100 Years 1915–2015

PERSPECTIVE



# Ergodic theorem, ergodic theory, and statistical mechanics

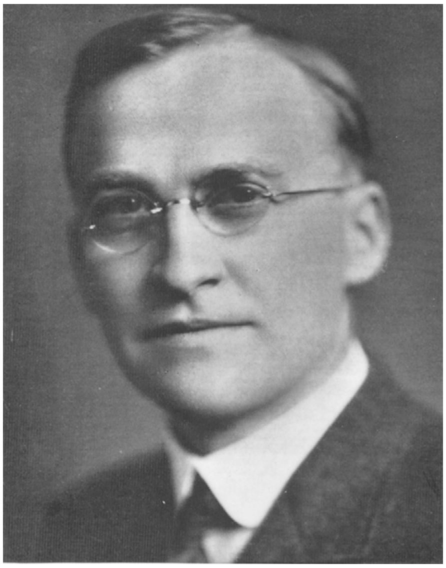
**Calvin C. Moore**[1]

*Department of Mathematics, University of California, Berkeley, CA 94720*

Edited by Kenneth A. Ribet, University of California, Berkeley, CA, and approved January 9, 2015 (received for review November 13, 2014)

**This perspective highlights the mean ergodic theorem established by John von Neumann and the pointwise ergodic theorem established by George Birkhoff, proofs of which were published nearly simultaneously in PNAS in 1931 and 1932. These theorems were of great significance both in mathematics and in statistical mechanics. In statistical mechanics they provided a key insight into a 60-y-old fundamental problem of the subject—namely, the rationale for the hypothesis that time averages can be set equal to phase averages. The evolution of this problem is traced from the origins of statistical mechanics and Boltzman's ergodic hypothesis to the Ehrenfests' quasi-ergodic hypothesis, and then to the ergodic theorems. We discuss communications between von Neumann and Birkhoff in the Fall of 1931 leading up to the publication of these papers and related issues of priority. These ergodic theorems initiated a new field of mathematical-research called ergodic theory that has thrived ever since, and we discuss some of recent developments in ergodic theory that are relevant for statistical mechanics.**

George D. Birkhoff (1) and John von Neumann (2) published separate and virtually simultaneous path-breaking papers in which the two authors proved slightly different versions of what came to be known (as a result of these papers) as the ergodic theorem. The techniques that they used were strikingly different, but they arrived at very similar results. The ergodic theorem, when applied say to a mechanical system such as one might meet in statistical mechanics or in celestial mechanics, allows one to conclude remarkable results about the average behavior of the system over long periods of time, provided that the system is metrically transitive (a concept to be defined below). First of all, these two papers provided a key insight into a 60-y-old fundamental problem of statistical mechanics, namely the rationale for the hypothesis that time averages can be set equal to phase averages, but also initiated a new field of mathematical research called ergodic theory, which has thrived for more than 80 y. Subsequent research in ergodic theory since 1932 has further expanded the connection between the ergodic theorem and this core hypothesis of statistical mechanics.



George D. Birkhoff. Image courtesy of the American Mathematical Society (www.ams.org).

The justification for this hypothesis is a problem that the originators of statistical mechanics, J. C. Maxwell (3) and L. Boltzmann (4), wrestled with beginning in the 1870s as did other early workers, but without mathematical success. J. W. Gibbs in his 1902 work (5) argued for his version of the hypothesis based on the fact that using it gives results consistent with experiments. The 1931–1932 ergodic theorem applied to the phase space of a mechanical system that arises in statistical mechanics and to the one-parameter group of homeomorphisms representing the time evolution of the system asserts that for almost all orbits, the time average of an integrable function on phase space is equal to its phase average, provided that the one-parameter group is metrically transitive. Hence, the ergodic theorem transforms the question of equality of time and phase averages into the question of whether the one-parameter group representing the time evolution of the system is metrically transitive.

