{u} \cong \frac{3}{2 m^{2} v_{\mathrm{F}}^{2}}\left(\frac{\hbar}{b}\right)^{2}
$$

represents the distortion asymmetry: the velocity of the 'centre of mass' of the wave packet slows down somewhat. We can say that the collapse 'carves out' the slower part of the electron wave packet. The change in the wave function of the atom therefore corresponds to a longer atom-electron interaction.

Let us now turn back to Fig. 22b. We can see that the asymmetry seems to correspond to an effective delay of the reflected packet by a time interval $\alpha \Delta t=\alpha l / \nu_{\mathrm{F}}$. It is evident from Fig. $22 b$ that after reflection at the points $M^{\prime}$ and $M$ the wave packets are not quite symmetric relative to the atom. After a displacement by $\Delta z=v_{0} \Delta t$ during the time $\Delta t=l / v_{\mathrm{F}}$ needed to cross the interaction layer the packet $\mathrm{M}^{\prime}$ is somewhat further from the atom than the packet M. In other words, at the point $\mathrm{M}^{\prime}$ there is apparently an effective charge and at the point M there is an equal (in the absolute sense) opposite charge. If we now bear in mind that in the interaction layer of thickness $l=n^{1 / 3}$ there are $n l=n^{2 / 3}$ electrons per unit area, we obtain an estimate of the effective density of the surface charge

$$
\begin{equation*}
\sigma_{*} \propto \alpha n e^{1 / 3} v_{0} v_{\mathrm{F}}^{-1} l_{0}^{-1} \tag{195}
\end{equation*}
$$

We can assume here approximately that $N^{1 / 3}=a_{0}^{-1}\left(T / E_{\mathrm{F}}\right)^{1 / 3}$. Knowing this surface charge density we can estimate [38] the phase shift $\Delta a_{P}$ for a finite aperture

$$
\begin{equation*}
\Delta a_{P} \propto 3 \alpha \frac{e^{2}}{\hbar v_{\mathrm{F}}}\left(\frac{T}{E_{\mathrm{F}}}\right)^{1 / 3} \tag{196}
\end{equation*}
$$

A comparison of Eqn (196) with experiments shows that $\alpha \approx 10^{-2}$. In accordance with the above representations, $\alpha \approx \alpha_{0}^{2} / b^{2}$ and, consequently, the size $b$ of an electron packet is $\sim 10 a_{0}$, i.e. it is slightly larger than $l$.

We can say that an effective field $E_{*}$ represents the net effect on an atom. However, $E_{*}$ is not a real electric field: it cannot be measured with a macroscopic instrument. Probably the most appropriate name for $E_{*}$ is the daemon's field. The effective field $E_{*}$ acts only on an excited atom flying past. The field $E_{*}$ is closest to the field acting in insulators, when the effective field experienced by each specific microscopic dipole differs from the average field $E$ because of a mutual correlation between the interacting dipoles.

In our case the field $E_{*} \neq E \approx 0$ appears because of a correlation in the evolution of the wave function of an atom and of the collapsing wave functions of the conduction electrons. We meet here with an effect such as the EPR correlation, not in the variant of isolated EPR pairs, but under conditions when an atom is the only primary partner interacting with an enormous number of secondary partners, which are the conduction electrons. After the interaction with the atom in the conduction layer the electrons fly back into the interior of the metal, where in the process of collapse they create correlation responses which then accumulate at the atom in the form of the phase shift $\Delta a_{P}$. In the final analysis it is the collapse that results in a dipole deformation of the atom and in a gradual appearance of the $2 P$ amplitude from the
initial $2 S$ amplitude. The Sokolov effect is a completely new type of irreversible interaction in the microworld. It is based on well-understood (in principle) microscopic processes, but it has been observed for the first time in the variant of fine correlations of the EPR pairs.

## 37. Quantum telegraph

It would be desirable to carry out additional test experiments to provide the final proof that the above discussion does indeed give the correct explanation of the Sokolov effect. However, if we assume that everything is correct, we can consider what consequences follow from the possibility of hidden collapse on a mass scale. If the polarisation of an atom is in a final analysis owing to the collapse, we can try to estimate the velocity at which the collapse effect propagates. A more practical formulation of the problem is whether it is possible to utilise the collapse to create fundamentally new means for information transfer. For example, let us consider a device shown in Fig. 24, which could be called the 'quantum telegraph'.


Figure 24. Schematic representation of the quantum telegraph based on the Sokolov effect. The conduction electrons in a sample M made of a pure semimetal or a semiconductor fly, after the interaction with an excited atom A, away from the interaction surface to a scattering region R. Their wave functions collapse in this region and simultaneously an atom $\mathrm{A}^{\prime}$ located at a distance L from the sample acquires a 2 P amplitude.

This is in fact a more complex modification of Sokolov's experiment. Once again metastable $2 S$ hydrogen atoms fly above the surface of a sample M. If this sample is made of a very pure metal, semimetal, or a semiconductor, then at liquid helium temperatures it is possible to establish conditions under which the mean free path of an electron before the scattering is $\lambda \approx 1 \mathrm{~cm}$. Let us assume that the sample has exceptionally smooth, i.e. perfectly reflecting, side walls. Then, an electron e that has interacted with an atom may travel freely a distance $\lambda$ until it reaches a 'turbid medium' R , where it begins to scatter strongly and collapses. We can assume that before the collapse of electrons the $2 P$ amplitude of the $2 S$ atom is not excited. Therefore, in the time taken by electrons to travel the distance $\lambda$ the atom moves a distance $L$ away from the sample M.

Let $v_{\mathrm{c}}$ denote the average velocity of the electrons inside the sample. In a metal this velocity is $\nu_{\mathrm{e}}=\nu_{\mathrm{F}} \approx 10^{8} \mathrm{~cm} \mathrm{~s}^{-1}$, but in a semimetal or a semiconductor at liquid helium temperatures the value of $v_{\mathrm{c}}$ may be reduced to $\nu_{\mathrm{c}} \approx$ $10^{7} \mathrm{~cm} \mathrm{~s}^{-1}$. We shall moreover assume that atoms A have a very high energy, so that their velocity $v_{0}$ is of the order of $10^{10} \mathrm{~cm} \mathrm{~s}^{-1}$. During the time $\tau=\lambda / v_{\mathrm{e}}$ taken by an electron
to travel the mean free path an atom flies away from the sample to a distance $L=\nu_{0} \tau=\lambda\left(v_{0} / v_{\mathrm{e}}\right)$.

Let us assume also that the electron scattering region R is much less than $\lambda$; for example, that it amounts to $\Delta \lambda=0.1 \lambda$. Then the collapse duration $\Delta \tau$ is also ten times shorter than $\tau$; that is, information about the appearance of the $2 P$ amplitude is transferred from the point $R$ to the point $A$ apparently at a velocity ten times higher than $v_{0}$, i.e. at $v \approx 10^{11} \mathrm{~cm} \mathrm{~s}^{-1}$. We can see that this is three times the velocity of light.

The most interesting question is: can the collapse of electrons on a massive scale be controlled? In fact, it is easy to imagine a structure in which inhomogeneities in the region R can be altered deliberately either under the influence of deformations or with the aid of an external magnetic field acting on magnetic scattering centres. In principle, it is possible to use any methods for ordering or disordering of a medium. A periodic or an aperiodic variation of the rate of the electron collapse would seem to be transferred to an atom A at the velocity

$$
\frac{L}{\Delta \tau}=v_{0} \frac{\lambda}{\Delta \lambda} .
$$

We can say that the daemon's field $E_{*}$ acts on an atom flying above the sample not immediately but after a delay time $\tau=\lambda / v_{\mathrm{c}}$. The field begins to act only at a distance $L=v_{0} \tau$ $=\lambda v_{0} / v_{\mathrm{e}}$ and creates decaying $2 P$ atoms. Lyman-alpha photons produced by such decay can be detected and the time dependence of the decay rate should reproduce duration of the collapse in the scattering zone R. We meet here a completely new principle of information transfer based on collapse control. Single events of the collapse of EPR pairs cannot be controlled, but the rate of collapse on a massive scale when electrons are scattered by impurities would seem to be controllable.

We can see from the description of the device in Fig. 24 that its not so easy to construct a practical working quantum telegraph because it is necessary to simultaneously satisfy several contradictory requirements. For example, the sample M must be very pure with perfectly reflecting side walls. The density of the conduction electrons must be sufficiently high [according to Eqn (195) the Sokolov effect is proportional to $\left.n^{1 / 3}\right]$ and the scattering zone R must be sufficiently narrow and, most important, controllable. Moreover, the beam of atoms must be sufficiently strong so that fluctuations of the Lyman-alpha photons do not suppress the useful signal.

However, at present much more interesting is not the device itself, but the principle of information transfer. Since it is based on the collapse of wave functions without any motion of matter or wave propagation, the rate of information transfer should not similarly be limited by the velocity of light. However, superluminal information transfer is so unusual, because it affects the main principles of modern physics, that the feasibility (or otherwise) of signal (not wave!) transfer at a speed greater than that of light will be considered in greater detail.

## 38. Superluminal communication

We shall denote the velocity of signal transfer by $V$ and the velocity of light by $c$. It follows from the theory of relativity that no material body and no wave can travel at a velocity higher than $c$. Therefore, superluminal communication characterised by $V>c$ cannot involve energy transfer over
a distance, i.e. it should be of completely different nature. Nevertheless, let us assume that signal transfer at a superluminal velocity is possible and consider the consequences of this assumption. For simplicity, we shall consider the case of one-dimensional propagation of signals and introduce the time $t$ and the coordinate $x$ along which the signal propagates. Let the signal be emitted from the point $x=0$ at time $t=0$. Then at subsequent times the coordinate $x$ is $x=V t$. If $V>0$, the signal travels to the right, but for $V<0$ it travels to the left.

Let us consider the question: what is seen by an observer moving at a velocity $v$ ? To answer this question we shall go over to the coordinate system of the observer and look at the external world together with the observer.

