Journal of Statistical Physics, Vol. 78, Nos. 5/6, 1995

Ergodicity, Ensembles, Irreversibility in Boltzmann and Beyond

Giovanni Gallavotti[1]

Received March 18, 1994; final September 22, 1994

The contents of a not too well-known paper by Boltzmann are critically examined. The etymology of the word ergodic and its implications are discussed. A connection with the modern theory of Ruelle is attempted.

KEY WORDS: Boltzmann; ergodicity; irreversibility; Ruelle principle; SRB measures; chaos; nonequilibrium.

1. THE ETYMOLOGY OF THE WORD ERGODIC AND THE HEAT THEOREMS

Trying to find the meaning of the word ergodic, one is led to an 1884 paper by Boltzmann.[(6), 2] This paper is seldom quoted[3] and no English translation is available, but I think that it is one of Boltzmann's most interesting papers: it is a precursor of the work of Gibbs[(20)] on ensembles, containing

---

This is an expanded and revised version of a paper read at a conference celebrating the 150th anniversary of the birth of Boltzmann, Vienna, 24 February 1994. This paper is archived in *mp_arc@math.utexas.edu*, #94-66; updated copies (in Postscript) can also be obtained by sending a request to the author by E-mail.

[1] Dipartimento di Fisica, Università di Roma, 00185 Rome, Italy. E-mail 40221::gallavotti.

[2] See the footnote by S. Brush in his edition[(9)] of the *Lectures on Gas Theory*, on p. 297 (§32): here Boltzmann's paper is quoted as the first place where the word is introduced, although the etymology is taken from the Ehrenfests' paper, which is incorrect on this point: see ref. 19, note 93, p. 89 (where also the first appearance of the word is incorrectly dated and quoted).

[3] I found only Brush's reference cited here in footnote 2 and a partial account in ref. 11, pp. 242 and 368, prior to my own etymological discussion which appeared in print in ref. 21 after several years of lectures on the subject. My discussion was repeated in refs. 22 and 23. More recently the paper has been appropriately quoted in ref. 38, unaware of my analysis. The paper was discussed also in ref. 37; see footnote 11 below.

0022-4715/95/0300-1571$07.50/0

it almost entirely [if one recalls that the equivalence of the canonical and microcanonical ensembles was already established (elsewhere) by Boltzmann himself, at least in the free case[1,2]], and I will try to motivate such a statement.

The paper stems from the fundamental, not too well-known work of Helmholtz,[25,26] who noted that *monocyclic* systems[4] could be used to provide models of thermodynamics in a sense that Boltzmann undertakes to extend to a major generalization.

After an introduction, whose relative obscurity has been probably responsible for the little attention this paper has received, Boltzmann introduces the notion of "stationary" probability distribution on the phase space of N interacting particles enclosed in a vessel with volume V. He calls a family $\mathscr{E}$ of such probabilities a *monode*, generalizing an "analogous" concept of monocyclic systems.[5]

In fact the orbits of a monocyclic system can be regarded as endowed with a probability distribution giving an arc length a probability proportional to the time spent on it by the motion: hence their family forms a family of stationary probability distributions.

Etymologically this undoubtedly[6] means a family of stationary distributions with a "unique nature" (each consisting of systems with a "unique nature," differing only by the initial conditions), from μόνοζ and ε' ιδοζ with a probable reference to Plato and Leibnitz.[7]

Then the following question is posed. Given an element μ of a monode $\mathscr{E}$, also called a monode by Boltzmann, we can compute the average values of various observables, e.g., average kinetic energy, average total energy, average momentum transfer per unit time and unit surface in the collisions

[4] This is what we call today a system whose phase space contains only periodic orbits, or cycles: i.e., essentially a one-dimensional conservative system.

[5] In fact Boltzmann first calls a monode just a single stationary distribution regarded as an ensemble. But sometimes he later implicitly, or explicitly, thinks of a monode as a collection of stationary distributions parametrized by some parameters: the distinction is always very clear from the context. Therefore, for simplicity, I take here the liberty of calling a "monode" a collection of stationary distributions, and the individual elements of the collection will be called "elements of the monode." The etymology that follows, however, is more appropriate for the elements of the monodes, as they are thought of as consisting of many copies of the same system in different configurations. By reading Boltzmann's analysis one can get the impression (see p. 132 of ref. 6) that the word monode had been already introduced by Maxwell in ref. 36; however, the reference to Maxwell is probably meant to refer to the notion of stationarity rather than to the word monode, which does not seem to appear in ref. 36.

[6] Of course one can doubt (this as well as many other things).

[7] The concept appears, in fact, in some of Plato's dialogues; see the entry μονοειδήζ ("one in kind") in ref. 35.

with the vessel walls, average volume occupied, and density, denoted, respectively,

$$T=\frac{1}{N}\langle K\rangle_\mu, \qquad U=\langle K+\Phi\rangle_\mu, \qquad p, \qquad V, \qquad \rho=\frac{N}{V} \tag{1.1}$$

where Φ denotes the potential interaction energy and K the total kinetic energy. We then imagine varying μ in the monode $\mathscr{E}$ by an infinitesimal amount (this means changing any of the parameters which determine the element).

Question. Is it true that the corresponding variations dU and dV are such that

$$\frac{dU+p\,dV}{T} \text{ is an exact differential } dS? \tag{1.2}$$

In other words, is it true that the above quantities, defined in purely mechanical terms, satisfy the same relation that would hold between them if, for some thermodynamic system, they were the thermodynamic quantities bearing the same name, with the further identification of the average kinetic energy with the absolute temperature?[8] If so, the monode would provide a "mechanical model of thermodynamics" extending, by far, the early examples of Helmholtz on monocyclic systems.

Thus Boltzmann is led to the following definition:

Definition. A monode $\mathscr{E}$ is called an *orthode* if the property described by (1.2) holds.

Undoubtedly the etymology of "orthode" is ὀρθός and εἶδος, i.e., "right nature."

