The dynamic origin of increasing entropy

Michael C. Mackey

Departments of Physiology and Physics, McGill University, Montreal, Quebec, Canada H3G 1Y6

Thermodynamic states are assumed to be characterized by densities. Recent ergodic-theory results on the evolution of densities are used to give a unified treatment of the origin of classical nonequilibrium thermodynamic behavior. Asymptotic periodicity is sufficient for the existence of at least one state of (metastable) thermodynamic equilibrium and for the evolution of the entropy to a relative maximum that depends on the way the system is prepared. Ergodicity is necessary and sufficient for a unique state of thermodynamic equilibrium to exist. Exactness, a property of chaotic semidynamical (irreversible) systems, is necessary and sufficient for the global evolution of the entropy to its unique maximum for all initial states. Since all of the laws of physics are formulated as (reversible) dynamical systems, it is unclear why entropy is observed to approach a maximum. Setting aside the possibility that all of the laws of physics are incorrectly formulated, it is demonstrated that either observation of a subset of the complete dynamics (trivial coarse graining) or interactions with an external heat bath (addition of noise) may induce exactness with a consequent evolution of entropy to a maximal state.

CONTENTS

> *It is not very difficult to show that the combination of the reversible laws of mechanics with Gibbsian statistics does not lead to irreversibility but that the notion of irreversibility must be added as a special ingredient . . . .*
> *. . . the explanation of irreversibility in nature is to my mind still open.*
>
> Bergmann (1967)

I. INTRODUCTION

This work examines the dynamical foundations of the evolution of entropy to maximal states. This is accomplished through an application of recent results in ergodic theory to so-called chaotic dynamical systems. As such it is an extension of the work of Khinchin (1949), Dugas (1959), Kurth (1960), Truesdell (1960), Farquhar (1964), O. Penrose (1970, 1979), Lebowitz (1973), Mackey (1974), Wehrl (1978), and Prigogine (1980). An easily accessible and enjoyable account of the issues of ergodicity, mixing, and coarse graining in classical statistical mechanics can be found in Lebowitz and Penrose (1973).

In no sense is it the intent of this paper to develop statistical mechanics as a subject. This is done rather nicely from several points of view in a variety of graduate-level texts [Kestin and Dorfman (1971), Pathria (1972), Reichl (1980), Ma (1985), and Grandy (1988) are representative of some of the more interesting and thought-provoking of these]. Rather, the aim is to understand the basic issues that must be answered before any satisfactory statistical mechanics can be constructed.

What are the basic issues? Although moderately easy to state, they are not so easy to resolve. First is the fundamental question of what guarantees a state of thermodynamic equilibrium, and what is required for this state to be unique. Second, what is the proper mathematical analog of the experimentally determined entropy? What properties does the mathematical entropy have at states of thermodynamic equilibrium? Third, what dynamical properties must exist to ensure that, starting from any initial condition, system entropy will evolve to a state of thermodynamic equilibrium (either relative or absolute)? And finally, what must be assumed concerning the dynamics of models of real physical systems in order that they display this evolution of system entropy as do the experimental systems they are designed to mimic?

It is the thrust of this paper that one of the primary issues that any successful statistical mechanics must address is the origin of the second law of thermodynamics. The second law of thermodynamics comes in so many forms that it is often confusing to understand precisely what a given author intends by the use of this term. To make matters explicit I distinguish four versions of the second law of thermodynamics.

Let $S_{\mathrm{TD}}(t)$ denote the thermodynamic entropy at time t. The weakest form of the second law is the zero-order second law,

$$S_{\mathrm{TD}}(t)=S_{\mathrm{TD}}(t') \quad \text{for all times } t,t' \text{ ;}$$

so the entropy difference $\Delta S=S_{\mathrm{TD}}(t')-S_{\mathrm{TD}}(t)$ satisfies $\Delta S\equiv 0$.

In this form the system entropy remains forever fixed at the value with which the system is prepared, be it by nature or by an investigator.

The next strongest form is called the first-order second law,

$$S_{\mathrm{TD}}(t)\geq S_{\mathrm{TD}}(t') \quad \text{for all times } t>t' \text{ .}$$

Thus with this form the system entropy *may* increase and $\Delta S\leq 0$.

Following the first-order form we have the stronger assertion, the second-order second law,

$$S_{\mathrm{TD}}(t)\geq S_{\mathrm{TD}}(t') \quad \text{for all times } t>t' \text{ ,}$$

and at least one

$$\lim_{t\to+\infty} S_{\mathrm{TD}}(t)=S^*_{\mathrm{TD}}$$

exists. Thus $\Delta S(t)=S_{\mathrm{TD}}(t)-S^*_{\mathrm{TD}}\leq 0$ and

$$\lim_{t\to+\infty} \Delta S(t)\leq 0 \text{ .}$$

In this case we may definitely assert that system entropy converges to a steady-state value S^*_{TD}, which may not be unique. If it is not unique it characterizes a metastable state.

The final, and strongest, form of the second law of thermodynamics is the most interesting and is the third-order second law,

$$S_{\mathrm{TD}}(t)\geq S_{\mathrm{TD}}(t') \quad \text{for all times } t>t' \text{ ,}$$

and there is a unique

$$\lim_{t\to+\infty} S_{\mathrm{TD}}=S^*_{\mathrm{TD}}$$

for all initial system preparations. Under these circumstances,

$$\lim_{t\to+\infty} \Delta S(t)=0 \text{ .}$$

In this case we know that the system entropy evolves to a unique maximum value irrespective of the way in which the system was prepared.

Because of the approach taken here, as contrasted with most work on foundational questions in statistical mechanics, and because of the nature of the material presented, a brief outline of the main points that are made will be helpful. Much of the background material for the current work is in Lasota and Mackey (1985).

Section II introduces densities and the definition of the Boltzmann-Gibbs entropy of a density. It is a simple demonstration that the entropy of a density will assume a maximal value if and only if the density is (in the terminology of Gibbs) either the density of the microcanonical or canonical ensemble. Further, it is demonstrated that the Boltzmann-Gibbs entropy of a density will attain its maximum if and only if the density is that of the microcanonical ensemble. Then it is shown that the mathematical definition of the Boltzmann-Gibbs entropy of a density can be argued plausibly to coincide with the thermodynamic entropy of a system characterized by that density.

Section III introduces Markov operators. These are linear integral operators that describe the evolution of densities by dynamical or semidynamical systems. Fixed points of Markov operators, known as stationary densities, define states of relative or absolute thermodynamic equilibrium depending on whether there are multiple or unique stationary densities. Thus a central question that must be answered in any treatment of thermodynamics is under what circumstance will the entropy change from its original value, determined by the way in which the system was prepared, to a final state corresponding to one of these states of relative or absolute equilibrium. Following this, the conditional entropy, a generalization of the Boltzmann-Gibbs entropy, is introduced and identified with ΔS. Under particular conditions the conditional entropy is shown to have its maximal value of zero if the stationary density of the state of thermodynamic equilibrium is that of the canonical ensemble. Then the distinction between reversible and irreversible systems is made. This is used to provide the not too surprising proof that entropy is constant for reversible systems (the zero-order form of the second law). It is only in irreversible systems that the entropy *may* increase (first-order form of the second law). Thus irreversibility is necessary but not sufficient for the entropy to increase. The section closes with a consideration of a particular form of dynamical behavior, called asymptotic periodicity, that allows the evolution of the entropy to at least a metastable state of relative equilibrium (second-order form of the second law).

Section IV introduces a special type of Markov operator, the Frobenius-Perron operator. Following illustrative material demonstrating its utility in studying the evolution of densities by a variety of dynamical and semidynamical systems, we turn to a consideration of the conditions that guarantee the existence of a *unique* state of thermodynamic equilibrium. The necessary and sufficient condition for this existence is the property of ergodicity, which may be shared by both reversible and irreversible systems.

Section V presents the concept of mixing, introduced in a qualitative sense by Gibbs, which is a stronger property than ergodicity though it still may be shared by irreversible and reversible systems. However, it is not sufficient to permit the entropy of a system to change from its initial value.

Section VI is, in a sense, the core of this paper. There we present results showing that, for systems with the density of the microcanonical ensemble defining their unique state of thermodynamic equilibrium, it is *necessary and sufficient* that the system have a property known as uniform exactness in order for there to be a global evolution of the entropy to its maximal value of zero. Analogously, for systems whose thermodynamic equilibrium is characterized by more complicated (nonuniform) densities, the *necessary and sufficient* condition for the global evolution of the entropy is a property known as f_* exactness. Both uniform and f_* exactness are properties that only irreversible systems may display and are necessary and sufficient for the strongest (third-order) form of the second law of thermodynamics to operate.

In a very real sense the results of Sec. VI raise as many questions as they answer. Although providing totally clear criteria for the global evolution of the entropy of a system, at the same time these criteria suggest that all currently formulated physical laws may not be at the foundation of the thermodynamic behavior we observe daily. This is simply because these laws are formulated as (reversible) dynamical systems.

One possibility is that the current reversible, dynamical system statements of physical laws are incorrect and that more appropriate formulations in terms of irreversible semidynamical systems await discovery. Alternately, other phenomena may mask the operation of these reversible systems so they appear to be irreversible to the observer. Sections VII and VIII explore this latter possibility.

Section VII examines two such phenomena that are closely related. First we examine the effects of nontrivial *coarse graining* of phase space, due either to measurement error or to an inherent graininess of space that is imposed by nature. It is easy to show that if we observe a system with mixing dynamics, but operating in a coarse-grained phase space, then the entropy of the coarse-grained density will evolve to a maximum as time goes either forward ($t \rightarrow +\infty$) or backward ($t \rightarrow -\infty$). Thus, though coarse graining induces entropy increase to a maximum, it fails to single out any unique direction of time for this to occur, illustrating that the origin of irreversible behavior is not to be found in reversible dynamics operating in a coarse-grained phase space.

Second we explore the consequence of taking a *trace*, in which we observe only some of the important dynamical variables of a dynamics operating in a higher dimensional space. In this case the total dynamics may be reversible and, consequently, have a constant entropy while the entropy of the trace system may smoothly evolve to a maximum (first-, second-, or third-order form of the second law).

Section VIII examines the effects of external perturbations on discrete and continuous time dynamics. This situation is usually called interaction with a heat bath. Under very mild assumptions concerning the nature of the perturbation, discrete time systems with the most uninteresting dynamics in the unperturbed situation will become either asymptotically periodic or f_* exact in the presence of perturbations. Thus they will display evolution of entropy toward states of thermodynamic equilibrium (either second- or third-order forms of the second law). In the latter part of this section we continue this theme by examining the effects of white-noise perturbations of continuous time systems whose dynamics are described by systems of ordinary differential equations. Again these perturbations induce f_* exactness and the consequent increase of the conditional entropy to its maximum value of zero (third-order form of the second law).

Although the material presented here is completely confined to fundamental issues in classical statistical mechanics, many of the results have quantum statistical mechanical analogs derived from noncommutative ergodic theory. Ruelle (1969), Thirring (1983), and Bratteli and Robinson (1987), all touch on aspects of these problems.

II. MAXIMAL ENTROPY PRINCIPLES

A. Dynamics and densities

We start by considering a general system operating in a *phase space* X. On this phase space we consider that the evolution in time of our system is described by a *dynamical law* S_t that maps points in the phase space X into new points, i.e., $S_t : X \rightarrow X$, as time t changes. In general X may be a d-dimensional phase space, either finite or not, and therefore x is a d-dimensional vector. Time t may be either continuous ($t \in R$) as, for example, it would be for a system whose dynamics were governed by a set of differential equations, or discrete (integer valued, $t \in Z$) if the dynamics are determined by discrete time maps.

We only consider autonomous processes in which the dynamics S_t are not an explicit function of the time t, so it is always the case that $S_t(S_{t'}(x)) = S_{t+t'}(x)$. Thus the dynamics governing the evolution of the system are the same on the intervals $[0,t']$ and $[t,t+t']$.

Two types of dynamics will be important in our considerations. First we introduce the concept of a *dynamical system* $\{S_t\}_{t \in R}$ (or, alternately, $t \in Z$ for discrete time systems) on a phase space X, which is simply any group of transformations $S_t : X \to X$ having the two properties

$$\text{(a)} \quad S_0(x) = x$$

and

$$\text{(b)} \quad S_t(S_{t'}(x)) = S_{t+t'}(x) \quad \text{for } t,t' \in R \; .$$

Since, from the definition, for any $t \in R$, we have

$$S_t(S_{-t}(x)) = x = S_{-t}(S_t(x)) \; ,$$

it is clear that dynamical systems are *invertible* or *reversible* since they may be run either forward or backward in time. Systems of ordinary differential equations are examples of dynamical systems because of their reversibility. All of the laws of classical and quantum physics are invertible and describe the behavior of reversible systems.

The second type of dynamics that is important to distinguish are those of *semidynamical systems* $\{S_t\}_{t>0}$, which is any semigroup of transformations $S_t : X \to X$ for which

$$\text{(a)} \quad S_0(x) = x$$

and

$$\text{(b)} \quad S_t(S_{t'}(x)) = S_{t+t'}(x) \quad \text{for } t,t' \in R^+ \text{ (or } N) \; .$$

The essential difference between the definition of dynamical and semidynamical systems lies solely in the restriction of t and t' to values drawn from the positive real numbers, or the positive integers, for the semidynamical systems. Thus, in sharp contrast to dynamical systems, semidynamical systems are *noninvertible* or *irreversible* and may not be run backward in time in an unambiguous fashion. Examples of semidynamical systems are given by noninvertible maps, delay differential equations, and some partial differential equations.

In spite of the enormous significance of distinguishing between dynamical and semidynamical systems later, initially no assumption is made concerning the invertibility or noninvertibility of the system dynamics.

B. Densities and measures

For every initial point x^0, the sequence of successive points $S_t(x^0)$, considered as a function of time t, is called a *system trajectory*. The *state* of a system at time t corresponding to an initial point x^0 is given by the value of $S_t(x^0)$. The usual way of examining the dynamics of systems is by studying the properties of individual trajectories, but in keeping with the ergodic-theory approach adopted here we opt instead to study the way in which the system dynamics operate on an infinite number of initial points.

More specifically we will examine the way in which the dynamics alter densities. What do we mean by a density? If f is an L^1 function in the space X, i.e., if

$$\int_X |f(x)| dx < \infty \; ,$$

then it is a *density* if $f(x) \geq 0$ and $\|f\| = 1$. As usual, $\|f\|$ denotes the L^1 norm of the function f,

$$\|f\| = \int_X |f(x)| dx \; .$$

The examination of the evolution of densities by system dynamics is, as mentioned above, equivalent to examining the behavior of an infinite number of trajectories. This viewpoint is fundamental and important to the understanding of the foundations of thermodynamics because of our assumption that densities characterize thermodynamic systems.

Postulate A. A thermodynamic system is a system that has, at any given time, states distributed throughout the phase space X, and the distributions of these states is characterized by a density $f(x)$.

Given a density f, then the *f-measure* $\mu_f(A)$ of the set A in the phase space X is defined by

$$\mu_f(A) = \int_A f(x) dx \; ,$$

and f is called the *density* of the measure μ_f. The usual Lebesgue measure of a set A is denoted by $\mu_L(A)$, and the density of the Lebesgue measure is the uniform density, $f(x) = 1/\mu_L(X)$ for all points x in the phase space X. We always write $\mu_L(dx) = dx$.

It is instructive to compare the approach used here with that of Boltzmann and Gibbs in their treatments of statistical mechanics. Both started from the assumption that they were dealing with systems of dimension $d = 2s$, whose dynamics were described by s position variables x_i and s momentum variables p_i.

Boltzmann considered the basic phase space to be a $2s$ dimensional space, which is usually called *μ space.* He then considered the evolution of a large number N of identical particles, each with the same dynamics, in μ space. N is large and typically on the order of Avogadro's number 6×10^{23}. The limiting case of $N \to \infty$ is the *thermodynamic limit* in which case the Boltzmann approach considers the evolution of a density in μ space.

Gibbs also considered N identical particles operating with these $2s$ dimensional dynamics in a phase space (commonly called the Γ *space*) of dimension $2sN$. He then considered an infinite number of copies of this original system and gave this construct the name *ensemble*. Thus Gibbs studies the evolution of the ensemble density.

Later it will become clear what types of systems may be described by the evolution of densities. For now, however, if we accept Postulate A that such systems exist, then it will be easy to examine the consequences of this postulate. Virtually the entire rest of this paper is devoted to the study of systems by the evolution of densities, how the system properties determine the character of the density evolution, and how this is translated into the behavior of entropy.