Let $x^{\prime}$ and $t^{\prime}$ be the space-time coordinates of the moving observer. It is well known they are related to $x$ and $t$ by the Lorentz transformation:

$$
\begin{equation*}
x^{\prime}=\frac{x-v t}{\sqrt{1-v^{2} / c^{2}}}, \quad t^{\prime}=\frac{t-x v / c^{2}}{\sqrt{1-v^{2} / c^{2}}} \tag{197}
\end{equation*}
$$

The point from which the signal originates in a coordinate system at rest, i.e. $x=0, t=0$, is seen from the moving coordinate system as the point $x^{\prime}=0, t^{\prime}=0$. Let us consider now how the signal propagates. If we substitute $x=V t$ into Eqn (197), we obtain

$$
\begin{equation*}
x^{\prime}=\frac{V-v}{\sqrt{1-v^{2} / c^{2}}} t, \quad t^{\prime}=\frac{1-v V / c^{2}}{\sqrt{1-v^{2} / c^{2}}} t \tag{198}
\end{equation*}
$$

Dividing one relationship by the other, we find the signal velocity $V^{\prime}=x^{\prime} / t^{\prime}$ in the moving coordinate system:

$$
\begin{equation*}
V^{\prime}=\frac{V-v}{1-v V / c^{2}} \tag{199}
\end{equation*}
$$

If $V<c$, the above relationship shows that for $v \geqslant V$ there is a change in the sign of $V^{\prime}$, which is to be expected: if the observer overtakes the signal, it sees it as lagging, i.e. as propagating in the opposite direction. If we are dealing with an electromagnetic wave, then $V=c$ and according to Eqn (199) we simply have $V^{\prime}=c$. This is the familiar result: light propagates at its constant velocity in any coordinate system. This is the portulate that underlies the theory of relativity.

However, let us assume that $V>c$. Eqn (199) directly yields a coordinate system in which the signal velocity $V^{\prime}$ is infinite. This coordinate system moves at a velocity $v=c^{2} / V<c$. It is evident that the coordinate system with the infinite velocity $V$ is for some reason special. We shall therefore assume that it is a coordinate system at rest and, going to the limit $V \rightarrow \infty$, we find from Eqn (199) that

$$
\begin{equation*}
V^{\prime}=-\frac{c^{2}}{v} \tag{200}
\end{equation*}
$$

We can see that in the coordinate system moving to the right the signal moves to the left and for $v<0$ the signal moves to the right, i.e. $V^{\prime}>0$. In the limit $v \rightarrow 0$, we have $V^{\prime} \rightarrow \pm \infty$, i.e. the signal propagates at an infinite velocity in both directions at the same time. Naturally, these relationships are somewhat simplified and idealised, because the signal emission time is assumed to be zero.

At first sight the relationship (200) seems to be in clear conflict with the principle of causality. Let us assume that indeed the propagation of a signal at the velocity $V^{\prime}>0$ from the coordinate system with $v<0$ is observed. For example, the signal sender may be at the point $x^{\prime}=0, t^{\prime}=0$, and
then the signal reaches the recipient at the point $x^{\prime}=L$ somewhat later, i.e. at $t^{\prime}=L / V^{\prime}>0$. It is quite clear here which is the cause and which is the effect. However, the observer travelling at a velocity $v>0$ sees the signal at a velocity $V^{\prime}<0$, i.e. the cause and the effect are interchanged. It would seem that because of this absurdity the superluminal information transfer is impossible. However, we must not be too hasty! Passive observation by itself means little. The real conflict with the principle of causality occurs only if the information recipient can send the signal back to the sender before the emission of the first signal and therefore the cause of the effect may be changed. Let us consider whether this can occur and if so, what additional constraints ensure that the principle of causality is obeyed.

Let there be two events A and B related by the cause and effect link. Let us assume that $A$ is the cause and $B$ is the effect. F or example, in the quantum telegraph of Fig. 24 the event $A$ is the collapse of the wave functions of the conduction electrons and the event $B$ is the appearance of the $2 P$ amplitude in the $2 S$ atom and the emission of Lymanalpha photons. Let us select a coordinate system in which the signal transfer rate is infinite. Thus both events A and B occur at the same time $t=0$. This means that they are in the same spatial segment $A B$ (Fig. 25). The selected coordinate system with $V= \pm \infty$ is clearly special. Therefore, the time $t$ in this system has certain features of absolute time, but this point will be considered later.

Let us assume that these events are observed from the coordinate system $x^{\prime}, t^{\prime}$ moving at a velocity $v>0$. Then the result obtained from Eqn (197) for $t=0, x=L$ is

$$
t^{\prime}=t_{\mathrm{B}}^{\prime}=-\frac{L v}{c \sqrt{c^{2}-v^{2}}}<0
$$

In other words, this observer sees the signal $B$ earlier than $A$ and undoubtedly is very surprised. Even more surprising is that at the point $x=0$ the time $t_{\mathrm{B}}^{\prime}$ corresponds to the time of the event $A^{\prime}$, which is

$$
t_{\mathrm{A}^{\prime}}=-\frac{v}{c} \frac{L}{c}
$$

Bearing this in mind, the observer B may want to influence the event A by, for example, attempting to remove the sample $M$ in the experiment of Fig. 24. We can imagine for example that near the sample $M$ there is an automaton which removes the sample immediately on receiving an order from outside. In particular, the order may be a signal sent by the observer in the primed coordinate system. Naturally, if no additional


Figure 25. Two events A and B occur simultaneously at $t=0$ and are separated by a spatial segment of length $L$. An observer, moving at a velocity $v>0$, sees the event B earlier than the event A . The dashed lines represent $t^{\prime}=$ const.
constraints are imposed on the signals, the causality principle is lost immediately.

We shall therefore formulate three additional principles:
(I) each new coordinate system $x^{\prime}, t^{\prime}$ corresponds to a macroscopic body which exists in reality;
(II) a superluminal telegraph associated with this body transmits in the body a signal at the velocity $V^{\prime}= \pm \infty ;$
(III) a superluminal control signal cannot propagate in the opposite direction of the time $t$ of the main coordinate system if the two coordinate systems are informationcoupled.

Let us return to Fig. 25. It is clear that any signal cannot be transferred by a material carrier from the point B to the point A faster than in $\Delta t=L / c$ because the velocity of such signals is less than the velocity of light in any coordinate system. Therefore, $\Delta t>-t_{\mathrm{A}^{\prime}}$, i.e. a combination of just one superluminal signal from A to B together with an ordinary light signal from $B$ to $A$ does not violate the principle of causality. In other words, it is sufficient to consider just the superluminal telegraph satisfying the principles (I)-(III).

We shall assume that we have a telegraph sending signals at the velocity $V^{\prime}=x^{\prime} / t^{\prime}$ in the primed coordinate system. With the aid of Eqn (197) we can easily find that in the laboratory coordinate system the signal velocity $V$ is

$$
\begin{equation*}
V=\frac{V^{\prime}+v}{1+v V^{\prime} / c^{2}} \tag{201}
\end{equation*}
$$

In the limit $V^{\prime} \rightarrow \infty$ this velocity is

$$
\begin{equation*}
V=\frac{c^{2}}{v} \tag{202}
\end{equation*}
$$

We can see that if $v>0$, the velocity $v$ is positive. If we adopt the principle (III), this signal can only propagate to the right. This means that the telegraph in the primed coordinate system cannot send a signal back from the point $B$ to the point A, i.e. it cannot influence the cause although it can 'see' it as occurring after $B$. The signal can be sent to the left, i.e. from $B$ to $A$, if there is a 'doubly primed' superluminal telegraph moving at the velocity $v^{\prime \prime}=-\left|v^{\prime \prime}\right|$. According to Eqn (202) the velocity $V^{\prime \prime}$ of the signal from this telegraph is $V^{\prime \prime}=-c^{2} /\left|v^{\prime \prime}\right|$. However, this telegraph sees the event B at the moment

$$
t^{\prime \prime}=\frac{L / v^{\prime \prime}}{c \sqrt{c^{2}-v^{2}}}>0
$$

[we are using Eqn (197) again but with the substitution $\left.v \rightarrow\left|v^{\prime \prime}\right|\right]$. In other words, for this telegraph the event B is indeed the consequence of the cause $A$ and even for a signal travelling at an infinite velocity it cannot influence the cause A.

A decisive role in our discussion is played by the principle (III). This principle imparts the features of absolute time to $t$, which - strictly speaking - is in conflict with the principle of relativity. Therefore, we should consider in greater detail the physical meaning of the constraint set by this principle, which seems to be quite natural for an observer at rest.

We shall therefore approach this problem from a somewhat different standpoint. We shall implicitly assume that there is no way that superluminal signals can be generated except on the basis of the wave collapse and we shall consider a certain wave function $\psi=\exp (-\mathrm{i} \omega t+\mathrm{i} k x)$. Let this wave function apply to a particle of mass $m$ and a momentum $\hbar k$. The quantity $\hbar \omega$ represents the energy of this
particle. We shall allow for the kinetic energy of the particle as well as for the rest energy, so that in the nonrelativistic limit the energy is $\varepsilon=\hbar \omega=m c^{2}+\frac{1}{2} m v^{2}$ and $p_{k}=\hbar k=m v$, where $v$ is the particle velocity. Let $V=\omega /$ $k$ represent the phase velocity. As is known, the phase velocity is the velocity of motion of a point with a constant phase $\varphi=\omega t-k x=$ const. Let us consider how this point moves in a primed coordinate system which is travelling at a relative velocity $v$ relative to a coordinate system at rest. This can be done by using the relationships of Eqn (197) to express $x, t$ in terms of $x^{\prime}, t^{\prime}$; these expressions are then substituted into $\varphi=\omega t-k x=$ const. The result is an expression for a new frequency $\omega^{\prime}$, a new wave vector $k^{\prime}$, and a new phase velocity $V^{\prime}$ :

$$
\begin{align*}
\omega^{\prime} & =\frac{\omega-k v}{\sqrt{1-v^{2} / c^{2}}} \\
k^{\prime} & =\frac{k-\omega v / c^{2}}{\sqrt{1-v^{2} / c^{2}}}  \tag{203}\\
V^{\prime} & =\frac{V-v}{1-V v / c^{2}}
\end{align*}
$$

We can readily see that the expression for $V^{\prime}$ is exactly the same as that given by Eqn (199). We can therefore directly draw the conclusion that there is a special coordinate system moving at the velocity $v=p_{k} / \varepsilon=c^{2} / V=c^{2} k / \omega$. In this coordinate system the phase velocity $V^{\prime}$ becomes infinite and the wave number $k^{\prime}$ vanishes, so that the wave is converted simply into oscillations homogeneous in $x^{\prime}$ and of frequency $\omega^{\prime}$. At the selected frequency we have $\hbar \omega^{\prime}=m c^{2}+\frac{1}{2}\left(m v^{2}\right)$ and in the nonrelativistic limit we obtain $\hbar \omega^{\prime}=m c^{2}$. This means that we are dealing simply with a particle at rest and its wave number is zero. It is usual to regard the phase velocity as not having much physical meaning: the energy, and consequently the wave information, propagate at the group velocity and not at the phase velocity. However, the phase velocity plays a major role in the collapse.