I find it almost unbelievable that such a deep definition has not been taken up by the subsequent literature. This is more so as Boltzmann, in the same paper, proceeds to discuss "examples" of mechanical models of thermodynamics, i.e., examples of orthodic monodes.

It has certainly not escaped the reader's attention that an orthodic monode (or orthode) is what we call today an *equilibrium ensemble.* The

[8] That the temperature should be identified with the average kinetic energy per particle was quite well established (for free gases) since the paper by Clausius[13] and the paper on the equipartition of kinetic energy by Boltzmann[2] (in the interacting cases); see the discussion of it in Maxwell's last scientific work.[36] The latter paper is also very interesting, as Maxwell asks there whether there are other stationary distributions on the energy surface, and tries to answer the question by putting forward the ergodic hypothesis.

above concept is far more ambitious than the ones involved in the previous proof of the heat theoem in ref. 3. And the above orthodicity concept is still attributed to Gibbs; see, however, ref. 11, p. 242, where it is recognized that ref. 6 amounts to a general theory of ensembles (curiously attributing a particular importance from the point of view of generality to the later contribution of Gibbs on the grand canonical ensemble).

The examples of orthodes discussed by Boltzmann in his paper are the *holode* and the *ergode*, which are two ensembles whose elements are parametrized with two parameters β, N or U, N, respectively. Their elements are

$$\mu_{\beta,N}(d\mathbf{p}\,d\mathbf{q}) = \frac{d\mathbf{p}_1 \cdots d\mathbf{p}_n\,d\mathbf{q}_1 \cdots d\mathbf{q}_n}{\text{const}}\,e^{-\beta(K+\Phi)} \tag{1.3}$$

and

$$\mu_{U,N}(d\mathbf{p}\,d\mathbf{q}) = \frac{d\mathbf{p}_1 \cdots d\mathbf{p}_n\,d\mathbf{q}_1 \cdots d\mathbf{q}_n}{\text{const}}\,\delta(K(\mathbf{p}) + \Phi(\mathbf{q}) - U) \tag{1.4}$$

Boltzmann proves that the above two ensembles are both orthodes! thus establishing that the canonical and the microcanonical ensembles (using our modern terminology) are equilibrium ensembles and provide mechanical models of thermodynamics.[9]

Boltzmann's proof makes use of the auxiliary (with respect to the above definition) notion of heat transfer: in the canonical case it yields exactly the desired result; in the microcanonical case it is also very simple, but somehow based on a different notion of heat transfer. An analysis of the matter easily shows[(24)] that the correct[10] statement becomes exact only in the limit as $N, U \to \infty$, keeping of course U/V, N/V constant, i.e., in what we call today the "thermodynamic limit."

Undoubtedly the word holode has the etymological origin of ὅλοζ and εἶδοζ, while ergode is a shorthand for ergomonode and it has the etymological root of ἔργον and εἶδοζ, meaning a "monode with given

[9] He also studies other ensembles; for instance, in a system in which angular momentum is conserved, e.g., a gas in a spherical container, he considers the stationary distributions with fixed energy and fixed total angular momentum $\mathbf{L}$. Such monodes are called by Boltzmann *planodes* (from the "area law"); and he remarks that in general they are not orthodic (in fact one needs the extra condition that $\mathbf{L} = \mathbf{0}$).

[10] There is a problem only if one insists in defining in the same way the notion of heat transfer in the two cases: this is a problem that Boltzmann does not even mention, possibly because he saw as obvious that the two notions would become equivalent in the thermodynamic limit.

energy."[(21), 11] The word holode is probably a shorthand for holomonode, meaning a "global monode" (perhaps a monode involving states with arbitrary energy, e.g., spread over the whole phase space).

This is not what is usually believed to be the etymology of ergode: the usual belief comes from the Ehrenfests' statement that the etymology is ἔργον and ὁδόζ, with the meaning of "unique path on the surface of constant energy"; see ref. 19, note 93. This etymology has been taken up universally and has been attached to the subject of ergodic theory, which is instead a theory dealing with time evolution properties.

2. THE ERGODIC HYPOTHESIS, CONTINUOUS AND DISCRETE PHASE SPACE

The etymological error of the Ehrenfests could be just an amusing fact: but it had a rather deep negative influence in the development of 20th century physics. They present their etymology in connection with the discussion (amounting to a *de facto* rejection) of the ergodic hypothesis of Boltzmann. In fact, Boltzmann had come to the ergodic hypothesis in his attempts to justify *a priori* that the ergode, as a model of thermodynamics, had to produce *the* thermodynamics of a system with the given Hamiltonian function (and not just a model).

Boltzmann had argued that the trajectory of any initial datum evolves on the surface of constant energy, visiting all phase space points and spending equal fractions of time in regions of equal Liouville measure.

The Ehrenfests criticize such a viewpoint on surprisingly abstract mathematical grounds: basically they say that one can attach to each different trajectory a different label, say a real number, thus constructing a function on phase space constant on trajectories. Such a function would of course have to have the same value on points on the same trajectory (i.e., it would be a constant of motion). This is stated in note 74, p. 86, where the number of different paths is even "counted," and referred to in note 94, p. 89. Therefore, they conclude, it is impossible that there is a single path on the surface of constant energy, i.e., the ergodic hypothesis is inconsistent

[11] The word "ergode" appears for the first time on p. 132 of ref. 6, but this must be a curious misprint, as the concept is really introduced on p. 134. On p. 132 the author probably meant to say "holode" instead: this has been correctly remarked in ref. 38. See also below, footnote 15. The above etymology was proposed probably for the first time by myself in various lectures in Rome, and was included in the first section of ref. 21. A year later a reference to the same new etymology appeared.[(28, 37)]

(except for the monocyclic systems, for which it trivially holds).[12] Having disposed of the ergodic hypothesis of Boltzmann, the Ehrenfests proceed to formulate a new hypothesis, the rather obscure (and somewhat vague, as no mention is made of the frequency of visits to regions in phase space) "quasiergodic hypothesis"; see notes 98 and 99, p. 90, in ref. 19: it led physicists away from the subject and it inspired mathematicians to find the appropriate definition giving birth to ergodic theory and to its first nontrivial results.