C. Boltzmann-Gibbs entropy

Having postulated the existence of thermodynamic systems that have a distribution of system states with density f, we are now in a position to develop the physically useful concept of entropy as both Boltzmann and Gibbs introduced the term.

In his celebrated work Gibbs, assuming the existence of a density f of system states over the phase space X, introduced the concept of the *index of probability* given by $\log f(x)$, where "log" denotes the natural logarithm. He then introduced a quantity $H(f)$, which is the negative of the phase-space average of the index of probability weighted by the density f, i.e.,

$$H(f)=-\int_X f(x)\log f(x)dx \ .$$

This is now known as the *Boltzmann-Gibbs entropy* of a density f, since the same equation appears in Boltzmann's work (with the opposite sign), but the phase space is different for Boltzmann (μ space) and for Gibbs (Γ space). It is easily shown that the only function for the index of probability that gives the requisite additive property to make the entropy an extensive quantity is the logarithmic function, and that it is unique up to a multiplicative constant (Khinchin, 1949; Skagerstam, 1974).

We formalize this by saying that if f is a density, then the *Boltzmann-Gibbs entropy* of the density f is defined by

$$H(f)=\int_X \eta(f(x))dx \ , \tag{2.1}$$

where the function $\eta(w)$ is defined by

$$\eta(w)=-w\log w \quad \text{for } w>0, \quad \text{and } \eta(0)=0 \ . \tag{2.2}$$

The function $\eta(w)$ is strictly concave; so $\eta''(w)<0$ for all $w>0$. From this it is an immediate consequence that

$$\eta(w)\le(w-v)\eta'(v)+\eta(v) \tag{2.3}$$

for all $w,v>0$. Combining Eq. (2.2) defining η with inequality (2.3) gives the so-called *Gibbs inequality*

$$w-w\log w\le v-w\log v \quad \text{for } w,v>0 \ . \tag{2.4}$$

A useful integral inequality is available from the Gibbs inequality. If f and g are two densities, and $\eta(f(x))$ and $\eta(g(x))$ are both integrable, then we have directly from (2.4) that

$$-\int_X f(x)\log f(x)dx\le-\int_X f(x)\log g(x)dx \ . \tag{2.5}$$

Only when $f=g$ does the equality hold.

D. Maxima of the Boltzmann-Gibbs entropy

Armed only with the integrated version of the Gibbs inequality (2.5) we may immediately understand the origin of the classical Gibbs microcanonical and canonical ensembles as reflecting simple manifestations of extremal properties of the entropy.

1. Microcanonical ensembles

Consider a given space X with finite Lebesgue measure, $\mu_L(X)<\infty$ [temporarily forego the normalization $\mu_L(X)=1$], and all possible densities f. Then the only density that will make the entropy maximal is the uniform density given by the inverse of the Lebesgue measure of X. This is stated more precisely by the following.

Theorem 2.1. When $\mu_L(X)<\infty$, the density that maximizes the Boltzmann-Gibbs entropy is the uniform density

$$f_*(x)=\frac{1}{\mu_L(X)} \ . \tag{2.6}$$

For any other density $f\ne f_*$, $H(f)<H(f_*)$.

Proof. Pick an arbitrary density f; so, by definition, the entropy of f is just

$$H(f)\le-\int_X f(x)\log g(x)dx \tag{2.7}$$

However, with $g(x)=1/\mu_L(X)$ the integrated Gibbs inequality (2.5) gives

$$\begin{aligned}H(f)&\le-\int_X f(x)\log g(x)dx\\&=-\log\left[\frac{1}{\mu_L(X)}\right]\int_X f(x)dx \ ,\end{aligned}$$

or $H(f)\le\log[1/\mu_L(X)]$, since f is a density; the equality holds if and only if $f=f_*$. However, the entropy of f_* is easily calculated to be

$$\begin{aligned}H(f_*)&=-\int_X\frac{1}{\mu_L(X)}\log\left[\frac{1}{\mu_L(X)}\right]dx\\&=-\log\left[\frac{1}{\mu_L(X)}\right] ;\end{aligned}$$

so $H(f)\le H(f_*)$ for any density f, or $H(f)<H(f_*)$ for $f\ne f_*$. Clearly, if X is normalized so $\mu_L(X)=1$, then $H(f)\le 0$.

The uniform density f_* defined by Eq. (2.6) is a generalization of the density of the microcanonical ensemble introduced by Gibbs in his work on the foundations of thermodynamics.

In the derivation of the density (2.6) maximizing the entropy on a finite space, there was no need to refer to the nature of the dynamics of the system generating the density. This is in sharp contrast to the usual approach in thermodynamics in which the dynamics are quite specifically used to argue for the plausibility of the microcanonical density (2.6). The fact that a generalization of this density appears in such a natural way merely illustrates (1) the generality of both the density and the method used to obtain it and (2) the fact that the existence of the density of the microcanonical ensemble is quite independent of the system dynamics.

2. Canonical ensembles

Even more interesting consequences can emerge from the extremal properties of entropy that offer insight into the basic foundation of thermodynamics of both classical and quantum systems. In this section we state and prove a simple theorem that allows us to derive all of the conventional reversible thermodynamics in Sec II.D.

Theorem 2.2. Assume that a non-negative measurable function $\alpha(x)$ is given as well as an average or expectation value $\langle \alpha \rangle$ of that function over the entire space X, weighted by the density f:

$$\langle \alpha \rangle = \int_X \alpha(x) f(x) dx \ . \tag{2.8}$$

(Note that $\langle \alpha \rangle$ may be time dependent.) Then the maximum of the entropy $H(f)$, subject to the constraint (2.8), occurs for the density

$$f_*(x) = Z^{-1} e^{-\nu\alpha(x)} \ , \tag{2.9}$$

where Z is defined by

$$Z = \int_X e^{-\nu\alpha(x)} dx \tag{2.10}$$

and ν is implicitly determined from

$$\langle \alpha \rangle = Z^{-1} \int_X \alpha(x) e^{-\nu\alpha(x)} dx \ . \tag{2.11}$$

Proof. The proof again uses the integrated Gibbs inequality. From (2.5), for densities f and f_*,

$$\begin{aligned} H(f) &\le -\int_X f(x) \log f_*(x) dx \\ &= -\int_X f(x)[-\log Z - \nu\alpha(x)] dx \\ &= \log Z + \nu \int_X f(x)\alpha(x) dx \\ &= \log Z + \nu \langle \alpha \rangle \ . \end{aligned}$$

However, it is equally trivial to show that

$$H(f_*) = \log Z + \nu \langle \alpha \rangle \tag{2.12}$$

and therefore $H(f) \le H(f_*)$, with the equality holding if and only if $f = f_*$.

The choice of notation in Eqs. (2.9) and (2.10) was intentional to draw the connection with the density of the Gibbs canonical ensemble. Thus if X is the conjugate position-momentum (q,p) space, $\alpha(x)$ is the system energy functional and $\langle \alpha \rangle$ the average (over the phase space) energy of the ensemble, then Z, as given by (2.10), can be identified directly with the *partition function*; the density f_* given in (2.9) that maximizes the entropy is to be identified with the density of the *Gibbs canonical ensemble.*

In deriving the density of the Gibbs canonical ensemble, it is implicit in the writing of the average $\langle \alpha \rangle$, Eq. (2.8), over the entire phase space that if the density (2.9) is to describe a state of thermodynamic equilibrium, then the quantity $\alpha(x)$ must eventually be independent of time at long times. An even more restrictive requirement would be that $\alpha(x)$ be a constant of the motion. If $\alpha(x)$ is identified with the system energy U, then we are dealing with a system in which the energy is conserved.

It is easy to extend this theorem to the case in which there are multiple known averages $\langle \alpha_i \rangle$. The resulting density maximizing the entropy is a generalization of the density of the grand canonical ensemble. Grad (1952) and Jaynes (1957) have followed this procedure.

E. The thermodynamic connection

All of conventional equilibrium thermodynamics can be derived from the density (2.9). Let us briefly see how.

The following is a fundamental assumption of thermodynamics.

Postulate B. There exists a one-to-one correspondence between states of thermodynamic equilibrium and states of maximum entropy.

If there is but one state of thermodynamic equilibrium that is attained regardless of the way in which the system is prepared, then this is called a globally stable equilibrium and is associated with a globally stable state of maximal entropy (third-order form of the second law). If, however, there are multiple states of thermodynamic equilibrium, each corresponding to a state of locally maximal entropy and dependent on the initial preparation of the system, then we say that these are local or metastable states of equilibrium (second-order, second law).

Given these observations, it is natural to examine the consequences of associating the equilibrium thermodynamic entropy S^*_{TD} with the maximal Boltzmann-Gibbs H given by (2.12):

$$H(f_*) \leftrightarrow S^*_{\mathrm{TD}} \ .$$

Since we have not specified units for H, a multiplicative constant c^{-1} may be necessary, viz.,

$$H(f_*) = S^*_{\mathrm{TD}}/c \ ,$$

which implies from Eq. (2.12) that the equilibrium thermodynamic entropy is given by

$$S^*_{\mathrm{TD}} = c \log Z + c\nu\langle \alpha \rangle \ .$$

If, as before, we make the association of $\langle \alpha \rangle$ with the internal thermodynamic energy U,

$$\langle \alpha \rangle = U$$

and if ν is associated with the inverse temperature

$$\nu = \frac{1}{kT} \ ,$$

where k is Boltzmann's constant, then the entropy expression in (2.12) becomes

$$S^*_{\mathrm{TD}} = c \log Z + cU/kT$$

or

$$TS^*_{\mathrm{TD}} = cT \log Z + (c/k)U \ . \tag{2.13}$$

If the constant c is taken to be identical with Boltzmann's constant $c = k$, then (2.13) immediately becomes

$$F = U - TS^*_{\mathrm{TD}} \ , \tag{2.14}$$

the fundamental equation of thermodynamics relating the *Helmholtz free energy*

$$F = -kT \log Z$$

to the internal energy U, temperature T, and equilibrium entropy S^*_{TD}.

Thus, by the simple use of Postulates A and B in conjunction with the identification of certain parameters and functions with corresponding quantities of thermodynamic interest, the fundamental relationship on which all of the equilibrium thermodynamics is based is the result. This illustrates the general nature of the assumptions (Postulates A and B) that are sufficient to derive thermodynamics and thus highlights the importance of understanding the basis of these two postulates.

F. Maximal entropy principles

As attempts to justify Postulate B on dynamical grounds increasingly met with failure, more and more authors tried to enshrine this hypothesis as a basic principle, often known as the *maximal entropy principle.* Tolman (1938) was the first to forcefully and clearly argue for this point of view. His argument was that since the techniques being used in thermodynamics were statistical in nature, one had to have some principle that would guide the selection of the proper density out of the unlimited number of possibilities. The maxmimal entropy principle certainly offers one such guide. Jaynes (1957), Scalapino (1961), and Katz (1967) have written extensively on the use of the maximal entropy principle in reformulating classical and quantum statistical mechanics, and Lewis (1967) has tried to justify it on dynamical grounds.

It may appear that the use of the maximal entropy principle gives a great deal (thermodynamics) for very little. Such is surely not the case. In actuality the hardest aspect of the understanding of thermodynamics is contained in Postulate A (which systems can reasonably be described by densities) and Postulate B (which of these systems, described by densities, will evolve toward equilibrium in such a way that the entropy is maximized). The advantage of the maximal entropy principle is that it strips away all of the apparent mystery concerning the definitions of the various ensembles and their corresponding densities and allows one to focus attention on the nature of the system dynamics necessary for guaranteeing the evolution of the system entropy to its maximal value. The remaining sections of this paper consider this problem.

III. REVERSIBLE AND IRREVERSIBLE SYSTEMS: GETTING THE ENTROPY TO INCREASE

A. Markov operators

In every situation considered by theoretical physics, as developed to this point in time, the evolution of densities may be studied by the use of the linear Markov or Frobenius-Perron operators. This is in spite of the fact that the underlying system dynamics responsible for the evolution of the density may be highly nonlinear.

The Frobenius-Perron operator, which is adjoint to the more familiar Koopman operator (Koopman, 1931), describes the evolution of densities in systems for which the dynamics are totally deterministic; i.e., the dynamics evolve according to a very specific law that permits the accurate specification of a system state at any point in time.

As a special and familiar example in which the evolution of the density is described by the Frobenius-Perron operator, consider a set of ordinary differential equations operating in R^d:

$$x_i' = F_i(x), \quad i = 1, \ldots, d \ . \tag{3.1}$$

As discussed in Lasota and Mackey (1985), starting from an initial density f, the evolution of the time-dependent density $f(t,x) \equiv P^t f(x)$ (here P is a Frobenius-Perron operator) is described by the *generalized Liouville equation*

$$\frac{\partial f}{\partial t} = -\sum_i \frac{\partial (fF_i)}{\partial x_i} \ . \tag{3.2}$$

The Frobenius-Perron operator is a special type of the more general Markov operator, which may be used in the description of both deterministic and stochastic systems. Since the first results on reversibility and irreversibility that are of importance to an understanding of thermodynamics can be stated for Markov operators, we start with them and defer the formal introduction of the Frobenius-Perron operator to the next section.

Any linear operator $P^t : L^1 \to L^1$ that satisfies

$$\text{(a) } P^t f \geq 0 \text{ and (b) } \|P^t f\| = \|f\|$$

for all $t \in R$ and $f \geq 0$, $f \in L^1$ is called a *Markov operator.* If we restrict ourselves to only considering densities f, then any operator P which when acting on a density again yields a density is a Markov operator.

As an example of a system in which the evolution of the density is governed by the operation of a Markov operator, consider the stochastic differential equations (Lasota and Mackey, 1985)

$$x_i' = F_i(x) + \sigma(x)\xi_i, \quad i = 1, \ldots, d \tag{3.3}$$

obtained when the system of ordinary differential equations (3.1) is perturbed by white noise ξ_i of amplitude $\sigma(x)$. Then, as discussed in Sec. VIII, starting from an initial density f the evolution of the time-dependent density $f(t,x) \equiv P^t f(x)$ (now P is a Markov operator) satisfies a modified form of the generalized Liouville equation,

$$\frac{\partial f}{\partial t} = -\sum_i \frac{\partial(fF_i)}{\partial x_i} + \frac{1}{2}\sum_{i,j} \frac{\partial^2(\sigma^2 f)}{\partial x_i \partial x_j}, \tag{3.4}$$

known as the *Fokker-Planck equation.*

Markov operators have a number of useful properties. The most important of these is that for all $f \in L^1$ (and not restricted to $f \geq 0$)

$$\|P^t f\| \leq \|f\|,$$

which is known as the *contractive property* of P^t. This contractive property implies that during the iteration of two functions f_1, f_2 by a Markov operator P, the distance $f = f_1 - f_2$ between them can only decrease and will never increase.

With the concept of the Markov operator we can introduce the important notion of a stationary density. If some density f_* satisfies $P^t f_* = f_*$ for all t, then f_* is called a *stationary density* of the Markov operator P^t. For the system of ordinary differential equations (3.1), the stationary densities f_* are given by the solutions of

$$-\sum_i \frac{\partial(f_* F_i)}{\partial x_i} = 0,$$

while for the stochastic differential equations (3.3) the f_* are the solutions of

$$-\sum_i \frac{\partial(f_* F_i)}{\partial x_i} + \frac{1}{2}\sum_{i,j} \frac{\partial^2(\sigma^2 f_*)}{\partial x_i \partial x_j} = 0.$$

The importance of stationary densities comes from the fact that the existence of a stationary density may be associated with a state of thermodynamic equilibrium.

In precise analogy with the definitions of dynamical and semidynamical systems in the last section, we may discuss reversible and irreversible Markov operators. Given a Markov operator P^t, then P^t is a *reversible Markov operator* if

$$\text{(a) } P^0 f = f$$

and

$$\text{(b) } P^t(P^{t'} f) = P^{t+t'} f \text{ for all } t, t' \in R \text{ (or } Z).$$

Clearly, allowing $t, t' \in R$ or Z is the origin of the reversibility.

As an example of a reversible Markov operator, again consider the evolution of the density f described by the Liouville equation (3.2) for dynamics determined by the system of ordinary differential equations (3.1). It is clear in (3.1) that replacing t by $-t$ gives dynamics described by $x_i' = -F_i(x)$, but that the corresponding Liouville equation is identical with (3.2).

