Journal of Statistical Physics, Vol. 99, Nos. 5/6, 2000

Macroscopic Determinism in Noninteracting Systems Using Large Deviation Theory

Brian R. La Cour[1,2] **and William C. Schieve**[1]

Received September 24, 1999; final January 11, 2000

We consider a general system of n noninteracting identical particles which evolve under a given dynamical law and whose initial microstates are *a priori* independent. The time evolution of the n-particle average of a bounded function on the particle microstates is then examined in the large-n limit. Using the theory of large deviations, we show that if the initial macroscopic average is constrained to be near a given value, y, then the macroscopic average at time t converges in probability as $n \to \infty$ to a value $\psi_t(y)$ given explicitly in terms of a canonical expectation. Some general features of the graph of $\psi_t(y)$ versus t are examined, particularly in regard to continuity, symmetry, and convergence.

KEY WORDS: determinism; causality; large deviation theory; many-particle systems; fluctuations; nonequilibrium statistical mechanics; kinetic theory.

1. INTRODUCTION

The emergence of determinism in the macroscopic variables of a system from its underlying microscopic dynamics has been a subject of great importance in the field of statistical mechanics. If one supposes the microscopic variables are rigidly deterministic, then the time evolution of the macroscopic variables as a function of the initial *microstate* is of course rigidly deterministic as well. Considered as a function of the initial *macrostate*, however, its time evolution fails to be deterministic due to the multiplicity of microstates consistent with the given macrostate. This leads to the familiar ensemble description, as described by a conditional *a priori* measure. Given the highly irregular behavior of many macroscopic systems on the

[1] Ilya Prigogine Center for Studies in Statistical Mechanics and Complex Systems and Department of Physics, University of Texas at Austin, Austin, Texas 78712.
[2] To whom correspondence should be addressed; e-mail: blacour@physics.utexas.edu.

0022-4715/00/0600-1225$18.00/0

microscopic level, however, it may appear that all deterministic behavior is utterly lost on the macroscopic level.

By contrast, many macroscopic systems are described quite well by deterministic models, even if they are not deterministic in the strict sense. Well known examples include thermal conduction, diffusion, hydrodynamics, and chemical reactions. Characteristic of many such systems is the presence of extensive macroscopic variables which are sums or averages over a great many microscopic quantities. The task of deriving differential equations for such variables has been taken up by several researchers,[1–3] and here Markov processes have played an important role. In this approach, one derives a coupled set of differential equations for the moments of the macroscopic variable, where the deterministic behavior is given by the first moment when the dispersion goes to zero. Since few physical observables are truly Markovian, delicate scaling between vanishing interactions and dilating time scales must be used to obtain an asymptotically Markovian process.[4, 5] However, the consistency of such assumptions with the underlying microscopic dynamics and the relevance of strong interactions have been repeatedly called into question.[6, 7] In particular, the relaxation of macroscopic systems to a state of equilibrium may seem inconsistent with the underlying reversible microscopic dynamics.

We take a somewhat different and more general view upon this problem. Instead of attempting to derive differential equations of motion for a class of macroscopic variables, we consider instead the existence and character of an emergent form of determinism for large systems, which we call macroscopic determinism. By this we mean that if the macrostate is initially constrained to be near a given value, y, then there exists a map ψ_t such that the probability that the macrostate is near $\psi_t(y)$ at time t approaches one as the number of particles approaches infinity. For simplicity, we restrict attention to systems of dynamically noninteracting particles and macroscopic variables which take the form of averages over these microscopic quantities. The mathematical theory of large deviations, an extension of the law of large numbers, provides an useful tool for addressing such questions and is used in Section 3 to obtain the main result. Our primary contribution has been to apply well-known equilibrium results to systems initially out of equilibrium.

The collection of macrostates $\{\psi_t(y): t \in T\}$ for a given y and set of times T constitutes a collection of highly probable states, which we call the deterministic curve, akin to the concentration curve of P. and T. Ehrenfest in their discussion of Boltzmann's H-theorem.[8] In Section 4 we show that the macroscopic average converges to this curve in probability for any given finite, and in some cases infinite, set of times, though large deviations from this curve will persist whenever the number of particles is finite. For certain reversible microscopic dynamics which preserve the *a priori* measure, the deterministic curve of a restricted class of macroscopioc variables is symmetric in time about the initial specification of the macroscopic variable. Furthermore, if the microscopic dynamical law is mixing, then the deterministic curve converges as $t \to \pm\infty$ to a value independent of y, a situation corresponding to equilibration of the macrostate.

2. PROBLEM DESCRIPTION

We begin with a mathematical description of the relevant physical quantities. The microstate space of a single particle is denoted by X, while the microstate space of the total n-particle system is given by the Cartesian product $X^n = X \times \overset{n}{\cdots} \times X$. The projection map π_i is defined such that, if $(x_1,..., x_n) \in X^n$ is the microstate of the system, then $\pi_i(x_1,..., x_n) = x_i$ is the microstate of the ith particle. To apply the techniques of large deviation theory we shall need to make the mild assumption that (X, d) is a Polish space, i.e., a complete separable metric space, with respect to some given metric d, and denote by $\mathscr{A}$ the set of Borel subsets of X generated by the metric topology.

Let $\mathscr{P}(X)$ denote the set of Borel probability measures on X and note that $\mathscr{P}(X)$ is itself a Polish space under the Prohorov metric, ρ, induced by the metric d on X [9, p. 317]. Denote by $\mu \in \mathscr{P}(X)$ the *a priori* probability measure for the initial microstate of a given particle. In other words, for $C \in \mathscr{A}$, $\mu[C]$ is the probability that a given particle's initial microstate is in $C \subseteq X$, *before* any conditioning on the macrostate has taken place. All particle microstates therefore have equal *a priori* probability in the sense that they are uniformly distributed with respect to μ. Usually, μ is taken to be an invariant measure. The system microstates are assumed to be *a priori* independent and identically distributed (i.i.d.) with marginal μ. Thus, the *a priori* distribution on X^n is given by the product measure $\mu^n = \mu \times \overset{n}{\cdots} \times \mu$.

Let $g: X \to Y$ be a measurable function which is bounded and continuous μ-almost everywhere (a.e.). For a given particle microstate $x_i \in X$, $g(x_i)$ gives the corresponding single particle macrostate. For a collection of n particles, the macroscopic average $G: X^n \to Y$ of g is given by

$$G := \frac{1}{n} \sum_{i=1}^{n} g \circ \pi_i \tag{1}$$

It is to this macroscopic variable that we will focus our attention. We shall take Y to be a set of real numbers equipped with the Euclidean norm. To accommodate all possible values of G as well as any accumulation points, we shall assume $Y = [y_{\min}, y_{\max}]$, where $y_{\min} = \inf g(X)$, $y_{\max} = \sup g(X)$, and $g(X)$ is the image of X under g. Note that G is indeed a measurable function since it is a finite sum of measurable functions.

The dynamics of the system are described by a family of measurable transformations Φ_t on X^n indexed by the time parameter $t \in T \subseteq \mathbb{R}$, where Φ_0 is the identity map. The macroscopic average at time t is thus given by $G_t := G \circ \Phi_t$. We shall suppose that the particles are dynamically independent and identical in the sense that

$$\Phi_t = (\varphi_t \circ \pi_1, ..., \varphi_t \circ \pi_n) \tag{2}$$

where φ_t is a measurable transformation on X. We shall further suppose that φ_t is continuous μ-a.e. on X for each $t \in T$, though not necessarily continuous on T for μ-a.e. $x \in X$.

The macrostate G_t may be considered a sample mean over the *a priori* i.i.d. random variables $\{\varphi_t \circ \pi_i\}_{i=1}^n$ with common marginal $\mu \circ (g \circ \varphi_t)^{-1}$. Since g is bounded, the weak law of large numbers implies G_t converges in probability to the expectation value $\int_X g \circ \varphi_t \, d\mu$ as $n \to \infty$; in other words,

$$\lim_{n \to \infty} \mu^n[G_t \in B] = \begin{cases} 0, & \text{if } \int_X g \circ \varphi_t \, d\mu \notin \bar{B}, \\ 1, & \text{if } \int_X g \circ \varphi_t \, d\mu \in B^\circ \end{cases} \tag{3}$$

for any Borel set $B \subseteq Y$, where B° is the interior of B and $\bar{B}$ is its closure. No limit is specified for points on the boundary, ∂B, of B.

If the initial microstates are restricted so as to satisfy some initial macroscopic constraint, then the variables $\{\varphi_t \circ \pi_i\}_{i=1}^n$ may no longer be independent due to the correlations imposed by this conditioning. A direct application of the law of large numbers is therefore no longer valid. Suppose, in particular, that the initial macrostate $G_0 = G$ is constrained to be in a small region B_δ containing a given macrostate y. We will show that G_t converges in conditional probability to some $\psi_t(y)$; in other words,

$$\lim_{n \to \infty} \mu^n[G_t \in B \mid G \in B_\delta] = \begin{cases} 0, & \text{if } \psi_t(y) \notin \bar{B}, \\ 1, & \text{if } \psi_t(y) \in B^\circ \end{cases} \tag{4}$$

where $\psi_t(y)$ is the expectation value of $g \circ \varphi_t$ with respect to a new probability measure determined by y. Of course, if $y = y_* := \int_X g \, d\mu$, the *a priori* expectation, then this result clearly holds with $\psi_t(y_*) = y_*$.