To be more specific about statistical mechanics systems, consider a typical situation in gas dynamics where one has a macroscopic quantity of a dilute gas enclosed in a finite container. The molecules are in motion, colliding with each other and with the hard walls of the container. The molecules can be assumed for instance to be hard spheres (billiard balls) bouncing off each other or alternately may be assumed to be polyatomic molecules with internal structure and where collisions are governed by short-range repelling potentials. One may also choose to include the effects of external forces, such as gravity on the molecules. We assume that the phase space $M$ consists of a surface of constant energy. This assumption, together with the finite extent of the container, ensures that $M$ is compact and that the invariant measure derived from Liouville's theorem is finite. The equations of motion, say in Hamiltonian form, can be written in local coordinates as a first-order system of ordinary differential equations

$$dx_i/dt = X_i(x_1, \ldots x_n).$$

First, the number of variables in the equations is enormous, perhaps on the order of Avogadro's number, and the equations are quite complex. The system is perfectly deterministic in principle; hence, given the initial positions and momenta of all of the molecules at an initial time, the system evolves

Author contributions: C.C.M. wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

This article is part of the special series of PNAS 100th Anniversary articles to commemorate exceptional research published in PNAS over the last century. See the companion articles, "Proof of the ergodic theorem" on page 656 in issue 12 of volume 17 and "Proof of the quasi-ergodic hypothesis" on page 70 in issue 1 of volume 18, and see Core Concepts on page 1914.

[1]Email: ccmoore@math.berkeley.edu.

John von Neumann. Image courtesy of the Shelby White and Leon Levy Archives Center, Institute for Advanced Study, Princeton.

deterministically. However, there is no chance of knowing these initial conditions exactly and little chance of integrating these equations to find the solutions. Given therefore that we only have partial information about the system, a statistical approach to the analysis of such systems is appropriate and necessary. Maxwell (6) and Boltzmann (7) began such a project, which was further developed and elaborated by Gibbs (5).

The system of differential equations above generates a flow, which we denote $P_t(x)$ where $x=(x_1,\dots x_n)$ is a point in the phase space $M$ representing the system at time $t=0$. $P_t(x)$ is the position to which the system moves after time $t$ so that $P_t(x)$ is the solution of the differential equation with initial value $x$ at time $t=0$. $P_t$ is a homeomorphism of $M$ onto itself defined for all $t$ (as the system is time reversible) and satisfies the group property $P_tP_s=P_{t+s}$. It also preserves the Liouville measure $\mu$, which is finite in this case.

Now if $f$ is an integrable function on the phase space $M$, one may argue that if one makes a physical measurement of $f$ on a system that is in state $x\in M$, what one gets, at least for all but a negligible number of states $x$, is a time average

$$1/2T\int_{-T}^{T} f(P_t(x))dt.$$

This is a time average of the values of $f$ over a part of the path of the solution of the equations with value $x$ at time $t=0$. The argument that it is reasonable to assume that a physical measurement takes a short period to perform, but this short time period is a long period for the physical system because, for instance, each molecule will on average experience billions of collisions per second in a typical gas. In fact, this line of reasoning leads to the idea that the result of the measurement is best represented by the limit as $T$ tends to infinity of the time average above. The first problem is, of course, to know that this limit exists, except of course for a negligible set of $x$. Then the assumption of equality of time averages and phase averages would assert that the limit of the time average above is independent of $x$ and equal to the phase average

$$\int_M f(v)d\mu(v),$$

for all but a neglible set of $x\in M$ where $\mu$ is the Liouville invariant measure. The significant and useful point is this phase average can be calculated in many cases, whereas the time average cannot be calculated.

Another way of phrasing this equality is to use for $f$ the indicator function of a measurable subset $A$ of $M$ (a function that is equal to 1 on $A$ and 0 outside of $A$). Then the time averages above record the fraction of time that the orbit spends in $A$, and the basic hypothesis of statistical mechanics asserts that this is equal for almost all orbits to the Liouville measure of $A$ (assuming that the measure of the total space $M$ has been normalized to 1). However, another way of expressing this equality is to assert that for all but a negligible set of states of the gas, the observed value of a function $f$ will be equal to the average value of $f$ taken over $M$ that is an average value of $f$ taken over an ensemble (to use Gibbs's language) of all possible states with the same energy. The constant energy surface $M$ with its invariant volume element here is what Gibbs called the microcanonical ensemble.