The modern notion of ergodicity is not the quasiergodicity of the Ehrenfests. It is simply based on the remark that the Ehrenfests had defined a nontrivial constant of motion very abstractly, by using the axiom of choice. In fact, from the definition, consisting in attaching a different number, or even $6N-2$ different numbers, to each distinct trajectory, there is *in principle* no way to construct a table of the values of the function defined in order to distinguish the different trajectories. In a system ergodic in the modern sense the Ehrenfests' construction would lead to a nonmeasurable function; and to a physicist endowed with common sense, *such a function, which in principle cannot be tabulated, should appear as nonexistent, or as noninteresting.* Thus the motion on the energy surface is called ergodic if there are no *measurable* constants of motion: here measurable is a mathematical notion which essentially states the possibility of a tabulation of the function.

It is surprising that a generation of physicists could be influenced (in believing that the ergodic hypothesis of Boltzmann had to be abandoned as a too naive viewpoint) by an argument of such an exquisitely abstract nature, resting on the properties of a function that could not be tabulated (and not even defined if one did not accept the sinister axiom of choice).[13] Therefore it is perhaps worth trying to understand what Boltzmann could have possibly have meant when he formulated the ergodic hypothesis. Here one cannot fully rely on published work, as the question was never really directly addressed by Boltzmann in a critical fashion (he might have thought, rightly, that what he was saying was clear enough). The following analysis is an elaboration of refs. 21 and 22: in some points it gets quite close to ref. 38. It will not escape the reader that ref. 38 has a somewhat

[12] The abstract mathematical nature of this argument (see also below for a critique) was apparently remarked only by a mathematician (see ref. 38, p. 86), although a great one (Borel, 1914), but it escaped many physicists. It is worrying to note how seriously the mathematicians took the ergodic hypothesis and how easily they disposed of it, taking for granted that the Ehhenfests' formulation was the original formulation by Boltzmann and Maxwell (see ref. 11, p. 383).

[13] We recall, as it is quite an irony, the coincidence that the recognition and the development of the axiom of choice was due essentially to the same Zermelo who was one of the strongest opponents of Boltzmann's ideas on irreversibility; see also ref. 43.

different point of view on several key issues, although we seem to share the main thesis that ref. 19 is responsible for most of the persisting misunderstandings on Boltzmann's work, including the exclusive attribution to Gibbs of Boltzmann's ideas on ensembles, so clearly elaborated in ref. 6. This is so even though, by carefully reading the literature, it is possible to realize that many were aware of the connection of Gibbs's work with Boltzmann's; see, for instance, ref. 11, p. 242, and the translator's (Brush') introduction to ref. 9, p. 12.

My point of view is that Boltzmann always conceived the phase space as a discrete space, divided into small cells; see ref. 4, p. 346. He always stressed that the continuum must be understood as a limit; see ref. 11, p. 371, and refs. 30–32 and 15. The book of Dugas[15] is particularly illuminating also in this respect (see, for instance, Chapter 1 and the quotations of Boltzmann presented there, where he appears to identify the discrete viewpoint with the atomistic conceptions).

Although Boltzmann seems to have been sometimes quite apologetic about such a viewpoint (even calling it a "mathematical fiction"; ref. 10, p. 18, from ref. 4; see also ref. 38, p. 75), he took advantage of it to a point that one can say that most of his arguments are based on a discrete conception of phase space, followed at the end by a passage to the continuum limit. It should be understood, however, that the discretization that Boltzmann had in mind is by no means to be identified with the later concept of coarse graining: see Section 4, where a modern version of Boltzmann's discretization is considered and where a distinction has to be made between cells and volume elements; see also refs. 38 and 23.

It is easier for us, by now used to numerical simulations, to grasp the meaning of a cell: in the numerical simulations a cell is nothing but an element of the discrete set of points in phase space, each represented within computer precision (which is finite). One should always discuss how much the apparently harmless discreteness of the phase space affects the results. This is, however, almost never attempted: see ref. 23 for an attempt. A volume element has, instead, a size much larger than the machine resolution, so that it appears as a continuum (for some purposes).

Hence one can say that an essential characteristics of Boltzmann's thought is to have regarded a system of N atoms, or molecules, as described by a *cell* of dimension δx and δp in position and momentum coordinates. He always proceeded by regarding such quantities as very small, avoiding to enter into the analysis of their size, but every time this had some importance he seems to have regarded them as positive quantities.

A proof of this is when he refutes Zermelo's paradoxes by counting the number of cells of the energy surface of 1 cm^3 of normal air,[7] a feat that can only be achieved if one considers the phase space as discrete.

In particular this point of view must have been taken when he formulated the ergodic hypothesis: in fact, conceiving the energy surface as discrete makes it possible to assume that the motion on it is "ergodic," i.e., it visits *all* the phase space points, compatible with the given energy (and possibly with other "trivial" constants of motion) behaving as a monocyclic system (as all the motions are necessarily periodic).

The passage to the continuum limit, which seems to have never been made by Boltzmann, of such an assumption is of course extremely delicate, and it does not lead necessarily to the interpretation given by the Ehrenfests. It can easily lead to other interpretations, among which is the modern notion of ergodicity. But it should not be attempted here, as Boltzmann himself did not attempt it.