However, if property (b) of a reversible Markov operator is replaced by

$$\text{(b}'\text{) } P^t(P^{t'} f) = P^{t+t'} f \text{ for all } t, t' \in R^+ \text{ (or } N),$$

then P^t is an *irreversible Markov operator.*

The Fokker-Planck equation (3.4) associated with the stochastic differential equation (3.3) is an example of an equation governing the evolution of an irreversible Markov operator. Going through the same exercise of replacing t by $-t$ in the stochastic differential equation yields a new stochastic differential equation different from (3.3) and a corresponding Fokker-Planck equation,

$$\frac{\partial f}{\partial t} = -\sum_i \frac{\partial(fF_i)}{\partial x_i} - \frac{1}{2}\sum_{i,j} \frac{\partial^2(\sigma^2 f)}{\partial x_i \partial x_j},$$

that is different from the original equation (3.4) since the sign of the term arising from the noise is reversed. Therefore the effect of noise in the reversible system (3.1) is to create a system in which the density evolution is governed by an irreversible Markov operator (cf. Sec. VIII.B).

To illustrate some of the concepts of this section, pick the simple discrete time transformation

$$S(x) = \begin{cases} 2x, & 0 \leq x < \frac{1}{2}, \\ 2(1-x), & \frac{1}{2} \leq x \leq 1, \end{cases} \tag{3.5}$$

so $x_{t+1} = S(x_t)$, $t \geq 0$. This noninvertible transformation is commonly known as the *tent map* after the appearance of its graph.

To investigate how the tent map (3.5) transforms densities we first derive an expression for the operator P that corresponds to this transformation. To do this, suppose first of all that we started with some initial density f that was transformed by one application of the map S into a new density Pf. Then the fraction of the density Pf contained in some interval $[0,x]$ is given by

$$\int_0^x Pf(s)\,ds.$$

However, the points in $[0,x]$ after one application of the map S, and which have contributed to Pf, had their origin in the counterimage of the interval $[0,x]$ under the action of S. This is denoted by $S^{-1}([0,x])$ and defined by $S^{-1}([0,x]) = \{y : S(y) \in [0,x]\}$. With these comments it must be the case that the fraction of the density Pf in the interval $[0,x]$ is equal to the fraction of the original density f in the set $S^{-1}([0,x])$, or that

$$\int_0^x Pf(s)ds = \int_{S^{-1}([0,x])} f(s)ds \ .$$

For the tent map it is a simple calculation to show that the counterimage of an interval $[0,x]$ is given explicitly by the union of two intervals:

$$S^{-1}([0,x])=[0,\tfrac{1}{2}x]\cup[1-\tfrac{1}{2}x,1] \ .$$

Substituting this into the previous integral expression between f and P gives the result

$$\int_0^x Pf(s)ds = \int_0^{x/2} f(s)ds + \int_{1-x/2}^{1} f(s)ds \ .$$

Finally, differentiating both sides of this equation with respect to x gives an explicit form for Pf in terms of f:

$$Pf(x)=\tfrac{1}{2}[f(\tfrac{1}{2}x)+f(1-\tfrac{1}{2}x)] \ . \tag{3.6}$$

This equation for the Frobenius-Perron operator Pf for the tent map (3.5) is also an irreversible Markov operator. It is a straightforward exercise to show that the uniform density $f_*=1$ is a stationary density for the operator P defined by (3.6), since it satisfies $Pf_*=f_*$.

Markov operators, reversible or irreversible, are quite general. In many situations it is possible to talk about the evolution of densities in physical systems by studying the properties of special types of Markov operators such as the Frobenius-Perron operator.

B. Markov operators and conditional entropy

Before starting an examination of the behavior of the entropy of a density under the action of a Markov operator, we introduce a generalization of the Boltzmann-Gibbs entropy, the *conditional entropy.* The definition can be motivated by examining the behavior of the Boltzmann-Gibbs entropy when the system dynamics are governed by the quadratic map

$$S(x)=rx(1-x) \tag{3.7}$$

operating on the phase space $[0,1]$ when $r=4$. (It is precisely for this value of r, and no other, that the map is onto.) This transformation, as the tent map, is not invertible and indeed can be obtained from the tent map (3.5) from $S_Q=h^{-1}\circ S_T\circ h$, where S_Q and S_T denote, respectively, the quadratic and tent maps, and the function h is given by

$$h(x)=\frac{1}{2}-\frac{1}{\pi}\sin^{-1}(1-2x) \ ;$$

so

$$h^{-1}(x)=\tfrac{1}{2}-\tfrac{1}{2}\cos(\pi x) \ .$$

To construct the Frobenius-Perron operator P governing the evolution of densities by the quadratic transformation (3.7), proceed as for the text map. Given a set $[0,x]$ and $r=4$, it is straightforward to show that the counterimage of this set consists of the union of two disjoint intervals:

$$S^{-1}([0,x])=[0,\tfrac{1}{2}(1-\sqrt{(1-x)})]\cup[\tfrac{1}{2}(1+\sqrt{(1-x)}),1] \ .$$

Thus the operator P corresponding to the quadratic map (3.7) is

$$Pf(x)=\frac{1}{4\sqrt{(1-x)}}[f(\tfrac{1}{2}-\tfrac{1}{2}\sqrt{(1-x)}) + f(\tfrac{1}{2}+\tfrac{1}{2}\sqrt{(1-x)})] \ . \tag{3.8}$$

It is straightforward to verify that the nonuniform density

$$f_*(x)=\frac{1}{\pi\sqrt{x(1-x)}} \tag{3.9}$$

is a stationary density of P defined by (3.8).

Now consider the Boltzmann-Gibbs entropy of a density evolving under the action of the operator P defined by Eq. (3.8). Pick as an initial density $f=1$; so clearly $H(f)=0$. Further, with this choice of initial density, then from (3.8)

$$Pf(x)=\frac{1}{2\sqrt{1-x}} \ ;$$

so the Boltzmann-Gibbs entropy of Pf is given by

$$H(Pf)=-\int_0^1 \frac{1}{2\sqrt{1-x}}\log\left[\frac{1}{2\sqrt{1-x}}\right]dx$$

$$=\log 2-1 \ .$$

Obviously, $H(Pf)<H(f)=0$, and the Boltzmann-Gibbs entropy has decreased. This is clearly unacceptable if we are to be able to draw a connection between the behavior of the Boltzmann-Gibbs entropy and the behavior of thermodynamic entropy.

The way out of this difficulty is simple and merely requires the definition of a generalization of the Boltzmann-Gibbs entropy called the conditional entropy.

If f and g are two densities such that supp$f \subset$ supp g (supp$f$ denotes the support of f), then the *conditional entropy* of the density f with respect to the density g is

$$H_c(f|g)=\int_X g(x)\eta\left[\frac{f(x)}{g(x)}\right]dx$$

$$=-\int_X f(x)\log\left[\frac{f(x)}{g(x)}\right]dx \ . \tag{3.10}$$

The conditional entropy, we note, is always defined; i.e., H_c is finite or equal to $-\infty$, since g is a density and η is bounded from above. As is evident from the defining equation (3.10), $H_c(f|g)$ measures the deviation of f from the density g.

There are two important properties of $H_c(f|g)$.

(1) Since f and g are both densities, the integrated Gibbs inequality (2.5) implies that $H_c(f|g)\le 0$. It is only when $f=g$ that the equality holds.

(2) If g is the constant density of the microcanonical

ensemble, i.e., $g=1/\mu_L(X)$ throughout the phase space X, then $H_c(f|g)=H(f)-\log\mu_L(X)$. If the space is normalized, then $g=1$ and $H_c(f|g)=H_c(f|1)=H(f)$. This illustrates how the conditional entropy is a generalization of the Boltzmann-Gibbs entropy.

From the definition of $H_c(f|g)$ it follows that

$$H_c(f|g)=H(f)+\int_X f(x)\log g(x)dx \ .$$

An elementary calculation using property 1 of $H(f|g)$ shows that the second term in the rewritten form of $H_c(f|g)$, with $f=g=f_*$, is just $-H(f_*)$ and that $H_c(f|f_*)=0$ when $f=f_*$. These observations, in conjunction with our formulations of the second- and third-order forms of the second law, immediately suggest that the conditional entropy H_c can be interpreted as the entropy difference ΔS. For example, under the conditions of Theorem 2.2, the Boltzmann-Gibbs entropy $H(f)$ is maximized by the density f_* given by Eq. (2.9) and

$$H(f_*)=\log Z+\nu\langle\alpha\rangle \ . \tag{3.11}$$

Thus within the context of Theorems 2.1 and 2.2 we conclude that the conditional entropy will be zero whenever the Boltzmann-Gibbs entropy is at its maximum value.

With only the few tools developed so far and our identification of H_c with ΔS, the behavior of the entropy of a sequence of densities $\{P^tf\}$ evolving under the action of a Markov operator may be examined.

The first result along these lines is a weak (first-order) form of the second law of thermodynamics stating that the conditional entropy is never decreasing, and is stated more precisely as follows.

Theorem 3.1 (Voigt, 1981). Let P^t be a Markov operator. Then

$$H_c(P^tf|P^tg)\ge H_c(f|g)$$

for all densities f and g.

Notice that in this theorem if g is a stationary density of P^t, $g=f_*$ where $P^tf_*=f_*$, then

$$H_c(P^tf|f_*)\ge H_c(f|f_*) \ .$$

Thus the conditional entropy with respect to a stationary density is always a nondecreasing function bounded above by $H_{\max}=H_c(f_*|f_*)=0$. In examining the behavior of $H_c(P^tf|f_*)$ we therefore know that it converges as $t\to\infty$, though more information about P^t is required to define the limiting value.

Furthermore, in the special case that the system is operating on a finite space and the Markov operator P^t has the density of the microcanonical ensemble as a stationary density, i.e., $P^t1=1$, then this theorem implies that

$$H(P^tf)\ge H(f)$$

for all non-negative f. Coupling this with the observation from Sec. II that on a finite space the maximum entropy is $H_{\max}=-\log[1/\mu_L(X)]$ we have

$$H_{\max}=-\log\left[\frac{1}{\mu_L(X)}\right]\ge H(P^tf)\ge H(f)$$

for all t; so, once again, we have convergence of $H(P^tf)$ to a maximum as $t\to\infty$.

To this point nothing has been said about the reversibility or irreversibility of the Markov operator P^t with respect to the behavior of the entropy. However, this distinction turns out to be crucial, since the entropy for a reversible Markov operator is constant. More precisely, for reversible Markov operators, we can state the weakest (zero-order) form of the second law of thermodynamics in the following theorem.

Theorem 3.2. If P^t is a reversible Markov operator, then the conditional entropy is absolutely constant for all times t and equal to the value determined by the choice of the initial densities f and g. That is,

$$H_c(P^tf|P^tg)=H_c(f|g)$$

for all t.

Proof. Since P^t is reversible, by the previous theorem it follows that

$$\begin{aligned}H_c(P^{t+t'}f|P^{t+t'}g)&=H_c(P^{t'}P^tf|P^{t'}P^tg)\\&\ge H_c(P^tf|P^tg)\ge H_c(f|g)\end{aligned}$$

for all t,t'. Pick $t'=-t$ so that for all times t

$$H_c(f|g)\ge H_c(P^tf|P^tg)\ge H_c(f|g)$$

and therefore

$$H_c(P^tf|P^tg)=H_c(f|g)$$

for all t.

From this theorem any system whose evolution of densities is described by a reversible Markov operator has an entropy that is forever fixed at a value determined by the initial state. Or, put another way, the entropy is uniquely determined by the method of preparation of the system.

In particular, for the system of ordinary differential equations (3.1) whose density evolves according to the Liouville equation (3.2), we can assert that the entropy of the density P^tf will be constant for all time and will have the value determined by the initial density f with which the system is prepared. This result can also be proved directly by noting that from the definition of the conditional entropy we may write

$$H_c(f|f_*)=-\int_{R^d}f(x)\left[\log\left[\frac{f}{f_*}\right]+\frac{f_*}{f}-1\right]dx$$

when the stationary density is f_*. Differentiating with respect to time gives

$$\frac{dH_c}{dt}=-\int_{R^d}\frac{df}{dt}\log\left[\frac{f}{f_*}\right]dx \tag{3.12}$$

or, after substituting from (3.2) for $(\partial f/\partial t)$ and integrating by parts under the assumption that f has compact support,

$$\frac{dH_c}{dt}=\int_{R^d}\frac{f}{f_*}\sum_i\frac{\partial(f_*F_i)}{\partial x_i}dx \ .$$

However, since f_* is a stationary density of P^t, it is clear from (3.2) that

$$\frac{dH_c}{dt}=0 \ ,$$

and we conclude that the conditional entropy $H_c(P^tf|f_*)$ does not change from its initial value when the dynamics evolve in this manner.

Thus, not too surprisingly, we arrive at the general conclusion that *irreversibility in system dynamics,* as reflected in an evolution of densities via an irreversible Markov operator, *is necessary for the entropy to increase* as the system evolves. We cannot, however, assert that irreversibility is sufficient to guarantee this, and indeed it is not the case.

Based on much more specific assumptions, this result was well known to the founders of modern thermodynamic theory. Clausius (1879) and Boltzmann (1909) tried to circumvent this clear problem associated with the use of reversible (Hamiltonian) dynamics by their *Stosszahlansatz* (molecular chaos) postulate. Furthermore, the argument used in the proof of Theorem 3.2 was used, in a much more specialized form, by Loschmidt (1876) in his *Umkehreinwand* (objection based on time reversal) argument against the Boltzmann approach to statistical mechanics.

Spohn (1978) has put forward a slightly different interpretation of the conditional entropy defined in (3.10) when g is taken to be a unique stationary density f_*. This is done by writing a (local phenomenological) balance equation for the entropy density $\mathcal{S}$

$$\frac{d\mathcal{S}}{dt}=-J_{\mathcal{S}}+\sigma \ ,$$

where $J_{\mathcal{S}}$ is interpreted as an energy flow and σ (≥ 0) as an entropy production. In the special case of a system coupled to a reservoir at a single temperature he identifies $\mathcal{S}$ with the Boltzmann-Gibbs entropy $H(f)$, the entropy production σ with

$$\sigma=\frac{dH_c(f|f_*)}{dt} \ ,$$

and the energy flow $J_{\mathcal{S}}$ with

$$J_{\mathcal{S}}=\frac{d}{dt}\int_X f(x)\log f_*(x)dx \ .$$

These relations follow immediately as a special case of the definition of the conditional entropy if one takes $g=f_*$, writes (3.10) in the form

$$H(f)=-\int_X f(x)\log f_*(x)dx+H_c(f|f_*) \ ,$$

and takes the time derivative. Thus in this interpretation the conditional entropy $H_c(f|f_*)$ is to be viewed as the entropy production integrated over time. It is a trivial consequence of Theorem 3.1 that $\sigma\geq 0$ and that system irreversibility is necessary (though not sufficient) for $\sigma>0$.

C. Smoothing Markov operators and asymptotic periodicity

Next we turn to an investigation of a fascinating property that may be displayed by the evolution of densities in discrete time systems. This behavior, called asymptotic periodicity, is the statistical analog for densities of the more common periodicity found in some time series and will allow us to prove a stronger (second-order) form of the second law.

First, we must define what is meant by a smoothing Markov operator. A Markov operator P^t is said to be *smoothing* if there exists a set A of finite measure and two constants $k<1$ and $\delta>0$ such that for every set E with $\mu_L(E)<\delta$ and every density f there is some integer $t_0(f,E)$ for which

$$\int_{E\cup(X\setminus A)}P^tf(x)dx\leq k \quad \text{for } t\geq t_0(f,E) \ .$$

This definition of a smoothing Markov operator just means that any initial density, no matter how small a region of the phase space X it is concentrated on, will eventually be smoothed out by P^t. Smoothing operators are important because of a theorem of Komornik and Lasota (1987), first proved in a more restricted situation by Lasota *et al.* (1984).

Theorem 3.3, spectral decomposition theorem (Komornik and Lasota, 1987). Let P^t be a smoothing Markov operator. Then there is an integer $r>0$, a sequence of non-negative densities g_i and a sequence of bounded linear functionals λ_i, $i=1,\ldots,r$, and a bounded linear operator $Q:L^1\rightarrow L^1$ such that for all densities f, Pf has the form

$$Pf(x)=\sum_{i=1}^{r}\lambda_i(f)g_i(x)+Qf(x) \ . \tag{3.13}$$

The densities g_i and the operator Q have the following properties.

(1) The g_i have disjoint support (i.e., are mutually orthogonal); so $g_i(x)g_j(x)=0$ for all $i\neq j$.

(2) For each integer i there is a unique integer $\alpha(i)$ such that $Pg_i=g_{\alpha(i)}$. Furthermore, $\alpha(i)\neq\alpha(j)$ for $i\neq j$. Thus the operator P permutes the densities g_i.