The ergodic theorems of Birkhoff and von Neumann assert first of all of the existence of the time limit for $T\to\infty$ for any one parameter measure preserving group, and then, assuming that $P_t$ is metrically transitive, they assert the equality

$$\lim_{T\to\infty}(2T)^{-1}\int_{-T}^{T} f(P_t(x))dt=\int_M f(v)d\mu(v).$$

The difference between the two theorems is that Birkhoff proved that the convergence of the functions of $x$ on the left side is pointwise almost everywhere (the limit in general can be identified as the conditional expectation of $f$ onto the sigma field of invariant sets, to use the language of probability theory). In the case of metric transitivity, this function is just the constant function equal to the integral of $f$ over $M$. von Neumann proved that these functions of $x$ converge in mean square [that is in $L^2(M,\mu)$] to the orthogonal projection of the function $f$ onto the closed subspace of invariant functions, which in the metrically transitive case is one dimensional consisting of the constant functions. Birkhoff assumes the function $f$ is bounded and measurable, whereas von Neumann assumes the more general condition that the function $f$ is square integrable. Although both theorems were originally formulated and proved for measure preserving one parameter groups generated by first-order differential equations on compact manifolds, subsequent work has shown using the same arguments that these results are valid for a much broader class of dynamical systems including one-parameter families of measure preserving transformations of a finite measure space, which may not necessarily be defined by systems of differential equations. Later work also showed that Birkhoff's theorem holds for an integrable function $f$. Thus, these theorems are theorems about one-parameter groups of automorphisms of measure spaces with no mention of topology. The theorems also clarify what is meant by the informal term in statistical mechanics, negligble set, namely, it is a set of $\mu$ that measures zero. It should be added that the time average does not have to be taken from $-T$ to $T$, but can be taken over any intervals from $T_1$ to $T_2$, as long as the difference $T_2-T_1$ tends to infinity.

Before moving on to subsequent developments in ergodic theory, it is worth pausing to examine the sequence of events leading to the proofs and publication of the two ergodic theorems: the pointwise ergodic theorem of Birkhoff and the mean ergodic theorem of von Neumann. Much of this was laid out in a subsequent paper of Birkhoff and Koopman (8) in the March 1932 issue of the PNAS. von Neumann was very much aware of the results of M. H. Stone (9) on spectral theory of one-parameter groups of unitary operators and the results of Koopman (10) that used Stone's results to analyze one-parameter groups of measure preserving transformations. Koopman had indeed suggested to von Neumann that he might use these results to resolve the problem of equality of time and phase averages, and von Neumann writes that Andre Weil had made the very same suggestion to him. von Neumann seized on the notion of metric transitivity, introduced, somewhat ironically, by

Birkhoff and Smith (11) a few years earlier in 1928, and proved his mean ergodic theorem under the hypothesis of metric transitivity. See the article by Mackey (12) for more details.

According to Birkhoff and Koopman, von Neumann communicated his result personally to both of them on October 22, 1931, and pointed out to them that his result raised the important question of whether a pointwise result might be valid. Birkhoff then went to work and, by different methods, quickly established his pointwise ergodic theorem. He submitted his paper to PNAS on December 1, 1931, for appearance in the December 1931 issue. One presumes that he sent copies to Koopman and von Neumann, who would have noticed that Birkhoff had not given von Neumann adequate credit and recognition for his result. von Neumann evidently planned to include his ergodic theorem and its proof in a much longer paper he was writing for the *Annals of Mathematics*, but he then apparently quickly drafted a short paper for PNAS with his proof of the mean ergodic theorem and submitted it to PNAS on December 10, 1931. It appeared in the January 1932 issue. One suspects that these events led Koopman and Birkhoff to write and publish their paper in PNAS 2 months later, which set matters straight and clearly acknowledged von Neumann's priority. It should also be noted that E. Hopf (13) presented a slightly different proof of the mean ergodic theorem and some improvements on the Birkhoff theorem in a paper, which appeared in the January 1932 issue of PNAS. For whatever reason, the Birkhoff paper and its result has over time become the better known of the two papers, but in light of these historical details, the von Neumann paper deserves at least equal billing.

There are also ergodic theorems for a single measure-preserving map $P$ and its iterates $P^n$. They assert the existence of the time limit below and that it converges to the phase average if $P$ is metrically transitive; that is

$$\lim_{N\to\infty}(N+1)^{-1}\sum_{n=0}^{N} f(P^n x) = \int_M f(v)d\mu(v).$$

The convergence is pointwise for almost all $x$ for integrable $f$, and in the mean for $f$ square summable. One way to conceive of metric transitivity and the ergodic theorem for a single transformation is that for almost all points $x$, the $n$ iterates under $P$ of $x$ is distributed in some sense evenly throughout the space so that taking the average of a function $f$ over these points gives a result that is a good approximation to the integral of $f$ over the space and that the more iterates one includes in the average, the better the approximation. Therefore, it is like a numerical integration scheme.