In general, one can hardly conceive that studying the continuum problem could lead to really new information that cannot be obtained by taking a discrete viewpoint. Of course, some problems might be easier if studied in the continuum[(42)]: the few results on ergodicity of physical systems do in fact rely explicitly on continuum models. However, I interpret such results rather as illustrations of the complex nature of the discrete model: for instance, the ergodicity theory of a system like a billiard is very enlightening, as it allows us to get some ideas on the question of whether there exist other ergodic distributions (in the sense of ergodic theory) on the energy surface, and which is their meaning.[(12)]

And the theory of the continuum models has been essential in providing new insights in the description of nonequilibrium phenomena.[(41, 14)]

Finally the fruitfulness of the discrete models can be even more appreciated if one notes that they have been the origin of the quantum theory of radiation: it even can be maintained that already Boltzmann had obtained the Bose–Einstein statistics.[(10)]

The latter is a somewhat strong interpretation of the 1877 paper.[(5)] The most attentive readers of Boltzmann have, in fact, noted that in this discretizations he uses, eventually, the continuum limit as a device to expedite the computations, manifestly not remarking that this would lead to important differences in some extreme cases. In fact he *does not discuss* the two main "errors" that one commits in regarding a continuum formulation as an approximation (based on replacing integrals with sums): they were exploited for the first time by Planck, much later,[14] with respect to a discrete one.

[14] And which amount to the identification of the Maxwell–Boltzmann statistics and the Bose–Einstein statistics and to neglecting the variation of physically relevant quantities over the cells; see the lucid analysis in ref. 29, p. 60; for a technical discussion see refs. 23 and 24.

The above "oversight" might simply be a proof that Boltzmann never took the discretization viewpoint to its extreme consequences, among which is the one that the equilibrium ensembles are *no longer* orthodic in the sense of Boltzmann[23, 24] (although they still provide a model for thermodynamics provided the temperature is no longer identified with the average kinetic energy, a remark that very likely was not made by Boltzmann *in spite of his consideration and interest in the possibility of finding other integrating factors for the heat transfer dQ*; see the footnote on p. 152 in ref. 6.[15] The necessity of understanding this "oversight" has been particularly clearly advocated by Kuhn, referring to Boltzmann's "little studied views about the relation between the continuum and the discrete"; see ref. 29, for instance.

3. THE ERGODIC HYPOTHESIS AND IRREVERSIBILITY

The reaction of the scientific world to the ergodic hypothesis was, "on the average," a violently negative one, also as it was intended to provide further justification for the irreversibility predicted by the Boltzmann equation, derived earlier.

The great majority of scientists saw absurd and paradoxical consequences of the hypothesis, without apparently giving any importance to the "unbelievable" fact that on the basis of a maximal simplicity assumption (i.e., only one cycle on the energy surface) Boltzmann was obtaining not only the possibility of explaining mechanically the classical equilibrium thermodynamics, but also of explaining it in a quantitative way. It allowed for the first time the theoretical calculation of the equations of state of many substances (at least in principle) such as imperfect gases and even other fluids and solids.

The success of the highly symbolic but very suggestive formula of Boltzmann (ref. 19, p. 25)

$$\lim_{T \to \infty} \frac{dt}{T} = \frac{\sigma\, ds}{\int \sigma\, ds} \tag{3.1}$$

(where σ is the microcanonical density on the energy surface, whose area element is ds) in the calculation of the equilibrium properties of matter quickly led physicists to accept it in the "minimal interpretation." Such an interpretation demanded that the r.h.s. be used to compute the equilibrium averages and the l.h.s. ignored, together with the atomic hypothesis. This is

[15] I have profited, in checking my understanding of the original paper as partially exposed in ref. 21, from an English translation that Dr. J. Renn kindly provided while my student in Rome (1984). I was able to cite this footnote in ref. 6 and insert a few new remarks in the present paper because of his translation (unfortunately still unpublished).

regarded as a *law of nature*, in spite of the persistent skepticism (or deep doubts) on its deducibility from the laws of mechanics, a point of view usually attributed to Gibbs, referring to ref. 20, and which is still around, although with, since the mid-fifties, a slow but inexorable inversion of this tendency.

Immediately after the first critiques, Boltzmann elaborated answers often very clear and simple by our modern understanding: but they were very frequently ill understood not only by the opponents of Boltzmann and their epigones, but also by those who were closest to him. The above-quoted critique of the ergodic hypothesis by the Ehrenfests is an example.

Another example is the recurrence paradox, based on the simple theorem of Poincaré. Boltzmann was finally led to the calculation of the number of cells on the energy surface,[7] thus to a superastronomical estimate of the recurrence time, which, nevertheless, did not seem to impress many.

It is also clear that Boltzmann himself became aware of the fact that, after all, the ergodic hypothesis might have been unnecessarily strong and perhaps even useless to explain the approach to equilibrium in physical systems. The latter in fact reach equilibrium normally within times which are microscopic times, not at all comparable with the recurrence time. He asserted repeatedly that the (very few) macroscopic observables of interest had essentially the *same* value in most of the energy surface, and the time spent in the "anomalous phase space cells" is therefore extremely small: a quantitative understanding of this is provided by the Boltzmann equation. This remark also frees (3.1) from the ergodic hypothesis: it might well be that the r.h.s. can be used to evaluate the average values in equilibrium of the few observables which are of interest, although there might be observables (i.e., functions on phase space) for which (3.1) fails.

It is well known that Boltzmann went quite far in this direction, by providing a concrete method to estimate the true times of approach to equilibrium: the Boltzmann equation (developed well before the 1880s).

Finally it is worth noting that the methods used by Boltzmann in deriving the theory of ensembles and the ergodic hypothesis are quite modern and in fact are most suited to illustrate the new developments on nonequilibrium theory, as I shall try to prove in the next section.

4. NONEQUILIBRIUM. RUELLE'S PRINCIPLE. OUTLOOK

I cannot resist the temptation of at least mentioning some recent new developments which look exciting and very likely will be important progress in the field.[16] Equation (3.1), in its minimal interpretation of providing, via the r.h.s. (i.e., the microcanonical distribution), the law for the evaluation of the "relevant" macroscopic observables, starting from the energy function of the system, "solves" the problem of equilibrium theory—completely, as far as we know (in classical physics).

