

A. W. RICHARDS

# Modern ergodic theory

There is much more to the mathematical study of Gibbs ensembles than the question of whether or not time averages and ensemble averages are equal

Joel L. Lebowitz and Oliver Penrose

The founding fathers of statistical mechanics, Boltzmann, Maxwell, Gibbs and Einstein, invented the concept of ensembles to describe equilibrium and nonequilibrium macroscopic systems. In trying to justify the use of ensembles, and to determine whether the ensembles evolved as expected from nonequilibrium to equilibrium, they introduced further concepts such as "ergodicity" and "coarse graining." The use of these concepts raised mathematical problems that they could not solve, but like the good physicists they were they assumed that everything was or could be made all right mathematically and went on with the physics.

Their mathematical worries, however, became the seeds from which grew a whole beautiful subject called "ergodic theory." Here we describe some recent (and some not so recent) developments that partially solve some of the problems that worried the Founding Fathers; these results appear not to be widely known in the physics community, and we hope to remedy the situation. (For a more thorough review of the subject, see reference 1.)

Joel L. Lebowitz is chairman of the physics department at the Belfer Graduate School of Science, Yeshiva University, in New York. Oliver Penrose, a member of the mathematics faculty at the Open University, Walton, UK, worked on this article while on leave at the Applied Physics Department, Stanford.

At the top of the page we see Boltzmann, Maxwell, Gibbs and Einstein, the "Founding Fathers" of statistical mechanics.

Although results are not yet well enough developed to answer all the questions in this area that are of interest to physicists today, such as the derivation of kinetic equations or the general problem of irreversibility, they do make a start.

Ergodic theory is concerned with the time evolution of Gibbs ensembles. It has revealed that there is more to the subject than the simple question of whether a dynamical system is ergodic (which means, roughly, whether the system, if left to itself for long enough, will pass close to nearly all the dynamical states compatible with conservation of energy). Instead there is a hierarchy of properties that a dynamical system may have, each property implying the preceding one, and of which ergodicity is only the first (see figure 1). The next one, called "mixing," provides a clear-cut formulation of the type of irreversible behavior that people try to obtain by introducing coarse-grained ensembles. At the top of our hierarchy is a condition (the Bernoulli condition) ensuring that in a certain sense the system, though deterministic, may appear to behave as randomly as the numbers produced by a roulette wheel.

Some of the mathematical results we shall be discussing have established the positions of some model physical systems in this hierarchy. Of particular interest to physicists is the work of Ya. Sinai<sup>2</sup> on the hard-sphere system, which shows that this system is both ergodic and mixing. We shall also discuss some work by A. N. Kolmogorov,

V. I. Arnold and J. Moser on systems of coupled anharmonic oscillators, which shows that, contrary to a common assumption, these systems may not even reach the "ergodic" rung on the ladder. (G. H. Walker and J. Ford have described this work for physicists.<sup>3</sup>)

All the physical systems we shall discuss obey classical mechanics, have a finite number of degrees of freedom and are confined to a finite region of physical space. The reason for our restriction to classical mechanics is that a finite quantum system can never exhibit any of the properties higher than simple ergodicity in our hierarchy (although, of course, a large quantum system may approximate closely the behavior characterized by these concepts). This is because the spectrum of a finite quantum system is necessarily discrete, whereas for a finite classical system the spectrum (of the Liouville operator) can be continuous. The reason for our restriction to finite systems is that very little is known about the ergodic properties of infinite systems. This lack of knowledge is regrettable because only for an infinite system (by which term we mean the limit of a finite system as its size becomes infinite) can one expect to find strictly irreversible behavior in quantum mechanics. Moreover, the distinction between microscopic and macroscopic observables, which appears essential to any complete theory of irreversibility and kinetic equations, can only be formulated precisely for infinite systems. Much research remains

## History of ergodic theory

The "ergodic hypothesis" was introduced by Boltzmann in 1871. To quote Maxwell "... (it) is that the system, if left to itself in its actual state of motion, will, sooner or later, pass through every phase which is consistent with the equation of energy." In our notation "phase" means dynamical state and the original ergodic hypothesis means that if  $y$  and  $x$  are any two points on the energy surface  $S_E$ , then  $y = \phi_t(x)$  for some  $t$ . The ergodic hypothesis thus stated was proven to be false, whenever  $S_E$  has dimensionality greater than one, by A. Rosenthal and M. Plancherel in 1913. S. G. Brush gives a nice account of the early work on this problem (see reference 5).

The definition of an ergodic system given in equation 1 (page 25) can be shown to be equivalent to what is sometimes called the "quasi-ergodic" hypothesis, which replaces "every phase" in Maxwell's statement by "every region  $R$  on  $S_E$  of finite area," with the further

qualification that this is true for "almost all" dynamical states. Indeed as we point out in this article, the fraction of time that the system will spend in  $R$  is equal, for an ergodic system, to the fraction of the area of  $S_E$  that is occupied by  $R$ .

It was shown by G. D. Birkhoff in 1927 that ergodicity is equivalent to the energy surface being "metrically transitive." Stated precisely this means that a system is ergodic on  $S$  if and only if all the regions  $R$  on  $S$  left invariant by the time evolution,  $\phi_t(R) = R$ , either have zero area or have an area equal to the area of  $S$ .

The difficult part of Birkhoff's Theorem is to show that  $f^*(x)$ , which involves taking the time average over infinite times, actually exists for almost all  $x$  when  $f(x)$  is an integrable function. It is then relatively easy to show that  $f^*(x)$  is time invariant; that is,  $f^*[\phi_t(x)] = f^*(x)$ , and that ergodicity is equivalent to  $S$  being metrically intransitive.

to be done on the ergodic theory of infinite systems, both classical and quantum, but we can be sure that the concepts to be discussed here will play an important part in it.