Finally, we need to define metric transitivity, a concept, as previously noted, that was introduced by Birkhoff and Smith (11). A one-parameter group of measure preserving transformation $P_t(x)$ (or a single transformation $P$) on a measure space $M$ is metrically transitive provided that any $\mu$ measurable set invariant under $P_t$ for all $t$ (or $P$) must have zero measure or its complement must have zero measure. This means that the flow is indecomposable or irreducible in the sense that one cannot decompose it into a union of two disjoint subflows. It also means that there are no measurable functions invariant under the flow (or the transformation $P$).

It is heuristically reasonable to argue, owing to the molecular chaos in gas dynamics, that there are no nonconstant continuous invariants or so-called first integrals of the motion. However, more is required for metric transitivity—namely no nonconstant measurable invariants of the motion. In the example from gas dynamics, the total energy is clearly an invariant of the motion, but we have restricted the flow to a surface of constant energy. In addition, total momentum is normally preserved, but although momentum is preserved in collisions between molecules, collisions with the walls do not preserve momentum, so this possible invariant of the motion disappears. Although the term metric transitivity is still in use, current terminology, due to von Neumann, is that any flow or single transformation with this property is simply called ergodic.

It is worth observing that metric transitivity is a necessary and sufficient condition for the validity of the ergodic theorem. To see this, assume the ergodic theorem holds and then apply the statement of the theorem to the indicator function $f$ of a supposed invariant measurable set $A$—that is, $f$ is equal 1 on $A$ and 0 on the complement of $A$; the left side of time averages is always equal to $f$, but the right side is a constant function. Hence, $f$ is a constant function, and the alleged invariant set is of measure zero or its complement is of measure zero.

It is interesting to look back at the early history of statistical mechanics to see how the founders of the subject handled the topic of time averages and space averages. Boltzmann (4) coined the terms ergoden or ergodische (which we translate as ergodic) from the Greek $\epsilon\rho\gamma o\nu$ (energy) and $o\delta o\sigma$ (path) or energy path. He put forth what he called the ergodic hypothesis, which postulated that the mechanical system, say for gas dynamics, starting from any point, under time evolution $P_t$, would eventually pass through every state on the energy surface. Maxwell and his followers in England called this concept the continuity of path (3). It is clear that under this assumption, time averages are equal to phase averages, but it is also equally clear to us today that a system could be ergodic in this sense only if phase space were one dimensional. Plancherel (14) and Rosenthal (15) published proofs of this, and earlier, Poincare (16) had expressed doubts about Boltzmann's ergodic hypothesis. Certainly part of the problem Maxwell and Boltzmann faced was that the mathematics necessary for a proper discussion of the foundations of statistical mechanics, such as the measure theory of Borel and Lebesgue, and elements of modern topology had not been discovered until the first decade of the 20th century and were hence unavailable to them.

In their influential 1911 article, Ehrenfest and Ehrenfest (17) summarized and discussed problems with the ergodic hypothesis and then proposed instead the quasi-ergodic hypothesis as a replacement. This hypothesis states that some orbit of the flow will pass arbitrarily close to every point of phase space, or in other words this orbit is topologically dense in the phase space. This hypothesis is a far more plausible one than the old ergodic hypothesis, and it does imply that any continuous function invariant under the flow is constant. Some authors [von Plato (18)] have argued that, despite what Boltzmann had written down most of the time in his articles about the ergodic hypothesis, that he probably really meant something like what was later termed the quasi-ergodic hypothesis. However, the quasi-ergodic hypothesis does not imply metric transitivity. For instance, it is not even true that a minimal flow (every orbit is dense) with an invariant measure is metrically transitive [see Furstenberg (19)]. for examples. Therefore, although the original ergodic hypothesis was too strong to be plausible, the quasi-ergodic hypothesis was too weak to establish equality of time and phase averages. Further mathematical, progress had to await the concept of metric transitivity and the ergodic theorems of 1931 and 1932. For more details, see the survey article of Mackey (20).