We shall consider a typical example of an EPR pair in the Bohm variant: two particles with spin $\frac{1}{2}$ fly apart in opposite directions and their total momentum is zero. The two partners of such a pair have equal and oppositely directed momenta and exactly the same phases at the same distance from the point where their motion begins. A measurement carried out on one particle immediatelly collapses the wave function of the second particle to the spin value corresponding to the oppositve direction. It is natural to assume that this process is instantaneous in a coordinate system in which the centre of mass is at rest. In other words, the velocity of the signal about the collapse is $V=\infty$. However, in a moving coordinate system we find from Eqn (200) that $V^{\prime}=-c^{2} / v$. Depending on the sign of $v$, this quantity can be positive or negative. It means that for some observers the signal of collapse to the second particle arrives after a delay, which is quite natural. F or other observers it arrives ahead of the collapse, i.e. in the reverse direction in time from the future to the present. This seems to be a fantasy, but there is no violation of the principle of causality: the collapse of correlated functions is a purely random process, i.e. it is uncontrollable. Therefore, in the collapse of isolated EPR pairs there is in fact no information transfer: this is simply a unique process without cause or effect. In this sense the signal about a single collapse resembles the phase velocity: it exists, but carries no information.

The analogy has in fact a deeper meaning. After all, the collapse appears when the coherence is lost. The coherence between two points in a wave can be destroyed by a very small external perturbation most easily if for a considerable time there is a constant phase difference between the points. In other words, they travel at the phase velocity of the wave. The phase velocity itself carries no information, but makes it possible to impose 'markers' on those parts of the wave which can subsequently collapse.

We shall now consider an EPR pair moving at a velocity $v$. In its own coordinate system the collapse corresponds to the transmission of a signal at an infinite velocity $V^{\prime}=\infty$. In the laboratory coordinate system the velocity is $V=c^{2} / v$, as is clear from Eqns (201) and (202). Once again, depending on which of the partners the first measurement is carried out, the signal of collapse can be sent to the second partner either into the future or into the past. There is again no violation of the principle of causality, because the measured result is purely random: recipes from quantum theory can be used to calculate the probability of the results of measurements, but the results themselves cannot be predicted a priori.

Therefore, although the collapse represents an irreversible process, there is a great variety of such processes: they occur absolutely spontaneously and cannot be controlled from outside if we consider only separate elementary events. The question arises whether in principle it is possible to consider any forms of controlled quantum collapse. The example of a chain reaction in an atomic reactor suggests that there is some hope of a positive answer. After all, the whole reaction is based on elementary quantum transitions, each of which cannot be controlled. However, if the probabilities of the transitions are controlled, they are multiplied by the large number of nuclei participating in the process, which makes them automatically the corresponding macroscopic variables of nuclear kinetics. Control then becomes possible. It therefore follows that many participants of a process are needed if the process is to be controlled.

We shall go back to the quantum telegraph of Fig. 24. We shall first consider an elementary event: an excited atom A flies above a sample which contains the conduction electrons and then the electrons travel deeper into the metal and participate there in the collapse event, whereas the atom A acquires a $2 P$ amplitude which may create a photon. If this photon is detected, this represents a 'measurement' in which the collapse of an atom into the $2 P$ state with a subsequent transition to the $1 S$ state takes place and at the same time in the region R of the sample M the occurrence of numerous events of the collapse of wave functions of the conduction electrons is confirmed. At first sight this seems to be the sole random collapse process: a detector of Lyman-alpha radiation records a photon and inside the metal numerous wave functions of the electrons collapse. It would seem that such a collapse should be detectable from outside. The process has no cause or effect. Therefore, the signals of the collapse may be transferred between the electrons and the atom at an infinite velocity and observers moving outside may observe such collapse events in different time sequences.

Let us, however, assume that we have a whole beam of excited atoms and, consequently, we can carry out a long series of identical measurements. A change in the structure of the region R makes it possible to control the probabilities of the collapse of the electrons in this region, i.e. to alter the coefficient $\alpha$ in Eqn (196) for the daemon's field $E_{*}$. In this
way we seem to control the rate of decay of excited hydrogen atoms at a large distance from the sample M. The control signal travels together with the collapse signal and, therefore, is superluminal. In actual fact this represents control of the probabilities of real quantum 'measurement' processes, i.e. of irreversible information processes in their relationship to external classical devices and objects (in this case, observers). This process is clearly irreversible and determines the direction of the 'arrow of time'. The time $t$ inside the sample M assumes the nature of the 'absolute time' and can no longer be reversed: the principle of causality is not compatible with reversibility in time.

It is obvious that a superluminal control signal cannot propagate in the reverse direction of time $t$ in that system of coordinates in which the telegraph is at rest. For the same reason of irreversibility of the processes of measurement and control of the probabilities in the case of the telegraph in Fig. 24 it is quite clear that the cause of the changes in the rates of decay of the atoms A can only be a controlled change in the collapse probability occurring in $R$. The cause is $R$ and the effect is $\mathrm{A}^{\prime}$, so that the control signal travels from R to $\mathrm{A}^{\prime}$ (possibly with a short delay $\Delta t$ required to change the probabilities in R).

More difficult is the question how the telegraph of Fig. 24 acts when it is set in motion and observed from a coordinate system at rest. In proper time this telegraph receives a control signal at the velocity $V^{\prime}=\infty$. Formally, the signal in a coordinate system at rest may propagate along $t$ or against $t$. On the other hand, however, this signal is irreversible. The question now is how to resolve this paradox.

We have decided earlier to impose an additional constraint (III) the meaning of which is that in the usual (conventional) time $t$ of immobile bodies the irreversible processes should occur only from the past to the future, i.e. there should exist the 'arrow of time'. However, further analysis is needed to understand better the meaning of this principle.

We shall return to this problem later. At this stage we note that the spatially correlated collapse events simply must lie on spatially similar world lines. Otherwise we would have collapse events consecutive in time (in one of the mobile coordinate systems) and this would have led to even greater conflicts with the principle of causality if there are indeed no hidden variables. Thus, the most natural assumption is that the correlated collapse events occur 'instantaneously' when they are, for example, located on lines of the phase velocity $v_{\mathrm{ph}}=\varepsilon / p=$ const of the centre of mass of the correlated system. The velocities $\varepsilon / p$ are always superluminal and random measurements $K$ take place along such velocities. However, in order to understand how this occurs, we must consider how the collapse events agree with the principle of relativity.

## 39. Present, past, future

What is the present? Each of us has a ready answer: this is our world, because all that has happened earlier has taken place in the past. Until the appearance of the theory of relativity such an answer would have been in full agreement also with a more rigorous mathematical, i.e. quantitative, treatment (Fig. 26a).

In the simplest case of one measurement of $x$ we have a space which travels uniformly along the time axis $t$. In this approach time is absolute; at all points $x$ it flows at the same


Figure 26. (a) Before the theory of relativity the structure of space-time was considered as uniform motion of the whole space along the world time coordinate $t$. (b) Since the velocity of light is constant, the real 'visible' present lies in the shaded layer and the line $t=$ const here represents an arbitrary time.
rate and absolutely identically for immobile and moving bodies.

In the theory of relativity everything is more complicated. Its main principle states that there are no signals propagating faster than the velocity of light and that this velocity is the same in immobile or moving coordinate systems. To people living on Earth this means that simultaneity on the Earth's surface has been determined to within $2 R_{0} / c \approx 40 \mathrm{~ms}$, where $R_{0}$ is the radius of the Earth. And if we look at the starry sky, we can see stars along the $t=-L / c$ line, where $L$ is the distance to the star and $t$ is the time in the past measured from the moment of observation. Hardly any of those watching the stars will agree that what is seen is not 'now'. This 'present visible' corresponds to the lines $x= \pm c t$, extending to negative times. For all the inhabitants of the Earth the present is defined to within $2 R_{0} / c$.

What does $t=$ const mean? Strictly speaking this isochronous line, i.e. a line with the same time, is purely arbitrary. It can be obtained as follows. First, a set of clocks is placed along the $x$ axis. The clocks are identical physical bodies or mechanisms that are in periodic motion such that the period is constant. These clocks are then synchronised by, for example, a light signal travelling from a given clock to the neighbouring one and vice versa. All these effects have taken place in the past and, therefore, an extended periodic object with the same instantaneous phase is formed: at $t=$ const all the 'arrows of time' point in the same direction.

If we consider each of the clocks as an oscillator with a period $T=12 \mathrm{~h}$, we obtain a wave with wave vector $k=0$ and frequency $\omega=2 \pi / T$. Clearly, such a classical object with an infinite phase velocity can be formed only by evolution from the past and at this moment $t$ there is no physical link between the clocks.

The arbitrary nature of such 'simultaneity' can be seen quite clearly by considering the example of ordinary clocks on the Earth. They can be set so that they would show exactly 12.00 hours at midday. Then, viewed from space, we can see a wave constructed from the phases of the clocks. The wavelength is $2 \pi R_{0}$ and the phase velocity of the wave is exactly equal to the velocity of rotation of the Earth and oppositely directed. The constant phase equal to 12 h will always point to the Sun. There is no internal physical link between such clocks: they are set on the basis of external information, i.e. they are set by the Sun!

We shall now consider a single observer at rest at the point $x=0$ at a time $t$ (Fig. 27).

The world line of an observer at rest corresponds to the half-axes of the ordinates, arriving from $t=-\infty$ at the point


Figure 27. Observer O at the point $x=0$, at the moment $t=0$ finds the real 'past' in the shaded region bounded by the lines $x=c t$. The lines A and $B$ correspond to moving bodies.
$t=0$. At the point O with the coordinates $x=0, t=0$ there may also be moving bodies, for example A with a positive velocity and $B$ with a negative velocity. According to the theory of relativity the maximum velocity at which any material objects can travel, and this includes the waves, is equal to the velocity of light. Therefore, at the point $O$ it is possible to collect only that information which existed earlier, inside the shaded region in Fig. 27. This is the past, which can be directly seen, heard, touched, or reconstructed on the basis of all the monuments and 'fossils' located at the point O .