### Surfaces and ensemble densities

Before we go on to discuss the new results, we review some mathematical definitions.<sup>4</sup> If our dynamical system has  $n$  degrees of freedom, we can think of its possible dynamical states geometrically, as points in a  $2n$ -dimensional space (phase space), with  $n$  position coordinates and  $n$  momentum coordinates. If the energy of the system is  $E$ , then its dynamical state  $x = (q_1 \dots q_n, p_1 \dots p_n)$  must lie on the energy surface  $H(x) = E$ , where  $H$  is the Hamiltonian function. We denote the energy surface, which is  $(2n - 1)$ -dimensional, by  $S_E$  or just  $S$  and assume that  $S$  is smooth and of finite extent; for example in the case of a system of harmonic oscillators, for which the Hamiltonian is a quadratic form, the energy surfaces are  $(2n - 1)$ -dimensional ellipsoids.

The time evolution of the system causes  $x$  to move in phase space, but since we are assuming our system to be conservative the point  $x$  always stays on the energy surface. If the system is in some state  $x$  at some time  $t_0$  then its state at any other time  $t_0 + t$  is uniquely determined by  $x$  and  $t$  (only). Let us call the new state  $\phi_t(x)$ . This defines a transformation  $\phi_t$  from  $S$  onto itself. There is one such function for each value of  $t$ .

We shall want to integrate dynamical functions (that is, functions of the

dynamical state) over  $S$ . When doing this it is convenient not to measure  $(2n - 1)$ -dimensional "areas" on the surface  $S$  in the usual way but to weight the areas in such a way that the natural motion of the system on  $S$  carries any region  $R$  (after any time  $t$ ) into a region  $\phi_t(R)$  of the same area. This can be accomplished by defining the weighted area of a small surface element near  $x$ ,  $dx$ , to be such that  $dx dE$  is the correct Euclidean  $2n$ -dimensional volume element of a "pill box" with base  $dx$  and height  $dE$ .

By a *Gibbs ensemble* we mean an infinitely large hypothetical collection of systems, all having the same Hamiltonian but not necessarily the same dynamical state. We shall only consider ensembles whose systems all have the same energy, so that their dynamical states are distributed in some way over some energy surface  $S$ . It may happen that this distribution can be described by an *ensemble density*; by this we mean a real-valued function  $\rho$  on  $S$  such that the fraction of members of the ensemble whose dynamical states lie in any region  $R$  on the surface  $S$  is

$$\int_R \rho(x) dx$$

with  $dx$  the weighted area defined above. The simplest ensemble density on  $S$  is given by

$$\rho(x) = C \quad (\text{all } x \text{ in } S)$$

where  $C$  is a constant, which can be determined from the normalization condition  $\int_S C dx = 1$ . This is called the *microcanonical ensemble* on  $S$ .

The systems constituting the ensemble

evolve with time, so that the ensemble density will depend on time. The rule connecting the ensemble densities  $\rho_t$  describing the same ensemble at different times  $t$  is Liouville's theorem, which can be written

$$\rho_t(x) = \rho_0[\phi_{-t}(x)] \quad (\text{all } t, \text{ all } x \text{ in } S)$$

where  $\rho_0(x)$  is the ensemble density at time zero. For example, Liouville's theorem shows that the density of the microcanonical ensemble does not change with time: If

$$\rho_0(x) = C$$

for all  $x$  in  $S$ , then Liouville's theorem gives, for any  $t$ ,

$$\rho_t(x) = C$$

for all  $x$  in  $S$ .

### The ergodic problem

The principal success of ensemble theory has been in its application to equilibrium. To calculate the equilibrium value of any dynamical function we average it over a suitable ensemble. The same ensemble also enables us to estimate the magnitude of the fluctuations of our dynamical function. To ensure that the calculated averages are independent of time, we use an *invariant ensemble*; that is, one for which the fraction of members of the ensemble in every region  $R$  on the energy surface  $S$  is independent of time. We already know one invariant ensemble: the microcanonical, whose ensemble density is uniform on  $S$ . Before we can use it confidently to calculate equilibrium values, however, we would like to be sure that this is the only invariant ensemble: If other invariant ensembles exist then, in principle, they could do just as well for the calculation of equilibrium properties, and we would have to choose which to use in a particular situation.

There are two questions to settle: the first is whether there are any invariant ensembles on  $S$  that do not have an ensemble density. In general there are; for example in the case of a hard-sphere system in a box one could have an invariant ensemble where every particle moves on the same straight line being reflected at each end from a perfectly smooth parallel wall (see figure 2).

The obviously exceptional character of this motion is reflected mathematically in the fact that this ensemble, though invariant, is confined to a region of zero "area" on  $S$  and therefore has no ensemble density. To set up such a motion would presumably be physically impossible because the slightest inaccuracy would rapidly destroy the perfect alignment. It is therefore natural to rule out such exceptional ensembles by adopting the principle that any ensemble corre-

sponding to a physically realizable situation must have an ensemble density.

There remains the second part of the question: Are there any invariant ensembles on  $S$  that do have a density but differ from the microcanonical ensemble? This is equivalent to the *ergodic problem*<sup>5</sup> in which one compares the time averages of a dynamical function  $f$ ,

$$f^*(x) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(\phi_t(x)) dt$$

with its microcanonical ensemble average

$$\langle f \rangle = \int_S f(x) dx / \int_S dx$$

A system is said to be *ergodic* on its energy surface  $S$  if time averages are in general equal to ensemble averages; that is, if for every integrable function  $f$  we have

$$f^*(x) = \langle f \rangle \quad (1)$$

for almost all points  $x$  on  $S$ . "Almost all" means that if  $M$  is the set of points  $x$  for which equation 1 is false, we have  $\int_M dx = 0$ . The answer to our second question is given by a theorem,<sup>6</sup> which we shall not prove: the microcanonical ensemble density is the only invariant ensemble—that is, the only one satisfying  $\rho[\phi_t(x)] = \rho(x)$  for all  $x$  in  $S$ —if and only if the system is ergodic on  $S$ .