If we take $y = y_{\min}$ and suppose $\mu[\{g = y_{\min}\}] > 0$, then the law of large numbers may in fact be applied since $\{G = y_{\min}\} = \{g = y_{\min}\} \times \overset{n}{\cdots} \times \{g = y_{\min}\}$ implies

$$\mu^n[A_1 \times \cdots \times A_n \mid \{G = y_{\min}\}] = \prod_{i=1}^n \mu[A_i \mid \{g = y_{\min}\}] \tag{5}$$

for any $A_1,..., A_n$ in $\mathscr{A}$. Thus, conditioned on $\{G = y_{\min}\}$, G_t is a sample mean of i.i.d. random variables with common marginal $\mu[(g \circ \varphi_t)^{-1}(\cdot) \mid \{g = y_{\min}\}]$. The weak law of large numbers then implies

$$\lim_{n \to \infty} \mu^n[G_t \in B \mid \{G = y_{\min}\}] = \begin{cases} 0, & \text{if} \quad \psi_t(y_{\min}) \notin \bar{B}, \\ 1, & \text{if} \quad \psi_t(y_{\min}) \in B^\circ \end{cases} \tag{6}$$

for any Borel set $B \subseteq Y$, where

$$\psi_t(y_{\min}) := \int_X g \circ \varphi_t \, d\mu[\cdot \mid \{g = y_{\min}\}] \tag{7}$$

A similar result holds for conditioning on $\{G = y_{\max}\}$, provided $\mu[\{g = y_{\max}\}] > 0$.

For $y \in (y_{\min}, y_{\max}) = Y^\circ$ we employ a different approach using large deviation theory to study the behavior as $n \to \infty$ of the empirical measures $L_n\colon X^n \to \mathscr{P}(X)$ defined by

$$L_n(x_1,..., x_n) := \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \tag{8}$$

Using the fact that $G = E_g \circ L_n$, where

$$E_g(P) := \int_X g \, dP \qquad \text{for} \quad P \in \mathscr{P}(X) \tag{9}$$

we may then deduce convergence properties of G_t from those of L_n. Note that, since g is bounded and continuous μ-a.e., E_g is continuous at P (in the weak topology) for any $P \prec \mu$.[9, 10] As will be shown, by allowing g to be discontinuous at most on a set of measure zero we are able to incorporate both continuous observables and physically relevant discrete-valued observables within a single framework using the standard weak topology on $\mathscr{P}(X)$. (One may obtain similar results using the stronger τ-topology,[11] for which E_g is continuous everywhere by definition.)

3. LARGE DEVIATION THEORY APPROACH

We have seen in the previous section that the weak law of large numbers is insufficient for evaluating limiting conditional probabilities for an arbitrary choice of y. Large deviation theory provides a tool for evaluating such limits and, through its application, provides an explicit expression for $\psi_t(y)$. This approach is a refinement of the weak law of large numbers for cases in which the probabilities converge exponentially fast at a rate given in terms of a function I, the so-called "rate function." The ground work for this theory was established by Boltzmann[12] in his study of the asymptotic properties of multinomials and later applied by Einstein[13] in his analysis of fluctuations. Recent years have seen great development of this relatively new field of mathematical probability, including applications in equilibrium statistical mechanics, stochastic processes, and mathematical statistics.[14, 15, 11, 16] Ruelle[17] and Lanford[18] have developed similar techniques for studying the equilibrium distributions of dynamical maps, where the (negative) Kolmogorov–Sinai entropy serves as a rate function.[19]

In Section 3.1 we define rate functions and the large deviation principle. The main result is Theorem 1 regarding convergence in probability for conditional probabilities. In Section 3.2 we take up the notion of Gibbs Conditioning, which will allow us to apply Theorem 1 to cases in which we condition on an arbitrary initial macrostate y. Our results are similar to those of Dembo and Zeitouni,[11] who consider Gibbs conditioning using the stronger τ-topology on $\mathscr{P}(X)$, and follow from an approach adapted from Ellis,[20] who considers discrete macrostates. Our final results are somewhat more general in that g may be any bounded, μ-a.e. continuous function and Eq. (16) holds for any fixed $\delta > 0$.

3.1. Large-Deviation Theory

For a topological space $(X, \mathscr{T})$, a function I mapping X into $[0, \infty]$ is called a *rate function* iff $\inf I(X) = 0$ and I is lower semicontinuous, i.e., the preimage $I^{-1}([0, \alpha])$ is closed for all $\alpha \in [0, \infty)$. A *good* rate function is one in which these preimages are compact. The property of being lower semicontinous guarantees that I attains its infimum on any compact set, from which it follows that a good rate function attains its infimum over any closed set [11, pp. 4, 308]. Any point, x_*, at which $I(x_*) = 0$ is called an *equilibrium point*, since it corresponds to a state of maximum probability.

A sequence $(P_n)_{n \in \mathbb{N}}$ of probability measures on the Borel subsets of X is said to satisfy a *large deviation principle* with rate function I iff there exists a sequence $(a_n)_{n \in \mathbb{N}}$ of positive numbers tending to infinity such that, for any Borel set $A \subseteq X$,

$$\begin{aligned} -\inf I(A^\circ) &\leqslant \liminf_{n \to \infty} \frac{1}{a_n} \log P_n[A] \\ &\leqslant \limsup_{n \to \infty} \frac{1}{a_n} \log P_n[A] \\ &\leqslant -\inf I(\bar{A}) \end{aligned} \tag{10}$$

A set A for which $\inf I(A^\circ) = \inf I(\bar{A})$ is called an *I-continuity set*. Clearly for such sets we have

$$\lim_{n\to\infty} \frac{1}{a_n} \log P_n[A] = -\inf I(A) \tag{11}$$

This result should be compared with the Boltzmann relation, $S = k_{\mathrm{B}} \log W$, relating the thermodynamic entropy, S, to a volume, W, in phase space, where k_{B} is Boltzmann's constant. The quantity $-a_n I$ serves as an entropy, while P_n may be viewed as a normalized volume measure.

If $0 < \inf I(A) < \infty$, Eq. (11) implies that $P_n[A]$ converges to zero at least exponentially fast. More generally, from Eq. (10) we may deduce the following: Given any Borel set $A \subseteq X$ and $\varepsilon > 0$, then for all n sufficiently large

$$\exp[-a_n(1+\varepsilon) \inf I(A^\circ)] \leqslant P_n[A] \leqslant \exp[-a_n(1-\varepsilon) \inf I(\bar{A})] \tag{12}$$

provided $\inf I(A^\circ) > 0$ and $\inf I(\bar{A}) < \infty$. Equality for the lower bound may hold only if $\inf I(A^\circ) = \infty$, where $e^{-\infty} := 0$, while equality for the upper bound may hold only if $\inf I(\bar{A}) = 0$.

Conditioning on a set, B, of nonzero probability may give rise to a new equilibrium point, $x_B \in \bar{B}$, different from x_*. (Of course, x_B will equal x_* if $x_* \in B^\circ$.) The following theorem, which falls short of a general conditional large deviation principle, gives sufficient conditions for convergence in conditional probability to such a point. The proof is deferred to Appendix A.

Theorem 1. Suppose $(P_n)_{n\in\mathbb{N}}$ satisfies a large deviation principle with a good rate function I and a unique equilibrium point x_*. Let $B \subseteq X$ be an I-continuity set such that $x_* \notin \partial B$ and there exists a unique $x_B \in \bar{B}$ such that $I(x_B) = \inf I(\bar{B}) < \infty$. Given any Borel set $A \subseteq X$,

$$\lim_{n\to\infty} P_n[A \mid B] = \begin{cases} 0, & \text{if} \quad x_B \notin \bar{A}, \\ 1, & \text{if} \quad x_B \in A^\circ \end{cases} \tag{13}$$

In the next section, we shall consider a large deviation principle for the sequence $(\mu^n \circ L_n^{-1})_{n\in\mathbb{N}}$ of distributions of empirical measures. Although $L_n \to \mu$ almost surely [9, p. 313], conditioning on $E_g^{-1}(B_\delta)$ gives rise to a new equilibrium probability measure, P_λ, which is canonical in form. We shall show that under these conditions Theorem 1 is satisfied, thus establishing convergence of L_n in conditional probability to P_λ.

3.2. Gibbs Conditioning

In his 1877 paper, Boltzmann proved that the asymptotically most probable configuration for a gas of n particles with a finite number of macrostates is given by a multinomial distribution. Sanov's theorem [14, p. 70], a modern refinement of this classic result, states that the sequence $(\mu^n \circ L_n^{-1})_{n \in \mathbb{N}}$ of distributions of empirical measures satisfies a large deviation principle with rate function $I_\mu: \mathscr{P}(X) \to [0, \infty]$. Here $I_\mu(P)$ is the (negative) Gibbs entropy of P with respect to μ defined by

$$I_\mu(P) := \begin{cases} \displaystyle\int_X \frac{\mathrm{d}P}{\mathrm{d}\mu} \log \frac{\mathrm{d}P}{\mathrm{d}\mu} \, \mathrm{d}\mu, & \text{if} \quad P \prec \mu, \\ \infty & \text{otherwise} \end{cases} \tag{14}$$

where $0 \log 0 := 0$. It can be shown that I_μ is a good, strictly convex rate function [11, p. 240] which attains its infimum uniquely at μ [16, pp. 32–34].