(3) $\|P^tQf\|\rightarrow 0$ as $t\rightarrow\infty$, $t\in N$.

Notice from Eq. (3.13) that P^tf may be immediately written in the form

$$P^tf(x)=\sum_{i=1}^{r}\lambda_i(f)g_{\alpha^t(i)}(x)+Q_tf(x), \quad t\in N \ , \tag{3.14}$$

where $Q_t=P^{t-1}Q$, $\|Q_tf\|\rightarrow 0$ as $t\rightarrow\infty$, and $\alpha^t(i)=\alpha(\alpha^{t-1}(i))=\cdots$. The terms in the summation of (3.14) are just permuted by each application of P^t.

Since r is finite the series

$$\sum_{i=1}^{r} \lambda_i(f) g_{\alpha^t(i)}(x) \tag{3.15}$$

must be periodic with a period $T \leq r!$. Further, as

$$\{\alpha^t(1), \ldots, \alpha^t(i)\}$$

is just a permutation of $1, \ldots, r$ the summation (3.15) may be written in the alternative form

$$\sum_{i=1}^{r} \lambda_{\alpha^{-t}(i)} g_i(x) ,$$

where $\alpha^{-t}(i)$ is the inverse permutation of $\alpha^t(i)$.

This rewriting of the summation portion of (3.14) makes the effect of successive applications of P completely transparent. Each operation of P permutes the set of scaling coefficients associated with the densities $g_i(x)$ (remember that these densities have disjoint support).

Since f is finite and the summation (3.15) is periodic (with a period bounded above by $r!$), and $\|Q_t f\| \to 0$ as $t \to \infty$, we say that for any smoothing Markov operator the sequence $\{P^t f\}$ is *asymptotically periodic.*

One of the interesting interpretations of Eq. (3.14) is that any asymptotically periodic system is quantized from a statistical point of view. Thus if t is large enough, which simply means that we have observed the system longer than its relaxation time so $\|Q_t f\|$ is approximately zero, then

$$P^t f(x) \simeq \sum_{i=1}^{r} \lambda_i(f) g_{\alpha^t(i)}(x) .$$

Asymptotically, $P^t f$ is equal either to one of the densities g_i of the ith pure state or to a mixture of the densities of these states each weighted by $\lambda_i(f)$. It is important also to realize that the limiting sequence $\{P^t f\}$ is, in general, dependent on the choice of the initial density f.

How would the property of asymptotic periodicity be manifested in a continuous time system? If t is continuous, $t \in R^+$, then for every t we can find a positive integer T and a number $\theta \in [0,1]$ such that $t = T + \theta$. Then, asymptotically

$$P^t f(x) = P^T(P^\theta f) \simeq \sum_{i=1}^{r} \lambda_i(P^\theta f) g_{\alpha^T(i)}(x) .$$

Now, however, the discontinuous permutation of the densities g_i that occurred in the discrete time situation will not take place. This is because these densities have disjoint support and there is therefore no possibility of a continuous movement of the densities through a continuous permutation of their associated scaling coefficients. Thus in the continuous time case we expect that there will be a constant (in time) asymptotic limiting density, generally dependent on the initial density f, that continues to display the quantized nature characteristic of the discrete time situation.

Asymptotically periodic Markov operators always have at least one stationary density given by

$$f_*(x) = \frac{1}{r} \sum_{i=1}^{r} g_i(x) , \tag{3.16}$$

where r and $g_i(x)$ are as in the Komornik-Lasota theorem. It is easy to see that $f_*(x)$ is a stationary density, since by Property 2 of the Komornik-Lasota theorem we also have

$$Pf_*(x) = \frac{1}{r} \sum_{i=1}^{r} g_{\alpha(i)}(x) ;$$

thus f_* is a stationary density of P^t. Hence for any smoothing Markov operator the stationary density (3.16) is just the average of the densities g_i.

To examine the properties of asymptotically periodic systems, choose a simple generalization of the tent map (3.5),

$$S(x) = \begin{cases} ax - a + 2, & 0 \leq x < 1 - \dfrac{1}{a} , \\ a(1-x), & 1 - \dfrac{1}{a} \leq x \leq 1 , \end{cases} \tag{3.17}$$

where $1 < a < 2$.

To investigate how the map (3.17) transforms densities we must first derive an expression for the operator P that corresponds to this transformation. Proceeding exactly analogously as in the early part of this section when the tent map was introduced, it is a simple calculation to show that the Frobenius-Perron operator corresponding to (3.17) is given by

$$Pf(x) = \frac{1}{a} \left[f\left[1 - \frac{1}{a}x\right] + f\left[1 - \frac{1}{a}(2-x)\right] 1_B(x) \right] , \tag{3.18}$$

where the set $B = [2-a, 1]$. We set $1_B(x) = 1$ if $x \in B$ and $1_B(x) = 0$ if $x \notin B$. The sequence of densities $\{P^t f\}$ generated by the Frobenius-Perron operator corresponding to the tent map is asymptotically periodic with period $T = 2^n$, $n = 0, 1, \ldots,$ for $2^{1/T} > a \geq 2^{1/2T}$ (Dorfle, 1985; Lasota and Mackey, 1985). The quadratic map (3.7) is asymptotically periodic for a variety of values of $r < 4$ (Lorenz, 1980).

To illustrate analytically the eventual dependence of the sequence $\{P^t f\}$ on the initial density f for asymptotically periodic systems, pick $a = \sqrt{2}$, which is the upper boundary of the range of a values for which (3.17) is asymptotically periodic with period 2. For this value of a, the stationary density (3.16) satisfying $Pf_* = f_*$, where P is given by (3.18), takes the explicit form

$$f_*(x) = r 1_A(x) + s 1_B(x) , \tag{3.19}$$

where $r = \frac{1}{4}(2+\sqrt{2})$, $s = \frac{1}{2}(1+\sqrt{2})$, and the sets A and B are defined as $A = [0, 2-\sqrt{2}]$ and $B = [2-\sqrt{2}, 1]$, respectively. Picking this $f_*(x)$ as an initial density simply results in a string of densities all equal to the starting density. However, this is in sharp contrast to what happens with an initial density $f_0(x) = 1$. In this case, the first iterate $f_1=Pf_0$ is given by

$$f_1(x)=\frac{1}{a}1_A(x)+\frac{2}{a}1_B(x)\ , \tag{3.20}$$

and iteration of $f_1(x)$ leads, in turn, to an $f_2(x)=1$; thus the cycling of densities repeats indefinitely with period 2.

The fact that asymptotically periodic Markov operators have a stationary density given by (3.16) does not guarantee the uniqueness of this stationary density. Regardless of whether or not they have unique stationary densities, they have the important property that their conditional entropy is an increasing function that approaches a maximum. This result is formulated more precisely in the following.

Theorem 3.4. Let P be an asymptotically periodic Markov operator with stationary density $f_*=(1/r)\sum g_i$. Then the conditional entropy $H_c(P^tf|f_*)$ of P^tf with respect to f_* approaches a limiting value $H_{\max}(f,f_*)\le 0$, where

$$H_{\max}(f,f_*)=-\sum_i\int_X\lambda_i(f)g_i(x)\log\left[\frac{1}{f_*(x)}\sum_i\lambda_i(f)g_i(x)\right]dx\ .$$

Proof. Since P is asymptotically periodic, the representation of the spectral decomposition theorem 3.3 is valid, and more precisely Eq. (3.14) for P^tf. Write Eq. (3.14) in the form

$$P^tf(x)=S_t(f,x)+Q^tf(x)\ ,$$

where $S_t(f,x)$ denotes the summation portion of (3.14). Remember that since P is asymptotically periodic, for large times t, $\|Q^tf\|\simeq 0$ and thus $P^tf(x)\simeq S_t(f,x)$; so the long-time conditional entropy is given by

$$H_c(P^tf|f_*)\simeq-\int_X S_t(f,x)\log\left[\frac{S_t(f,x)}{f_*(x)}\right]dx=H_c(S_t(f)|f_*)\ .$$

However, $S_t(f,x)$ is periodic with finite period T. Since by Theorem 3.1 we also know that $H_c(P^tf|f_*)\ge H_c(f|f_*)$ (the conditional entropy can never decrease), it follows that the approach of $H_c(P^tf|f_*)$ to $H_c(S_t(f)|f_*)$ must be uniform; even though $S_t(f,x)$ is periodic with a finite period T, $H_c(S_t(f)|f_*)$ is a constant independent of t. In fact, we have

$$H_c(S_t(f)|f_*)=-\int_X\sum_i\lambda_i(f)g_i(x)\log\left[\frac{1}{f_*(x)}\sum_i\lambda_i(f)g_i(x)\right]dx\equiv H_{\max}(f,f_*)\le 0$$

for large t. The nonpositivity of $H_{\max}(f,f_*)$ is a consequence of (2.5).

This result is the strongest (second-order) form of the second law of thermodynamics that we have yet encountered. It demonstrates that as long as the density in a discrete time system evolves under the operation of a Markov operator that is smoothing, the conditional entropy of that density converges to a maximum. However, there are two important facets of this evolution that should be recognized.

(1) The convergence of the entropy is only due to the fact that $\|Q^tf\|\to 0$ as $t\to\infty$ in the representation (3.14) of the Komornik-Lasota theorem 3.3.

(2) The maximum value of the entropy $H_{\max}(f,f_*)$ as made explicit by the notation, is generally dependent on the choice of the initial density f and thus the method of preparation of the system.

To illustrate the evolution of the conditional entropy of an asymptotically periodic system we return to our example of the tent map (3.17) with $a=\sqrt{2}$. For this value of a, a stationary density f_* is given by Eq. (3.19). If we pick an initial density given by f_*, then the conditional entropy $H_c(P^tf_*|f_*)=0$, its maximal value, for all t. However, if we pick an initially uniform density $f_0=1$ then it is straightforward to show that

$$H_c(f_0|f_*)=H_c(f_1|f_*)\simeq-0.01479\ ,$$

where $f_1=Pf_0$ is given by Eq. (3.20). Thus by choosing an initial density $f_0(x)=1$ of $f_1(x)$ given by (3.20), the limiting conditional entropy approaches a value less than its maximal value of zero.

The effect of the choice of the initial density affecting the limiting value of the conditional entropy can be made even more dramatic, as illustrated by choosing an initial density

$$f_0(x)=\left[1+\frac{\sqrt{2}}{2}\right]1_A(x)$$

totally supported on the set A. In this case,

$$f_1(x)=Pf_0(x)=(1+\sqrt{2})1_B(x)\ ,$$

and $f_2=f_0$, $f_3=f_1$, etc.; so once again the densities cycle between f_0 and f_1. However, the limiting value of the conditional entropy is given by

$$H_c(f_0|f_*)=H_c(f_1|f_*)=-\log(2)$$
$$\simeq-0.69315\ .$$

Thus with three different choices of an initial density f_0 we have shown that the conditional entropy of the asymptotically periodic system (3.17) may have at least three different limiting asymptotic values.

IV. ERGODICITY: GUARANTEEING A UNIQUE STATE OF THERMODYNAMIC EQUILIBRIUM

A. The Frobenius-Perron operator

To formalize the concept of the Frobenius-Perron operator introduced in the previous section, a few concepts must be introduced. A transformation S_t is said to be *measurable* if $S_t^{-1}(A)\subset X$ for all $A\subset X$. Furthermore, given a density f_* and associated measure μ_*, a measurable transformation S_t is *nonsingular* if $\mu_*(S_t^{-1}(A))=0$ for all sets A such that $\mu_*(A)=0$.

If S_t is a nonsingular transformation, then the unique operator $P^t:L^1\to L^1$ defined by

$$\int_A P^t f(x)dx=\int_{S_t^{-1}(A)} f(x)dx \tag{4.1}$$

is called the *Frobenius-Perron operator* corresponding to S.

From our comments of Sec. III and the formal definition, if f is a density, then Eq. (4.1) defining the Frobenius-Perron operator has a simple intuitive interpretation. We start with an initial density f and integrate this over set B that will evolve into set A under the action of the transformation S_t. However, set B is $S_t^{-1}(A)$. This integrated quantity must be equal, since S_t is nonsingular, to the integral over set A of the density obtained after one application of S_t to f. This final density is $P^t f$.

The fact that the Frobenius-Perron operator is unique is a straightforward consequence of the Radon-Nikodym theorem. It is clear from the definition that the Frobenius-Perron operator is a Markov operator, and so P^t is a linear contracting operator. Moreover, if $f\geq 0$, then $P^t f\geq 0$ and $\|P^t f\|=\|f\|$. Finally, it is easy to show that if $S_{nt}=S_t\circ\cdots\circ S_t$, and P^{nt} and P^t are, respectively, the Frobenius-Perron operator corresponding to S_{nt} and S_t, then $P^{nt}=P^t\circ\cdots\circ P^t=(P^t)^n$.

Sometimes (as for the examples of the tent and quadratic maps of Sec. III) the implicit defining equation (4.1) for the Frobenius-Perron operator allows one to write P^t explicitly. A general formula, valid for higher-dimensional spaces $X=R^d$ when S_t is invertible, may be derived by a change of variables in the definition of the Frobenius-Perron operator to give

$$P^t f(x)=f(S_{-t}(x))J^{-t}(x)\ , \tag{4.2}$$

where J^{-t} is the Jacobian of S_{-t}.

Given a density f and associated measure μ_f, then a measurable transformation S_t is said to be *f-measure preserving* if

$$\mu_f(S_t^{-1}(A))=\mu_f(A)$$

for all sets A. Measure-preserving transformations are necessarily nonsingular. Since the concept of measure preservation is not only dependent on the transformation but also on the measure, we alternately say that the *measure μ_f is invariant* with respect to the transformation S_t if S_t is f-measure preserving. For example, we can say either that the tent map (3.5) preserves the Lebesgue measure or that the Lebesgue measure is invariant with respect to the tent map. In an entirely analogous fashion we say that the quadratic map (3.7) with $r=4$ preserves the measure defined by

$$\mu_*(x)=\int_0^x f_*(s)ds=\frac{1}{2}-\frac{1}{\pi}\sin^{-1}(1-2x)\ ,$$

where the density f_* is the stationary density (3.9) of the Frobenius-Perron operator P, given by (3.8), corresponding to the quadratic map. We also express this by saying that the measure μ_* is invariant with respect to the quadratic map. We note that the quadratic map does not preserve the Lebesgue measure and that μ_* is just the function h of Sec. III used in transforming between the quadratic and tent maps $S_Q=\mu_*^{-1}\circ S_T\circ\mu_*$.

It is possible to make an interesting connection among states of thermodynamic equilibrium, invariant measures, and stationary densities of the Frobenius-Perron operator through the following theorem.

Theorem 4.1 (Lasota and Mackey, 1985). Let S_t be a nonsingular transformation and P^t the Frobenius-Perron operator associated with S_t. Then there exists a state of thermodynamic equilibrium whose density f_* is a stationary density of P^t if and only if the measure μ_*,

$$\mu_*(A)=\int_A f_*(x)dx\ ,$$

is invariant.

This theorem is interesting in the following sense. Since the density maximizing the entropy must be an equilibrium density in a state of thermodynamic equilibrium, if the densities are evolving under the action of a Frobenius-Perron operator P^t then it is clear that the equilibrium density must also be a stationary density of P^t. In particular, the density $f_*=1$ of the microcanonical ensemble corresponds to a state of thermodynamic equilibrium if and only if the system dynamics preserve the Lebesgue measure. That is, systems preserving the Lebesgue measure may be appropriately described by the microcanonical ensemble. Of course it is important to realize that this theorem says nothing about either the uniqueness of this state of thermodynamic equilibrium or of the invariant measure corresponding to it.

B. Recurrence

For nonsingular transformations S_t with an invariant measure operating in a phase space X, a point x in a subset A of the phase space X is called a *recurrent point* if there is some time $t>0$ such that $S_t(x)$ is also in A. An important result, which deals with recurrent points, has become known as the Poincaré recurrence theorem.

Theorem 4.2 (Poincaré recurrence theorem). Let S_t be a transformation within an invariant measure μ_* operating in a finite phase space X, $\mu_*(X)<\infty$, and let A be a subset of X with positive f_* measure. Then there exists a point x in A that is recurrent.