The physical importance of ergodicity is that it can be used to justify the use of the microcanonical ensemble for calculating equilibrium values and fluctuations. Suppose  $f$  is some macroscopic observable and the system is started at time zero from a dynamical state  $x$ , for which  $f(x)$  has a value that may be very far from its equilibrium value. As time proceeds, we expect the current value of  $f$ , which is  $f[\phi_t(x)]$ , to approach and mostly stay very close to an equilibrium value with only very rare large fluctuations away from this value. This equilibrium value should therefore be equal to the time average  $f^*$  because the initial period during which equilibrium is established contributes only negligibly to the formula defining  $f^*(x)$ . The theorem tells us that this equilibrium value is almost always equal to  $\langle f \rangle$ , the average of  $f$  in the microcanonical ensemble, provided the system is ergodic.

To justify the use of the microcanonical ensemble in calculating equilibrium fluctuations we proceed in a similar way. For some observable event  $A$  (such as the event that the percentage of gas molecules in one half of a container exceeds 51%) let  $R$  be the region in phase space consisting of all phase points compatible with the event  $A$ ; that is the event  $A$  is observed if and only if the phase point is in  $R$  at time  $t$ . If the system is observed over a long

period of time, the fraction of time during which event  $A$  will be observed is given by the time average  $g^*(x_0)$ , where  $x_0$  is the initial dynamical state and  $g$  is defined by

$$g(x) = \begin{cases} 1 & \text{if } x \text{ is in } R, \\ 0 & \text{if not} \end{cases}$$

The ergodic theorem tells us that for almost all initial dynamical states this fraction of time is equal to the ensemble average of  $g$ , which is

$$\langle g \rangle = \int_R dx / \int_S dx$$

This is just the "probability" of the event  $A$  as calculated in the microcanonical ensemble on  $S$ .

Another way of defining ergodicity is to say that any integrable invariant function is constant almost everywhere. That is to say, if  $f$  is an integrable function satisfying the condition that

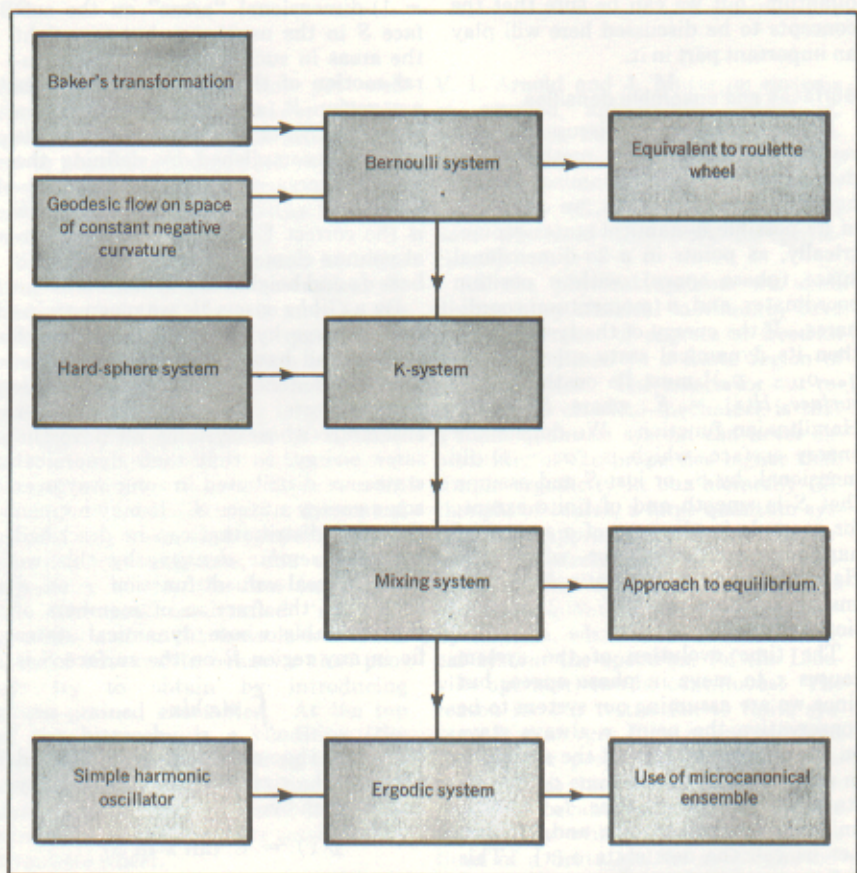
$$f[\phi_t(x)] = f(x)$$

for all  $x$  in  $S$ , then there is a constant  $c$  such that  $f(x)$  equals  $c$  for almost all  $x$ : In other words, the set  $M$  of points  $x$

for which  $f(x)$  does not equal  $c$  satisfies  $\int_M dx = 0$ . This has the physical interpretation that for a Hamiltonian system ergodic on  $S$  every integrable constant of the motion is constant on  $S$ . Furthermore if ergodicity holds on each  $S_E$  then there are no integrable constants of the motion other than functions of the energy  $E$ . Indeed, if there were other constants of the motion (for example angular momentum if the Hamiltonian had an axis of symmetry) we could construct invariant densities that were not microcanonical by taking  $\rho(x)$  to be a function of one of these other constants of the motion, and so the system would clearly be nonergodic. When such additional constants of the motion exist they must be taken into account in the statistical mechanics and thermodynamics of the system; the standard methods, based on the microcanonical ensemble, must then be generalized for these systems.

### Some special systems

We now illustrate the idea of ergodicity by considering some specific mechanical systems. The simplest of these is the harmonic oscillator, whose



**Hierarchy of systems.** Arrows denote implication: Every mixing system is ergodic, every Bernoulli system equivalent to a roulette wheel, and so on. The middle column lists the three systems discussed in this article—with the "strongest" at the top—as well as the  $K$ -system (after A. N. Kolmogorov), which is not discussed. At the left are examples of the systems and at the right physical interpretations or implications.