Given y and $\delta > 0$, let $A_\delta = E_g^{-1}(B_\delta)$, where, $B_\delta = (y - \delta, y]$ when $y < y_*$, $B_\delta = [y, y + \delta)$ when $y > y_*$, and $B_\delta = (y_* - \delta, y_* + \delta)$ when $y = y_*$. (Note that $\mu \notin \partial A_\delta$, since $y_* \notin \partial B_\delta$ and E_g is continuous at μ.) We wish to consider the asymptotic behavior of the conditional probability $(\mu^n \circ L_n^{-1})[A \mid A_\delta] = \mu^n[\{L_n \in A\} \mid \{L_n \in A_\delta\}]$, as $n \to \infty$, for any Borel set $A \subseteq \mathscr{P}(X)$. For example, if G is the macroscopic average energy of the system, then $\mu^n[\cdot \mid \{G \in B_\delta\}]$ is the microcanonical distribution on the "thickened" energy shell with energy y.

We will show that, conditioned on $\{G \in B_\delta\}$, the empirical measures $(L_n)_{n \in \mathbb{N}}$ converge in probability to the canonical Gibbs measure P_λ, where λ satisfies the constraint $y = \int_X g \, \mathrm{d}P_\lambda$ and

$$\mathrm{d}P_\lambda(x) := \frac{\mathrm{e}^{\lambda g(x)}}{Z(\lambda)} \, \mathrm{d}\mu(x) \tag{15}$$

The normalization factor $Z(\lambda) := \int_X \mathrm{e}^{\lambda g} \, \mathrm{d}\mu$ is the *partition function*, and the quantity $\Psi(\lambda) := \log Z(\lambda)$ is the *generalized free energy*. Note that, if G is the macroscopic average energy, then $-k_{\mathrm{B}} T \, \Psi(-1/(k_{\mathrm{B}} T))$ is the familiar Helmholtz free energy at temperature T.

Denote by θ the map $\theta: [-\infty, +\infty] \to Y$ which associates a given λ with a certain value of y and is defined by $\theta(\lambda) := \int_X g \, \mathrm{d}P_\lambda$ for $\lambda \in \mathbb{R}$, $\theta(-\infty) := y_{\min}$, and $\theta(+\infty) := y_{\max}$. The following lemmas will be needed. The proofs are deferred to Appendix B.

Lemma 1. If $\Psi''(\lambda)$ is nonzero for all $\lambda \in (-\infty, +\infty)$, then the map θ is one-to-one. Furthermore, $y < y_*$ iff $\lambda < 0$, $y = y_*$ iff $\lambda = 0$, and $y > y_*$ iff $\lambda > 0$.

Lemma 2. Let $A_\delta = E_g^{-1}(B_\delta)$ and $\lambda = \theta^{-1}(y)$. Then $I_\mu(P_\lambda) = \inf I_\mu(\overline{A_\delta}) < \infty$ and $I_\mu(P_\lambda) < I_\mu(P)$ for all $P \in \overline{A_\delta} \setminus \{P_\lambda\}$, where $\lambda = \theta^{-1}(y)$.

Lemma 3. The set $A_\delta = E_g^{-1}(B_\delta)$ is an I_μ-continuity set.

Using the above lemmas we may deduce that $A_\delta = E_g^{-1}(B_\delta)$ is such that $\mu \notin \partial A_\delta$ and, for $\lambda = \theta^{-1}(y)$, P_λ is the unique measure in $\overline{A_\delta}$ such that $I_\mu(P_\lambda) = \inf I(\overline{A_\delta}) = \inf I(A_\delta^\circ) < \infty$. By Theorem 1 we conclude that for any Borel set $A \subseteq \mathscr{P}(X)$,

$$\lim_{n \to \infty} \mu^n[\{L_n \in A\} \mid \{L_n \in A_\delta\}] = \begin{cases} 0, & \text{if } \ P_\lambda \notin \bar{A}, \\ 1, & \text{if } \ P_\lambda \in A^\circ \end{cases} \tag{16}$$

In particular, for any Borel set $B \subseteq Y$,

$$\lim_{n \to \infty} \mu^n[\{G_t \in B\} \mid \{G \in B_\delta\}] = \begin{cases} 0, & \text{if } \ \psi_t(y) \notin \bar{B}, \\ 1, & \text{if } \ \psi_t(y) \in B^\circ \end{cases} \tag{17}$$

where

$$\psi_t(y) := \int_X g \circ \varphi_t \, \mathrm{d}P_{\theta^{-1}(y)} \tag{18}$$

Note that $\psi_0(y) = y$ since φ_0 is the identity on X. Conditioning on $\{G_{t_0} \in B_\delta\}$ merely shifts the time axis, in which case G_t converges to $\psi_{t-t_0}(y)$ in probability.

4. THE DETERMINISTIC CURVE

We have shown that, conditioned on $G \in B_\delta$, the macrostate G_t at a given time, t, converges in probability to the expectation value $\psi_t(y)$, where y is contained in B_δ. Hence, of the initial microstates consistent with the initial macrostate, y, "most" will be such that the actual macrostate realized will be near this value. Now, each microstate, $(x_1,..., x_n)$, gives rise to a collection, $\{G_t(x_1,..., x_n): t \in T\}$, of macrostates constituting a single trajectory. Likewise, each macrostate, y, gives rise to a collection, $\{\psi_t(y): t \in T\}$, of expected macrostates, which we shall call the *deterministic curve*. This graph represents the asymptotically deterministic behavior of the macrostate. In this section we shall investigate in what sense the deterministic curve is representative of a typical trajectory and consider several properties of the deterministic curve itself, considered as a function of time.

In general, there may be striking qualitative differences between the deterministic curve and a particular macrostate trajectory. Suppose μ is invariant under φ_t and $\mu^n[G^{-1}(B_\delta)]>0$. The Poincaré recurrence theorem tells us that for some unbounded sequence $\tau_1, \tau_2,\ldots$ of times

$$\mu^n[\{G_{\tau_1}\in B_\delta, G_{\tau_2}\in B_\delta,\ldots\} \mid \{G\in B_\delta\}]=1 \tag{19}$$

i.e., the macrostate returns infinitely often to a neighborhood of its initial value, y, almost surely. Now suppose the map $t\mapsto\psi_t(y)$ has an attracting set A with domain of attraction D and take $B_\delta\subseteq D\backslash U$, where U is a neighborhood of A. For all t sufficiently large, $\psi_t(y)$ will be in U, but G_t will almost surely fall outside U on the recurrence times $\tau_1, \tau_2,\ldots$. These recurrence times will depend upon B_δ and φ_t, of course, but typically increase rapidly with n. On a time scale small compared to τ_1, one then expects the deterministic curve to be quite representative of a typical trajectory. On time scales larger than τ_1, however, the deterministic curve will be qualitatively quite different from a typical trajectory; the former converges to an attracting set while the latter exhibits quasi-periodic behavior. This highlights the importance of a clearer understanding of the correspondence between the very distinct $n<\infty$ and $n=\infty$ cases.

4.1. Convergence to the Deterministic Curve

Consider a finite set, $\{t_1,\ldots, t_m\}$, of times and for each time t_i let $B_i\subseteq Y$ be a Borel set. The set of microstates in which $G_{t_i}\in B_i$ for all i is given by

$$\bigcap_{i=1}^{m}\{G_{t_i}\in B_i\}=\left\{L_n\in\bigcap_{i=1}^{m}A_i\right\}$$

where $A_i=E^{-1}_{g\circ\varphi_{t_i}}(B_i)$. According to Eq. (16), with $A=\bigcap_{i=1}^m A_i$, the limiting conditional probability will be zero if $P_\lambda\notin\overline{\bigcap_{i=1}^m A_i}$, where $\lambda=\theta^{-1}(y)$, and it will be one if $P_\lambda\in(\bigcap_{i=1}^m A_i)^\circ$. Since the intersection is over a finite number of sets, we may use the fact that $(\bigcap_{i=1}^m A_i)^\circ=\bigcap_{i=1}^m A_i^\circ$ and $\overline{\bigcap_{i=1}^m A_i}\subseteq\bigcap_{i=1}^m\overline{A_i}$ to obtain the following:

$$\lim_{n\to\infty}\mu^n\left[\bigcap_{i=1}^{m}\{G_{t_i}\in B_i\}\,\middle|\,\{G\in B_\delta\}\right]=\begin{cases}0, & \text{if}\quad \psi_{t_j}(y)\notin\overline{B}_j \text{ for some } j,\\ 1, & \text{if}\quad \psi_{t_i}(y)\in B_i^\circ \text{ for all } i\end{cases} \tag{20}$$

For a countably infinite set of times, we may again conclude that the conditional probability goes to zero if $\psi_{t_j}(y)\notin\overline{B}_j$ for some value of j, since $\bigcap_{i=1}^\infty\{G_{t_i}\in B_i\}\subseteq\{G_{t_j}\in B_j\}$. However, even if $\psi_{t_i}(y)\in B_i^\circ$ for all i and hence $P_\lambda\in\bigcap_{i=1}^\infty A_i^\circ$, this does not assure us that $P_\lambda\in(\bigcap_{i=1}^\infty A_i)^\circ$. For the latter to be true, there must be a neighborhood of P_λ which is sufficiently small so that it is contained in every A_i°. Let us consider, then, conditions for which this is true.