Is a similar theory possible for systems in nonequilibrium, but in a stationary state? What (if anything) replaces the microcanonical distribution in such cases? As an example of "cases" we mean the motion of a gas of particles subject to a constant force ("electric field") setting them in motion, while the energy produced is dissipated into a reservoir.

The answer seems positive, at least in some cases. The problem lies in the fact that the motion of such systems is dissipative, hence the volume element of the energy surface is not conserved even in the simple case in which the thermostat is such that it keeps the total energy of the system constant (as I shall suppose, to simplify the discussion), i.e., the microcanonical distribution cannot describe the stationary state. Taking the continuum viewpoint, we can imagine that the motion is essentially concentrated, after a transient time, on a set A which has zero measure with respect to the Liouville measure on the energy surface.

To avoid giving the impression that the discussion is abstract (hence possibly empty), let me define explicitly one model, among many, that one should have in mind. We consider a system of N particles interacting with a potential energy Φ and subject to an external constant force field $\mathbf{E}$ (e.g., electric field):

$$\dot{\mathbf{q}}_i = \frac{1}{m}\mathbf{p}_i, \qquad \dot{\mathbf{p}}_i = -\partial_{\mathbf{q}_i}\Phi + \mathbf{E} - \alpha(\mathbf{p})\mathbf{p}_i \tag{4.1}$$

where $\mathbf{E}$ is the external constant force and α is defined so that the energy $\sum_{i=1}^{N} \mathbf{p}_i^2/2m + \Phi$ is constant (i.e., $\alpha = \mathbf{E}\cdot\sum \mathbf{p}_i/\sum \mathbf{p}_i^2$). The term $\alpha\mathbf{p}_i$ is a model of a thermostat (this should be called a *Gaussian thermostat*, as it is related to Gauss' principle of "least constraint."[14] Here Φ can be a short-range pair potential plus an external potential: we think of an external potential such that no particle trajectory can avoid interacting with it ("finite horizon"). The system is considered enclosed in a box with periodic boundary conditions: hence we expect that a current parallel to $\mathbf{E}$ will be established and the system will reach a stationary state. The volume in phase space contracts at a rate $(3N-1)\alpha$ (which is positive, on the average): hence the motion will asymptotically develop on some "attractor"

[16] I like to think that Boltzmann is listening to the celebration of his birthday: he would certainly be bored by hearing a (presumably poor) exposition dealing only with things that he knew far better.

which is a set of zero Liouville measure and which we wish to identify with an unstable manifold of the motion (e.g., the unstable manifold of a fixed point or of a periodic orbit) on which the trajectories separate at an exponential rate.

What follows will lead to a unified theory of equilibrium as well as nonequilibrium for system (4.1).

The discrete viewpoint is also possible: the energy surface consists of cells which are relevant (for the study of the asymptotic properties) forming a set A in phase space, and of cells which are irrelevant. The motion can be regarded to develop on the set of cells which are in A, which is strictly smaller than the set of all the cells: in fact far smaller (and in the continuum limit the fraction of cells in A approaches 0).

Since the volume of the cells is not conserved, care must be exercised in regarding the dynamics as a permutation of the cells of A. This is in fact also true in the equilibrium case because, even if the cells do not change in volume, they are deformed, being squeezed in some directions and dilated in others. In equilibrium it is possible to conceive situations in which the deformation can be neglected (this leads to restrictions on the region of temperature and density in which the consideration of the dynamics as a cell permutation is acceptable; a discussion which we have not begun above and which we avoid here as well; see ref. 23 for a quantitative analysis). A similar analysis can be carried in the present case.

Basically one has to think that the system is observed at time intervals τ_0 which are not too small (so that something really happens) and not too large (so that the cell's deformations can be either neglected or controlled, at least for a large majority of cells): see ref. 23 for a quantitative analysis of what this means in the equilibrium cases and of when this might lead to inconsistencies. Let S_{τ_0} denote the transformation of A describing the dynamics on A over the time τ_0. By making the cells small enough we can take τ_0 larger.

We shall imagine the set A as a set of cells around a surface in phase space of dimension roughly $6N/2$ at least if the external force is small (so that the friction α, i.e., the phase space volume contraction, is also small): in fact, if there is no external force the dimension of A should be $1+(6N-2)/2$.[17] The surface A can fold itself on the energy surface filling

[17] Because there are as many contracting directions as expanding ones (the volume being conserved in the $6N$-dimensional phase space), and there are two "neutral" directions (the direction orthogonal to the energy surface and the direction of the phase space motion), one of which lies on the energy surface (the direction of motion).[16,17,44] Of course the existence of other conserved quantities, as in (4.1) when the linear momentum is conserved (e.g., in the trivial case of no external potential), can affect this calculation: in (4.1), when $\mathbf{E}=\mathbf{0}$, this

it up completely (in the $\mathbf{E}=\mathbf{0}$ case) or not (in the general case).[18] We can assume the following extension of the ergodic hypothesis: on A *the dynamics is a one-cycle permutation of the cells.*

Then the motion of a randomly chosen initial datum, randomly with respect to a distribution with some density on the energy surface, will simply consist in a fast approach to the surface A; at the same time data which are on A itself and close to each other will separate from each other at some exponential rate, because on A all the directions are dilated, by definition. To fix the ideas we take the initial data with constant density in some little ball U. If we assume, for simplicity, the above ergodic hypothesis, the layer is, over times multiples of the recurrence time, a set of cells each visited with equal frequency. However, the surface A will, in general, not be a monolayer of cells, but it will have a large "width," i.e., a (macroscopic) area element $d\sigma$ will contain many (microscopic) cells.[19]

The number of cells per unit area can be deduced by remarking that after a time $\tau = M\tau_0$ the density of cells around $x \in A$, initially distributed with constant density in the region U (where the initial data are randomly chosen), has to be proportional to the inverse of the area expansion rate of the transformation S_τ. This means that we expect that the distribution on A which has to be used to compute the stationary averages is described by a suitable density with respect to the area element on A.