Figure 1

Hamiltonian (in some suitable units) is

$$H(q, p) = \frac{1}{2} \omega (p^2 + q^2)$$

where  $\omega$  is the angular frequency. The transformation  $\phi_t$  for this system is a rotation through angle  $\omega t$  in the  $(q, p)$  plane. The trajectories, which here coincide with the energy surfaces  $S_E$ , are circles of radius  $(2E)^{1/2}$ . (The surface element  $dx$  is here proportional to the ordinary length of an arc segment.) To be invariant under the transformation  $\phi_t$  an ensemble density on  $S$  must be unaffected by rotations and is therefore a constant. It follows, then, that the only invariant density is the microcanonical density and so the simple harmonic oscillator is ergodic.

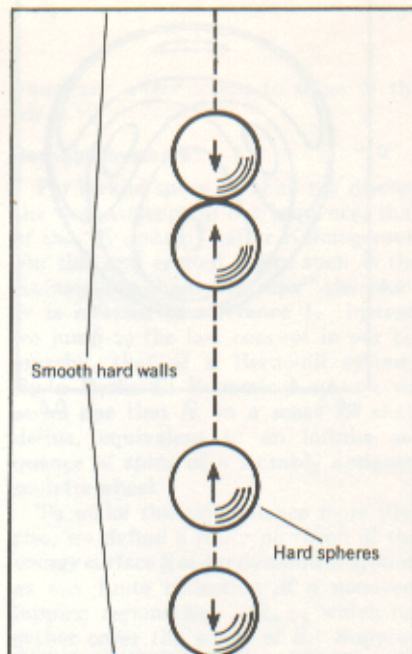
Almost as simple is the multiple harmonic oscillator (physically, say, an ideal crystal), that is, a system with two or more degrees of freedom whose potential energy is a quadratic form in the position coordinates. Unlike the simple harmonic oscillator it cannot be ergodic, because it has constants of the motion (the energies of the individual normal modes) that are not constant on the energy surfaces (the surfaces of constant total energy).

It used to be thought that this lack of ergodicity was an accident and that any small anharmonicity (such as would inevitably be present in a real system) must make the system ergodic by permitting transfer of energy from one mode to another. In 1954, however, Kolmogorov announced results that contradicted this belief.<sup>7</sup> In 1955, Enrico Fermi, J. Pasta and S. W. Ulam<sup>8</sup> carried out a computer simulation of such a system. Initially, they excited one mode only, and instead of the equipartition of the energy between all modes predicted by the microcanonical ensemble they found that most of it appeared to remain concentrated in a few modes; this indicated that anharmonic oscillator systems may not be ergodic.

The lack of ergodicity was proved rigorously by Kolmogorov, Arnold and Moser.<sup>9</sup> They found that if the frequencies of the unperturbed oscillators are not "rationally connected" (that is, if no rational linear combination of them is zero) then, in general, adding to the Hamiltonian an anharmonic perturbation sufficiently small compared to the total energy does not make the system ergodic. The unperturbed trajectories (possible paths of the phase point) all lie on  $n$ -dimensional surfaces in  $S$  (which has  $2n - 1$  dimensions) called "invariant tori," and "KAM" prove that under a weak perturbation most trajectories continue to lie within smooth  $n$ -dimensional tori, so that the perturbed system is also nonergodic. The trajectories that do not lie on the new invariant tori are, on

**Ensemble with no ensemble density.** Hard spheres move up and down the dotted line, which meets the perfectly smooth hard walls at right angles. Collisions between particles and collisions with the walls do not deflect the particles from the line if they are perfectly aligned at the start. An ensemble of such systems has no ensemble density because it is concentrated on a region on the energy surface with zero area (just as the area of a line or of a line segment in a plane is zero).

Figure 2



the other hand, very erratic indeed and may fill some  $(2n - 1)$ -dimensional region densely.<sup>9</sup> One consequence of this very complicated behavior is that even though the system is not ergodic the motion can no longer be decomposed into independent normal modes.

Similar results probably hold also for rationally connected frequencies (which cannot be treated rigorously, although they are of more physical interest); thus Michael Hénon and Carl Heiles<sup>10</sup> carried out computer calculations for the Hamiltonian

$$H = \frac{1}{2} (p_1^2 + p_2^2 + q_1^2 + q_2^2) + (q_1^2 q_2 - \frac{1}{3} q_2^3)$$

whose unperturbed frequencies  $\omega_1 = 1$ ,  $\omega_2 = 1$ , are rationally connected since  $1 \cdot \omega_1 - 1 \cdot \omega_2 = 0$ . They found that the energy surfaces with  $E$  equal to  $1/12$ ,  $1/8$  and probably also  $1/6$  are not ergodic (see figure 3). As seen in the diagrams the fraction of the area corresponding to smooth curves (which are responsible for the nonergodic behavior) decreases as the energy increases.