We thus reach a somewhat paradoxical conclusion that the future of the observer O lies not at $t>0$ but throughout the unshaded part of Fig. 27. In any case, this should be the situation as seen by a single observer. As far as the past is concerned, it should be a region accessible to passive acquisition of information, which arrives with electromagnetic waves or which is stored in any moving material carriers (for example, sound, mail, etc.).

We shall now consider a region of bidirectional information communication between an observer O and any other observer in the past (or in the 'present').

Let us assume that three observers A, O, and B are at rest (Fig. 28a). In the state $\mathrm{O}^{\prime}$, i.e. at the moment $-t$, an observer may send signals with the maximum velocity $c$. These signals reach the world lines $\mathrm{A}^{\prime} \mathrm{A}$ and $\mathrm{B}^{\prime} \mathrm{B}$ of the other two observers at the points $A^{\prime \prime}$ and $B^{\prime \prime}$, which in turn send their own 'responses' to the point O. It is evident from Fig. 28 that the observers A and B are separated by a distance $L= \pm c t / 2$. Inside these world lines there is a region (shown shaded) of bidirectional communication between the observer O at rest and the other observers also at rest. In particular, there may be clocks at the points A, O, and B and then bidirectional communication may be used to synchronise clocks at rest, i.e. to plot the lines $t=$ const.

Mobile observers may also congregate at the point $x=0$ at the time $t=0$. If they also are in bidirectional communication with other objects, then all the information on bidirectional communication is collected at the point O . For example, in Fig. 28b a mobile observer at the position $\mathrm{O}^{\prime}$ sends, at a time $-t$, information at the velocity of light to the point $\mathrm{A}^{\prime \prime}$ on the world line $\mathrm{A}^{\prime} \mathrm{A}$ of an observer at rest, and


Figure 28. Space - time region of bidirectional communication between an observer $O$ and other observers: (a) observer $O^{\prime}$ at rest; (b) observer $O^{\prime}$ moving along the line $\mathrm{O}^{\prime} \mathrm{O}$.
then the response from the second observer appears at the point O .

We can see that if all the mobile observers congregate at the point $O$, then all the information on bidirectional communication in the region between the lines $x= \pm c t$ is collected. In other words, in addition to the results of passive observations, it is possible to collect at this point the information on bidirectional communication.

We shall now consider two observers A and B at rest and separated by a distance $L$ (Fig. 29).

Neither of the observers has any special features to distinguish him from the other. However, one of them, for example A, sees B at a somewhat earlier time at the point $\mathrm{B}^{\prime}$, whereas B experiences an opposite effect: it sees A somewhat earlier at the point $\mathrm{A}^{\prime}$. Naturally, the two results should match, because at any given moment they correspond to $t=$ const and because of the signal delay they do not have instantaneous information on one another. The region covered by dots in Fig. 29 is the shared past of the observer A and the observer B.

Obviously, to within an error $\Delta t=L / c$, the observers A and $B$ can be regarded as sharing the time even from the point of view of their mutual information coupling. This shared time they naturally regard as the present. For shorter characteristic times of physical processes the observers are more likely agree with the delayed signals than with their own existence in different times.

In the theory of relativity the concept of simultaneity is even more complex. Specifically, the Lorentz transformation to the system of coordinates $x^{\prime}, t^{\prime}$ moving at the velocity $v$ is described by the expressions in Eqn (197). Lines correspond-


Figure 29. Observers A and B separated by a distance $L$ from one another. A sees $\mathrm{B}^{\prime}$ and B sees $\mathrm{A}^{\prime}$; the shaded region is their common past.
ing to the same time $t^{\prime}=$ const do not coincide with the lines corresponding to $t=$ const (Fig. 30). However, the boundary of the past $x= \pm c t$ remains in its place also in the new system, i.e. $x^{\prime}= \pm c t^{\prime}$.

When the velocity $v$ is increased the slope of the line $t^{\prime}=$ const relative to the $x$ axis increases, but it never reaches the position $x=c t$ and approaches it only in the limit $v \rightarrow c$. We can now consider the concept of the future. In the framework of the Galilean invariance characterised by the absolute time $t$ the future relative to $t=0$ is defined by the simple relationship $t>0$. When the time increases by a small amount $\Delta t$ the realisation of the future can be viewed simultaneously for all values of the coordinate $x$. If we bear in mind that there are no bodies or wave signals travelling at a velocity exceeding $c$, then the 'active' future for the point $x=0, t=0$ lies in the range $t \geqslant|x| / c$ which can be reached by light signals or moving bodies from the point $x=0$, $t=0$. Only in this region the material point $x=0, t=0$ can act to influence the evolution of its environment. This region is identified by the same properties also in the case of moving bodies with world lines passing through the point $x^{\prime}=0, t^{\prime}=0$.

All the discussions in the present section are based on the ideas of the theory of relativity and classical field theory. We shall now consider what new features are introduced by the collapse of wave functions. As established earlier, the collapse of functions of correlated particles occurs on world lines of the phase velocity of their centre of mass. When the


Figure 30. Coordinates $x^{\prime}$ and $t^{\prime}$ of a moving coordinate system in an $(x, t)$ plane.
centre of mass is fixed, this line corresponds to $t=$ const and for the centre moving at the velocity $v$ it is the line $t^{\prime}=$ const. According to Figs 28 and 29, the collapse event seems to occur in the future. This is the reason why they are random and causeless. However, the collapse events are irreversible. This becomes particularly clear when the collapse of a wave function is accompanied by the collapse of probabilities, and it is true in measurements. In the case of this process the information about a quantum system increases and the entropy should increase in the external world. However, even in the absence of the collapse of probabilities, the wave function collapse is irreversible: in an ensemble of many particles it converts a pure ensemble into a mixed one. In the case of just one partner a pure state 'collapses' into a random 'representative' of a mixed ensemble. The irreversible process of collapse is related to the information interaction of a given system with the external world. If this is true, then the time $t$ in the coordinate system associated with the external environment becomes the absolute time. On Earth, and more precisely in the solar system, this time is linked to a system of coordinates in which the centre of mass of the solar system is at rest.

The line $t^{\prime}=\operatorname{const}$ (Fig. 30) can be regarded as the world line of a constant phase velocity on which the collapse of corre-lated particles may take place and the centre of mass of these particles travels at a velocity $v$. If a system of particles and bodies with the centre of mass at rest in the coordinate system $x^{\prime}$ and $t^{\prime}$ is completely (also in respect of information) isolated from the external world, the collapse events can occur in the system simultaneously along the line $t^{\prime}=$ const.

From the point of view of absolute time $t$ all looks exactly the same except that the collapse events associated with the external world (for example those observed from the outside) should occur consecutively in the direction of the increasing time $t$. The collapse events can occur in the opposite direction of $t$ along the line $t^{\prime}=$ const only in the absence of observation or information coupling with the external world. Only in that case the two directions of time in a closed reversible quantum systems are fully equivalent. In reality even a weak information coupling to the external world leaves only $t$ as the absolute time.

We can now return to the quantum telegraph. It utilises objectively occurring collapse events. A change in the configuration of a physical system imposed from outside, which thus influences the probabilities of random natural collapse events, makes superluminal information transfer possible. However, this information transfer is based on irreversible processes in the absolute time $t$ and cannot occur in the opposite direction along $t$ : the arrow of time is quite clearly encountered here. In this way we reach in a natural manner the principle (III), i.e. that propagation of superluminal information signals from the future to the past cannot occur in the absolute time $t$. This time variable corresponds to the reference frame where the centre of mass of the combined system is at rest.

In order to make it clear why the existence of the quantum telegraph should not be rejected a priori, it is useful to consider the following analogy. The main means for information transfer in the animal world, including mankind before the invention of radio, have been sound and light. Sound is used actively: our speech is formed by sound and we receive it by hearing. In the case of light we receive passively the solar radiation scattered by objects. Since the invention of radio, electromagnetic waves are generated actively by transmitters
and reach receivers, i.e. they are used in the same way as sound. From this point of view the superluminal telegraph is an analogue of light and colour observed in daylight. The wave functions of the microworld are under conditions of a continuing process of consecutive collapse. Variation of details of such a collapse at one point in space makes it possible then (i.e. after a small shift in the absolute time $t$ ) to change the probabilities of the collapse at other points in space. The collapse itself cannot be transferred to another point in space. However, it is quite feasible to vary in a controlled manner the classical environment of a set of collapsing systems at one point in space and to detect this influence on the collapse at another distant point in space.

## 40. Free will

Before returning to the collapse events, let us make a trip to a completely different field. Specifically we shall try to understand how to deal with the phenomenon of free will within the framework of physics.

By free will we understand here the freedom of action or of selection between two or more possibilities. It is usual to assume without question that human beings have a free will because they are free in their conduct. Naturally, man frequently has to act under the force of external circumstances, but even then he retains the final choice.

It is not true that absolutely everybody accepts these statements as true. For example, following Schopenhauer, one could say that man analyses only his desires and at the last moment the decision may be outside his control. In other words, the exercise of his will may be imposed from outside.

However, we shall retain a more naive point of view and assume that man is free in his deeds and is therefore responsible for them. In accepting the freedom of action for man we must not offend the animal world. The paradox of Buridan's ass has been passed on to us from ancient times: an ass exactly half-way between two bundles of hay dies of hunger because it cannot decide from which one to begin its meal. We all know that this does not happen in reality and that the ass chooses one hay bundle without any difficulty. But the logical paradox remains. Its simplest solution is based on small perturbations: an accidental breeze may bring a stronger smell of one of the bundles and this is sufficient to ensure that the ass makes its choice. Observations of animals lead us to the conclusion that the ass anyway is free to make its choice unrelated to any external motivation: all that it needs to know is the existence of two bundles of hay.