Suppose each B_i is an open interval of radius δ' centered at $\psi_{t_i}(y)$. Now, each corresponding A_i is an open set of probability measures which give an expectation of $g \circ \varphi_{t_i}$ in the open ball B_i. Loosely speaking, if $g \circ \varphi_{t_i}$, and hence $E_{g \circ \varphi_{t_i}}$, varies too rapidly, then A_i will be small as a result. If, therefore, the functions $\{g \circ \varphi_{t_i}\}_{i \in \mathbb{N}}$ are somehow limited in how rapidly they may vary, then one might expect the sizes of $A_1, A_2,...$ to be bounded from below, thus giving a nonempty interior for their intersection. For a general metric space, (X, d), one way to characterize how rapidly a function, f, may vary by its Lipschitz norm, $\|f\|_{\mathrm{L}}$, define as follows:

$$\|f\|_{\mathrm{L}} := \sup_{x \neq x'} \frac{|f(x) - f(x')|}{d(x, x')} \tag{21}$$

Differentiable functions will be Lipschitz if they have bounded derivatives; discontinuous functions, such as indicator functions, have infinite Lipschitz norm provided $d(x, x')$ may be made arbitrarily small.

To ensure well-defined expectations we require the functions to be bounded as well, so consider instead the bounded Lipschitz norm, $\|\cdot\|_{\mathbf{BL}}$, defined simply by $\|f\|_{\mathbf{BL}} := \|f\|_{\mathrm{L}} + \|f\|_\infty$. (For a discussion of this norm, see Dudley[9]). A set, $\{f_i\}_{i \in \mathbb{N}}$, of functions will be called *uniformly bounded Lipschitz* if there exists a number, K, such that $\|f_i\|_{\mathbf{BL}} \leqslant K$ for all $i \in \mathbb{N}$. The following theorem states that for such observables the macroscopic trajectories converge in conditional probability to the deterministic curve on any countable set of times.

Theorem 2. If $\{g \circ \varphi_{t_i}\}_{i \in \mathbb{N}}$ are uniformly Lipschitz bounded functions, then for any $t_1, t_2,...,$

$$\lim_{n \to \infty} \mu^n[\{|G_{t_i} - \psi_{t_i}(y)| < \delta', \forall i \in \mathbb{N}\} \mid \{G \in B_\delta\}] = 1 \tag{22}$$

As noted above, this result may be proven by finding an open ball, $B_\rho(P_\lambda, \varepsilon)$, about P_λ which is contained in every A_i. Consider an arbitrary probability measure, P, in $B_\rho(P_\lambda, \varepsilon)$ and observe that, since we have assumed $\|g \circ \varphi_{t_i}\|_{\mathbf{BL}} \leqslant K$,

$$\begin{aligned}
|E_{g \circ \varphi_{t_i}}(P) - \psi_{t_i}(y)| &\leqslant \sup\{|E_f(P) - E_f(P_\lambda)| \colon \|f\|_{\mathbf{BL}} \leqslant K\} \\
&= K \sup\{|E_f(P) - E_f(P_\lambda)| \colon \|f\|_{\mathbf{BL}} \leqslant 1\} \\
&\leqslant 2K\rho(P, P_\lambda) < 2K\varepsilon
\end{aligned}$$

where the last inequality follows from [9, pp. 310, 322], since (X, d) is a separable metric space. Taking $\varepsilon \leqslant \delta'/(2K)$ shows that $E_{g \circ \varphi_{t_i}}(P) \in B_i$ for all $P \in B_\rho(P_\lambda, \varepsilon)$ and hence that $B_\rho(P_\lambda, \varepsilon) \subseteq A_i$. Since ε is independent of i, we conclude that $P_\lambda \in B_\rho(P_\lambda, \varepsilon) \subseteq (\bigcap_{i=1}^{\infty} A_i)^\circ$.

For discrete time maps with suitable observables, Theorem 2 shows that the macrostate trajectories converge in conditional probability *everywhere* to the deterministic curve. This result may seem surprising, since we have seen that the long-time behavior of a typical trajectory may differ radically from that of the deterministic curve. However, this simply means that the time scale on which the macroscopic behavior appears deterministic grows rapidly with the number of particles.

The case in which the set of times, T, takes on a continuum of values is somewhat more problematic owing to the fact that an arbitrary intersection of measurable sets need not be measurable. If, however, $\{G_t : t \in T\}$ is sample continuous on an interval, T, i.e., if $t \mapsto G_t(x_1,..., x_n)$ is continuous for every $(x_1,..., x_n)$, then we may consider $\{G_t : t \in T\}$ to be a random process on the metric space of bounded continuous functions with the supremum norm. For physical observables this is quite reasonable to suppose. Much as in the theory of Brownian motion, we may then consider events of the form $\{\sup_{t \in T} |G_t - \psi_t(y)| \geqslant \delta'\}$ and their corresponding conditional probabilities as $n \to \infty$. It may then be possible to derive a large deviation principle on the set of bounded continuous functions on T, much as is done for Brownian motion in Schilder's theorem,[11] but here we do not pursue this matter further.

4.2. Properties of the Deterministic Curve

In the previous section we considered the probabilistic convergence of macrostate trajectories to the asymptotic deterministic curve. In this section we consider properties of the deterministic curve, $\{\psi_t(y) : t \in T\}$ as a function of t in its own right. Of course, we do not expect these properties to necessarily carry over to those of typical trajectories. Nevertheless, they do give a clue to the behavior of these trajectories for large n and relatively small t.

If g is a discontinuous function, then specific realizations of $\{G_t : t \in T\}$ will also be discontinuous. Since G is the average of a bounded function, however, the size of these discontinuities will vanish as $n \to \infty$. It is then reasonable to suppose that the deterministic curve, $\{\psi_t(y) : t \in T\}$, will be continuous in t.

Recall that g and φ_t are continuous μ-almost everywhere. Since g is bounded μ-a.e., clearly $g \circ \varphi_t$ is so as well. If we suppose $t \mapsto g(\varphi_t(x))$ is continuous for μ-a.e. $x \in X$, then of course $g(\varphi_{t'}(x)) \to g(\varphi_t(x))$ as $t' \to t$ for μ-a.e. $x \in X$. By Eq. (18) and Lebesgue dominated convergence, this implies

$$\lim_{t' \to t} \psi_{t'}(y) = \lim_{t' \to t} \int_X g \circ \varphi_{t'} \, \mathrm{d}P_\lambda = \int_X g \circ \varphi_t \, \mathrm{d}P_\lambda = \psi_t(y) \tag{23}$$

where $\lambda = \theta^{-1}(y)$. Thus, $t \mapsto \psi_t(y)$ will be continuous provided $t \mapsto g(\varphi_t(x))$ is continuous for μ-a.e. $x \in X$.

Now suppose only that $t \mapsto \varphi_t(x)$ is continuous for μ-a.e. $x \in X$ and that φ_t is μ-nonsingular, i.e., $\mu \circ \varphi_t^{-1} \prec \mu$, for all $t \in T$. Since g is continuous μ-a.e., there exists an open set, A, with null complement such that g, restricted to A, is continuous. Furthermore, since φ_t is continuous μ-a.e., there is a set, B, with null complement such that $t \mapsto \varphi_t(x)$ is continuous for all $x \in B$. Thus, if $x \in \varphi_t^{-1}(A) \cap B$, then the composite map $t \mapsto \varphi_t(x) \mapsto g(\varphi_t(x))$ is continuous. Since φ_t is μ-nonsingular,

$$\mu[X \backslash (\varphi_t^{-1}(A) \cap B)] \leqslant \mu[X \backslash \varphi_t^{-1}(A)] + \mu[X \backslash B] = \mu[\varphi_t^{-1}(X \backslash A)] = 0$$

Thus, $t \mapsto g(\varphi_t(x))$ is continuous for μ-a.e. $x \in X$, and we conclude the following:

Theorem 3. If $t \mapsto \varphi_t(x)$ is continuous for μ-a.e. $x \in X$ and φ_t is μ-nonsingular for all $t \in T$, then $t \mapsto \psi_t$ is continuous.