One reaction to the Birkhoff and von Neumann ergodic theorems might be that they do not really solve the problem of equating time average and phase averages but only reduce it to a possibly equally difficult problem of proving metric transitivity. For instance, how can one prove that a one-parameter flow is metrically transitive and indeed how do you know metrically transitive systems exist at all. At this point, let us transfer to current terminology and simply call metrically transitive transformations ergodic, as von Neumann suggested.

As to the existence of ergodic transformations, Oxtoby and Ulam (21) showed that on a compact polyhedron $M$ equipped with a finite Lebesgue–Stieljes measure, the set of all ergodic measure preserving homeomorphisms is a dense $G_\delta$, or a residual set, among all measure preserving homeomorphisms. Hence, not only do ergodic maps exist, but almost all measure preserving homeomorphisms are ergodic in a topological sense. However, there is a different answer for Hamitonian dynamical systems. Here almost all systems are nonergodic, in the same sense as above (22). The space and the topology are different in these two cases so there is no contradiction.

von Neumann in his *Annals of Mathematics* paper (23) provides an intriguing and powerful answer to the existence problem. He shows that any one-parameter flow or any single transformation on a finite measure space can be written as a possibly continuous sum of ergodic ones. To make this precise, let $(M,\mu)$ be a finite measure space and form the product of $M$ with the unit interval $I$ with Lebesgue measure $I\times M=M^*$. Let $p$ be the projection to the first factor $I$ and denote the part of $M^*$ that lies over a point $r$ in $I$ by $M^r$. Then suppose that $P_t^r$ is an ergodic flow on $M^r$. Then assuming some regularity in $r$, one can piece together these $P_t^r$ into a measure preserving flow $P_t^*$ on $M^*$ by setting $P_t^*(r,m)=[r,P_t^r(m)]$. This flow is clearly nonergodic as it has many invariant sets, namely $J\times M$ for any measurable subset $J$ of $I$, but it is displayed as a continuous sum of ergodic ones, the $P_t^r$. von Neumann's theorem states that any general measure preserving flow has the structure like $P_t^*$ on $M^*$ in which it is displayed as the (possibly) continuous sum of its ergodic components $P_t^r$ on $M^r$.

Irrational rotations on a torus provide important examples of ergodic flows. We represent points on a 2D torus $T$ by pairs of complex numbers $(w,z)$ each of absolute value one, and we consider the flow defined by

$$P_t(w,z)=[\exp(it)w,\exp(itb)z],$$

where $b$ is irrational. All orbits are dense so this flow is quasi-ergodic, and it is also ergodic. To see this, assume there is an invariant measurable set $A$ and let $f$ be the indicator function of this set (1 on $A$ and 0 on its complement). Expand $f$ in a double Fourier series and use the invariance of $A$ under $P_t$ to see that all Fourier coefficients except for the constant term vanish. Therefore, $f$ is constant, and this establishes ergodicity.

An important set of examples for the subsequent development of ergodic theory is the shift transformations. Let $F$ be a finite set of $n$ elements and assign a probability measure to $F$; that is nonnegative numbers $p_1,\ldots,p_n$, whose sum is 1. Then form the infinite product $M$ of the measure space $F$ with itself using the integers as the index set. Thus, $M$ consists of doubly infinite sequences of points of $F$, $s=[\ldots,s(n-1),s(0),s(1),\ldots]$, with $s(i)$ in $F$. The shift transformation is defined by $P[s(n)]=s(n+1)$. $P$ has the effect of shifting a sequence $s$ one place. $P$ also preserves the product measure on $M$, and it is easily seen to be ergodic. Indeed any invariant set in the language of probability theory is statistically independent from the family of subsets of $F$ that are defined by any finite set of indices. Hence, it is a so-called tail event, and by the Borel–Cantelli Lemma of probability theory, it has probability zero or one. Hence, shifts are ergodic.

Specializing to the case $n=2$ and $p_1=p_2=1/2$, one sees that $M$ is the probability model of doubly infinite sequences of the results of tossing a fair coin. Letting $f(s)=1$ if $s(0)=1$ (heads) and $f(s)=0$ if $s(0)=0$(tails), the time average of $f$ in the ergodic theorem is

$$(1+N)^{-1}\sum_0^N f[P^n(s)],$$

which is simply the proportion of heads in the first $N+1$ coin tosses. The ergodic theorem says that this converges for almost all $s$ to the integral of $f$ over $M$, which is $1/2$. This is, of course, just the strong law of large numbers, a fundamental theorem in probability theory, and, therefore from the perspective of probability theory, the ergodic theorem emerges as a far-reaching generalization of the strong law of large numbers.