Proof [adapted from Petersen (1983)]. Assume the contrary, i.e., that there are no recurrent points in A. This then implies that $S_t^{-1}(A)\cap A=\varnothing$ for all times $t>0$, and thus that $S_t^{-1}(A)\cap S_{t'}^{-1}(A)=\varnothing$ for all positive times $t\neq t'$. However, since S_t is measure preserving, this implies that $\mu_*(S_t^{-1}(A))=\mu_*(S_{t'}^{-1}(A))$ and this, coupled with the pairwise disjoint nature of the sets $S_t^{-1}(A)$ and $S_{t'}^{-1}(A)$, leads to

$$\sum_{t=0}^{\infty}\mu_*(A)=\sum_{t=0}^{\infty}\mu_*(S_t^{-1}(A))=\mu_*\left[\bigcup_{t=0}^{\infty}S_t^{-1}(A)\right]$$

$$\leq\mu_*(X)\leq\infty\ .$$

The only way in which this inequality can be satisfied is for $\mu_*(A)$ to be zero, which is a contradiction. Thus we conclude that A contains recurrent points.

It is interesting that the proof of the Poincaré recurrence theorem also demonstrates that the set of nonrecurrent points has measure zero; so almost every point is recurrent. Furthermore, repeated application of the theorem tells us that a recurrent point of A will return to A infinitely often.

Following Loschmidt's (1876) objections to Boltzmann's attempt to justify thermodynamics, this recurrence result was used by Zermelo (1896) as the basis for an attack on Boltzmann's celebrated H theorem concerning the behavior of the entropy. In what has become known as the *Wiederkehreinwand* (objection based on recurrence), Zermelo argued that, because of recurrence, almost all points would constantly revisit the same areas of phase space and thus it would be impossible for the entropy to ever monotonically increase to its maximum.

Zermelo was right in his assertion that the entropy of a system whose dynamics are governed by Hamilton's equations, or any set of ordinary differential equations for that matter, cannot change, as we have proved in Sec. III.B. He was wrong, however, to base his argument on the result of the Poincaré recurrence theorem. The fallacy in the argument is to be found in his implicit assumption that densities (on which the behavior of the entropy depends) will behave as points and also be recurrent as, perhaps, in the spectral decomposition theorem of Sec. III. Just because points are recurrent, densities need not be; indeed, in Sec. VI we give a necessary and sufficient condition for the entropy of a system to increase to its maximum that is completely compatible with the Poincaré recurrence theorem.

C. Ergodicity

We are ready to begin consideration of the characteristics any dynamical system or semidynamical system S_t must have to guarantee that there exists a unique state of thermodynamic equilibrium that maximizes the entropy. The density maximizing the entropy should also be an equilibrium density, so our search is really one for the properties of S_t necessary to guarantee that a density f_* is a stationary density of the Frobenius-Perron operator corresponding to S_t, i.e., $P^tf_*=f_*$, and that f_* is the only stationary density.

We start by defining a few new terms, given, as usual, dynamics described by a transformation S_t. First, any set A such that $S_t^{-1}(A)=A$ is called an *invariant set.* Given a density f_* on a space, X, any invariant set A such that $\mu_*(A)=0$ or $\mu_*(X\setminus A)=0$ is called *trivial.* A nonsingular transformation S_t is said to be f_* *ergodic* if every invariant set A is a trivial subset of the phase space X; i.e., either $\mu_*(A)=0$ or $\mu_*(X\setminus A)=0$. If the phase space is finite and f_* is the uniform density of the microcanonical ensemble, then we say that S_t is *uniformly ergodic* instead of f_* ergodic.

Probably the simplest example of a physical system that is ergodic is a one-dimensional harmonic oscillator (O. Penrose, 1979); but to illustrate the property of ergodicity, we consider a simple dynamical system in which momentum (p) and position (q) evolve according to

$$\frac{dp}{dt}=\alpha,\quad \frac{dq}{dt}=\beta,\quad \alpha\neq\beta\ , \tag{4.3}$$

on a unit torus (doughnut). Imagine cutting the torus apart in both directions and laying it down on the plane so that we have a phase space $X=[0,1)\times[0,1)$. Then, starting from any initial point (q_0,p_0) in X, the trajectory emanating from that initial position is given by

$$p(t)=p_0+\alpha t,\quad q(t)=q_0+\beta t\ (\mathrm{mod}1)\ , \tag{4.4}$$

as depicted by the solid lines in Fig. 1.

If either α or β is irrational, then (4.3) is an example of a uniformly ergodic transformation (Arnold and Avez, 1968). If, instead of examining the behavior of a single trajectory we look at the motion of many initial conditions located in some subset of the phase space X (as an approximation to the behavior of a density), then we obtain the behavior of Fig. 2 in which is shown the first six interactions of an initial 10^4 points located in the subset $[0,\frac{1}{10}]\times[0,\frac{1}{10}]$ by the uniformly ergodic transformation

$$S(x,y)=(\sqrt{2}+x,\sqrt{5}+y)\ (\mathrm{mod}1)\ .$$

obtained from Eq. (4.4) by restricting the time t to the set of non-negative integers and picking $\alpha=\sqrt{2}$, $\beta=\sqrt{5}$. This clearly illustrates an important property that ergodic transformations *may* display; namely, nearby initial conditions always remain close to one another in spite of the fact that in the limit as $t\rightarrow\infty$ the entire phase space has eventually been visited.

This example of an ergodic transformation (if α and/or

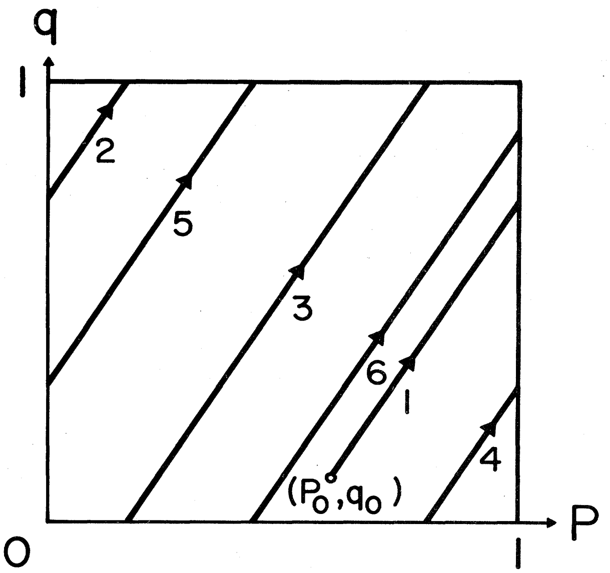

FIG. 1. Single trajectory of the ergodic system (4.3) in (p,q) phase space, starting from the initial point (p_0,q_0); $\alpha=\sqrt{2}$ and $\beta=\sqrt{5}$, thus ensuring uniform ergodicity. The numbers indicate successive portions of the trajectory and the arrows show the direction.

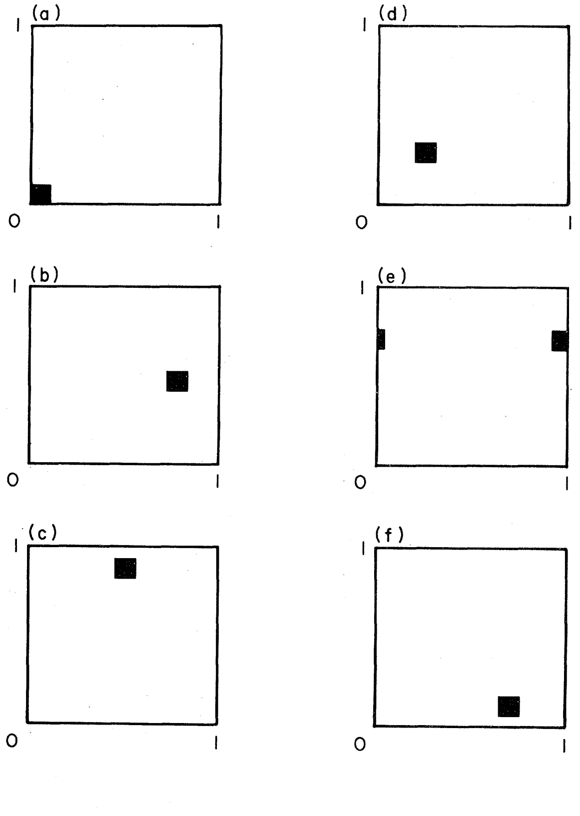

FIG. 2. Iteration of 10^4 initial points concentrated on a small region of phase space by a discrete time version of (4.4); values for α and β are the same as in Fig. 1.

β are irrational) can also be viewed as a Hamiltonian system

$$\frac{dp}{dt}=-\frac{\partial H}{\partial q},\quad \frac{dq}{dt}=\frac{\partial H}{\partial p}kh\ ,$$

where the Hamiltonian H is given by $H=-\alpha q+\beta p$. However, that substitution of the explicit expressions for $p(t)$ and $q(t)$ from Eq. (4.4) into the Hamiltonian gives $H=\beta p_0-\alpha q_0$, and thus the "energy" of this ergodic system is forever determined by its initial values.

By themselves, neither the definition of ergodicity nor its illustration is terribly instructive with respect to the question of how one might prove ergodicity and the relation of ergodicity to the existence of stationary densities and to states of thermodynamic equilibrium. However, the following result makes the connection between the property of ergodicity and the uniqueness of a state of thermodynamic equilibrium clear.

Theorem 4.3 (Lasota and Mackey, 1985). Let S_t be a nonsingular transformation and P^t the corresponding Frobenius-Perron operator. S_t is f_* ergodic if and only if P^t has a unique state of thermodynamic equilibrium with associated stationary density f_*, $P^t f_*=f_*$.

What does this result, in conjunction with Theorem 4.1, tell us? First consider the microcanonical ensemble with its uniform density. Then a given dynamical system or semidynamical system S_t will be measure preserving with respect to the Lebesgue measure if and only if the uniform density of the microcanonical ensemble is a stationary density of the Frobenius-Perron operator corresponding to S_t. Furthermore, from Theorem 4.3 the uniform density of the microcanonical ensemble will be the unique stationary density of P^t if and only if the system S_t is uniformly ergodic. Hence the existence of a unique state of thermodynamic equilibrium, characterized by the uniform density of the microcanonical ensemble that maximizes the Boltzmann-Gibbs entropy to zero, is totally dependent on the operation of a uniformly ergodic dynamical system or semidynamical system that preserves the Lebesgue measure.

In the more general case, the nonuniform density f_* of the canonical ensemble that maximizes the conditional entropy will be the unique density corresponding to a state of thermodynamic equilibrium if and only if it is the stationary density of the Frobenius-Perron operator corresponding to an f_* ergodic system S_t with respect to which the measure

$$\mu_*(A)=\int_A f_*(x)dx$$

is invariant.

Thus in complete generality *ergodicity is necessary and sufficient to guarantee the existence of a unique state of thermodynamic equilibrium characterized by maximal entropy.* That this unique state exists is, of course, only half of the picture, for we must also understand what kind of systems can evolve to that state.

To conclude this section we state one last theorem concerning necessary and sufficient conditions for the f_* ergodicity of a transformation S_t.

Theorem 4.4 (Lasota and Mackey, 1985). Let S_t be a nonsingular transformation and P^t the corresponding Frobenius-Perron operator with stationary density $f_* > 0$ for all points in the phase space X. Then S_t is f_* ergodic if and only if $\{P^t f\}$ is Cesàro convergent to f_* for all densities f, i.e., if

$$\lim_{t\to\infty} \frac{1}{t}\sum_{k=0}^{t-1} \langle P^k f, g\rangle = \langle f_*, g\rangle$$

in the discrete time case, or if

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \langle P^t f, g\rangle dt = \langle f_*, g\rangle$$

in the continuous time case, for all bounded measurable functions g.

Our identification of states of thermodynamic equilibrium with densities, in conjunction with the contents of Theorems 4.3 and 4.4, shows that f_* ergodicity, Cesàro convergence of $\{P^t f\}$ to f_*, and the uniqueness of a state of thermodynamic equilibrium with density f_* are all equivalent.

V. MIXING: BEYOND ERGODICITY BUT NOT FAR ENOUGH

A. Mixing

Gibbs (1902) realized that ergodicity, while necessary and sufficient to guarantee the existence of a unique state of thermodynamic equilibrium characterized by a stationary density f_*, was inadequate to guarantee the approach of a system to that equilibrium. As a consequence he qualitatively discussed a property stronger than ergodicity, which is now called (*strong*) *mixing*. This was subsequently developed mathematically by Koopman (1931), Hopf (1932), and von Neumann (1932).

Let S_t be an f_*-measure-preserving transformation operating on a finite normalized $[\mu_*(X)=1]$ space. Then S_t is called f_* *mixing* if

$$\lim_{t\to\infty} \mu_*(A \cap S_t^{-1}(B)) = \mu_*(A)\mu_*(B) \tag{5.1}$$

for all sets A and B. If f_* is the uniform density of the microcanonical ensemble, then in analogy with our definition of ergodicity we say that S_t is *uniformly mixing*.

To see how f_* mixing works examine Fig. 3 where the evolution of 10^4 points by the uniformly mixing transformation

$$S(x,y) = (x+y, x+2y) \pmod 1$$

is shown in the phase space $[0,1]\times[0,1]$. In contrast to the uniformly ergodic transformation of Fig. 1, the mixing transformation here acts to very quickly spread the initial set of points throughout phase space in threadlike structures.

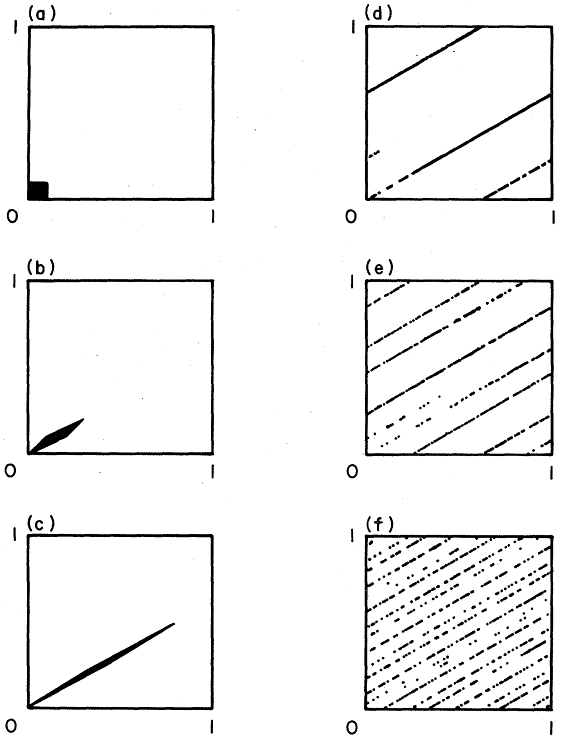

FIG. 3. Iteration of 10^4 initial points under the action of a uniformly mixing transformation. Note the development of the threadlike structures.

As a second example of a uniformly mixing transformation we introduce the *baker transformation*, so called because of the similarity of its operation to the kneading and folding operations involved in the preparation of pastry dough. We take the phase space X to be the unit square $X=[0,1]\times[0,1]$ and define the baker transformation by

$$S(x,y) = \begin{cases} (2x, \tfrac{1}{2}y), & 0 \le x \le \tfrac{1}{2}, \\ (2x-1, \tfrac{1}{2}+\tfrac{1}{2}y), & \tfrac{1}{2} < x \le 1 . \end{cases} \tag{5.2}$$

To see pictorially how the baker transformation works, consult Fig. 4(a) where the space X is indicated as a hatched region. The first operation involved in the application of S, shown in Fig. 4(b), involves a compression of X in the y direction and a concomitant stretching in the x direction by a factor of 2. Finally, in Fig. 4(c), this deformed area is divided vertically at $x=1$, and the right-hand portion is placed on top of the left-hand portion to give the final result of one application of S to the entire space X. Since it is clear from this geometric construction that the counterimage of any rectangle A in X is again a rectangle or a pair of rectangles with the same total area as A, the baker transformation S is measurable and preserves the Lebesgue measure.

To further illustrate the operation of the baker transformation and to show that it is in fact uniformly mixing, consider Fig. 5(a) where two sets A and B of the phase space X are indicated with $\mu_L(B)=\frac{1}{2}$. By taking repeated counterimages of B with the baker transformation we find that after t counterimages (three are shown), $S^{-t}(B)$ consists of 2^{t-1} vertical rectangles of equal area whose total area is $\mu_L(B)$. Clearly the measure of the set $A \subset S^{-t}(B)$ approaches $\mu_L(B)/2$ in the limit as t becomes large and the condition (5.1) defining f_* mixing is satisfied. (A more mathematical proof with less insight is possible by analytically calculating the successive counterimages.)

If, instead of running the baker transformation backward we run it forward in time [Fig. 5(b)], then the behavior is remarkably similar to that found for counterimages. Namely, after t applications of the baker transformation $S^t(B)$ consists of 2^t horizontal rectangles whose measure is equal to $\mu_L(B)$. It is this characteristic of mixing transformations that gives rise to the thready behavior seen in Fig. 3.