Moreover, any animal is alive because it has to make decisions how to feed itself and not to become food for a beast or bird of prey. The higher a given species is on the evolution tree, the greater the spectrum of decisions that it has to take. However, we cannot accept the hypothesis that the freedom of action appears abruptly at some stage of the evolution: even the most primitive members of the animal world retain the freedom of action. Moreover, it is very difficult to set the boundary for the freedom of will between living and inani-mate worlds. Its much more natural to assume that the free-dom of will is an immanent, i.e. an intrinsic property of the universe. Only on the basis of this initial assumption can we escape from the thoughtless completely deterministic mech-anistic universe to the universe which is alive and developing.

We shall therefore assume that the universe as a whole has the freedom of will, i.e. the ability to take decisions and freely
act within those restraints which are imposed by the laws of physics, including those of classical physics. This freedom of action is realised in the form of an enormous range of small free actions and each of them should fit within the framework of the physical laws. This means that the freedom of action can be realised only at bifurcation points where the laws of mechanics and physics permit multipath development of a process.

Let us begin with classical physics. A typical example of bifurcation can be found in Fig. 15b, which shows an unstable position of a material point at the top of a hump between two potential wells. The instability of the initial state causes the material point to roll down into one of the wells. The result is spontaneous symmetry breaking. The process can be regarded as the result of evolution under the influence of an initial perturbation. The perturbation itself can be regarded as completely random, unrelated to any cause. However, equally well we can say that the universe as a whole (including a large number of small links, i.e. perturbations) 'makes a decision' on the subject of this spontaneous symmetry breaking. Similarly, random bifurcations can be regarded as occurring in a causeless and spontaneous manner, i.e. as if they have been adopted in a volitional manner from outside the system.

Let us now turn to the microworld. Large-scale bifurcations can occur in systems of many particles. However, of much greater interest are the smallest bifurcation transitions, i.e. the collapse of wave functions. It is naturally tempting to consider the collapse events as microinstabilities of a certain kind. One could assume, for example, that the wave function which splits into incoherent parts is unstable and that a particle may be in one coherent subpacket. This approach is close to the idea of hidden variables and can hardly serve as the basis for complete understanding of microprocesses. Moreover, we can easily give a counterexample which contradicts this hypothesis. Specifically, let us again divide one wave function between two potential boxes and then heat the boxes, so that the coherence between the two parts of the wave function is destroyed. The collapse of the wave function occurs in one of the boxes and this is where the particle is located. The particle cannot jump from one box to the other.

Therefore, it is more logical to assume that the wave function is above all a 'measurement index'. This can be seen particularly clearly in Schwinger's formalism when measurement symbols are introduced. These symbols are simply possibilities or intentions of a particle to manifest some value of a physical quantity. The real physical quality is found only by measurement when the intention collapses, i.e. when the wave function collapses, and this is accompanied by the collapse of probabilities, which is the recording of the relevant quantity by a measuring instrument.

If we now turn back to Section 24, we can demonstrate once again that the evolution of the state vector, i.e. of the wave function, in time is in a natural way inscribed in the formalism of time evolution of the measurement symbols. Exaggerating somewhat, we can say that the whole quantum theory represents the formalism for the description of the time evolution of measurements in the microworld. Even in quantum theory the field operators evolve with time only to be able to act on the stationary state vector, which is the quintessence of the intentions in the microworld. We can now understand at a qualitative level the evolution of the nature on the Earth, as observed by us. The most general approach to the description of atoms, molecules, and fields which
form our environment is based on the use of the wave function $\psi$-i.e. of the state vector-which obeys the following equation:

$$
\begin{equation*}
\mathrm{i} \hbar \frac{\partial \psi}{\partial t}=H \psi+M \psi \tag{204}
\end{equation*}
$$

Here, $H$ is the Hamiltonian operator and $K$ is the collapse operator. In the absence of collapse events, Eqn (204) describes the evolution of a pure state for the whole matter on the Earth including the biosphere. We then have

$$
\psi(t)=\exp \left(-\frac{\mathrm{i} H t}{\hbar}\right) \psi(0)
$$

where $\psi(0)$ is the initial state vector. Instead of $\psi(t)$ we can use the Heisenberg representation by introducing timedependent operators. These operators act on the stationary state vector $\psi(0)$. All the relationships then have Lorentzinvariant form.

However, $\psi(0)$ has an enigmetic form and in order to describe real processes allowing for thermal motion we have to adopt additional statistical averaging. Such averaging in fact implicitly implies the occurrence of collapse events which create a mixed ensemble from a pure state. If statistical averaging is not carried out, but the existence of collapse events is assumed, then the operator $M$ should be regarded as random because it is an analogue of a set of causeless volitional events. Naturally, the state vector $\psi(t)$ also becomes a random quantity.

It is natural to assume that the collapse events $K$ realise a nonequilibrium evolution of a system. This is particularly clear in the processes similar to measurements when the collapse of wave functions is accompanied by the collapse of probabilities. As demonstrated in the preceding sections, the negentropy of a collapsing system should then increase because of an increase in the entropy of the environment. It is clear that this can only give rise to a nonequilibrium system. On Earth the most powerful source of nonequilibrium is solar radiation and, therefore, the rate of collapse (of the state vector and probabilities) is governed by the negentropy flux from the Sun.

However, the collapse of wave functions need not be accompanied by the collapse of probabilities because of, for example, thermal motion. Nevertheless even in this case a nonequilibrium environment plays an important role. The situation is similar to that of molecular chaos. We have seen that even a weak coupling to the external world considerably alters the evolution of a system of many particles: in a closed system the reversible Liouville equation applies and the coupling with the environment destroys the reversibility of time. A similar situation also arises in the quantum case: a closed system evolves as a pure state and the coupling to the external environment disturbs the coherence and leads to collapse events.

The structure of Eqn (204) suggests that a real physical system includes both the cause-and-effect Lorentz-invariant evolution of the state vector, -i.e. the evolution of 'intentions'-and the random 'volitional' sequence of actions, i.e. the collapse events $M$. The collapse of the wave functions on the Earth is in the final analysis due to the collapse of solar radiation photons in the cascades of their transformations into thermal motion of atoms and molecules. The rate of collapse, i.e. the absolute value of the nonlinear operator $M$, is governed by the level of the solar negentropy flux. The collapse operator need not be Lorentz-
invariant. It acts primarily in a coordinate system linked rigidly to Earth. In a system of correlated particles at rest the collapse operator acts simultaneously in the whole space, i.e. it acts 'instantaneously' in terms of the absolute time $t$ to within a constant time of the process $\Delta t$. The collapse operator for mobile correlated systems of particles, not isolated from the environment, acts consecutively in time $t$ on a world line of the phase velocity of the centre of mass.

It follows that even in terms of the microscopic structure of matter and fields we have to distinguish the cause-andeffect and 'volitional' (i.e. spontaneously acting) aspects of the evolution of the world. Without the collapse event the loss of coherence of the separate parts of the wave function would have resulted in a branched scenario of the development of the world: according to Everett III [27], we can imagine many worlds developing in parallel. In fact, however, we live in one natural world: the spontaneously occurring 'volitional' collapse events create a unique and unrepeatable line of evolution and development of the world. The addition of macroscopic bifurcations does not alter the qualitative picture, but greatly extends the range of possible scenarios out of which the history chooses just one.

We shall go back again to the problem of the past, present, and future, but we shall now assume that superluminal signals may propagate because of the collapse of wave functions. We shall select a coordinate system linked to the solar system. We can assume that the bulk of matter in such a coordinate system is at rest because the motion of all the macroscopic bodies takes place at velocities much less than the velocity of light. The time $t$ in such a system will be regarded as 'absolute'. Then, in the simplified pattern of onedimensional motion, a simplified graph of the past, present, and future for a particle at $x=0, t=0$ has the form shown in Fig. 31. In this figure the hatched region P corresponds to the past: an observer at the point $x=0, t=0$ can receive signals transported by material carriers (waves, particles) from the whole of this region. The boundary N of this region corresponds to the present: this is what the observer at $x=0$,


Figure 31. Graph of the past, present, and future for a point $x=0$ at the time $t=0$. The shaded region P covers the past events and its boundary N corresponds to the present. The region F is the dynamic future: the point $x=0, t=0$ can be the cause of the events in this region. In the regions $C$ an information coupling is possible with other points of the spatial intervals: the continuous line $t=0$ corresponds to the collapse in the 'absolute time' and the dashed lines represent the collapse in moving bodies which are information-coupled to the main body at rest. The arrows indicate possible directions of control signals.
$t=0$ sees around himself, including the stars in the distant galaxies. All that this is outside $P$ is the future: if the observer moves uniformly along the $t$ axis, then sooner or later any point outside P passes through the 'present' crossing the moving boundary N. However, the future divides naturally into two regions F and C .

The region F is the dynamic (i.e. active) future for the point $x=0, t=0$ : an observer at this point can actively influence the events in F with the aid of signals transported by material carriers. The region C is the passive future. This future sooner or later crosses the 'present' N , but the observer at $x=0, t=0$ cannot influence this future by material signals. The reverse is also true: no object in the future C can send a signal by a material carrier to the point $x=0, t=0$. In other words, within the framework of the dynamical interactions the points in the region $C$ cannot be in the cause-and-effect relationship with the point $x=0, t=0$.

We shall now assume that the quantum collapse events occur in such a way that they can transport superluminal signals because of the quantum correlations of the 'entangled' states, i.e. we shall admit the feasibility of the 'quantum tele-graph'. Then the collapse events involving particles travelling at subrelativistic velocities correspond to the 'instantaneous' $(V=\infty)$ transmission of a signal, which is identified by the double-headed arrows on the $x$ axis. The dashed lines in Fig. 31 correspond to moving bodies for which the signal velocity is $V^{\prime}=\infty$ in their own coordinate systems. If the wave functions of the particles of these bodies have correla-tion links with the external environment, then only that direc-tion of propagation of the control (i.e. carrying information) signals is possible which corresponds to an increase in the 'absolute time' $t$. Directions of this kind are identified by arrows of the dashed lines in Fig. 31. It is implied that we can imagine a moving body isolated absolutely from the external world, and this includes isolation in respect of information. If nonequilibrium irreversible processes occur in this body, then the collapse of wave functions also occur in it. In this case it is sufficient just to lock the coordinate system of Fig. 31 to a moving body and the rest of the reasoning can be applied in the new coordinate system. In other words, the full Lorentz invariance includes information isolation of the system.

## 41. Hierarchical systems

The term 'hierarchical system' reminds us a little of bureaucracy. In fact, the principle of hierarchical organisation is used to construct any management structure, which may be management of a small enterprise, a ministry, or the State as a whole. The principle of such a structure is illustrated in Fig. 32.