With this intuitive picture in mind,[41, 18] we see that a little ball U in phase space evolves, becoming a thin layer around A: the density of the layer, after a large time T, is proportional to the expansion rate of the surface area on A under the transformation S_T that generates the time evolution over the given time.

In the case of no external forces one has that the surface A folds itself on the energy surface coming back to a given phase space volume element V_0 (not to be confused with a cell, which has to be thought of as much smaller); just enough times, and with enough volume around, so that the fraction of the volume initially in U and falling in the volume element V_0 is proportional to V_0 itself (this is consistent because of the equality of the

brings down the dimension to $1+(6N-8)/2$. Furthermore, we are assuming here that there are no "neutral" directions other than the ones possibly provided by the obvious conservation laws: i.e., that our system has strong instability properties (hence this does not *directly* apply to the free gas, for instance).

18 .

.

.

19 This can perhaps be clarified if one thinks of the numerical experiments in which the computer representatives of the phase space points are regarded as cells, while the unstable manifolds of the motion are regarded as surfaces built with computer points, i.e., cells.

total expansion rate and the total contraction rate, due to the Hamiltonian nature of the equations of motion). But in general the fraction of volume U falling into a volume element will be far different from the volume element fraction of the energy surface.

One is thus led to the following unified "principle" to describe the stationary states of nonequilibrium systems.[41]

Principle. The average values of the observables in the stationary state describing the asymptotic behavior of systems like (4.1) are computable from a probability distribution on A which has a density with respect to the surface element of A.[20]

This principle can be more mathematically stated (a problem we refrain from entering here), and is due to Ruelle,[41] who based his work also on the results of Anosov, Sinai, and Bowen on the theory of a class of dynamical systems known as "hyperbolic systems" (which play in some sense, for nonequilibrium statistical mechanics, the role of the monocyclic systems of Helmholtz). The probability distributions selected by the above principle (which in "good cases" are unique) are called SRB measures.[41]

What is the predictive value of the above statements? In cases without external forces we have already mentioned that this principle leads to the microcanonical distribution and therefore implies classical thermodynamics.[6] Life is made easy by the fact that although A may be very difficult to identify, still the stationary distribution is just the microcanonical ensemble because A folds on the energy surface filling it up completely, with no gaps.

In the dissipative cases it seems that we have little control on A and hence on the stationary distribution.

Yet this might not be really so: we simply have to learn how to extract information from such an abstract principle. After all, it now seems natural that the Gibbs distribution predicts all the phenomena of equilibrium statistical mechanics (from the phase coexistence, to the critical point, to crystallization). But this was far from clear only a few decades ago, and many decades after the original formulations of Maxwell, Gibbs, and Boltzmann (as many of us certainly recall).

That the principle might have predictive value is indicated by the first attempts as its use in problems of statistical mechanics; see ref. 18 (see also ref. 14), who were somewhat inspired by previous papers; see also ref. 27. In fact, only recently has the principle started being considered in the

[20] It is extremely important to think, to avoid trivial contradictions, that the cells on A must be regarded as much smaller than the surface elements of A that we consider in talking about the density.

theory of nonequilibrium, as it was developed originally by Ruelle mainly as an attempt at a theory of turbulent phenomena. This is not the appropriate place to discuss ref. 18 in the perspective of the above principle: the discussion is rather delicate (as ref. 18 should be regarded as a pioneering work): see ref. ??.

A simpler example of a quantitative (yet quite abstract) consequence of the above principle is the determination of the density function mentioned in the principle: the latter is in fact essentially determined. If we are interested in stationary distributions, phenomena which are observable by measurements that take place in a fixed time τ, we can just take averages over A with respect to a distribution with density over A proportional to $\Lambda_{\tau'}^{-1}(x)$, with $\tau' = M'\tau_0 \gg \tau$ [where the expansion rate is the Jacobian determinant of the transformation $S_{\tau'}$ at x, i.e., $\Lambda_{\tau'}^{-1}(x) \equiv \prod_{-M'}^{0} \Lambda_{\tau_0}^{-1}(S_{\tau_0}^{j} x)$]. So two equal-area elements of A around x and y have a *relative probability* of visit equal to $\Lambda_{\tau'}^{-1}(x)/\Lambda_{\tau'}^{-1}(y)$.

Of course τ' cannot be taken too large: if τ' is taken of the order of the recurrence time, the ratio becomes 1. The natural upper bound on τ' has to be such that the cells in U ending in the considered area elements are still large in number. This sets an upper limit to the values of τ for which the above remark applies.[21]

The example (4.1) is very special.

It is, however, generalizable: many generalizations have already been considered in the literature.[40] Still, it should be stressed that the models to which the above principle can be applied form a rather small class of deterministic models. It is not immediately clear how it can be applied to stationary nonequilibrium phenomena in which the thermostat is realized in a different way, e.g., by some stochastic boundary conditions. Nor it is obvious that the different thermostats are physically equivalent.

In my opinion there is also some misunderstanding in the literature about the fact that the set A has zero measure (in the nonequilibrium cases this has been sometimes associated with questions related to irreversibility) and about the fact that A, regarded as a folded surface on the equal-energy manifold, has a fractal dimension (thereby representing a "strange attractor"). Such facts may be quite misleading. The above analysis shows that A should be more conveniently regarded as a smooth nonfractal surface of dimension $1 + (6N-2)/2 \sim 3N$: its fractal dimension arises from the

[21] This means that the ratio between the linear dimension of U and the linear dimension of the cells has to be large compared to the maximal linear expansion rate over the time τ, a condition that can be expressed in terms of the largest Lyapunov exponent.

folding of A on the surface of constant energy (rising from $\sim 3N$ to about $6N$ if $\mathbf{E}$ is small).[22]

Furthermore, in the assumption that the stochastic thermostats and the Gaussian thermostat (or other thermostat[40]) are equivalent one sees clearly a problem related to attaching importance to the set A as a fractal with zero measure. In fact, we expect that stochastic thermostats lead to stationary distributions which have a density in phase space, and hence which cannot be concentrated on a set of zero measure.