The behavior of the uniformally mixing baker transformation as time runs either backward [Fig. 5(a)] or forward [Fig. 5(b)] simply reflects the fact that the relation (5.1) could equally well be written as

$$\lim_{t\to\infty} \mu_*(A \cap S_t(B)) = \mu_*(A)\mu_*(B) .$$

This is a consequence of the fact that S_t is f_*-measure preserving

A third example of a mixing transformation in a continuous time system is given by a model of an ideal gas in which the position of the ith particle is denoted by a position vector x_i and a velocity vector v_i; so $y_i=(x_i,y_i)$ is a point in R^6. It is assumed that the particles in the gas are physically indistinguishable from one another, and that the gas is so dilute that for any bounded region of the phase space there are at most a finite number of particles present at any given time. Under the assumption that the particles move with constant speed and do not

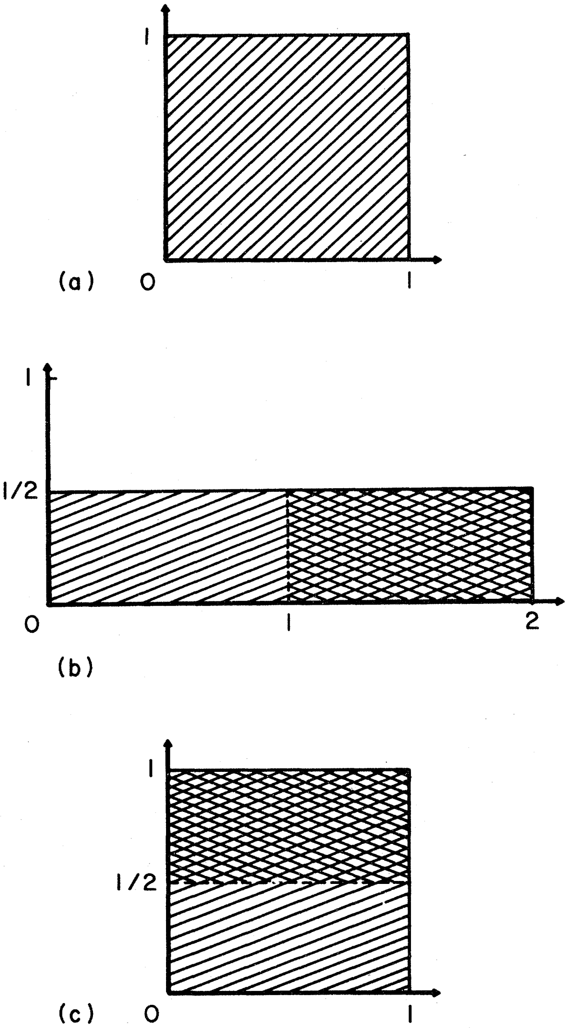

FIG. 4. Operation of the baker transformation. See the text for details.

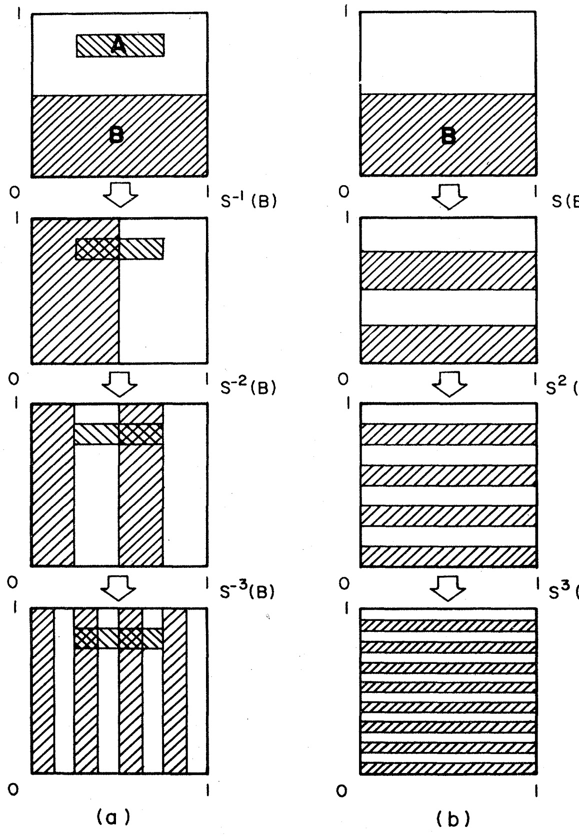

FIG. 5. Behavior of an initial set B under backward (a) and forward (b) iteration by the uniformly mixing baker transformation, and the generation of the thready behavior as seen in Fig. 3. See the text for more discussion.

interact, the transformation describing the evolution of this system is given by

$$S_t(x,v)=(x+vt,v) \ . \tag{5.3}$$

The surprising result is that this simple system of noninteracting particles is in fact f_* mixing (Losota and Mackey, 1985).

There is a whole class of continuous time systems of importance in classical mechanics that are mixing. These systems, known as geodesic flows on manifolds with negative curvature, have been intensively studied by Anosov (1963, 1967) and discussed in Arnold and Avez (1968) and Abraham and Marsden (1978). Sinai (1963, 1970) has shown that the Boltzmann-Gibbs model of a dilute gas (spherical molecules with totally elastic collisions) is an example of this type of system and is thus mixing.

an der Heiden (1985) has studied the delay differential equation

$$\frac{dx}{dt}=a[f(x(t-1))-x(t)]$$

with a suitably defined initial function $\phi(t')$, $t'\in[-1,0]$. It is known that this equation, with a mixed feedback function of the form $f(z)=z/(1+z^n)$, $n>0$, may have numerical solutions that are stable, periodic, or chaotic (Mackey and Glass, 1977; an der Heiden and Mackey, 1982). an der Heiden uses a piecewise constant mixed feedback function f of the form

$$f(z)=\begin{cases} 0, & z<1 \ , \\ c, & 1\le z\le b \ , \\ d, & b<z \ , \end{cases}$$

identifies the sequence of times t_i such that $x(t_i)=1$, and derives an analytic map S relating successive differences in crossing times, $\Delta_i=t_{2i}-t_{2i-1}$,

$$\Delta_{i+1}=S(\Delta_i) \ .$$

He has proved that for some parameter values (a,b,c,d) the Frobenius-Perron operator P corresponding to S is either asymptotically periodic or mixing.

It is a straightforward consequence of the definition that f_* mixing implies ergodicity. Furthermore, an f_*-measure-preserving transformation S_t, with associated Frobenius-Perron operator P^t and stationary density f_*, is f_* mixing if and only if the sequence $\{P^tf\}$ is weakly convergent to the density f_* for all initial densities f. If $f_*=1$, then P^t is uniformly mixing if and only if $\{P^tf\}$ is weakly convergent to the density of the microcanonical ensemble for all initial densities f.

This is expressed more formally in the following.

Theorem 5.1 (Lasota and Mackey, 1985). Let S_t be an ergodic transformation, with stationary density f_* of the associated Frobenius-Perron operator, operating in a phase space of finite f_* measure. Then S_t is f_* mixing if and only if $\{P^tf\}$ is weakly convergent to f_*, i.e.,

$$\lim_{t\to\infty}\langle P^tf,g\rangle=\langle f_*,g\rangle \ .$$

As usual, given a measure μ, the *scalar product* of an L^1 function f with a bounded measurable function g, denoted by $\langle f,g\rangle$, is defined by

$$\langle f,g\rangle=\int_X f(x)g(x)\mu(dx) \ .$$

Gibbs (1902), Krylov (1979), and many later authors have emphasized the importance of mixing for the understanding of thermodynamic behavior. Indeed, at first one might think that this weak convergence of the sequence $\{P^tf\}$ to the density f_* of the canonical ensemble, or to the density $f_*=1$ of the microcanonical ensemble, no matter what initial density f was chosen, would be exactly what is required to guarantee the approach of the entropy (conditional or otherwise) to its maximum. Such is not the case.

It is most certainly true that mixing is necessary for this convergence of the entropy, but it is also not sufficient, as we show in Sec. VI. As a simple illustration of this fact we again return to our example of the uniformly mixing baker transformation. The baker transformation (5.2) is clearly invertible, so

$$S^{-1}(x,y)=\begin{cases} (\frac{1}{2}x,2y), & 0\le y\le\frac{1}{2} \ , \\ (\frac{1}{2}+\frac{1}{2}x,2y-1), & \frac{1}{2}<y\le 1 \ , \end{cases} \tag{5.4}$$

and we may apply Eq. (4.2) directly to obtain an expression for the Frobenius-Perron operator corresponding to S. The only other fact that we must use is that S preserves the Lebesgue measure so $J^{-t}=1$, which is easily verified directly from (5.4). Thus

$$P_Sf(x,y)=\begin{cases} f(\frac{1}{2}x,2y), & 0\le y\le\frac{1}{2} \ , \\ f(\frac{1}{2}+\frac{1}{2}x,2y-1), & \frac{1}{2}<y\le 1 \ ; \end{cases}$$

so $P_S1=1$, and thus the density of the microcanonical ensemble is a stationary density of P_S.

Since the stationary density of P_S is the uniform density (it is easily proved that $f_*=1$ is the unique stationary density) we may calculate the entropy of P_Sf as

$$H(P_Sf)=-\int_0^{1/2}\int_0^1 f(\tfrac{1}{2}x,2y)\log f(\tfrac{1}{2}x,2y)\,dx\,dy-\int_{1/2}^1\int_0^1 f(\tfrac{1}{2}+\tfrac{1}{2}x,2y-1)\log f(\tfrac{1}{2}+\tfrac{1}{2}x,2y-1)\,dx\,dy \ .$$

A change of variables on the right-hand side in the two integrals gives

$$H(P_Sf)=H(f) \ ,$$

and we have proved that $H(P_S^tf)=H(f)$ for all positive times t and all densities f. A simple extension of this argument extends the result to negative times when the baker transformation is being run backward. Thus the uniformly mixing baker transformation has an entropy that is constant for all of time and which is equal to the

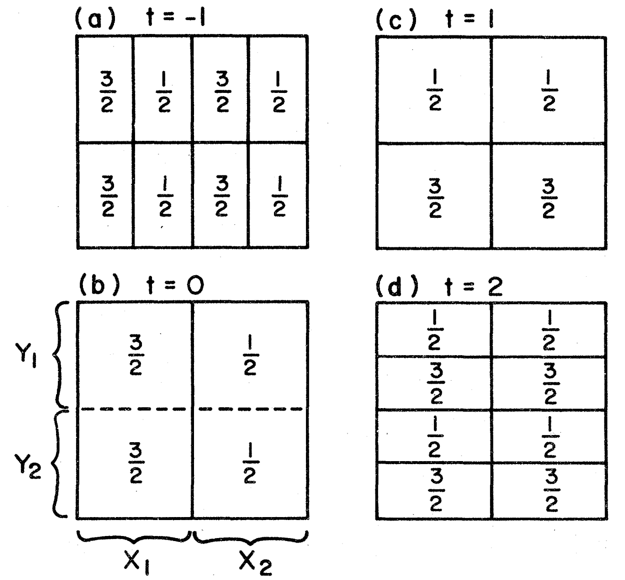

FIG. 6. Behavior of an initial nonuniform density for backward and forward iteration of the baker transformation. See Secs. V.A and VII.A for a discussion of the corresponding changes in entropy.

entropy of the initial density f with the system was prepared.

This is simply illustrated in Fig. 6. At $t=0$ we prepare the phase space with an initial density

$$f_0(x,y)=\tfrac{3}{2}1_{X_1}(x)+\tfrac{1}{2}1_{X_2}(x) \ , \tag{5.5}$$

where $X_1=[0,\frac{1}{2})$ and $X_2=[\frac{1}{2},1]$. Then the entropy of the initial state is

$$H(f_0)=-\tfrac{1}{4}(\log\tfrac{1}{2}+3\log\tfrac{3}{2})\simeq -0.13 \ .$$

After one application of the baker transformation [Fig. 6(c)] we have

$$f_1(x,y)=\tfrac{1}{2}1_{Y_1}(y)+\tfrac{3}{2}1_{Y_2}(y) \ ,$$

where $Y_1=[0,\frac{1}{2})$ and $Y_2=[\frac{1}{2},1]$. It is a trivial matter to show that $H(f_1)\equiv H(f_0)$. Further applications of the baker transformation yield $H(f_t)\equiv H(f_0)$ [Fig. 6(d) for $t=2$], which also holds for negative times [Fig. 6(a) for $t=-1$] since the baker transformation is invertible.

B. Kolmogorov automorphisms

In this section a concept that will be used later is briefly intoduced, namely, that of Kolmogorov automorphisms, or K automorphisms. We first introduce the notion of a σ algebra $\mathcal{A}$, which is a collection of subsets of X such that (1) if $A\in\mathcal{A}$, the $X\setminus A\in\mathcal{A}$; (2) given a sequence of subsets of X, $\{A_i\}$, $A_i\in\mathcal{A}$, then

$$\bigcup_i A_i\in\mathcal{A} \ ;$$

and (3) $X\in\mathcal{A}$.

We use the notation $S_t(\mathcal{A})=\{S_t(A):A\in\mathcal{A}\}$, $t=0,\pm1,\pm2,\ldots$, where $\mathcal{A}$ is a σ algebra. Then if S_t is an invertible transformation operating on a normalized space, and both S_t and S_{-t} are f_*-measure preserving, S_t is said to be a K *automorphism* if there is a σ algebra $\mathcal{A}_0\in\mathcal{A}$ such that (1) $S_{-t}(\mathcal{A}_0)\in\mathcal{A}_0$; (2) the σ algebra defined by

$$\bigcap_{t=0}^{\infty} S_{-t}(\mathcal{A}_0)$$

is trivial in the sense that it only contains sets of f_*-measure 0 or 1; and (3) the smallest σ algebra containing

$$\bigcup_{t=0}^{\infty} S_t(\mathcal{A}_0)$$

is identical to $\mathcal{A}$.

Kolmogorov automorphisms have behaviors stronger than mixing in that if a transformation is a K automorphism, then this also implies that it is f_* mixing (Walters, 1982). The other property of K automorphisms that is important for thermodynamic considerations is that, since they are f_*-measure preserving, they have a unique stationary density f_*. However, since they are invertible, the entropy of a K automorphism is identically equal to the initial entropy determined by the initial density with which the system is prepared. In fact, our familiar example of the baker transformation, which we have used to illustrate uniform mixing, is an example of a Kolmogorov automorphism (Lasota and Mackey, 1985).

VI. THE GLOBAL EVOLUTION OF ENTROPY TO ITS MAXIMUM

A. Exactness

To extend the catalog of irregular behaviors that transformations may exhibit we define the concept of exactness.

If S_t is an f_*-measure-preserving transformation operating on a normalized phase space X, then S_t is said to be f_* *exact* if

$$\lim_{t\to\infty}\mu_*(S_t(A))=1$$

for all sets A of nonzero measure. If f_* is the uniform density, $f_*=1$, then we say that S_t is *uniformly exact.*

To understand the nature of exactness, it is important to realize that reversible systems cannot be exact. To see this, note that for a reversible f_*-measure-preserving transformation S_t we have $\mu_*(S_t(A))=\mu_*[S_t^{-1}(S_t(A))]=\mu_*(A)$. Thus the definition of exactness is violated.

A demonstration of the operation of an exact transformation, similar to that for ergodicity and mixing, is helpful in showing how exact systems operate. Figure 7 shows the first six iterates of 10^4 points randomly distributed in $[0,\frac{1}{10}]\times[0,\frac{1}{10}]$ under the action of the uniformly exact transformation

$$S(x,y)=(3x+y,x+3y) \pmod 1 . \tag{6.1}$$

It is clear that the behavior is quite different from an f_* mixing transformation. Under the action of an exact transformation an initial set A is very quickly dispersed throughout the entire phase space X.

A second example of a uniformly exact transformation is given by the tent map originally introduced in Sec. III, Eq. (3.5). The tent map preserves the Lebesgue measure; if we start with an initial set $B=[0,b]$, then a simple geometric argument (try it) suffices to show that after a finite number of iterations $\mu_L(S_t(B))=1$ and the transformation is uniformly exact. A more precise proof can be carried out using the behavior of the evolution of densities by the Frobenius-Perron operator contained in Theorem 6.1 below.

Exact systems are important for an understanding of how convergence to a stationary density f_* of the canonical ensemble may be reached in a way that is an extension of mixing. To be specific, we have the following.