At the head of such a management structure there is one person $M$ who makes the final decisions on the most important problems in an enterprise D. Between M and D there is an intermediate management level (or several such levels). The level C reports to $M$ about the decisions made, receives orders from $M$, and controls sections of the enterprise D so as to ensure its optimal functioning.

It would seem that Fig. 32 is very far from the physical systems under discussion here. However, this is not true: in complex open physical systems there may be tendencies to splitting into information and dynamical subsystems, somewhat similar to the diagram in Fig. 32. Let us consider what is involved here. We recall that an open physical system is


Figure 32. Hierarchy of a management structure: $M$ is the head of the organisation, C are divisional heads, and D are sections of the organisation.
constructed in accordance with the principle illustrated in Fig. 33. An open system $X$ receives from outside an ordered energy of power $P$ and a negentropy flux $-\dot{S}_{i}$. If, for example, the temperature of the system X is $T$ and the arriving energy has an entropy which can be described by an effective temperature $T_{\text {eff }}$, then the negentropy flux is $-\dot{S}_{i}=$ $-P\left(T_{\mathrm{eff}}^{-1}-T^{-1}\right)$.

Irreversible processes inside the system X lead to entropy creation and a small part of the negentropy arriving in the system is dissipated in the maintenance and improvement of the internal structure of the system X. The entropy flux $\dot{S}_{\mathrm{e}}$, generated in the system $X$, is expelled, together with an excess mass, outside in the form of waste.

If the system X has an internal structure which is not too complex, it may behave as a homogeneous dynamical (mechanical) system. For example a mountain river receives mass (water) and an ordered potential energy from the melting of glaciers. In terms of hydrodynamics, the river has no internal structure with a memory, unless we are interested in the evolution of the river gorges or valleys.

However, in more complex systems with a complicated internally organised structure a single system may split into two closely coupled subsystems. One of them we shall still call dynamical or force subsystem and the second can be called the information or control subsystem. This is possible because of the great complexity of the 'phase portrait' of the system. If such a system is described by certain order parameters, which are generalised coordinates $Q_{i}$, then the evolution of $Q_{i}$ in time may be very complex because of


Figure 33. Open system $X$ receives energy from outside at a rate equal to the power $P$ and can receive matter at a rate $\dot{M}$. The energy should arrive in an organised form, so that together with the energy in the system a negative entropy $\left(\dot{S}_{i}\right)$ is introduced. Waste is ejected from the system $X$ to the external environment and this waste results in 'thermal pollution' of the medium in the form of an increase in the external entropy at a rate $\dot{S}_{\mathrm{e}}$; the waste also includes the mass $M$ flowing through the system.
nonlinear links between $Q_{i}$. Consequently, the path $Q_{i}$ in the phase space may be very sensitive to small perturbations and have many bifurcation points. Under these conditions the phase point may be easily transferred from one path to the other by small external perturbations or small changes in the structure elements of the system.

Those structure elements which can strongly influence the dynamics of a system by relatively small perturbations (signals) naturally form a control structure. In this way complex dynamical systems can themselves split into two hierarchical levels as shown in Fig. 34.


Figure 34. Hierarchical systems $X$ and $Y$ consist of dynamic systems $D_{X}$ and $D_{Y}$ and control subsystems $C_{X}$ and $C_{Y}$. Each of the 'control units', for example $\mathrm{C}_{\mathrm{X}}$, receives information on its dynamic subsystem and on other systems. It uses this information to form control signals $\mathrm{C}_{\mathrm{X}}$ and $\mathrm{D}_{\mathrm{X}}$.

Let us consider, for example, a system X. It has two hierarchical levels: a control subsystem $\mathrm{C}_{\mathrm{X}}$ and a dynamical subsystem $D_{X}$. Here $C_{X}$ is also a dynamical system, but it is more 'sensitive' than $\mathrm{D}_{\mathrm{X}}$. The block $\mathrm{C}_{\mathrm{X}}$ can 'work' with much weaker energy exchange processes, i.e. it can in fact work with 'signals'. If the subsystem $\mathrm{C}_{\mathrm{X}}$ is organised in a fairly complex manner, it may not respond to the intensities of the transmit-ted signals, but to their profiles, i.e. to their 'meaningful parts'. In other words, $\mathrm{C}_{\mathrm{X}}$ becomes an information system and its complex internal organisation makes it possible to construct a thesaurus, which is a set of internal archives which enables the system to process the incoming information and generate control signals addressed to the dynamical sub-system $\mathrm{D}_{\mathrm{x}}$. We shall refer to the hierarchical systems of the type shown in Fig. 34 as systems with information behaviour.

We shall now consider two systems with information behaviour, X and Y , as shown in Fig. 34. The highly sensitive block $\mathrm{C}_{\mathrm{X}}$ can receive signals not only from its dynamical subsystem $D_{X}$, but also from the subsystems $C_{X}$ and $D_{X}$ of the second system Y. The signals should carry information, i.e. the relevant negentropy should also be supplied from somewhere outside. The blocks $\mathrm{C}_{\mathrm{X}}$ and $\mathrm{C}_{\mathrm{Y}}$ can then utilise either part of that negentropy which reaches the blocks $D_{X}$ and $D_{Y}$ or other negentropy fluxes which exist in the nonequilibrium external world. If we consider living organisms, then the negentropy source for the D subsystems is food and information may be obtained with the aid of light received by the vision organs.

For any system X the second system Y may be the whole external world. In this case we should speak of 'immersion' of a given system in the external Universe and of its adaptation to the energy and information fluxes in the nonequilibrium Universe. It is in the case of the systems with information behaviour that this mutual relationship
between the dynam-ical and information processes, discussed in the preceding sections, is important.

It should be pointed out that the structural complexity and structural hierarchy, rather than the hierarchy of elementary levels (particles, atoms, molecules, bodies) are more important for information behaviour of complex physical systems. Elements of information behaviour are also exhibited by microparticles in the form of the collapse of the wave functions and as the structures become larger and more complex, they acquire nonequilibrium collective order parameters playing the role of dynamical variables. The collapse of wave functions and the bifurcations of their dynamical variables near branching points look like free deeds, i.e. as manifestations of the freedom of will. This is why nature as a whole can develop freely and this is realised by structural complications and development of its component elements which are complex physical systems.

## 42. Summary

Let us now summarise briefly all that has been said above.
At a school level knowledge of mechanics an unwitting impression is gained that this branch of physics is an exact science dealing with a set of clearly formulated and exactly solvable elegant dynamical problems. At least this is how mechanics is usually taught. This impression of mechanics leads quite naturally to a mechanistic approach to other phenomena in physics and Laplace's determinism then seems to be fully justified: it would seem that it is sufficient to find the forces acting between individual bodies and then use the initial conditions to predict the evolution of the world. We know that nature is built and develops in a much more complex manner. However, it may not yet be very widely known that the situation in mechanics itself has changed considerably in the last few years.

It has been found that all the exactly solvable, or integrable, problems belong to a class of specially selected, strongly simplified tasks. The majority of the problems dealing with mechanical systems are not integrable. This is not simply due to the inability of finding the final form of the solution, but is the result of a complex behaviour of dynamical systems which resembles chaos (random behaviour). This behaviour, known as the dynamical chaos, has been demonstrated and analysed for a large number of specific examples and it seems a fairly universal phenomenon. In this case, neighbouring paths of motion diverge in the phase space, i.e. the motion is locally unstable. Therefore, a description of such motion can be provided only by rigorous calculations of the paths on a computer, but also by statistical methods if our interest is in the behaviour of a system evolving over a fairly long period time.

However, this is not all. In discussing the behaviour of systems with stochasticity and the characteristics of the interaction of such systems we can use ideas borrowed from the theories of control and cybernetics that have been developed to describe control systems deliberately created by man, but which also have their own analogues in natural phenomena, particularly in biological processes.

Chaotic behaviour is typical of a large number of dynamical systems, both energy-conserving and dissipative. In Hamiltonian systems in which the phase volume is conserved the motion represents mixing in the phase space: an initial 'drop' of the phase space is of a size which is set by the uncertainty of the initial data and becomes deformed in a
complex manner during its motion. The 'drop' sprouts 'branches' which then become longer, deformed, and gradually penetrate the whole phase space while retaining their volume, so that the result resembles a ball of cotton wool. The close paths diverge rapidly from one another during such motion and the average range of their divergence is characterised by the Kolmogorov-Sinai entropy. In the course of this mixing process the paths may at any given point in space be approached arbitrarily close. These systems are known as ergodic: the temporal and spatial average values of a certain function of the coordinate of the phase space are identical.

In systems with dissipation the phase volume contracts during motion. In the simplest case such a system evolves to a state of equilibrium: the corresponding path in the phase space has the form of a stable focus. When energy is supplied from outside, a dissipative system may experience stable oscillations-they represent a stable cycle in phase space (and in the multidimensional phase they represent a torus) or it may exhibit complex stochastic motion known as a strange attractor. All the paths of a dissipative system in the phase space thus correspond to attractors: they represent an equilibrium, periodic oscillations, or a strange attractor. One of the attractors may be the destruction of the system.

If a dissipative system has many degrees of freedom, it can have many attraction zones in the phase space. If they are composed of stable foci, the system tends to one of the stable equilibrium points. In this case it is usual to speak of a multiequilibrium system and the simplest example is a computer memory. The limiting state may also be one of the limit cycles: such memories also exist (for example, a cyclic train of magnetic bubble domains running one after another). In a more general case a system may tend to one of many possible attractors, including strange attractors. When external action takes a system out of a given attractor state, the system will evolve to another attractor where the attraction zone covers the point of the initial state of the system. An ordering process known as self-organisation may develop in complex many-attractor physical systems.

In the simplest case the process of self-organisation represents the appearance of order in an initially homogeneous systems, i.e. it represents spontaneous symmetry breaking in an unstable homogeneous state. There are many examples of self-organisation of this type in physics, chemistry, and other natural sciences.

Second-order phase transitions can be described, following Landau, by introduction of an order parameter. A similar approach is used to describe the appearance of Bénard cells in a layer of a liquid heated from below. These examples demonstrate the appearance of order and of new symmetry is associated with an increase of a certain physical order parameter well above the initial thermal level. For an instability in a liquid the relevant physical parameters are the velocity, density, and temperature. Chemical self-oscillatory waves (self-waves) can be described in terms of the concentrations of the substances participating in the reaction.