The contradiction disappears if one thinks that, in a stationary state, there may be several distributions which, in the limit as $N \to \infty$, become equivalent. A distribution concentrated on a set of zero measure might well be equivalent to one distributed on the whole energy surface, or on the whole phase space, if $N \to \infty$. A much simpler, but very familiar, example of such a situation is provided by the microcanonical distribution, which is concentrated on a set of zero measure, but it is equivalent (in the thermodynamic limit) to the canonical distribution, which is concentrated on the whole phase space.

Finally it should be clear that the problem of approach to stationarity will show up exactly in the same terms as in the equilibrium cases. The "ergodicity" assumptions above cannot in any way justify the use of the distribution satisfying the Ruelle principle: the time necessary for a phase space point to visit the full set of cells building A will be of the order of magnitude of the recurrence time, and as in the equilibrium cases, we can expect that the rapidity of the approach to equilibrium is rather due to the fact that we are interested only in very few observables, and such observables have the same value in most of phase space.

I hope to have shown, or at least given arguments, that the point of view (see, e.g., ref. 38) according to which Boltzmann was a 19th century physicist judged by his interpreters with 20th century mathematical standards is not exactly correct: today's way of thinking is not too different from his, and most problems the physicists had with his work were problems with the understanding of his physics and *not* of his mathematics (see also ref. 33). The misunderstandings about his ideas are, in my opinion, largely due to unwillingness to study the original publications and to the unfounded belief that they were treated with fidelity by the reviewers who wrote about his achievements.

[22] This is shown also by the fact that the operation i mapping $x = (\mathbf{p}, \mathbf{q})$ to $ix = (-\mathbf{p}, \mathbf{q})$ is such that $t \to ix(-t)$ is a solution of the equation of motion if $t \to x(t)$ is such: a time-reversal symmetry. This has several implications, including the properties that both initial data x and ix evolve toward the same attractor A in the future and to the attractor iA in the past. In general A and iA are different, except in the case $\mathbf{E} = \mathbf{0}$ (because A is the full energy surface).

Recently a new book has appeared in which von Plato[39] not only develops the concepts he already gave in ref. 38, but also gives a very interesting discussion of the relation between Boltzmann's and Maxwell's points of view as well as a detailed and documented history of the ergodic hypothesis, before and after Boltzmann: I am indebted to a referee for pointing this reference out to me. The point of view of von Plato, discussed here with reference to ref. 38, is further developed.

ACKNOWLEDGMENTS

My father Carlo provided me essential help in the explanation of the etymology of the word ergodic. Part of the interpretation of Ruelle's principle presented here was developed in collaboration with E. Cohen in a joint effort to understand more deeply the results of ref. 18; while our analysis, which preceded this paper, will be published elsewhere, I wish to thank him for communicating to me his enthusiasm on the subject while I was visiting Rockefeller University, and for his thoughtful comments on this paper. I am indebted to J. Lebowitz for his hospitality at Rutgers University and for stimulating my interest in Gaussian thermostats. To him and to Y. Sinai I owe also the correction of several misconceptions and mathematical errors.

REFERENCES

1. L. Boltzmann, Über die mechanische Bedeutung des zweiten Haupsatzes der Wärmetheorie, in *Wissenschaftliche Abhandlungen*, F. Hasenöhrl, ed. (reprinted Chelsea, New York), Vol. I, pp. 9–33.
2. L. Boltzmann, Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten, in *Wissenschaftliche Abhandlungen*, F. Hasenöhrl, ed. (reprinted Chelsea, New York), Vol. I, pp. 49–96.
3. L. Boltzmann, Analytischer Beweis des zweiten Hauptsatzes der mechanischen Wärmetheorie aus den Sätzen über das Gleichgewicht des lebendigen Kraft, in *Wissenschaftliche Abhandlungen*, F. Hasenöhrl, ed. (reprinted Chelsea, New York), Vol. I, pp. 288–308.
4. L. Boltzmann, Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen, in *Wissenschaftliche Abhandlungen*, F. Hasenöhrl, ed. (reprinted Chelsea, New York), Vol. I, pp. 316–402 [English transl., in S. Brush, ed., *Kinetic Theory* (Pergamon Press, Oxford), Vol. 2, p. 88].
5. L. Boltzmann, Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung, respektive den Sätzen über das Wärmegleichgewicht, in *Wissenschaftliche Abhandlungen*, F. Hasenöhrl, ed. (reprinted Chelsea, New York, 1968), Vol. II, pp. 164–223.
6. L. Boltzmann, Über die Eigenschaften monzyklischer und anderer damit verwandter Systeme, in *Wissenschaftliche Abhandlungen*, F. P. Hasenöhrl, ed. (reprinted Chelsea, New York, 1968), Vol. III.