Theorem 6.1 (Lasota and Mackey, 1985). If S_t is an f_*-measure-preserving transformation operating on a finite normalized phase space X and P^t is the associated Frobenius-Perron operator corresponding to S_t, then S_t is f_* exact if and only if

$$\lim_{t\to\infty}\|P^tf-f_*\|=0 ,$$

i.e., $\{P^tf\}$ is strongly convergent to f_*, for all initial densities f.

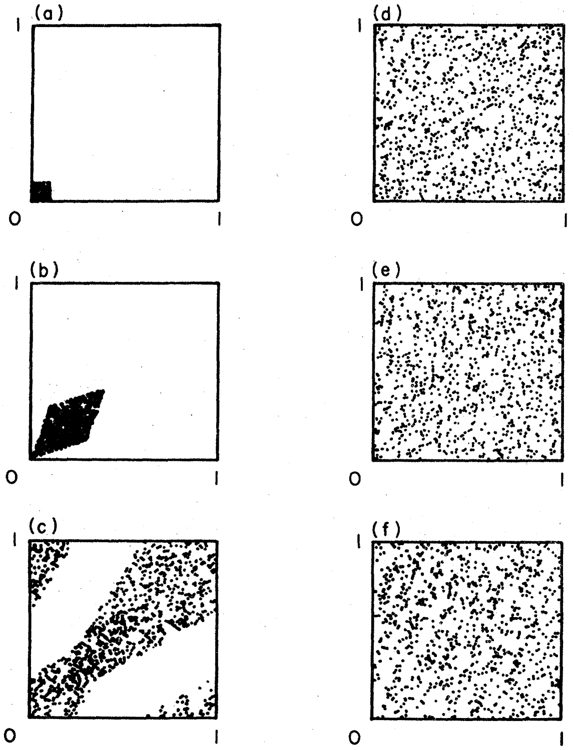

FIG. 7. Behavior of an initial 10^4 points under the action of the uniformly exact transformation (6.1). Note in particular how rapidly these points spread throughout the phase space without developing the threadlike structures seen in Fig. 3 for a uniformly mixing transformation.

This theorem offers a necessary and sufficient condition for the exactness of S_t in complete analogy with the previously presented necessary and sufficient ergodicity and mixing conditions.

As an illustration of Theorem 6.1, consider the dyadic transformation

$$S(x)=2x \pmod 1 , \tag{6.2}$$

which is a special case of the r-adic transformation $S(x)=rx \pmod 1$ when r is a positive integer. The dyadic transformation (6.2) is clearly not invertible. A simple calculation shows that it preserves the Lebesgue measure so the constant density $f_*=1$ is a stationary density. The counterimage of an interval $[0,x]$ under the action of the dyadic transformation is

$$S^{-1}([0,x])=[0,\tfrac{1}{2}x]\cup[\tfrac{1}{2},\tfrac{1}{2}+\tfrac{1}{2}x] ;$$

so the Frobenius-Perron operator corresponding S is given by

$$P_Sf(x)=\tfrac{1}{2}[f(\tfrac{1}{2}x)+f(\tfrac{1}{2}+\tfrac{1}{2}x)] .$$

It is obvious that $P_S1=1$. By an inductive argument, P_S^tf can be written in the form

$$P_S^tf(x)=\frac{1}{2^t}\sum_{i=0}^{2^t-1} f\left[\frac{i+x}{2^t}\right] .$$

In the limit as $t\to\infty$, the right-hand side of this equation simply becomes the Riemann integral of the density f over $[0,1]$, i.e.,

$$\lim_{t\to\infty}P_S^tf(x)=\int_0^1 f(x)dx=1 .$$

Therefore the dyadic transformation is uniformly exact, and precisely the same argument generalized in a trivial way extends to the r-adic transformation when $r>2$ is an integer.

We can use this same technique to show that the tent map (3.5) is uniformly exact. Further, the uniform exactness of the tent map implies the f_* exactness of the quadratic transformation (3.7), with f_* given by Eq. (3.9), through the following result for one-dimensional maps.

Theorem 6.2 (Lasota and Mackey, 1985). Let $T:[0,1]\to[0,1]$ be a nonsingular transformation, f_* be a positive density on (a,b), and $S:[0,1]\to[0,1]$ be a second transformation defined by $S=h^{-1}\circ T\circ h$ where

$$h(x)=\int_a^x f_*(y)dy,\quad a<x<b .$$

Then T is uniformly exact if and only if S is f_* exact.

The f_* exactness of a transformation implies that it is f_* mixing. To show this it suffices to remember that, from the Cauchy-Holder inequality, for a sequence of functions $\{f_t\}$ we have

$$|\langle f_t - f, g\rangle| \le \|f_t - f\| \cdot \|g\| \ .$$

Thus the convergence of $\|f_t - f\|$ to zero implies that $\langle f_t - f, g\rangle$ also converges to zero; strong convergence therefore implies weak convergence. As a consequence f_* exactness implies f_* mixing, just as f_* mixing implies f_* ergodicity.

We have so far defined three different types of irregular behaviors that f_*-measure-preserving transformations may display: ergodicity, mixing, and exactness. Furthermore, in each case we have been able to make strong statements about the types of convergence of iterates of densities by the Frobenius-Perron operator toward the stationary density characterizing a state of thermodynamic equilibrium that one should find. Since Frobenius-Perron operators are specialized Markov operators, there is a certain logic to extending these concepts to Markov operators. Thus let P^t be a Markov operator and assume that P^t has a stationary density f_*. Then we say the following.

(a) P^t is f_*. *ergodic* if $\{P^t f\}$ is Cesàro convergent to f_* for all initial densities f.

(b) P^t is f_* *mixing* if $\{P^t f\}$ is weakly convergent to f_* for all initial densities f.

(c) P^t is f_* *exact* if $\{P^t f\}$ is strongly convergent to f_* for all initial densities f.

Operators P^t that are f_* exact have been called strong Markov operators by Misra *et al.* (1979) and monotonic Markov operators by Goldstein *et al.* (1981). Since strong convergence implies weak convergence, which, in turn, implies Cesàro convergence, for Markov operators we have immediately that f_* exactness implies f_* mixing implies f_* ergodicity.

B. The local evolution of entropy to a maximum

This section is devoted to a statement of a sufficient condition for the entropy of a sequence of densities evolving under the action of a Markov operator to approach a relative maximal value. That is, we have a sufficient condition for the second-order form of the second law of thermodynamics.

Theorem 6.3 (Lasota and Mackey, 1985). Let P be a Markov operator working in a normalized measure space, and assume that there is a stationary density f_* of P. If there is a constant $c > 0$ such that for every bounded initial density f

$$H_c(P^t f | f_*) \ge -c$$

for sufficiently large t, then P^t is asymptotically periodic and

$$\lim_{t \to \infty} H_c(P^t f | f_*) = H_{\max}(f, f_*) \le 0 \ .$$

This theorem assures us that if we are able to find some time t_1 such that the conditional entropy is bounded below for times $t > t_1$, then the entropy is evolving under the action of an asymptotically periodic Markov operator and, as a consequence of Theorem 3.4, the conditional entropy of $P^t f$ approaches a maximum that is generally dependent on the initial density with which the system was prepared.

C. The strong (third-order) version of the second law of thermodynamics

Our next result is much stronger in that we consider a Markov operator P that has a stationary density f_* that is not necessarily constant, thus corresponding to the density of the canonical ensemble, and define a simple necessary and sufficient condition for the conditional entropy to go to its absolute maximal value of zero. Thus we are offering necessary and sufficient conditions for the second law of thermodynamics to operate in its strongest possible (third-order) form. Namely, we have the following.

Theorem 6.4 (Lasota and Mackey, 1985). Let P^t be a Markov operator operating in a phase space X. Then the conditional entropy of $P^t f$ with respect to a density f_* goes to its maximal value of zero as $t \to \infty$,

$$\lim_{t \to \infty} H_c(P^t f | f_*) = 0 \ ,$$

if and only if P^t is f_* exact.

A simple example of this theorem is given by the quadratic map (3.7), which is f_* exact by Theorem 6.2 and thus has a smooth increase of the conditional entropy to zero at long times, irrespective of the initial system density.

This theorem is remarkable in that it sets forth *necessary and sufficient criteria for the operation of the strongest form of the second law of thermodynamics,* namely, for the entropy of a system to globally converge to its maximal value regardless of the way in which the system was prepared. The only requirement that the system must satisfy is that *the density must evolve under the action of an exact operator.* If this operator is a Frobenius-Perron operator, then the dynamics must be f_* exact. Since f_* exactness implies f_* ergodicity, the state of thermodynamic equilibrium characterized by the density f_* is unique.

There is an obvious simple corollary to this theorem concerning the behavior of the Boltzmann-Gibbs entropy of systems operating in a finite normalized phase space when the stationary density is that of the microcanonical ensemble $f_* = 1$. Namely, we have the following.

Corollary 6.5. If P is a Markov operator operating in a finite normalized phase space X, then the Boltzmann-Gibbs entropy of $P^t f$ approaches its maximal value of zero as $t \to \infty$; i.e.,

$$\lim_{t\to\infty} H(P^t f)=0 \ ,$$

if and only if P is uniformly exact.

Thus the Boltzmann-Gibbs entropy will converge to its maximal value of zero if and only if the density of the microcanonical ensemble is a stationary density and the system evolves under the action of a uniformly exact Markov operator P. As before, ergodicity of P guarantees that the uniform density of the microcanonical ensemble is the unique state of thermodynamic equilibrium, while the uniform exactness of P guarantees that the entropy will approach its maximum value of zero regardless of the way in which the system is prepared. Both the r-adic transformation and the tent map are uniformly exact by our previous observations, and thus display a smooth evolution of their Boltzmann-Gibbs entropy to zero for all initial system states.

Thus, for finite normalized spaces,

$$P^t \text{ is uniformly exact} \Longleftrightarrow \lim_{t\to\infty} H(P^t f)=0 \ ,$$

while, in general,

$$P^t \text{ is } f_* \text{ exact} \Longleftrightarrow \lim_{t\to\infty} H_c(P^t f|f_*)=0 \ .$$

These results point out a very interesting property of the entropy *vis-à-vis* the common notion that maximal entropy should be associated with maximal disorder or minimal structure. Experimentally, what we *measure* is that the entropy of a system evolving in time goes to a maximum. However, in the course of any experiment, the *dynamics* are the ultimate selector of the proper f_* with respect to which the conditional entropy is "computed" by the system evolution. This state of maximal entropy, in turn, corresponds to a state of thermodynamic equilibrium and in no way makes a judgment about whether this state is totally structureless ($f_*=1$) or highly ordered. Any apparent inconsistency between a state of maximal entropy and a nonuniform $f_*\neq 1$ comes exclusively from the erroneous identification of $f_*=1$ as the preferred state of thermodynamic equilibrium. This stems from the long historical preoccupation of trying to find a rational foundation for thermodynamics in the statistical mechanics of Hamiltonian systems, which do preserve the Lebesgue measure and for which the attendant density $f_*=1$ is a stationary density.

With the results of this section giving necessary and sufficient conditions for the approach of system entropy to a maximum, one might think that our quest for the dynamical foundations of thermodynamics and the functioning of the second law was at an end. However, this is far from the reality of the situation, as a moment's reflection reveals.

Here it has been demonstrated that it is only through the operation of *irreversible* f_* exact systems that the entropy will increase to its maximal value. Further, given the observation that dynamics are the ultimate determinant of the stationary density f_*, that this corresponds to a state of thermodynamic equilibrium, and that since states of thermodynamic equilibrium depend on a variety of parameters (temperature, pressure, etc.), we must conclude that the corresponding f_* must also depend on these parameters *as must the underlying dynamics*. Given these results, we are now faced with another problem, since all of the laws of physics are framed in terms of reversible or invertible dynamical (as opposed to irreversible or noninvertible semidynamical) systems that are independent of these external parameters.

This dilemma seems to have at least two solutions; either (1) or (2).

(1) The laws of physics are at present incorrectly formulated. R. Penrose (1979) has argued quite lucidly and simply for this point of view, basing his thesis on CPT violation in K^0 meson decay. Fer (1977) makes a similar point, basing his argument on the neglect of time delays in the usual formulations of physical laws. Davies (1974) and Gal-Or (1974, 1981) have extensively examined possible sources of time asymmetry in physics, primarily from a cosmological and electromagnetic perspective.

(2) There is some effect, neglected up to this point in our considerations, that alters the behavior of a dynamical system to give rise to the observed behavior.

The remainder of this paper is devoted to an exploration of the second of these possibilities only, as the first involves a drastic restructuring of the entire formulation of classical and quantum physics.

In the next section we start by examining the closely linked concepts of coarse graining and traces. Nontrivial coarse graining, in spite of the fascination that it has held for decades as a potential source of irreversible thermodynamic behavior, is unable to stand as a viable candidate because it induces an increase in entropy to its maximal value of zero regardless of the direction of time (forward or backward). However, a more extreme form of coarse graining, known as a trace and involving the existence of hidden variables, is able to induce truly irreversible thermodynamic behavior.

VII. COARSE GRAINING AND TRACES

A. Coarse-grained entropy

In calculating the entropy from the defining equations, it has been assumed that the dynamical variables were known with infinite precision. As a consequence, the density f corresponding to a given thermodynamic state would also be known precisely. While this is the situation when an analytic form for the density is available, in the world of experiment the reality is that the density f (or, more usually, some functional of f) is either measured or estimated.

Several consequences may ensue from this. The first and perhaps most obvious is that due to errors (arising, for example, from measurement impreciseness or numerical roundoff in computer experiments), f will not be known exactly but will be known only to some level of accuracy.

Alternately, it is entirely possible that nature may have introduced an inherent graininess to phase space, rendering the absolute determination of dynamical variables, and thus densities, impossible. Many have suggested that perhaps there is an elementary fundamental volume in position-momentum space whose measure corresponds to Planck's constant. This would be entirely in keeping with other apparently fundamental indivisible units in the real world.

To examine the effect of such imprecision in the measurement of dynamical variables on the estimation of the entropy, we introduce the concept of the entropy of a coarse-grained density, or more briefly, the *coarse-grained entropy.* This concept seems to have been qualitatively discussed first by Gibbs (1902) and quantified by Ehrenfest and Ehrenfest (1959). Denbigh and Denbigh (1985) have considered aspects of the effects of coarse graining on the behavior of entropy.

Coarse graining is carried out by first partitioning the phase space X into discrete cells A_i that satisfy

$$\bigcup_i A_i = X \quad \text{and} \quad A_i \bigcup_{i \neq j} A_j = \varnothing \, . \tag{7.1}$$

There is no unique way in which such a partition $\{A_i\}$ may be formed, but we require that the partition be *nontrivial* with respect to some measure μ so $0 < \mu(A_i) < \mu(X)$ for all values of i. For every density f, within each cell A_i of this partition, we denote the average of f over A_i by $\langle f \rangle_i$,

$$\langle f \rangle_i = \frac{1}{\mu(A_i)} \int_{A_i} f(x)\mu(dx) \; ; \tag{7.2}$$

so the density f coarse grained with respect to the partion A_i is given by

$$f^{\mathrm{cg}}(x) = \sum_i \langle f \rangle_i 1_{A_i}(x) \; . \tag{7.3}$$

Clearly $\sum_i \langle f \rangle_i \mu(A_i) = 1$, and it is important to realize that f^{cg} is constant within each cell A_i, having the value given by Eq. (7.2).

Therefore, if given a partition A_i satisfying (7.1) (nontrivial with respect to Lebesgue measure), a density f, and a coarse-grained density f^{cg} defined by Eqs. (7.2) and (7.3), then the Boltzmann-Gibbs entropy of the coarse-grained density f^{cg} is given by

$$H(f^{\mathrm{cg}}) = -\sum_i \langle f \rangle_i \mu_L(A_i) \log \langle f \rangle_i \; .$$

The demonstration is almost trivial since, from (7.3) and the definition of the entropy, we have

$$\begin{aligned}
H(f^{\mathrm{cg}}) &= -\int_X \left[\sum_i \langle f \rangle_i 1_{A_i}(x) \right] \log \left[\sum_i \langle f \rangle_i 1_{A_i}(x) \right] dx \\
&= -\sum_i \langle f \rangle_i \int_X 1_{A_i}(x) \log \left[\sum_i \langle f \rangle_i 1_{A_i}(x) \right] dx \\
&= -\sum_i \langle f \rangle_i \int_{A_i} \log \left[\sum_i \langle f \rangle_i 1_{A_i}(x) \right] dx \\
&= -\sum_i \langle f \rangle_i \mu_L(A_i) \log \langle f \rangle_i \; .
\end{aligned}$$

It is noteworthy that for any density f the Boltzmann-Gibbs entropy of the coarse-grained density may be greater than the entropy of f, or, more specifically, the following.