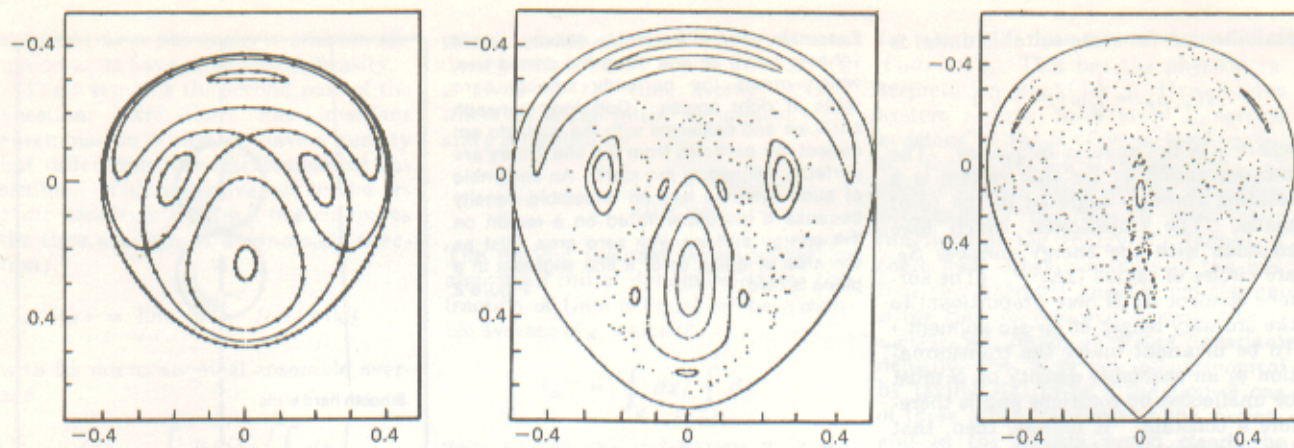
For a general system of anharmonic oscillators, such as a real crystal, we expect similar behavior, with the fraction of  $S_E$  corresponding to nonergodic behavior decreasing as  $E$  increases, and probably disappearing altogether at some critical energy, above which the system would be ergodic and perhaps also show the stronger properties that we shall discuss.<sup>9</sup> At present very little is known about the magnitude of this critical energy in a system with many degrees of freedom.

In the case of gases, the situation is

somewhat different. If there were no interaction at all between the molecules then the energy of each molecule would be an invariant of the motion, so that the system (an ideal gas) would be nonergodic. The KAM theorem would therefore lead us to expect nonergodicity to persist in the event of a sufficiently weak interaction between the particles. The actual interactions, however, are not weak because two molecules very close together repel each other strongly; consequently the theorem does not apply. A simple model of this type is the hard-sphere gas enclosed in a cube with perfectly reflecting walls or periodic boundary conditions. Sinai has outlined a proof that this system is ergodic; he has published a detailed proof, based on the same ideas, for a particle moving in a periodic box containing any number of rigid convex elastic scatterers. We shall refer again to this important result.

#### Mixing

We have seen how to formulate a condition to ensure that the equilibrium properties of a dynamical system are determined by its energy alone and can be calculated from a microcanonical ensemble. This ergodicity condition does not, however, ensure that if we start from a nonequilibrium ensemble the expectation values of dynamical functions will approach their equilibrium values as time proceeds. An illustrative example is given by the harmonic oscillator. For the harmonic-oscillator system, Liouville's theorem shows that the ensemble density repeats itself regularly at time intervals of  $2\pi/\omega$ ; therefore all ensemble



**Nonergodicity of an anharmonic oscillator system with rationally connected frequencies.** The Hamiltonian for this system is given on page 26. Michael Hénon and Carl Heiles (see reference 10) did computer calculations for this system and found that energy surfaces  $S_E$  with  $E$  equal to  $1/12$ ,  $1/8$  and probably  $1/6$  are not ergodic. The planes shown here are intersections of the surface  $q_1$  equal to zero with  $S_E$  for  $E$  equal to  $1/12$  (a),  $1/8$  (b) and  $1/6$  (c), and the points are the intersections of a trajectory with this plane. When the trajectory lies on a smooth two-dimensional invariant torus, the intersection points form a smooth curve, but the intersections of an "erratic" trajectory (one that does not lie on a smooth curve) are more or less random. Note that the fraction of area corresponding to smooth curves (which are responsible for the nonergodic behavior) decreases with increasing energy. Figure 3

averages also have this periodicity, and so cannot irreversibly approach their equilibrium values.

To ensure that our ensembles approach equilibrium in the way we would expect of ensembles composed of real systems, we need a stronger condition than ergodicity. To see what is required, let us start at  $t = 0$  with some ensemble density  $\rho_0(x)$  on  $S$ , which is supposed to represent the initial nonequilibrium state. At a later time  $t$  the ensemble density is, by Liouville's theorem,  $\rho_0[\phi_{-t}(x)]$ . The expectation value of any dynamical variable  $f$  at time  $t$  is therefore

$$\int_S f(x)\rho_0[\phi_{-t}(x)]dx \quad (2)$$

As  $t$  becomes large, we would like this integral to approach the equilibrium value of  $f$ , which is (for an ergodic system)  $\int_S f(x)dx / \int_S dx$ . A sufficient condition for this is that the system should satisfy the condition called *mixing*,<sup>21</sup> which is that for every pair of functions  $f$  and  $g$  whose squares are integrable on  $S$  we require

$$\lim_{t \rightarrow \pm\infty} \int_S f(x)g(\phi_{-t}(x))dx = \frac{\int_S f(x)dx \int_S g(x)dx}{\int_S dx}$$

The special case where  $g$  is  $\rho_0$  shows that integral 2 will approach the equilibrium value of  $f$  for large  $t$  when the system is mixing. Another way of looking at this condition is that it requires every equilibrium time-dependent correlation function such as  $\langle f(x)g[\phi_t(x)] \rangle$  to approach a limit  $\langle f \rangle \langle g \rangle$

as  $t$  approaches  $\pm\infty$ . The condition can be shown to be equivalent to the following requirement: if  $Q$  and  $R$  are arbitrary regions in  $S$ , and an ensemble is initially distributed uniformly over  $Q$ , then the fraction of members of the ensemble with phase points in  $R$  at time  $t$  will approach a limit as  $t$  approaches  $\infty$ ; this limit equals the fraction of the area of  $S$  occupied by  $R$ .

Mixing is a stronger condition than ergodicity: it can easily be shown to imply ergodicity but is not implied by it, as we have seen in the case of a simple harmonic oscillator.