Macroscopic quantities such as the velocity, density, temperature, and concentration of chemical substances are continuous functions of a point, i.e. they are physical fields. Therefore, such fields formally have an infinite number of the degrees of freedom. However, when order appears or structures develop, only a finite number of the degrees of freedom is excited. This is manifested particularly by the Bénard cells or the Taylor vortices. Therefore, systems with
ordering can frequently be regarded as having a finite number of the degrees of freedom and they can be modelled (at least numerically) by simple dynamical systems. It is worth recalling that strange attractors have been found specifically when describing convection in a liquid.

After transition to turbulence the number of strongly excited degrees of freedom can be very large. However, even then the number of effectively active free quantities need not be very large. The fractal dimensionality of the effective phase space is as a rule not high: small-scale degrees of freedom simply match the large-scale ones and play the role of stronger dissipation.

Summarising, we can say that self-organisation in homogeneous media and the formation of structures bring some collective degrees of freedom to a level which is far from thermal. These degrees of freedom can be described in terms of macroscopic parameters which play the role of generalised coordinates in the relevant dynamical models. Therefore, descriptions of self-organised systems resemble those of dynamical systems.

A more complex type of self-organisation appears in that case when the hierarchy may develop further in the newly formed structures because of the appearance and interaction of new structure elements. An example is provided by prebiological evolution of macromolecules and biological evolution. Adopting a simplified approach we can say that such self-organisation involves new degrees of freedom which, in terms of dynamical modelling, can be regarded as memory components in multistable states. For example, desert sand dunes can be treated as a slowly evolving dynamical system which stores the history of the interaction of the surface of the sand with the wind. A quantitative description of the corresponding development processes may be modelled by dynamical constructs with dissipation and memory and with a sufficiently large number of the degrees of freedom.

The very existence of stochastic systems with a small number of degrees of freedom significantly alters our ideas on the formulation of dynamical problems. In fact, it had hitherto been assumed that stochasticity is a property only of systems with a very large number of the degrees of freedom when it is natural to employ a statistical description, whereas systems with few degrees of freedom should be described exactly, i.e. 'exactly' in the framework of the adopted model, which may not correspond fully to reality. However, this can be replaced by an analysis of the response of a system to small perturbations: in a stable system the perturbations remain small.

If a system is unstable, then all the paths in the phase space diverge and, therefore, the initial data become very important. If we wish the path of real motion to be close to a selected one, i.e. that the end of the path should be in a certain region of the phase space, the initial data should be selected sufficiently precisely. In other words, it is necessary to specify the value of each of the coordinates $q$ to a sufficiently large number of decimal places. The number of decimal places is proportional to $+\ln (1 / \Delta q)$, where $\Delta q$ is the error with which a coordinate is specified. When all the coordinates and momenta are specified with a certain error, the result is a quantity proportional to $\ln (V / \Delta V)$, where $V$ is the total volume of the phase space and $\Delta V$ is the proportion of this volume in the initial state. The quantity $\ln (V / \Delta V)$ is known as the amount of information. Therefore, in the case of unstable systems we have to deal not only with the
dynamical part of the problem, associated with the forces acting between the bodies, but also with the information part which plays a major role and is associated with the specification of the initial values of the coordinates.

However, this is not all. Let us consider, for example, a Hamiltonian system and postulate that we are interested in the problem of a phase path reaching a certain finite 'drop' in the phase space. Let us project this 'drop' along the phase path from the future to the initial time. Then this 'drop' spreads out to resemble a 'ball of cotton wool' covering a large part of the phase volume in the manner of a highly porous body. A random point in this volume is unlikely to be located on a 'fibre' of the 'cotton-wool ball', but a small displacement in the phase space may result in placing this point on a path inside a 'fibre' ending in a 'drop'. Therefore, a very small displacement in the phase space can bring the point to the required path: all that is necessary to know is in which direction the displacement should be made. This knowledge is also information. The simplest method of obtaining this knowledge involves following the paths. This is exactly how the control systems are constructed: they follow the real path and apply small signals to alter it to the required one.

It follows that small signals can greatly alter the paths in unstable systems. In the case of these signals it is not their amplitude which is important, but the exact matching to the feasibility of transferring the initial path to the required one, i.e. the correct information on the structure of the phase portrait is necessary.

Let us now consider dissipative systems. As pointed out earlier, the phase space of open dissipative systems can be represented in a simplified manner as divided into regions of attraction to different attractors. The system can be transferred from one attractor to another if it is taken from one region of attraction to another. Again it is not the magnitude of the force acting that is of prime importance, but its information characteristics: a force is necessary to transfer the system to any point of attraction of the second attractor, i.e. the system has to be provided with a certain amount of information $\ln (V / \Delta V)$, where $\Delta V$ is the attraction volume of the second attractor. Obviously, in reality the transfer of a system from one attractor to another requires a certain amount of energy and there may be a minimum value of this energy below which the transfer is impossible and the appro-priate signal does not implement the available information.

Nevertheless, not only the dynamical but also the informational aspect is important in a transfer of this kind. It should therefore be described in terms of information processes. The important feature is not only the amount of the information which is provided, but also its content (sometimes called the semantic content), i.e. an indication to which specific attractor the system is being transferred.

Let us consider a system which is dynamically unstable, i.e. a system with diverging paths. The discussion can be made easier by considering a very much simplified discrete model in which the path passes consecutively through bifurcation points.

At each such bifurcation (branching) point a system forms a 'deed' or 'act': it selects one of the subsequent parts of the path. At every point the selection is random and there is no correlation between the selections, which means that we are dealing with a Markov chain. However, this aspect is of no interest to us.

We shall assume that our system in fact complex and, in addition its external dynamic variables, it has also internal degrees of freedom such as those corresponding to multistable states or internal attractors. Then the path selection at the bifurcation point is governed by the state of the internal degrees of freedom. If the system has an energy source, then in the selection process it is again important to have just an information hint which path is preferable. A system of this kind may be regarded as one with control, which is governed by the state of the internal part of the dynamical system.

Against the background of force dynamics the behaviour of this system begins to be governed by its information properties so that it can be called a system with information behaviour. If there is moreover a feedback between the dynamical and internal degrees of freedom, then a system of this kind can react in a complex manner to all external stimuli.

The behaviour of complex systems is undoubtedly a suitable topic for specific investigations. Nevertheless, some general ideas can be put forward.

In a comparison of simple and complex systems, for example simple mechanisms and biological systems with a complex hierarchical organisation, the principal difference between them strikes the eye: the motion of simple mechanical systems is fully predetermined, whereas in the case of living (even the simplest) organisms we can detect signs of free behaviour, which is not determined simply by external factors. Naturally, this freedom of action is manifested even more strongly by complex organisms.

The question arises whether it is possible to follow this trend starting with the ideas on self-organisation.

As mentioned earlier, self-organising systems can split into dynamical and informational (control) parts. If we are dealing with biological or prebiological systems, capable of replication and multiple repetition of a growth cycle, we can easily see that systems with information behaviour can, and in fact have, advantages in the fight to survive, i.e. in the attempt to extend the phase space of their stable existence. It follows that under competitive conditions the dynamical behaviour of a system and its development begin to be determined to a greater extent by its information properties, including the information coupling to the external world.

In addition to the usual exchange of energy and negentropy with the external environment, which is necessary to retain an open self-organised structure, systems of this kind are capable of information development, i.e. of increasing the complexity and quality of their control. This process is related to complications of the controlling part, i.e. by appearance of new degrees of freedom or order parameters because of increasing number of new bifurcations. The process may be characterised phenomenologically as requiring additional information from the nonequilibrium external world.

When systems with memory and information behaviour interact, processes of adaptation of some system to others may take place. This is an analogue of the appearance of 'knowledge' of the reaction of other systems to the behaviour of a given one. Further development of the mechanisms of adaptation and optimisation of the behaviour of a system leads to the possibility of data processing, i.e. to the appearance of elementary consciousness. However, in addition to freedom, these complex systems exhibit the beginnings of free will.

Freedom is the ability to select one of several possibilities and the act of volition is the selection itself. In unstable systems with diverging paths it is sufficient to select one of the paths by altering very slightly the initial data, i.e. to use for this purpose a very small signal with specific information. In a complex system with information behaviour this selection can be made by a signal from a control unit. This signal is generated by a control system in the process of its dynamic motion and using the accumulated memory, i.e. knowledge. Such knowledge is the product of a long development of a given system in the course of its interaction with other information systems and this knowledge is to some extent of permanent nature, i.e. it is not related solely to the events in the recent past. Therefore, an act of volition seems to be free: it is not predetermined by causes which are obvious at present. Moreover, the information integrated over a long time and accumulated in memory includes data on the evolution of the system in its environment with time, so that it is possible to predict events and make decisions with a definite aim. In other words, in contrast to the usual dynamics where the behaviour of a system is governed by its instantaneous configuration and by the forces corresponding to this configuration, in information processes the selection is made on the basis of long-term memory and with possible extrapolation to the future. Hence it is possible to select the goal. In this way the selection of the goal and its feasibility can be regarded as quite natural products of the development of complex systems with information behaviour.

It would seem that the more complex the system, the greater the degree to which it should manifest aspects of information behaviour. In nature as a whole the 'volitional spirit' is related to the structure and memory that have formed as a result of a long development. If in the course of investigation it is possible to isolate a system from its external environment, the situation becomes much simpler and more readily understood, but the situation is impoverished by information isolation from the external world.

In tackling increasingly complex systems it is the structural and information aspects of their behaviour and evolution that become of prime importance and the dynamics is simply the base for the development of the information aspect.

If we allow for quantum processes in the microworld the picture of development of the world becomes even more complex and richer in the sense of information behaviour. New nonequilibrium and strongly nonlinear process in the form of the collapse of wave functions now come into play. The structure information links and their influence on dynamical processes become correspondingly more complex.

The concept of the collapse of wave functions appears naturally in measurements of quantum values: it follows from the fundamental principle of quantum mechanics that measurement of this kind can provide with a certain probability only one of the discrete eigenvalues of the corresponding operator. We can see that when an 'information contact' is established between a quantum object and its classical nonequilibrium environment (for example, an instrument), the wave function of the object retains only one projection and all the other components are destroyed. They are annihilated. This process is also known as decoherence and in recent years it has been discussed extensively, together with the 'historicity', in the literature on fundamentals of quantum theory (see, for example, the review by Omnes [41]).