For many physical systems, the dynamical law φ_t is not only invertible but also time reversible in the sense that $\varphi_t^{-1} = \varphi_{-t}$. In such cases, we shall say that φ_t is *time reversible*. Time reversibility often appears in physical systems which have the added property that $\varphi_t \circ R = R \circ \varphi_{-t}$ for some involution $R = R^{-1}$, a property referred to as *time reversal invariance*. When such an R exists, every particle microstate x has a mirror point $R(x)$ such that $\varphi_t(x) = R(\varphi_{-t}(R(x)))$; hence, the trajectory of x is mirrored by the trajectory of $R(x)$ with the direction of time reversed. We may then partition X into disjoint sets $X_0 = R(X_0)$, X_1 and $R(X_1)$. If the observable and *a priori* measure are invariant under R, i.e., $g = g \circ R$ and $\mu \circ R^{-1} = \mu$, then a typical initial microstate $(x_1, ..., x_n) \in \{G \in B_\delta\}$ will include roughly equal numbers of points from X_1 and $R(X_1)$. On this basis, one expects the trajectories $\{G_t(x_1, ..., x_n): t \geqslant 0\}$ and $\{G_t(x_1, ..., x_n): t \leqslant 0\}$ to be similar, though not identical, when n is large. This suggests that the deterministic curve should be perfectly symmtric in time. Indeed, it is easy to see that, if φ_t is time reversal invariant under R and both g and μ are invariant under R, then

$$\psi_t(y) = \int_X (g \circ R \circ \varphi_t) \frac{e^{\lambda g \circ R}}{Z(\lambda)} \, d\mu$$

$$= \int_X (g \circ \varphi_{-t} \circ R) \frac{e^{\lambda g \circ R}}{Z(\lambda)} \, d(\mu \circ R^{-1})$$

$$= \psi_{-t}(y)$$

Thus, time reversal invariance is sufficient for time symmetry of the deterministic curve, provided both the observable and the *a priori* measure are invariant under R.

Suppose g is a simple function of the form $g = \sum_{i=1}^{m} a_i 1_{C_i}$, where each a_i is distinct and $C_1,..., C_m$ form a partition of X. Now, the deterministic curve for this g is given by

$$\psi_t(y) = \sum_{i=1}^{m} \sum_{j=1}^{m} a_i e^{\lambda a_j} \mu[\varphi_t^{-1}(C_i) \cap C_j] \Big/ \sum_{k=1}^{m} e^{\lambda a_k} \mu[C_k] \tag{24}$$

using Eq. (18). Since we have supposed $g \circ R = g$, notice that $R^{-1}(C_i) = C_i$ for each i. Furthermore, since μ is invariant under R,

$$\begin{aligned} \mu[\varphi_t^{-1}(C_i) \cap C_j] &= \mu[R^{-1}(\varphi_t^{-1}(C_i)) \cap R^{-1}(C_j)] \\ &= \mu[\varphi_t(R^{-1}(C_i)) \cap C_j] \\ &= \mu[\varphi_t(C_i) \cap C_j] \end{aligned} \tag{25}$$

Furthermore, if μ is invariant under φ_t, then

$$\mu[\varphi_t^{-1}(C_i) \cap C_j] = \mu[C_i \cap \varphi_t(C_j)] \tag{26}$$

The system therefore exhibits *strong detailed balance* in the sense that

$$\mu[\varphi_t^{-1}(C_i) \cap C_j] = \mu[C_i \cap \varphi_t^{-1}(C_j)] \qquad \text{for all } i, j, \text{ and } t \tag{27}$$

Conversely, if we suppose only that $\varphi_t^{-1} = \varphi_{-t}$ exists and preserves μ, then Eq. (27) implies time symmetry of the deterministic curve when g is simple. For example, suppose $g = a_1 1_{C_1} + a_2 1_{C_2}$ has only two possible states, with $C_1 = C$ and $C_2 = X \backslash C$. For any φ_t which preserves μ,

$$\begin{aligned} \mu[\varphi_t^{-1}(X \backslash C) \cap C] &= \mu[C] - \mu[\varphi_t^{-1}(C) \cap C] \\ &= \mu[\varphi_t^{-1}(C)] - \mu[C \cap \varphi_t^{-1}(C)] \\ &= \mu[(X \backslash C) \cap \varphi_t^{-1}(C)] \end{aligned} \tag{28}$$

Thus, g exhibits strong detailed balance. The corresponding deterministic curve will be time symmetric provided $\varphi_t^{-1} = \varphi_{-t}$ exists. In general, macroscopic averages of two-state single-particle observables are always time symmetric, provided the dynamics are time reversible and preserve the *a priori* measure. A system which exhibits strong detailed balance need not, however, be time reversal invariant. (Consider $C = [0, 1]$ and $\varphi_t(x) = x + t$ on $\mathbb{R}$. The only possible R is $R(x) = -x$, yet clearly $R(C) \neq C$.) Thus, time reversal invariant systems form a proper subset of all systems exhibiting time symmetry.

Finally, let us consider the asymptotic behavior of the deterministic curve for large t. Suppose once more that μ is invariant under φ_t. The collection, $\{\varphi_t : t \in T\}$, of maps is said to be *mixing* with respect to μ if, for any measurable subsets A and B of X,

$$\lim_{t \to \infty} \mu[\varphi_t^{-1}(A) \cap B] = \mu[A]\, \mu[B] \tag{29}$$

From this definition and the fact that $P_\lambda \prec \mu$, it follows that [21, p. 72]

$$\lim_{t \to \infty} \int_X g \circ \varphi_t \, \mathrm{d}P_\lambda = \int_X g \, \mathrm{d}\mu = y_* \tag{30}$$

If φ_t is time reversible, then this result clearly holds in the limit $t \to -\infty$ as well. By its definition, mixing is both a necessary and sufficient condition for $\psi_t(y)$ to converge to $E_g(\mu) = y_*$ for *any* g. In cases where the dynamics are not mixing, however, one may still have convergence in time for a restricted set of macroscopic functions.

As we have seen in our discussion of Poincaré recurrence, convergence in time for $\psi_t(y)$ need not imply convergence in time for G_t. In fact, such behavior is often quite unlikely. Nevertheless, on a short enough time scale, a scale which increases with n, and for large enough n, both $\psi_t(y)$ and G_t will appear to converge to the same limit along the same trajectory.

5. FRACTIONAL OCCUPATIONS

Consider the case $g = 1_C$, for which G_t is the fraction of points in C at time t. Since $\mathrm{e}^{\lambda g} = 1_{X \setminus C} + \mathrm{e}^{\lambda} 1_C$, the corresponding partition function is

$$Z(\lambda) = \mu[X \setminus C] + \mathrm{e}^{\lambda} \mu[C] \tag{31}$$

Recall that $\theta(\lambda) = \Psi'(\lambda) = \mathrm{e}^{\lambda}\mu[C]/Z(\lambda)$ for $\lambda \in (-\infty, \infty)$; thus, for $y \in (0, 1)$,

$$\theta^{-1}(y) = \log\left[\frac{y}{\mu[C]} \frac{\mu[X \setminus C]}{1 - y}\right] \tag{32}$$

Using Eqs. (7), (18), (31), and (32) we find, for $y \in [0, 1]$,

$$\psi_t(y) = y\mu[\varphi_t^{-1}(C) \mid C] + (1-y)\,\mu[\varphi_t^{-1}(C) \mid X \backslash C] \tag{33}$$

This result is easily understood as follows: The expected number of points in C at time t will be the number of points starting in C times the fraction of those points expected to be in C at time t plus the number initially outside C times the fraction of those points expected to be in C at time t.

When φ_t is μ-measure preserving and time reversible, the deterministic curve for fractional occupations is such that it is never further from the equilibrium value, $y_* = \mu[C]$, than the initial macrostate. To see this, suppose $y > y_*$ and note that

$$\begin{aligned}\psi_t(y) - y &= -y\mu[\varphi_t^{-1}(X \backslash C) \mid C] + (1-y)\,\mu[\varphi_t^{-1}(C) \mid X \backslash C] \\ &\leqslant -\mu[C]\,\mu[\varphi_t^{-1}(X \backslash C) \mid C] + \mu[X \backslash C]\,\mu[\varphi_t^{-1}(C) \mid X \backslash C] = 0\end{aligned} \tag{34}$$

by Eq. (28). Therefore, $\psi_t(y) \leqslant y$ for all t, with equality if and only if $\mu[\varphi_t^{-1}(X \backslash C) \cap C] = 0$ (e.g., when $t = 0$). Similary, $y < y_*$ implies $\psi_t(y) \geqslant y$ for all t. If $\psi_t(y)$ is differential in t (i.e., for the left and right sided derivatives), this furthermore shows that the deterministic curve *always* tends initially toward equilibrium, even if it does not approach equilibrium asymptotically.

Notice that the function θ is associated only with the initial macrostate. The time-evolved behavior of fractional occupations is contained in the two transition probabilities in Eq. (33). In general, these may be difficult to determine. For the baker map, ϕ, a discrete time map on the unit square,[22] these may be computed for rectangular cells with Lebesgue measure.[23] This allows one to calculate $\psi_t(y)$ for several time iterations and to compare this with Monte Carlo simulations. Although an abstract map, the baker map shares many of the relevant features of more realistic Hamiltonian dynamical systems. In particular, it is Lebesgue measure preserving, mixing, and time reversal invariant in the sense that $\phi \circ R = R \circ \phi^{-1}$, where R interchanges the ordinate and the abscissa.