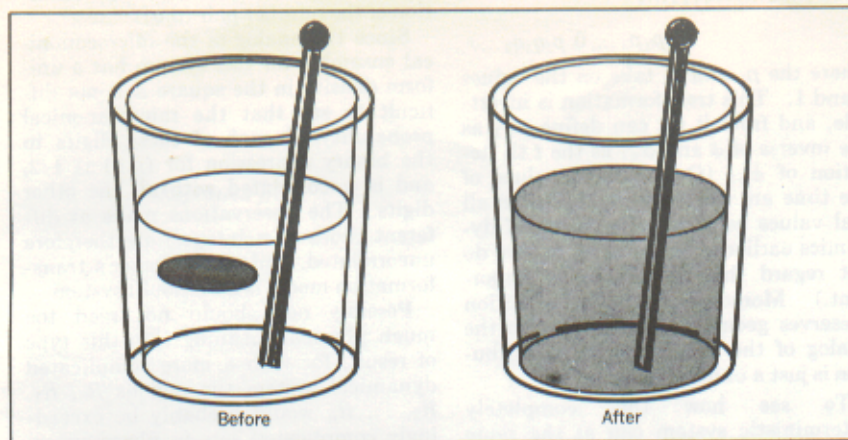
The mathematical definition of mixing was introduced by John von Neumann<sup>12</sup> in 1932 and developed by E. Hopf,<sup>13</sup> but goes back to J. Willard Gibbs,<sup>14</sup> who discusses it by means of an analogy: "... the effect of stirring an incompressible liquid ... Let us suppose the liquid to contain a certain amount of coloring matter which does not affect its hydrodynamic properties ... [and] that the coloring matter is distributed with variable density. If we give the liquid any motion whatever ... the density of the coloring matter at any same point of the liquid will be unchanged ... Yet ... stirring tends to bring a liquid to a state of uniform mixture."

Gibbs saw clearly that the ensemble density  $\rho_t$  of a mixing system does not approach its limit in the usual "fine-grained" or "pointwise" sense of  $\rho_t(x)$  approaching a limit as  $t \rightarrow \infty$  for each fixed  $x$ . Rather, it is a "coarse-grained" or "weak" limit, in which the average of  $\rho_t(x)$  over a region  $R$  in  $S$  approaches a limit as  $t \rightarrow \infty$  for each fixed  $R$ . (A similar distinction applies in defining the entropy. The fine-

grained entropy  $-k \int \rho_t(x) \log \rho_t(x)dx$ , where  $k$  is Boltzmann's constant, retains its initial value forever, but the coarse-grained entropy  $-k \int \rho_t^*(x) [\log \rho_t^*(x)]dx$ , where  $\rho_t^*(x)$  is a coarse-grained ensemble density obtained by averaging  $\rho_t(x)$  over cells in phase space, does change for a nonequilibrium ensemble, and approaches as its limit the equilibrium entropy value  $k \log \int_S dx$ .)

It is sometimes argued that one cannot have a proper approach to equilibrium for any finite mechanical system because of a theorem, due to Poincaré,<sup>15</sup> that every such system eventually returns arbitrarily close to its initial state. (The time involved, however, will be enormously large for a macroscopic system. Boltzmann, for example, estimated a typical Poincaré period for  $100\text{cm}^3$  of gas to be enormously long compared to  $10$  raised to the power  $10$  raised to the power  $10$  years.) Here, however, we are considering ensembles, not individual systems, and the mixing condition guarantees that the ensemble density eventually becomes indistinguishable from the microcanonical density and remains so forever after. It is true that individual systems in the ensemble will return to their initial dynamical states, as required by Poincaré's theorem, but this will happen at different times for different systems, so that at any particular time only a very small fraction of the systems in the ensemble are close to their initial dynamical states.

The reason for the irrelevance of Poincaré recurrences in mixing systems is that the motion of the phase point is very unstable. Dynamical states that start very close to each other in phase



**A familiar example of "mixing."** According to V. I. Arnold and A. Avez, the two liquids are rum (twenty percent) and cola (eighty percent), with the result of the mixing process known as a "Cuba libre." (See reference 1 for details of the process.) Figure 4

space become widely separated as time progresses, so that the recurrence time depends extremely sensitively on the initial conditions of the motion. (The importance of this instability in statistical mechanics was first recognized by N. S. Krylov, a Russian physicist who died in his twenties in 1947; a recent paper<sup>16</sup> gives a theorem that shows this statement about instability to be a rigorous consequence of mixing.) This type of instability appears to be characteristic of real physical systems, and leads to one sort of irreversibility: even if we could reverse the velocities of every particle in a real system that has been evolving towards equilibrium, the system would not necessarily return or even come close to its initial dynamical state with the velocities reversed because the unavoidable small external perturbations would be magnified. This instability is noticeable in molecular-dynamics calculations with hard-sphere systems: if we numerically integrate the equations of motion from time 0 to  $t$  and then try to recover

the initial state by integrating backwards from time  $t$  to time 0, we obtain instead a completely new state. This is because the numerical integration is unstable to small rounding-off errors made during the computation, which play the same role as external perturbations in a real system.