7. L. Boltzmann, Entgegnung auf die wärmetheoretischen Betrachtungen des Hrn. E. Zermelo, in S. Brush, ed., *Kinetic Theory* (Pergamon Press, Oxford), Vol. 2, p. 218 [English transl.].
8. L. Boltzmann, Zu Hrn. Zermelo's Abhandlung "Ueber die mechanische Erklärung irreversibler Vorgänge," in S. Brush, ed., *Kinetic Theory* (Pergamon Press, Oxford), Vol. 2, p. 238 [English transl.].
9. L. Boltzmann, *Lectures on Gas Theory* [annotated by S. Brush] (University of California Press, Berkeley, 1964).
10. A. Bach, Boltzmann's probability distribution of 1877, *Arch. History Exact Sci.* **41**:1–40 (1990).
11. S. Brush, *The Kind of Motion We Call Heat* (North-Holland, Amsterdam, 1976/Vol. II; 1986/Vol. I).
12. L. Bunimovitch, Y. Sinai, and N. Chernov, Statistical properties of two dimensional hyperbolic billiards, *Russ. Math. Surv.* **45**(3):105–152 (1990).
13. R. Clausius, The nature of the motion which we call heat, in *Kinetic Theory*, S. Brush, ed. (Pergamon Press, Oxford), pp. 111–147.
14. K. Chernov, G. Eyink, J. Lebowitz, and Y. Sinai, Steady state electric conductivity in the periodic Lorentz gas, *Commun. Math. Phys.* **154**:569–601 (1993).
15. R. Dugas, *La théorie physique au sens de Boltzmann* (Griffon, Neuchâtel, 1959).
16. U. Dressler, Symmetry property of the Lyapunov exponents of a class of dissipative dynamical systems with viscous damping, *Phys. Rev.* **38A**:2103–2109 (1988).
17. D. Evans, E. Cohen, and G. Morriss, Viscosity of a simple fluid from its maximal Lyapunov exponents, *Phys. Rev.* **42A**:5990–5997 (1990).
18. D. Evans, E. Cohen, and G. Morriss, Probability of second law violations in shearing steady flows, *Phys. Rev. Lett.* **71**:2401–2404 (1993).
19. P. Ehrenfest and T. Ehrenfest, *The Conceptual Foundations of the Statistical Approach in Mechanics* (Dover, New York, 1990) [reprint].
20. J. Gibbs, *Elementary Principles in Statistical Mechanics* (Ox Bow Press, 1981) [reprint].
21. G. Gallavotti, Aspetti della teoria ergodica qualitativa e statistica del moto, *Quaderni UMI* (Pitagora, Bologna) **21** (1982).
22. G. Gallavotti, L'hypothèse ergodique et Boltzmann, in *Dictionnaire Philosophique* (Presses Universitaires de France, Paris, 1989), pp. 1081–1086.
23. G. Gallavotti, Meccanica Statistica, in *Enciclopedia italiana delle scienze fisiche* (Rome, 1994); *Fisiche* (Rome, 1994); Equipartizione e critica della Meccanica Statistica Classica, in *Enciclopedia italiana delle scienze*; Teoria Ergodica, in *Enciclopedia del Novecento* (in press).
24. G. Gallavotti, Insiemi statistici, in *Enciclopedia italiana delle scienze fisiche* (Rome, 1994).
25. H. Helmholtz, Principien der Statik monocyklischer Systeme, in *Wissenschaftliche Abhandlungen* (Leipzig, 1895), Vol. III, pp. 142–162, 179–202.
26. H. Helmholtz, Studien zur Statik monocyklischer Systeme, in *Wissenschaftliche Abhandlungen* (Leipzig, 1895), Vol. III, pp. 163–172, 173–178.
27. B. Holian, W. Hoover, and H. Posch, Resolution of Loschmidt's paradox: The origin of irreversible behaviour in reversible atomistic dynamics, *Phys. Rev. Lett.* **59**:10–13 (1987).
28. K. Jacobs, Ergodic theory and combinatorics, *Contemp. Math.* **26**:171–187 (1984).
29. T. Kuhn, *Black Body Theory and the Quantum Discontinuity. 1814–1912* (University of Chicago Press, Chicago, 1987).
30. M. Klein, Maxwell and the beginning of the quantum theory, *Arch. History Exact Sci.* **1**:459–479 (1962).
31. M. Klein, Mechanical explanations at the end of the nineteenth century, *Centaurus* **17**:58–82 (1972).

32. M. Klein, The development of Boltzmann's statistical ideas, in *The Boltzmann Equation*, E. Cohen and W. Thirring, eds., *Acta Physica Austriaca*, Suppl. X, pp. 53–106.
33. J. Lebowitz, Boltzmann's entropy and time's arrow, *Phys. Today* **1993**(September):32–38.
34. R. Livi, A. Politi, and S. Ruffo (1986). Distribution of characteristic exponents in the thermodynamic limit, *J. Phys.* **19A**:2033–2040 (1986).
35. H. Liddell and R. Scott, *Greek–English Lexicon* (Oxford, University Press, Oxford, 1994).
36. J. Maxwell, On Boltzmann's theorem on the average distribution of energy in a system of material points, in *The Scientific Papers of J. C. Maxwell*, W. Niven, ed. (Cambridge University Press, Cambridge, 1890), Vol. II, pp. 713–741.
37. M. Mathieu, On the origin of the notion "Ergodic Theory," *Expositiones Math.* **6**:373–377 (1988).
38. J. von Plato, Boltzmann's ergodic hypothesis, *Arch. History Exact Sci.* **44**:71–89 (1992).
39. J. von Plato, *Creating Modern Probability* (Cambridge University Press, Cambridge, 1994).
40. H. Posch and W. Hoover, Nonequilibrium molecular dynamics of a classical fluid, in *Molecular Liquids: New Perspectives in Physics and Chemistry*, J. Teixeira-Dias, ed. (Kluwer, Dordrecht, 1992), pp. 527–547.
41. D. Ruelle, Measures describing a turbulent flow, *Ann. N.Y. Acad. Sci.* **357**:1–9 (1980); see also J. Eckmann and D. Ruelle, Ergodic theory of strange attractors, *Rev. Mod. Phys.* **57**:617–656 (1985); D. Ruelle, Ergodic theory of differentiable dynamical systems, *Publ. Math. IHES* **50**:275–306 (1980).
42. Y. Sinai, Dynamical systems with elastic reflections. Ergodic properties of dispersing billards, *Russ. Math. Surv.* **25**:137–189 (1970).
43. J. Schwartz, The pernicious influence of mathematics on science, in *Discrete Thoughts: Essays in Mathematics, Science, and Philosophy*, M. Kac, G. Rota, and J. Schwartz, eds. (Birkhauser, Boston, 1986), pp. 19–25.
44. S. Sarman, D. Evans, and G. Morriss, Conjugate pairing rule and thermal transport coefficients, *Phys. Rev.* **45A**:2233–2242 (1992).