In Fig. 1 we have plotted $\psi_t(y)$ for $t = -10,..., 10$ and $y = 0.4$ using the baker map. (Since the map is invertible, negative times refer to iterations of the inverse map.) The cell was chosen arbitrarily to be $C = [0.2, 0.6) \times [0.0, 0.5)$, for which $\mu[C] = 0.2$. The values of $\psi_t(y)$ are connected by straight solid lines in the figure. For comparison, a single realization of an ensemble of $n = 50{,}000$ points was generated which satisfied the initial macrostate $y = 0.4$. This was done by drawing the first $\lfloor ny \rfloor$ points uniformly from C and then drawing the rest from outside C. (Here,

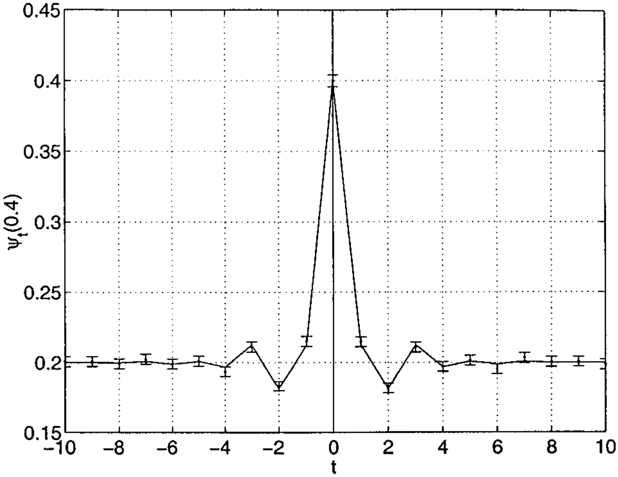

Fig. 1. Plot of the fractional occupation of C versus time for the baker map with $C = [0.2, 0.6) \times [0.0, 0.5)$ and $y = 0.4$. Straight lines are drawn between the values of $\psi_t(y)$ for each integer value of the iteration time t. The solid dots are the values of G_t for a single realization of an ensemble of $n = 50{,}000$ points. The error bars are 95% confidence intervals.

"uniformly" means with respect to Lebesgue measure.) Once generated, the known form of the map $\varphi_t := \phi^t$ was used to time evolve the initial ensemble for each value of t. The fractional occupation, G_t, was then computed for each time-evolution of the initial ensemble and is indicated by a solid dot in the figure.

The qualitative behavior of $\psi_t(y)$ in Fig. 1 is particularly notable in two regards. First, it is readily observed that the plot is symmetric about $t = 0$; in particular, $\psi_{-t}(y) = \psi_t(y)$ exactly, while G_{-t} and G_t are only approximately equal. This, as was shown in Section 4.2, is a general property of two-state systems for which φ_t is μ-measure preserving and time reversible. Hence, there is no distinction between the forward and reverse time directions. The second observation is that $\psi_t(y) \to \mu[C]$ as $t \to \pm\infty$, which is a direct consequence of the mixing property. Thus, the baker map provides a simple model of an equilibrating macroscopic quantity.

A second comment is that, while at each *given* time, t, the most probable macrostate is $\psi_t(y)$, for any finite n the set $\{\psi_t(y) : t \in \mathbb{Z}\}$ is itself an improbable realization of $\{G_t : t \in \mathbb{Z}\}$. This may be understood by observing that, given $\varepsilon > 0$, we have $|\psi_t(y) - \mu[C]| < \varepsilon$ for all $|t|$ sufficiently large, yet, by Poincaré recurrence theorem, $|G_t - \mu[C]| > \varepsilon$ for infinitely many values of t, almost surely.

The family of macroscopic maps, $\{\psi_t : t \in \mathbb{Z}\}$, does not form a group, or even a semigroup, in contrast to the family of microscopic maps, $\{\varphi_t : t \in \mathbb{Z}\}$.

Thus, while $x = \varphi_{-t}(\varphi_t(x))$, in general $y \neq \psi_{-t}(\psi_t(y)) = \psi_t(\psi_t(y))$, since ψ_t is time symmetric. Furthermore, $\{\psi_t : t \geqslant 0\}$ does not even form a semigroup, since this would imply $|\psi_t(y) - y_*| \geqslant |\psi_{t+s}(y) - y_*|$ for all $s, t \geqslant 0$, i.e., that all future macrostates are closer to equilibrium than their predecessors. To understand this note that, while $\psi_{t+s}(y)$ describes the state of an observable at time $t+s$ whose value was y at time zero, $\psi_s(\psi_t(y))$ describes the state of an observable at time $t+s$ whose value was $\psi_t(y)$ at time t. The latter corresponds to a rerandomization of the original distribution, which removes correlations that would otherwise be preserved by the dynamics and causes disagreement with the actual time evolution of the observable.

A more physical example is that of an ideal gas in a box, a paradigm in discussions of nonequilibrium and irreversibility.[24, 25] For a particle with position $x \in [0, L)$ and velocity $v \in \mathbb{R}$ in a box with rigid, reflecting walls of length L we may define the reduced parameters of position, $\xi = x/L$, and velocity, $\eta = v/L$. The dynamics are then given simply by

$$\varphi_t(\xi, \eta) = \begin{cases} (\xi + \eta t - \lfloor \xi + \eta t \rfloor, \eta) & \text{if } \lfloor \xi + \eta t \rfloor \text{ is even,} \\ (1 - \xi - \eta t + \lfloor \xi + \eta t \rfloor, -\eta) & \text{if } \lfloor \xi + \eta t \rfloor \text{ is odd} \end{cases} \tag{35}$$

The *a priori* measure is taken to be uniform in position and Gaussian in velocity; thus,

$$d\mu(\xi, \eta) = 1_{[0,1]}(\xi)\, \phi_\sigma(\eta)\, d\xi\, d\eta \tag{36}$$

where $\phi_\sigma(\eta) := (2\pi\sigma^2)^{-1/2}\, e^{-\eta^2/2\sigma^2}$ and $\sigma > 0$.

Let $C = [0, 1/2] \times \mathbb{R}$, so G_t is the fraction of particles on the left side of the box. We then have

$$\mu[\varphi_t^{-1}(C) \mid C] = \sum_{n=-\infty}^{\infty} I_n(t) \tag{37a}$$

$$\mu[\varphi_t^{-1}(C) | X \backslash C] = \sum_{n=-\infty}^{\infty} \bar{I}_n(t) \tag{37b}$$

where

$$I_n(t) = \begin{cases} 2 \int_{-\infty}^{\infty} \int_0^{1/2} 1_{[0,1/2]}(\xi - \eta t - n)\, \phi_\sigma(\eta)\, d\xi\, d\eta & \text{if } n \text{ is even,} \\ 2 \int_{-\infty}^{\infty} \int_0^{1/2} 1_{[0,1/2]}(1 - \xi + \eta t + n)\, \phi_\sigma(\eta)\, d\xi\, d\eta & \text{if } n \text{ is odd} \end{cases} \tag{38}$$

and similarly for $\bar{I}_n(t)$ with $1_{[0,1/2]}$ replaced by $1_{[1/2,1]}$. For $t>0$ and n even we find

$$I_n(t)=2(2\pi)^{-1/2}\,\sigma t[e^{-(n+1/2)^2/(2\sigma^2t^2)}+e^{-(n-1/2)^2/(2\sigma^2t^2)}-2e^{-n^2/(2\sigma^2t^2)}]$$
$$+\left(n+\frac{1}{2}\right)\operatorname{erf}\left(\frac{2n+1}{2\sqrt{2}\sigma t}\right)+\left(n-\frac{1}{2}\right)\operatorname{erf}\left(\frac{2n-1}{2\sqrt{2}\sigma t}\right)-2n\operatorname{erf}\left(\frac{n}{\sqrt{2}\sigma t}\right) \tag{39}$$

whereas for $t>0$ and n odd we find

$$I_n(t)=2(2\pi)^{-1/2}\,\sigma t[e^{-(n+1)^2/(2\sigma^2t^2)}+e^{-n^2/(2\sigma^2t^2)}-2e^{-(n+1/2)^2/(2\sigma^2t^2)}]$$
$$+(n+1)\operatorname{erf}\left(\frac{n+1}{\sqrt{2}\sigma t}\right)+n\operatorname{erf}\left(\frac{n}{\sqrt{2}\sigma t}\right)-(2n+1)\operatorname{erf}\left(\frac{2n+1}{2\sqrt{2}\sigma t}\right) \tag{40}$$

Similar expressions may be obtained for $\bar{I}_n(t)$: for n even the expression is the same as that for $I_n(t)$ when n is odd, while for n odd the opposite is true. (Note that, since μ is invariant under the time reversible map, φ_t, the deterministic curve for fractional occupations will necessarily be time symmetric.)

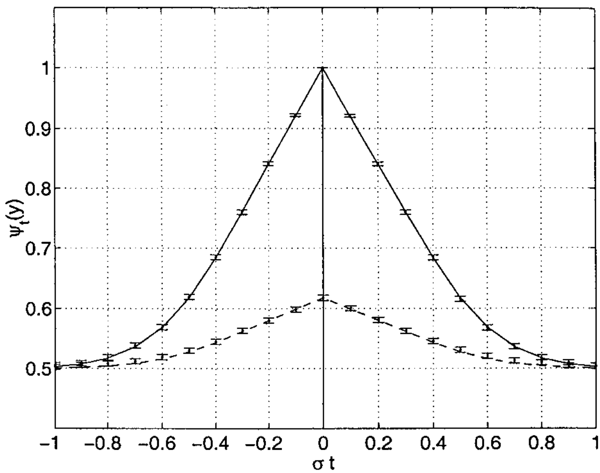

Fig. 2. Plot of the fraction of ideal gas particles on the left side of a rigid box vs. time. The time axis is in units of $1/\sigma$, where σ is the *a priori* dispersion of the velocity parameter, η. The solid line is the deterministic curve for $y=1$, which is time symmetric, while the solid dots are the values of G_t for a single realization of an ensemble of $n=50{,}000$ points. The error bars are 95% confidence intervals. The dashed curve is for $y=\psi_{0.5}(1)\approx 0.6181$, illustrating the lack of a semigroup property.