Only a few physical systems have been proven so far to be mixing. The most important is the hard-sphere gas, mentioned above. Sinai's proof that this system is ergodic<sup>2</sup> also gives the stronger result that it is mixing. Roughly, Sinai's method of proving mixing is to show that the hard-sphere system is unstable in the sense discussed above. Physically this instability comes from the fact that a slight change in direction of motion of any particle is magnified at each collision with the convex surface of another particle. The full proof for the simplest case of a "single" particle moving among fixed convex scatterers occupies some 70 pages of difficult mathematics. (G. Gallavotti has provided an el-

ementary introduction to some of the ideas.<sup>17</sup>)

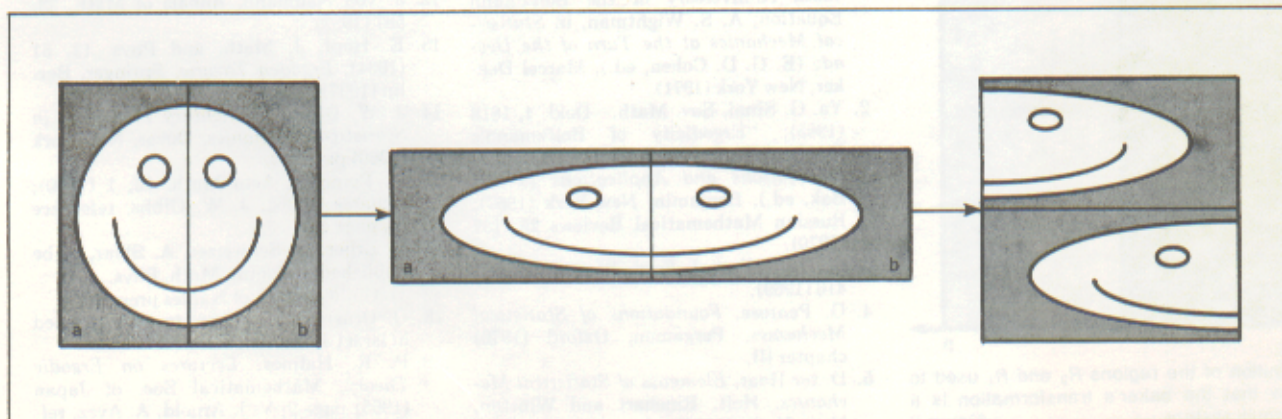
### Bernoulli systems

For lack of space we shall not discuss the next concept in our sequence, that of the "K-system" (after Kolmogorov). For this and related topics such as the Kolmogorov-Sinai "entropy" the reader is referred to reference 1. Instead we jump to the last concept in our hierarchy, that of a Bernoulli system. By a Bernoulli dynamical system we mean one that is, in a sense we shall define, equivalent to an infinite sequence of spins of a suitably designed roulette wheel.

To make this equivalence more precise, we define a *finite partition* of the energy surface  $S$  of our dynamical system as any finite collection of  $n$  nonoverlapping regions  $R_0, \dots, R_{n-1}$  which together cover the whole of  $S$ . Suppose that some device can be made that will determine which of these regions the phase point is in at any time, but gives no information whatever about which part of the region it is in. That is, every time we use this measuring device we obtain an outcome that is a positive integer—the label of the region the phase point of the system is in at that time.

Suppose we use the device repeatedly at intervals of, say, one second. Its outcome will be a sequence of positive integers from the set  $\{0, \dots, n-1\}$ , which can be extended indefinitely. In general, we would expect these integers to be correlated; that is, the microcanonical probability for each new observation depends on what has been observed before (as in a Markov process, for example). This correlation comes about because the dynamical states of the system at different times are deterministically related, through the equations of motion.

The astonishing fact is, however, that for a certain class of dynamical



**The baker's transformation recalls the kneading of a piece of dough.** We first squash the square to half its original height and twice its original width, and then cut the resulting rectangle in half and move the right half of the rectangle above the left. Figure 5

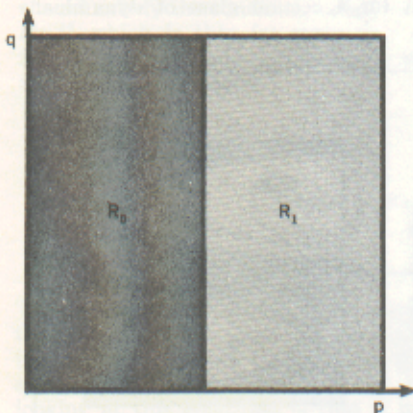
systems it is possible to choose the regions  $R_0, \dots, R_{n-1}$  in such a way that the observations made at different times are completely uncorrelated, just like the numbers shown at different times by a roulette wheel. At the same time, the regions so chosen give enough information to discriminate between dynamical states: if two systems have different dynamical states at some time, then the observations made on them cannot yield identical results for the observations at every time. When such regions can be chosen, we call the system a Bernoulli system.

Recently D. Ornstein and B. Weiss<sup>18</sup> showed that the class of Bernoulli systems includes a type of dynamical system (the geodesic flows on a space of constant negative curvature) whose ergodic properties are very similar to those of the hard-sphere system studied by Sinai, and it appears likely that the hard-sphere system too is a Bernoulli system.

As an illustration of a simple Bernoulli system, consider a system whose phase space is the square  $0 < p \leq 1$ ,  $0 < q \leq 1$  shown in figure 5, and whose (non-Hamiltonian) law of motion is given by a mapping known as the baker's transformation<sup>19</sup> because it recalls the kneading of a piece of dough. If the phase point is  $(p, q)$  at time  $t$ , then at time  $t + 1$  it is at the point obtained by squashing the square to a  $(1/2 \times 2)$  rectangle, then cutting and reassembling to form a new square as shown in the diagram. The formula for this transformation is

$$\phi(p, q) = \begin{cases} (2p, q/2) & \text{if } 0 \leq p \leq 1/2 \\ (2p-1, (q/2 + 1/2)) & \text{if } 1/2 < p \leq 1 \end{cases}$$

If  $p$  and  $q$  are written in binary notation ( $1/8$  in binary notation is  $0.00100\dots$ ,  $1/4$  is  $0.01000$ , and so on), the transformation removes the first



Definition of the regions  $R_0$  and  $R_1$  used to show that the baker's transformation is a Bernoulli system. Figure 6

digit after the binary point from  $p$  and attaches it to  $q$ , so that

$$\phi(0.p_1p_2\dots, 0.q_1q_2\dots) = (0.p_2p_3\dots, 0.p_1q_1q_2\dots)$$

where the  $p_i$  and  $q_i$  take on the values 0 and 1. This transformation is invertible, and from it we can define  $\phi_{-1}$  as the inverse of  $\phi$  and  $\phi_{\pm t}$  as the  $t$ th iteration of  $\phi_{\pm 1}$  (Only integer values of the time are used here, rather than all real values as in our discussion of dynamics earlier in this article, but we do not regard this distinction as important.) Moreover, the transformation preserves geometrical area, and so the analog of the microcanonical distribution is just a uniform density.