Another set of examples of significance for ergodic theory as well as statistical mechanics are geodesic flows, in particular geodesic flows on compact Riemannian manifolds of negative curvature. First, consider the 2D case of a surface. Geodesic flow take place on the unit tangent bundle $T$ of such a surface, which consists of pairs $(x,v)$, where $x$ is a point of the surface $M$ and $v$ is a unit tangent vector at $x$. Then $T$ consists of points of $M$ with a circle above each point consisting of unit tangent vectors, so $T$ is a 3D compact manifold. Geodesic flow $P_t$ on $T$ flows a point $(x,v)$ along the geodesic curve starting at $x$ with tangent vector $v$ a distance $t$ with $P_t(x,v)=(y,w)$, where $y$ is the endpoint of the geodesic of length $t$ and $w$ is its tangent vector at $y$. Hedlund (24) and Hopf (25) independently established that such flows were ergodic and also later extended the result to higher dimensional manifolds with negative curvature. A key part of the reasoning was the fact that negative curvature makes nearby geodesics diverge exponentially from each other so that the flow has sensitive dependence on initial conditions in that, if $u$ and $v$ are very close together, $P_t(u)$ and $P_t(v)$ will diverge away from each other exponentially in $t$. This property is called hyperbolicity, and it is known to be key in many proofs showing that flows or single transformations are ergodic. It is also a property that a system of colliding gas molecules will have. If one perturbs very slightly the position and momentum of a gas molecule, then in its next collision, it will bounce off the other molecule at a quite different direction, a situation that is repeated in further collisions, resulting in rapid divergence of its path from the path of the unperturbed molecule.

There is a substantial and fascinating amount of papers in ergodic theory pursuing a program using hyperbolicity to help prove that various approximations to a system of a volume of gas molecules confined in a container are ergodic. The program has not succeeded as yet in proving that an actual model of a gas is ergodic, but the results are getting close. The first paper was the path breaking paper of Sinai (26). Sinai redefined the problem from one where the gas molecules are contained in a cubical enclosure with reflecting walls to a model where the particles move in a cube with periodic boundary conditions. The molecules collide with each other but not with the walls. This is not physically realistic, but it can shed light on the original problem by analogy. Much progress has been made on this model, especially by Simanyi and Szasz (27) and Simanyi (28), where, using hyperbolocity as indicated above, a nearly complete resolution of ergodicity issues is established for such Sinai flows [see also the survey article by Szasz (29)].

Going back to the more realistic case of molecules in a container with reflecting walls, one should begin in dimension two with a single ball—a billiard flow in a planar region or table. If the table is a rectangle, the momenta of the ball over time can take on only a finite number of values so the flow on a phase space of fixed energy cannot be ergodic. However, if one has a billiard table of more complex geometry, the situation becomes more interesting. For instance, if the billiard table is a polygon, then Kerckhoff et al. (30) show that for topologically almost all polygons, the billiard flow on a phase space of constant energy is ergodic. In particular, one has to stay away from rational polygons where all of the angles are rational multiples of $\pi$. The authors point out an interesting corollary of this result, which is that a mechanical system of two particles of masses $m_1$ and $m_2$ moving along a finite track without friction, bouncing off each other elastically and bouncing off the fixed end points is ergodic on a phase space of constant energy for topologically almost all values of the ratio $m_1/m_2$. This is sort of like a one-dimensional gas.

Finally, Simanyi (31) has established that a system consisting of two hard spheres contained in a cube of any dimension at least two bouncing off each other and off the hard walls is ergodic on any surface of constant energy. This appears to be a rigorous ergodicity result in a situation that comes closest to an actual gas. However, this whole array of theorems, of which we have mentioned only a few, suggest that the hypothesis of ergodicity (or metric transitivity) for a physical system like that of gas dynamics mentioned at the outset of this essay is very plausible.


**1** Birkhoff GD (1931) Proof of the ergodic theorem. *Proc Natl Acad Sci USA* 17(12):656–660.

**2** von Neumann JV (1932) Proof of the quasi-ergodic hypothesis. *Proc Natl Acad Sci USA* 18(1):70–82.

**3** Maxwell JC (1879) On Boltzmann's theorem on the average distribution of energy in a system of material points. *Trans Camb Phil Soc* 12:547–575.