In Fig. 2 we have plotted the deterministic curve for $y = 1$ as a function of t (in units of $1/\sigma$). Near $t = 0$, the deterministic curve is linear with $\psi_t(y) \approx y - 2(2\pi)^{-1/2}(2y-1)\,\sigma\,|t|$. Thus, the initial tendency of the system is always to move toward the equilibrium value of $y_* = 1/2$. In fact, for the particular velocity distribution chosen, the entire curve happens to approach equilibrium monotonically as $|t|$ increases. (Other distributions, for example a uniform distribution, give an oscillatory approach to equilibrium.) Although monotonic, the family of maps $\{\psi_t : t \in T\}$ does not form a semigroup since, for example, $\psi_{0.6}(1) \approx 0.5687 \neq 0.5993 \approx \psi_{0.1}(\psi_{0.5}(1))$. This is illustrated in the figure by the secondary curve with initial ordinate $\psi_{0.5}(1) \approx 0.6181$.

It is worth remarking that similar behavior has also been ascribed to Boltzmann's famous H-function.[8] For a two-state system the H-function is a smooth convex function of the fractional occupation with a minimum at y_*; hence, it will always tend initially to decrease as $|t|$ increases, though it might not approach its minimum monotonically. (The Ehrenfests' interpretation of the H-theorem is therefore only partly correct.)

6. DISCUSSION

We have considered a general class of systems composed of identical constituents, here called "particles," that are dynamically noninteracting. For the microstates of the collective system, we supposed there is an *a priori* measure, typically an invariant measure, that describes the distribution of these microstates in the absence of any restrictions based on the given macrostate. We further supposed that the particles are statistically independent, that any correlations among them arise only by the need to satisfy the given macroscopic constraint. No attempt was made to justify these assumptions at a more fundamental level, though we believe they are quite reasonable for many physical systems.

What we have shown is that the time evolution of a particular macroscopic variable, namely the average over certain real-valued single-particle functions, is such that it converges in a probabilistic sense to a well defined curve as the number of particles tends to infinity. Specifically, we have derived a map ψ_t such that, if the macrostate at time 0 is constrained to be near a value y, then the macrostate at time t will be in a given neighborhood of $\psi_t(y)$ with a probability approaching one as the number of particles tends to infinity. The map ψ_t was defined in terms of an expectation with respect to a canonical distribution in which y plays the role of an average energy in the familiar thermodynamic formalism. The restrictions on the single-particle function were that it be bounded and continuous almost everywhere in a sense specified by the *a priori* measure. We found that the family of macroscopic maps, $\{\psi_t: t \in T\}$, in general forms neither a group nor a semigroup, even if the family of microscopic maps, $\{\varphi_t: t \in T\}$, has this property.

Having established this basic convergence result for a given time, we then considered how well the deterministic curve, the graph of $\psi_t(y)$ versus t, represented the behavior of a typical realization of the macrostates over all time. We found that the two may differ qualitatively quite substantially; while there may be good agreement on a finite set of selected times, there will typically be times at which they differ substantially. This was particularly true of mixing systems, for which $\psi_t(y)$ always converges in the long time limit, while a typical trajectory exhibits recurrences. Under some more restrictive conditions we proved convergence on any countably infinite set of times, but even then recurrences are possible when n is finite.

With these caveats on the correspondence between the finite and infinite particles cases, we considered some general properties of the expectation curve as a function of time. We found that, despite the fact that the macrostates may evolve discontinuously, the deterministic curve may be continuous in time. We also found that, for systems which are time reversal invariant, the deterministic curve is symmetric in time about $t=0$, the point at which conditioning of initial macrostates takes place. These properties were then related to familiar geometric properties attributed to Boltzmann's H-curve.

We have not considered here extensions of these results to macroscopic variables in, say, $\mathbb{R}^d$, which would involve issues of convexity that make such an extension nontrivial. The general problem of interacting particles poses a greater difficulty and requires a significant change of methodology, though we conjecture that similar results will hold if $\psi_t(y)$ is defined as a limit of n-particle expectations.

APPENDIX A. PROOF OF THEOREM 1

Proof. It suffices to consider $x_B \notin \bar{A}$ since, if $x_B \in A^\circ$, then $x_B \notin X \backslash A^\circ = \overline{X \backslash A}$.

Since $x_* \notin \partial B$, either $x_* \in B^\circ$ or $x_* \notin \bar{B}$. Suppose the former. By Eq. (12), $P_n[B] \to 1$ as $n \to \infty$, which implies $P_n[A \mid B] \to \lim_{n \to \infty} P_n[A]$. Now, $P_n[A] \to 0$ if $x_* \notin \bar{A}$, while $P_n[A] \to 1$ if $x_* \in A^\circ$. Since $x_B = x_*$ for this case, the result is proven.

Since $x_* \notin \bar{B}$, $0 < \inf I(\bar{B}) \leqslant \inf I(B^\circ)$. By Eq. (12) we have, for any $\varepsilon > 0$,

$$P_n[B] > \exp[-a_n(1+\varepsilon) \inf I(B^\circ)] > 0$$

for all n sufficiently large; thus, $P_n[A \mid B]$ is well defined. Suppose further that $\inf I(\overline{A \cap B}) < \infty$. From Eq. (12) we may also deduce that for all n sufficiently large,

$$P_n[A \mid B] \leqslant \frac{\exp[-a_n(1-\varepsilon) \inf I(\bar{A} \cap B)]}{\exp[-a_n(1+\varepsilon) \inf I(B^\circ)]}$$
$$= \exp[-a_n(1-\varepsilon) \inf I(\overline{A \cap B}) + a_n(1+\varepsilon) \inf I(B^\circ)]$$

Now suppose $x_B \notin \bar{A}$. To show that $P_n[A \mid B] \to 0$, it will suffice to show that we may choose ε such that

$$(1-\varepsilon) \inf I(\overline{A \cap B}) - (1+\varepsilon) \inf I(B^\circ) > 0$$

Since this means we must choose ε small enough so that

$$\varepsilon < \frac{\inf I(\overline{A \cap B}) - \inf I(B^\circ)}{\inf I(\overline{A \cap B}) + \inf I(B^\circ)}$$

we see that it will suffice to show that $\inf I(\overline{A \cap B}) > \inf I(B^\circ)$. (Note that, since $\inf I(B^\circ) > 0$, the denominator in the above inequality is indeed non-zero.)

We have assumed B is an I-continuity set, so $\inf I(B^\circ) = \inf I(\bar{B}) = I(x_B) < \infty$. Now, if $\overline{A \cap B} = \varnothing$ then $\inf I(\overline{A \cap B}) = \infty > \inf I(\bar{B}) = \inf I(B^\circ)$ and we are done. Suppose that $\overline{A \cap B} \neq \varnothing$. Then there exists an $x \in \overline{A \cap B}$ such that $I(x) = \inf I(\overline{A \cap B})$, since I is a good rate function. Notice that $x_B \notin \overline{A \cap B}$, since $\overline{A \cap B} \subseteq \bar{A} \cap \bar{B} \subseteq \bar{A}$ and $x_B \notin \bar{A}$. Clearly, then, $x_B \neq x$. Since $\overline{A \cap B} \subseteq \bar{B}$ as well, $\inf I(\overline{A \cap B}) \geqslant \inf I(\bar{B})$, or, equivalently, $I(x) \geqslant I(x_B)$. Equality cannot hold, however, since, if that were the case, then $I(x)$ would equal $\inf I(\bar{B})$, in violation of the assumed uniqueness of x_B. Therefore, $\inf I(\overline{A \cap B}) > \inf I(\bar{B}) = \inf I(B^\circ)$.

Now suppose $\inf I(\overline{A \cap B}) = \infty$ instead. By Eq. (12), this implies $P_n[A \cap B] \to 0$. Since $P_n[B] > 0$ for all n sufficiently large, this implies $P_n[A \mid B] \to 0$. ∎

APPENDIX B. PROOF OF GIBBS CONDITIONING LEMMAS

The proof of Lemmas 1 is adapted from that of Ellis,[20] who considers the case in which g is a simple function. The extension to a general bounded measurable function is similar and is included here for completeness. Dembo and Zeitouni [11, p. 294–297] prove Lemma 2 for the τ-topology, for which E_g is continuous by definition for any bounded g. (It is also continuous in the weak topology if one assumes further that g is continuous.) Since we need only show that $P \in \overline{A_\delta}$ implies $E_g(P) \in \overline{B_\delta}$ for $P \prec \mu$, we are able to prove a stronger result, assuming only that g is bounded and continuous μ-a.e., which is valid even in the standard weak topology. Similar comments apply to the proof of Lemma 3, which follows readily if one assumes g is continuous.