To see how this completely deterministic system can at the same time behave like a roulette wheel, we take the regions  $R_0$  and  $R_1$  to be the two rectangles  $0 \leq p < 1/2$ ,  $1/2 \leq p < 1$  shown in figure 6. Suppose the phase point at time 0 is

$$(p, q) = (0.p_1p_2\dots, 0.q_1q_2\dots)$$

If  $p_1$  is zero, the system at time 0 is in  $R_0$ ; if  $p_1$  is one, the system at time 0 is in  $R_1$ . At time 1 the phase point is

$$(0.p_2p_3\dots, 0.p_1q_1q_2\dots)$$

and so we observe the phase point in region  $R_{p_2}$ . At time 2 it is in  $R_{p_3}$  and so on. Likewise, at time  $-1$  it is in

## References

1. V. I. Arnold, A. Avez, *Ergodic Problems of Statistical Mechanics*, Benjamin, New York (1968); J. Ford, "The Transition from Analytic Dynamics to Statistical Mechanics," to be published in *Advances in Chemical Physics*; J. L. Lebowitz, "Hamiltonian Flows and Rigorous Results in Non-equilibrium Statistical Mechanics," in *Proceedings of the IUPAP Conference on Statistical Mechanics*, University of Chicago Press (1972); O. Lanford, "Ergodic Theory and Approach to Equilibrium for Finite and Infinite Systems," to be published in *Proceedings of the 100th Anniversary of the Boltzmann Equation*, Vienna, 1972; Ya. G. Sinai, "Ergodic Theory," to be published in *Proceedings of the 100th Anniversary of the Boltzmann Equation*; A. S. Wightman, in *Statistical Mechanics at the Turn of the Decade* (E. G. D. Cohen, ed.), Marcel Dekker, New York (1971).
2. Ya. G. Sinai, *Sov. Math.—Dokl.* 4, 1818 (1963); "Ergodicity of Boltzmann's Equations" in *Statistical Mechanics, Foundations and Applications* (T. A. Bak, ed.), Benjamin, New York (1967); *Russian Mathematical Reviews* 25, 137 (1970).
3. G. H. Walker, *J. Ford, Phys. Rev.* 188, 416 (1969).
4. O. Penrose, *Foundations of Statistical Mechanics*, Pergamon, Oxford (1970) chapter III.
5. D. ter Haar, *Elements of Statistical Mechanics*, Holt, Rinehart and Winston, New York (1964), Appendix 1, para-

$R_{q_1}$ , at time  $-2$  in  $R_{q_2}$ , and so on. Each observation is determined by a different digit in the binary representation of the number pair  $(p, q)$ .

Since the analog of the microcanonical ensemble for this system has a uniform density in the square it is not difficult to see that the microcanonical probability of each of these digits in the binary expression for  $(p, q)$  is  $1/2$ , and is uncorrelated with all the other digits. The observations made at different times  $t$  ( $=$  integer) are therefore uncorrelated, and so the baker's transformation model is a Bernoulli system.

Possibly one should not read too much physical meaning into this type of result, for with a more complicated dynamical system the regions  $R_0, R_1, R_2, \dots, R_n$  would probably be exceedingly complicated sets in phase space, but from a "philosophical" point of view it is very interesting to see how the same dynamical system can show perfect determinism on the microscopic level and at the same time perfect randomness on a "macroscopic" level.

It is the interplay of these two apparently incompatible levels of description that give the foundations of statistical mechanics their fascination.

\* \* \*

The work described here was supported by US Air Force O.S.R. Grant AFOSR-7-2430 and the US Army Research Office, Durham.

- graph 5; O. Penrose, *Foundations of Statistical Mechanics*, Pergamon, Oxford (1970), page 40; S. G. Brush *Transport Theory and Stat. Phys.* (1971).
6. V. I. Arnold, A. Avez, ref. 1, page 16.
7. A. N. Kolmogorov, "Address to the 1954 International Congress of Mathematicians," [translated in R. Abrahams, *Foundations of Mechanics*, Benjamin, New York (1967), Appendix D].
8. *Enrico Fermi: Collected Papers, Volume II*, University of Chicago Press, Chicago (1965) page 978.
9. G. H. Lunsford, *J. Ford, J. Math. Phys.* 13, 700 (1972).
10. M. Hénon, C. Heiles, *Astron. J.* 69, 73 (1964).
11. V. I. Arnold, A. Avez, ref. 1, page 25.
12. J. von Neumann, *Annals of Math.* 33, 587 (1932).
13. E. Hopf, *J. Math. and Phys.* 13, 51 (1934); *Ergoden Theorie*, Springer, Berlin (1937).
14. J. W. Gibbs, *Elementary Principles in Statistical Mechanics*, Dover, New York (1960) page 144.
15. H. Poincaré, *Acta Math.* 13, 1 (1890); *Oeuvres* 7, 262; J. W. Gibbs, reference 14, page 139.
16. T. Erber, B. Schweizer, A. Sklar, to be published in *Comm. Math. Phys.*
17. G. Gallavotti, U. of Naples preprint.
18. D. Ornstein, B. Weiss, to be published in *Israel J. Mathematics*.
19. P. R. Halmos, *Lectures on Ergodic Theory*, Mathematical Soc. of Japan (1965) page 9; V. I. Arnold, A. Avez, reference 1, page 8. □