We have adopted here a much simpler point of view according to which both the decoherence and the collapse of wave functions of our environment are the results of its very strong nonequilibium associated in particular with the negentropy flux from the Sun. Quantum transitions in the course of thermalisation of the solar radiation are, on the one hand, responsible for the monotonic increase in the entropy and, on the other, destroy the coherence of the wave functions of macroscopic bodies and thus impose on them the con-ditions of classical behaviour. In other words, the evolution of nonequilibrium nature by itself gives rise to classical behaviour of macroscopic bodies. The contact between the classical macroworld and the quantum microworld induces the collapse of wave functions. In quantum measurements such collapse is accompanied by the collapse of the prob-abilities, which creates information on a microobject in the external world.

The problem of measurement in quantum mechanics has always attracted and continues to attract the interest of many physicists, both theoreticians and experimentallists. There are two main points of view in this respect. One begins from Niels Bohr and states that the results of measurements are purely random and that quantum theory can only prescribe the probabilities proportional to the squares of the amplitudes of the wave function. The structure of a specific instrument and the macroscopic configuration determine which specific amplitudes are being measured. The other point of view was put forward by Albert Einstein. It postulates that behind the complex results of measurements lies a more complex physical picture of the universe. This second point of view has been developed in several variants of theories of 'hidden variables'. However, in recent years, particularly after an experimental confirmation of Bell's inequalities, we have seen the acceptance of Bohr's view. In spite of this, the problem is still under discussion and will continue to be discussed in the scientific literature.

The celebrated paradox of Einstein, Podolsky, and Rosen [2] is very frequently used in discussions of this kind. This paradox is based on a thought experiment in which two quantum particles become separated by a very large distance after they interact. The two particles are correlated and, therefore, the results of measurements on one particle are correlated with the results of measurements on the other particle. At first sight this looks like an instantaneous action-at-distance and it is not surprising that different authors have at times discussed the possibility of signal transmission faster than the velocity of light by means of correlated EPR pairs.

Shimony [5] and Ghirardi, Rimini, and Weber [42] have shown that superluminal signals are impossible within the framework of the standard quantum theory when all the results of measurement follow the $|\psi|^{2}$ law. It is found that the measurements carried out on one partner of an EPR pair cannot in any way control the results of measurements of the second partner. This would seem to solve the problem. However, as specially stressed by Shimony [5], the proof relies implicitly on the assumption that macroscopic bodies cannot have microcorrelations.

Sections 36-40 demonstrated that the problem is not so trivial. Section 36 described the effect discovered experimentally by Sokolov et al. [36,37] when the flight of an excited $2 P$ hydrogen atom through a slit in a metal screen creates continuously a correction to the $2 P$ amplitude. The simplest theoretical explanation of this effect [38] is based on the idea that the hydrogen atom is the first partner of a set of

EPR pairs and the second partner is a conduction electron. After interaction in the surface layer with metal the electrons travel into the interior of the metal and their wave functions collapse there by the decoherence mechanism. There is no external measurement in this collapse and therefore the $|\psi|^{2}$ law is not obeyed. However, it follows from energy conservation that each such collapse is very likely to release a slower part of a wave packet. There is in fact no measurement in the process of evolution of the complex wave function of many electrons. The collapse events have the effect that the slow part of a wave packet of each electron which has interacted with an atom is somewhat more likely to appear than the corresponding value of $|\psi|^{2}$. The individual effect is extremely small, but it has the same sign for the conduction electrons, so that the corresponding admixtures to the $2 P$ amplitude accumulate. We can say that an excited atom creates microcorrelations inside a macro-scopic body which gradually result in the accumulation of the $2 P$ amplitude of an atom from the $2 S$ amplitude.

Since the collapse of single wave functions of electrons does not necessarily obey the law of the probabilities $|\psi|^{2}$, we can discuss the possibility of control of the collapse of the secondary partners of EPR pairs. A suitable scheme was considered in Section 37. It is at present difficult to say how realistic such a scheme is: this is primarily a matter for experimental verification. However, we can formulate a purely theoretical question: do superluminal signals contradict the standard theory of relativity? This was discussed in Sections 38-40.

It is shown in these sections that superluminal signals associated with the collapse and not with material carriers do not contradict the theory of relativity. Moreover, we can imagine a scheme shown in Fig. 31 when the cause-and-effect closure loops are completely eliminated. In this case the causality is not violated even in the superluminal regions C . At a point $x=0, t=0$ such a region corresponds to the future and, therefore, the standard requirement of causality in a region C (effect after cause) may be too stringent.

All these topics require further investigation.
The ideas on self-organisation and formation of dissipative structures in open systems have proved very important in building bridges between physics and biology. It is, however, clear that one should go further and study hierarchisation of structures, formation of structures from memory, feasibility of acquiring information from outside, storing it in a memory, and using information to control and process the stored information with the aim of achieving optimal control. We can readily see that there are many features in common in the behaviour both of organic and of inorganic complex systems, and that inorganic systems with a complex structure are not simple in respect of the structure and their behaviour.

## References

1. Schwinger J Quantum Kinematics and Dynamics (New York: Benjamin, 1970)
2. Einstein A, Podolsky B, Rosen N Phys. Rev. 47777 (1935)
3. Bell J S Physics (Long Island City, N. Y. ) 1195 (1964)
4. Clauser J F, Horne M A, Shimony A, Holt R A Phys. Rev. Lett. 23880 (1969)
5. Shimony A, in Proceedings of International Symposium on Foundations of Quantum Mechanics in the Light of New Te chnology, To kyo, 1983 (Tokyo: Physical Society of Japan, 1985) p. 225
6. Ekert A K Phys. Rev. Lett. 67661 (1991)
7. Bennett C H, Brassard G, Mermin N D Phys. Rev. Lett. 68557 (1992)
8. Bennett C H, Wiesner S J Phys. Rev. Lett. 692881 (1992)
9. Bennett C H, Brassard G, Crepeau C, Jozsa R, Peres A, Wootters W K Phys. Rev. Lett. 701895 (1993)
10. Braunstein S L, Mann A, Revzen M Phys. Rev. Lett. 683259 (1992)
11. Vaidman L Phys. Rev. Lett. 703369 (1993)
12. Prigogine I From Being to Becoming: Time and Complexity in Physical Sciences (San Francisco: W H Freeman, 1980)
13. Lorenz E N Tellus 161 (1964)
14. Nicolis G, Prigogine I Self-Organization in Non-Equilibrium Sy stems: from Dissipative Str uctures to Order through Fluctuations (New York: Wiley, 1977)
15. Ebeling W Str ukt urbildung bei irreversiblen Prozessen (Stuttgart: Teubner, 1976)
16. Haken H Introduction to Synergetics: Non-Equilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology (Berlin: Springer, 1977)
17. Polak L S, Mikhailov A S Sam oorganizatsiya v Ne ravnovesnykh Fiziko-Khimicheskikh Sistemakh (Self-Organisation in Nonequilibrium Physicochemical Systems) (Moscow: Nauka, 1983)
18. Haken H Advanced Syn ergetics: Instability Hierarchies of Self-Organizing Sys tems and Devices (Berlin: Springer, 1983)
19. Nicolis J C Dynamics of Hierarchical Sys tems (Berlin: Springer, 1986)
20. Hartley R Information Transfer (Translated into Russian; Moscow: Inostr. Lit., 1957)
21. Shannon C E Works on Theory of Information and Cybernetics (Translated into Russian; Moscow: Inostr. Lit., 1963)
22. Szilard L Z. Phys. 53840 (1929)
23. Brillouin L Science and Information Theory 2nd ed. (New York: Academic Press, 1962)
24. Stratonovich R L Teoriya Informatsii (Theory of Information) (Moscow: Sovet-skoe Radio, 1975)
25. Poplavskii R P Termodinamika Informatsionnykh Protsessov (Thermodynamics of Information Processes) (Moscow: Nauka, 1981)
26. Schrodinger E Nat urwissenschaften 23 807, 844 (1935)
27. Everett H III Rev. Mod. Phys. 29454 (1957)
28. Yurke B, Stoler D Phys. Rev. Lett. 5713 (1986)
29. Davidovich L, Maalë A, Brune M, Raimond J M, Haroche S Phys. Rev. Lett. 712360 (1993)
30. Caldeira A O, Leggett A J Phys. Rev. A 311059 (1985)
31. Brune M, Haroche S, Raimond J M, Davidovich L, Zagury N Phys. Rev. A 455193 (1992)
32. Bohm D Phys. Rev. 85 166, 180 (1952)
33. von Neumann J Mat hematische Grundlagen der Quantenmechanik (Berlin: Springer, 1932)
34. Mermin D Rev. Mo d. Phys. 65803 (1993)
35. Zaslavskii G M, Sagdeev R Z Vvedenie v Nelineinuyu Fiziku (Introduction to Nonlinear Physics) (Moscow: Nauka, 1988)
36. Sokolov Yu L, Yakovlev V P, Pal'chikov V G Phys. Scr. 48000 (1993)
37. Sokolov Yu L, Yakovlev V P, Pal'chikov V G, Lin D N Nuovo Cimento D 14183 (1992)
38. Kadomtsev B B, Kadomtsev M B 'Sokolov effect' Phys. Scr. (in press)
39. Aspect A, Dalibard J, Roger G Phys. Rev. Lett. 491804 (1982)
40. Kiess T E, Shih Y H, Sergienko A V, Alley C O Phys. Rev. Lett. 713893 (1993)
41. Omnes R Rev. Mod. Phys. 64339 (1992)
42. Ghirardi G C, Rimini A, Weber T Lett. Nuovo Cimento 27293 (1980)

[^0]:    B B Kadomtsev Russian Research Centre 'Kurchatov Institute', Institute of Nuclear Fusion, pl. Kurchatova 46, 123182 Moscow Tel. (095) 196-98-14, F ax (095) 190-42-44
    E-mail: Kadomtsev@ufn.msk.su
    Received 28 December 1993; revised 5 March 1994
    Uspekhi Fizicheskikh Nauk 164 (5) 449-529 (1994)
    Translated by A Tybulewicz