B.1. Proof of Lemma 1

Proof. We first note that, since $\lambda \mapsto e^{\lambda g(x)}$ is bounded and differentiable to all orders for μ-a.e. $x \in X$, by Lebesgue dominated convergence Ψ' and Ψ'' are well-defined and continuous. Specifically, $\Psi'(\lambda) = \int_X g \, \mathrm{d}P_\lambda$ and $\Psi''(\lambda) = \int_X g^2 \, \mathrm{d}P_\lambda - (\int_X g \, \mathrm{d}P_\lambda)^2 \geqslant 0$ for $\lambda \in \mathbb{R}$. By assumption $\Psi''(\lambda)$ is nonzero, so in fact $\Psi''(\lambda) > 0$ and hence Ψ' increases monotonically. Since θ increases monotonically and $\theta(0) = y_*$, $\lambda > 0$ implies $y = \theta(\lambda) > y_*$, while $\lambda < 0$ implies $y = \theta(\lambda) < y_*$. Since θ is invertible, we have conversely that $y > y_*$ implies $\lambda > 0$, while $y < y_*$ implies $\lambda < 0$. Clearly, $\lambda = 0$ if and only if $y = y_*$. ∎

B.2. Proof of Lemma 2

Proof. Since $P_\lambda \prec \mu$ and g is bounded, we have $I_\mu(P_\lambda) = \lambda y - \Psi(\lambda) < \infty$. Now let $P \in \overline{A_\delta} \setminus \{P_\lambda\}$. If $P \not\prec \mu$, then $I_\mu(P) = \infty$ and $I_\mu(P_\lambda) < I_\mu(P)$, while for $P \prec \mu$ we find $I_\mu(P) > \lambda \int_X g \, \mathrm{d}P - \Psi(\lambda)$.

Since g is bounded and continuous μ-a.e. and $P \prec \mu$, E_g is continuous at P; thus, $P \in \overline{A_\delta}$ implies $E_g(P) = \int_X g \, \mathrm{d}P \in \overline{B_\delta}$. If $y < y_*$ then $\overline{B_\delta} = [y - \delta, y]$ and $\lambda < 0$ by Lemma 1. Now, $\lambda < 0$ and $\int_X g \, \mathrm{d}P \leqslant y$ imply $\lambda \int_X g \, \mathrm{d}P \geqslant \lambda y$. If, on the other hand, $y > y_*$, then $\overline{B_\delta} = [y, y + \delta]$ and $\lambda > 0$, while $\int_X g \, \mathrm{d}P \geqslant y$, so again $\lambda \int_X g \, \mathrm{d}P \geqslant \lambda y$. Finally, if $y = y_*$ then $\lambda = 0$ and $\lambda \int_X g \, \mathrm{d}P \geqslant \lambda y$ holds trivially. Thus, for all $P \in \overline{A_\delta}$,

$$I_\mu(P_\lambda) = \lambda y - \Psi(\lambda) \leqslant \lambda \int_X g \, \mathrm{d}P - \Psi(\lambda) < I_\mu(P)$$

Since $P_\lambda \in A_\delta \subseteq \overline{A_\delta}$, this completes the proof. ∎

B.3. Proof of Lemma 3

Proof. Suppose $y \leqslant y_*$ and let $\lambda_n = \lambda - 1/n$, where $\lambda = \theta^{-1}(y)$. (A similar argument may be applied if $y \geqslant y_*$.) Since θ is strictly monotonic,

$E_g(P_{\lambda_n}) = \theta(\lambda_n) < \theta(\lambda) = y$. We will first show that $E_g(P_{\lambda_n}) \in B_\delta^\circ$ for all n sufficiently large and that $E_g(P_{\lambda_n}) \to E_g(P_\lambda)$. Since E_g is continuous at P_λ, it will suffice to show that $P_{\lambda_n} \to P_\lambda$ in ρ.

Now, convergence in ρ is equivalent to the weak convergence of P_{λ_n} to P_λ [9, p. 310]. Thus, let h be an arbitrary bounded continuous function on X and define F such that $F(\lambda') = \int_X h e^{\lambda' g}/Z(\lambda')\, \mathrm{d}\mu$ for $\lambda' \in \mathbb{R}$. Since the integrand is bounded and continuous for μ-a.e. x, it follows that F is continuous everywhere and $F(\lambda_n) \to F(\lambda)$. This proves weak convergence and hence convergence in ρ.

We have shown that $E_g(P_{\lambda_n}) \in B_\delta^\circ$ for all n sufficiently large. From this it follows that $P_{\lambda_n} \in A_\delta^\circ \subseteq \overline{A_\delta}$ and hence $I_\mu(P_{\lambda_n}) = \lambda_n E_g(P_{\lambda_n}) - \Psi(\lambda_n) > I_\mu(P_\lambda)$ for all n sufficiently large. Now, $\inf I_\mu(A_\delta^\circ) \geqslant \inf I_\mu(\overline{A_\delta})$, so $I_\mu(P_{\lambda_n}) \geqslant \inf I_\mu(A_\delta^\circ) \geqslant \inf I_\mu(\overline{A_\delta}) = I_\mu(P_\lambda)$. Since $E_g(P_{\lambda_n}) \to E_g(P_\lambda)$ and $\Psi(\lambda_n) \to \Psi(\lambda)$ as $\lambda_n \to \lambda$, it is clear that $I_\mu(P_{\lambda_n}) \to I_\mu(P_\lambda)$. Hence $\inf I_\mu(A_\delta^\circ) = \inf I_\mu(\overline{A_\delta})$.

ACKNOWLEDGMENTS

The authors would like to thank D. Driebe for his helpful comments and constructive criticisms.

REFERENCES

1. N. G. van Kampen, A power series expansion of the master equation, *Canad. J. Phys.* **39**:551 (1961).
2. Thomas G. Kurtz, The relationship between stochastic and deterministic models for chemical reactions, *J. Chem. Phys.* **57**(7):2976–2978 (1972).
3. Ryogo Kubo, Kazuhiro Matsuo, and Kazuo Kitahara, Fluctuations and relaxation of macrovariables, *J. Stat. Phys.* **9**(1):51–96 (1973).
4. Herbert Spohn, *Large Scale Dynamics of Interacting Particles* (Springer-Verlag, New York, 1991).
5. Edward B. Davies, *The Theory of Open Systems* (Academic Press, London, 1976).
6. Lawrernce Sklar, *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics* (Cambridge University Press, Cambridge, 1993).
7. J. Mehra and E. C. G. Sudarshan, Some reflections on the nature of entropy, irreversibility and the second law of thermodynamics, *Il Nuovo Cimento* **11**B(2):215–256 (1972).
8. Paul Ehrenfest and Tatiana Ehrenfest, *The Conceptual Foundations of the Statistical Approach in Mechanics* (Cornell University Press, Ithaca, New York, 1959).
9. Richard M. Dudley, *Real Analysis and Probability* (Chapman & Hall, New York, 1989).
10. Robert B. Ash, *Real Analysis and Probability* (Academic Press, New York, 1972).
11. Amir Dembo and Ofer Zeitouni, *Large Deviations Techniques and Applications* (Jones and Bartlett, Boston, 1993).
12. Ludwig Boltzmann, Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmtheorie und der Wahrscheinlichkeitsrechnung respecktive den Sätzen über das Wärmegleichgewicht (On the relationship between the second law of the mechanical theory of heat and the probability calculus), *Wiener Berichte* **2**(76):373–435 (1877).

13. Albert Einstein, *Annalen der Physik* **22**:180 (1907).
14. Jean-Dominique Deuschel and Daniel W. Stroock, *Large Deviations* (Academic Press, San Diego, 1989).
15. Richard S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics* (Springer-Verlag, New York, 1985).
16. Paul Dupuis and Richard S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations* (John Wiley & Sons, New York, 1997).
17. David Ruelle, *Statistical Mechanics: Rigorous Results* (Benjamin, New York, 1969).
18. Oscar E. Lanford III, Entropy and equilibrium states in classical statistical mechanics, in *Statistical Mechanics and Mathematical Problems*, Lecture Notes in Physics, Vol. 20, pp. 1–111, A. Lenard, ed. (Springer-Verlag, 1973).
19. Yuri Kifer, Large deviations in dynamical systems and stochastic processes, *Trans. Amer. Math. Soc.* **321**(2):505–524 (1990).
20. Richard S. Ellis, The theory of large deviations: From Boltzmann's 1877 calculation to equilibrium macrostates in 2d turbulence, *Physica D* **133**(1–4):106–136 (1999).
21. Andrzej Lasota and Michael C. Mackey, *Chaos, Fractals, and Noise* (Springer-Verlag, New York, 1994).
22. Vladimir I. Arnol'd and Andre Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968).
23. Brian R. La Cour and William C. Schieve, Quasi Markovian behavior in mixing maps, *Physica D* **133**(1–4):309–320 (1999).
24. Arthur Hobson, Irreversibility in simple systems, *Amer. J. Phys.* **34**:411–416 (1966).
25. Joel L. Lebowitz, Microscopic origins of irreversible macroscopic behavior, *Physica A* **263**:516–527 (1999).