**4** Boltzmann L (1887) Über die mechanischen Analogien des zweiten Hauptsatzes der Thermodynamik. *J Mathematik* C(2): 201–212.

**5** Gibbs JW (1902) *Elementary Principles in Statistical Mechanics* (Scribners, New York).

**6** Maxwell JC (1860) Illustrations of the dynamical theory of gases. *Philos Mag* 19(1):19–51.

**7** Boltzmann L (1868) Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten, *Sitzungsberichte d.k. Akad Wien* 58:517–563.

**8** Birkhoff GD, Koopman BO (1931) Recent contributions to the ergodic theory. 18(1):279–283.

**9** Stone MH; Operational Methods and Group Theory (1930) Linear transformations on Hilbert space III. Operational methods and group theory. *Proc Natl Acad Sci USA* 16(2):172–175.

**10** Koopman BO (1931) Hamiltonian systems and transformations in Hilbert space. *Proc Natl Acad Sci USA* 17(5):315–318.

**11** Birkhoff GD, Smith PA (1928) Structure analysis of surface transformation. *J Math Pures Appl* 9(9):345–379.

**12** Mackey GW (1990) Von Neumann and the early days of ergodic theory. *Proc Symp Pure Math* 50:25–38.

**13** Hopf E (1932) On the time average in dynamics. *Proc Natl Acad Sci USA* 18(1):93–100.

**14** Plancherel M (1913) Beweis der Unmöglichkeit ergodisher mechanische Systeme. *Ann Physik* 42(5):1061–1063.

**15** Rosenthal A (1913) Beweis der Unmöglichkeit ergodischer Gassysteme. *Ann Physik* 42(4):796–806.

**16** Poincaré H (1894) Théorie Cinétique des Gaz, Rev. gen. des sciences pures et appl. (5):513–521.

**17** Ehrenfest P, Ehrenfest T (1911) *Begriffliche Grundlagen der statistischen Auffassung in der Mechanik Enc Math. Wiss* Vol 4, no 6 (Teubner, Leipzig, Germany); trans Moravcsik M (1959) [*The Conceptual Foundations of the Statistical Approach in Mechanics*] (Cornell Univ Press, Ithaca, NY).

**18** Von Plato J (1991) Boltzmanns ergodic hypthesis. *Arch Hist Exact Sci* 42(1):71–89.

**19** Furstenberg H (1961) Strict ergodicity and transformations of the torus. *Am J Math* 83(3):573–601.

**20** Mackey GW (1974) Ergodic theory and its significance for statistical mechanics and probability theory. *Adv Math* 12(2): 178–268.

**21** Oxtoby J, Ulam S (1941) Measure-preserving homeomorphisms and metrical transitivity. *Ann Math* 42(4):874–920.

**22** Markus L, Meyer K (1974) Generic Hamiltonian systems dynamical systems are neither integrable nor ergodic. *Mem Am Math Soc* 144.

**23** von Neumann J (1932) Zur Operator methoden in der Klassischen Mechanik. *Ann Math* 33(3):587–642.

**24** Hedlund GA (1939) The dynamics of geodesic flows. *Bull Am Math Soc* 45(4):241–260.

**25** Hopf E (1939) Statistik der geodätische Linen in Manigfaltigkeiten negativer Krümmung. *Ber Verh Sachs*. *Akad Leipzig* 91:261–304.

**26** Sinai Ya (1970) Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. *Uspehi Mat Nauk* 25(2): 141–192.

**27** Simanyi N, Szasz D (1991) Hard ball systems are completely hyperbolic. *Ann Math* 149(1):35–96.

**28** Simanyi N (2002) The complete hyperbolicity of cylindric billiards. *Ergodic Theory Dynam Systems* 22(1):281–302.

**29** Szasz D (1996) Boltzmann's ergodic hypothesis, a conjecture for centuries. *Studia Sci Math Hungar* 31(1-3):299–322.

**30** Kerckhoff S, Masur H, Smilie J (1986) Ergodicity of billiard flows and quadratic differentials. *Ann Math* 124(2):293–311.

**31** Simanyi N (1999) Ergodicity of hard spheres in a box. *Ergodic Theory Dynam Systems* 22(3):741–766.