Theorem 7.1. For any density f and any nontrivial partition A_i of the phase space X, $H(f) \leq H(f^{\mathrm{cg}})$.

Proof. This is quite straightforward to prove using the integrated form of the Gibbs inequality (2.5). First, by use of the indicator function $1_A(x)$, we may write

$$\begin{aligned}
H(f) &= -\int_X f(x) \log f(x) dx \\
&= -\sum_i \int_X f(x) 1_{A_i}(x) \log f(x) dx \\
&\leq -\sum_i \int_X f(x) 1_{A_i}(x) \log g(x) dx \; ,
\end{aligned}$$

for $g(x)$ an integrable density. Pick $g(x) = \langle f \rangle_i$ so that

$$\begin{aligned}
H(f) &\leq -\sum_i \log \langle f \rangle_i \int_X f(x) 1_{A_i}(x) dx \\
&= -\sum_i \log \langle f \rangle_i \int_{A_i} f(x) dx \\
&= -\sum_i \langle f \rangle_i \mu_L(A_i) \log \langle f \rangle_i \\
&= H(f^{\mathrm{cg}}) \; ,
\end{aligned}$$

and the assertion is proved.

Thus the effect of any imprecision in the estimation of a density f characterizing a system, no matter what the origin, will be to either increase the entropy of the estimated (coarse-grained) density $H(f^{\mathrm{cg}})$ above its actual value $H(f)$, or leave it unchanged.

Precisely analogously to the way in which the entropy of the coarse-grained density was derived, it is easy to show that the conditional entropy of f^{cg} with respect to a second density g, also coarse grained with respect to the partition A_i, is given by

$$H_c(f^{\mathrm{cg}} | g^{\mathrm{cg}}) = -\sum_i \langle f \rangle_i \mu_L(A_i) \log \left[\frac{\langle f \rangle_i}{\langle g \rangle_i} \right] \, .$$

It is equally easy to show that $H(f|g) \leq H_c(f^{\mathrm{cg}} | g^{\mathrm{cg}})$ for all densities f and g and nontrivial partitions A_i of the phase space X.

Therefore, in general, *coarse graining of the phase space,* and the consequent coarse graining of a density, *will either increase the entropy or leave it equal to its value before coarse graining.*

Now assume that the phase space X is finite and normalized and that an initial density f evolves under the action of a Markov operator to give the sequence $\langle P^t f \rangle$. In analogy with (7.3), the coarse-grained $P^t f$ is given by

$$(P^t f(x))^{\mathrm{cg}} = \sum_i \langle P^t f \rangle_i 1_{A_i} 1_{A_i}(x) \; ,$$

where

$$\langle P^t f \rangle = \frac{1}{\mu_L(A_i)} \int_{A_i} P^t f(x) dx \; .$$

It is important to realize that we are assuming that the Markov operator operates without any error on the density f, and that the coarse graining arises because of our inability to precisely measure dynamical variables and, consequently, densities, for whatever reason. The converse situation in which we may measure densities with infinite precision, but the dynamics always work with some error, are considered in Sec. VIII where we consider system interactions with a heat bath.

An example is helpful in illustrating the effect of course graining, and we return to our previous example of the baker transformation in Fig. 6, with f_0 given by Eq. (5.5).

Consider first of all the situation in which the partition $\{A_1, A_2\}=\{Y_1, Y_2\}$. Then we have $\langle f_0\rangle_1=\langle f_0\rangle_2=1$, and thus $H(f_0^{\mathrm{cg},Y})=0$. However, following one application of the baker transformation [Fig. 6(c)], $H(f_2^{\mathrm{cg},Y})\simeq -0.13$ and, finally [Fig. 6(d)], $H(f_2^{\mathrm{cg},Y})=0$. On the other hand, after one application of the inverse of the baker transformation [Fig. 6(a)], we have $H(f_{-1}^{\mathrm{cg},Y})=0$. Alternately, if we choose the partition to be $\{A_2, A_2\}=\{X_1,X_2\}$, then it is easy to show that $H(f_0^{\mathrm{cg},X})\simeq -0.13$ while $H(f_{-1}^{\mathrm{cg},X})=H(f_1^{\mathrm{cg},X})=H(f_2^{\mathrm{cg},X})=0$.

These simple computations illustrate the following effects of coarse graining in this reversible system.

(1) The entropy of the coarse-grained density approaches the equilibrium entropy for both positive and negative times.

(2) This approach any not be monotone [compare $H(f_0^{\mathrm{cg},Y})$ with $H(f_1^{\mathrm{cg},Y})$].

(3) The approach is not necessarily symmetric with respect to a reversal of time [compare $H(f_{-1}^{\mathrm{cg},Y})$ with $H(f_1^{\mathrm{cg},Y})$].

(4) The approach may be dependent on the partition chosen (compare X vs Y).

Coarse graining has interested numerous authors since the concept was first introduced by Gibbs (1902, pp. 148–150) with the observation that coarse graining of a mixing system should lead to an increase in the entropy to its maximal value, as we have just illustrated. To prove this, assume that P^t is an f_* mixing Markov operator and rewrite the definition of an f_* mixing Markov operator in the equivalent integral form

$$\lim_{t\to\infty}\int_X [P^t f(x)]g(x)dx=\int_X f_*(x)g(x)dx \ . \tag{7.4}$$

If we pick $g(x)=1_{A_i}(x)/\mu_L(A_i)$, then Eq. (7.4) becomes

$$\lim_{t\to\infty}\frac{1}{\mu_L(A_i)}\int_X [P^t f(x)]1_{A_i}(x)dx =\frac{1}{\mu_L(A_1)}\int_X 1_{A_i}(x)f_*(x)dx \ .$$

The left-hand side of this equality is $\langle P^t f\rangle_i$, while the right-hand side is identically equal to $\langle f_*\rangle_i$. Thus

$$\lim_{t\to\infty}\langle P^t f\rangle_i=\langle f_*\rangle_i$$

so

$$\lim_{t\to\infty}\sum_i \langle P^t f\rangle_i 1_{A_i}(x)=\sum_i \langle f_*\rangle_i 1_{A_i}(x) \ .$$

Finally, from the definitions of $(P^t f)^{\mathrm{cg}}$ and f_*^{cg} it follows that

$$\lim_{t\to\infty}(P^t f)^{\mathrm{cg}}=f_*^{\mathrm{cg}} \ .$$

Since the same result would hold if $g(x)$ were a simple function, it will also be true for all bounded measurable functions g. Finally, if P^t is also reversible (as for the baker transformation), we have proved the following.

Theorem 7.2. If P^t is an f_* mixing Markov operator with a unique stationary density f_* and A_i is a nontrivial partition of the phase space X, then

$$\lim_{t\to\pm\infty}(P^t f)^{\mathrm{cg}}=f_*^{\mathrm{cg}}$$

for all initial densities f, where

$$f_*^{\mathrm{cg}}(x)=\sum_i \langle f_*\rangle_i 1_{A_i}(x) \ .$$

As a consequence of this result we have the following.

Corollary 7.3. If P^t is a reversible f_* mixing Markov operator with a unique stationary density f_* and A_i is a nontrivial partition of the phase space X, then

$$\lim_{t\to\pm\infty}H_c((P^t f)^{\mathrm{cg}}|f_*^{\mathrm{cg}})=0$$

for all initial densities f.

For uniformly mixing systems operating in a normalized finite space, it is an easy consequence of these results that, after coarse graining of the phase space, $P^t f$ will approach the density of the microcanonical ensemble and the Boltzmann-Gibbs entropy will approach its maximum value of zero.

Given the fact that the result of Corollary 7.3 was clearly known by Gibbs, though not proved with rigor, it is surprising that the effects of nontrivial coarse graining have had, and continue to have, so much attention as possible sources of irreversible thermodynamic behavior.

Even setting aside the lack of irreversibility in the behavior of the coarse-grained entropy, it is important to realize that the rate of convergence of the entropy of the coarse-grained densities that Corollary 7.3 guarantees will, in general, depend on the way in which the coarse graining of the phase space is carried out. Experimentally, if entropy increases to a maximum only because we have reversible mixing dynamics and coarse graining due to measurement imprecision, then the rate of convergence of the entropy (and all other thermodynamic variables) to equilibrium should become slower as measurement techniques improve. Such phenomena have not been observed. Thus it seems likely that nontrivial coarse graining plays no role in determining thermodynamic behavior, even if the coarse graining is externally imposed by nature in the form of an inherent graininess or unitary cellularity of phase space.

B. Traces

As an alternative to course graining we now explore the consequences of a reversible dynamics in which not all dynamical variables are observable. Essentially this means that we have a dynamical system operating in an m-dimensional space, but are able to observe only $n<m$ of these variables. That is, we observe only a *trace* of its operation in an n-dimensional space because $(m-n)$ of the variables are hidden to us, e.g., because either we do not know about them or we do not have the technology to measure them.

To make these ideas more precise we return to a brief consideration of dynamical systems that were first formally introduced in Sec. II. If we have a dynamical system S_t operating in some phase space X, then only three possible types of trajectories may be observed. Either (1) the trajectory is a single fixed point x_* such that $x_*=S_t(x_*)$ for all times t; (2) the trajectory of S_t is a nonintersecting curve, which we express as $S_{t'}(x)\neq S_{t''}(x)$ for all times t' and t'' that are not equal; or (3) S_t has a periodic trajectory, so $S_t(x)=S_{t+\omega}(x)$ for all times t and all points x in some subset A of the phase space X. In this case we say that the trajectory has period ω.

The only possible type of intersecting trajectory that a dynamical system may have is a periodic one. To demonstrate this, we assume that for some point x in the phase space X we have $S_{t'}(x)=S_{t''}(x)$ for a time $t''>t'$. Apply $S_{t-t'}$ to this relation and use the definition of a dynamical system to obtain

$$S_{t-t'}\circ S_{t'}(x)=S_t(x)=S_{t-t'}\circ S_{t''}(x)=S_{t+(t''-t')}(x)\ .$$

With $\omega=t''-t'>0$, we have arrived at $S_t(x)=S_{t+\omega}(x)$ for all times t.

However, we often observe (apparently) nonperiodic intersecting trajectories, and it is not terribly difficult to understand how they might arise. As a simple example, a periodic or aperiodic trajectory of a three-dimensional oscillator might display an intersecting trajectory in a two-dimensional projection of its phase space.

As alluded to earlier, this projection is called a trace, a concept that is made more precise by the following. Let X and Y be two topological Hausdorff phase spaces, $\varphi:Y\rightarrow X$ a given continuous function, and $S_t:Y\rightarrow Y$ a dynamical system operating in the phase space Y. A function $h:R\rightarrow X$ is the *trace* of the dynamical system if there is a point y in the space Y such that $h(t)=\varphi(S_t(y))$ for all times t.

Given the fact that intersecting but nonperiodic trajectories are often observed in data, one is naturally led to wonder under what circumstances such a trajectory can be viewed as the trace of a higher-dimensional dynamical system. It is actually easy to give a surprising answer to a much more general question. Every continuous function in a space X is the trace of a *single* dynamical system operating in a higher-dimensional phase space Y. More precisely, we have the following theorem.

Theorem 7.4 (Lasota and Mackey, 1985). Let the phase space X be an arbitrary but topological Hausdorff space. Then there is a second phase space Y, also topological and Hausdorff, a dynamical system S_t operating in Y, and a continuous function $\varphi:Y\rightarrow X$ such that every continuous function $h:R\rightarrow X$ is the trace of S_t. That is, for every h there is a point y in the phase space Y such that $h(t)=\varphi(S_t(y))$ for all times t.

There are important consequences for the behavior of the entropy when one is considering the entropy of the trace of a dynamical system.

If we have a dynamical system S_t operating on Y, then the entropy is always identically equal to the entropy of the initial density, since it is impossible for the entropy of a reversible system to change (Theorem 3.2). However, this is not the case for the entropy of the density evolving under the action of a trace of a dynamical system if the trace is intersecting and nonperiodic. Thus we have the following result.

Theorem 7.5. Let the phase spaces X and Y be topological Hausdorff spaces and $h:R\rightarrow X$ be an intersecting and nonperiodic trace of a dynamical system $S_t:Y\rightarrow Y$. Then the entropy of densities evolving under the action of the trace h is either constant or increasing.

Proof. The proof rests on the trivial observation that if h is intersecting and nonperiodic, then at every intersection point x on the trajectory h the inverse $h^{-1}(x)$ is not unique. Therefore the trace h is the trajectory of a semidynamical system, and since semidynamical systems are irreversible, from Theorem 3.1 the entropy is either constant or increasing.

Thus the simple act of taking a trace of a dynamical system (with time-independent entropy) may be sufficient to generate a system in which the entropy is increasing. Of course, we do not know what the limit of this increase may be, and the entropy may certainly approach a limit considerably less than its maximal value.

However, there are certain types of traces for which much more can be said about the behavior of the entropy. To see how this works we introduce the notion of a factor of a transformation. Let X and Y be two different phase spaces with normalized measures μ_* and ν_* and associated densities f_* and g_*, respectively, and $T_t:X\rightarrow X$ and $S_t:Y\rightarrow Y$ be two measure-preserving transformations. If there is a transformation $\varphi:Y\rightarrow X$ that is also measure preserving, i.e., if $\nu_*(\varphi^{-1}(A))=\mu_*(A)$ for all subsets A of the phase space Y, and such that $T_t\circ\varphi=\varphi\circ S_t$, then T_t is called a *factor* of S_t.

From this definition the trajectory of the factor T_t is a trace of the system S_t. The connection between these concepts and the behavior of the entropy is furnished by the following theorem due to Rochlin (1964).

Theorem 7.6 (Rochlin, 1964). Every f_* exact transformation is the factor of a K automorphism.

The transformation φ involved in our discussion of factors and traces is precisely what Misra *et al.* (1979) and Misra and Prigogine (1981) refer to as a projection operator in their work on the generation of irreversible behavior from reversible dynamics. The Rochlin theorem 7.6 serves as the analytic link in their work between reversible K automorphism and f_* exact transformations (or strong or monotonic Markov operators).

As noted in Sec. V, since K automorphisms are invertible, their entropy is forever fixed at its initial value by Theorem 3.2. On the other hand, by Theorem 6.4 and Corollary 6.5, we know that the entropy of an f_* exact transformation, which by the above theorem is the factor of a K automorphism, increases smoothly to its maximum value of zero irrespective of the initial density with which the system was prepared. Hence for this particular case we have a much stronger result than covered by Theorem 7.5.

As a simple example of these points, again consider the baker transformation, which is a K automorphism. We have shown directly that the entropy of the baker transformation is absolutely constant and equal to the value determined by the choice of the initial density. However, if we simply take the x portion of the baker transformation, it is a factor and is identical to the dyadic transformation $T(x)=2x(\mathrm{mod}\,1)$, introduced in Sec. VI, which is uniformly exact and whose entropy smoothly increases to zero by Corollary 6.4. Goldstein and Penrose (1981) have called this entropy a type of coarse-grained entropy.

There is a connection between the effects of coarse graining the phase space and taking the trace of a dynamical system. In nontrivial coarse graining, we lose *some* information about the exact values of the dynamical variables. One could easily imagine the situation in which we have m variables, of which n were measured with infinite precision, and $(m-n)$ of them were measured with some error. Thus the act of taking a trace is just a more extreme case of coarse graining in which we are not able to measure *any* of the $(m-n)$ variables; i.e., the partition is trivial.

Viewed from this perspective, it is surprising that there is such a difference between the results of a nontrivial coarse graining of the phase space (no induction of irreversible thermodynamic behavior), and examining only the trace of a dynamical system operating in a higher-dimensional space than our observations permit. However, it is precisely trivial coarse graining of a phase space in which the evolution of densities is governed by the Liouville equation (3.4) that leads to the Boltzmann equation and its successful predictions of the behavior of dilute gases. The Boltzmann equation describes the behavior of a density evolving under the action of a factor of the original dynamics. Thus there is a strong analogy between the dynamics of the Liouville equation and of the Frobenius-Perron operator for the baker transformation, while the dynamics of the Boltzmann equation are analogous to those of the Frobenius-Perron operator corresponding to the dyadic transformation:

$$\begin{tikzcd}
\text{Liouville equation} \arrow[r, leftrightarrow, "\text{Reversible}"] \arrow[d, "\text{Trace}"'] & \text{baker transformation} \arrow[d, "\text{Trace}"] \\
\text{Boltzmann equation} \arrow[r, leftrightarrow, "\text{Irreversible}"'] & \text{dyadic transformation}
\end{tikzcd}$$

VIII. INTERACTIONS WITH A HEAT BATH

This section examines the consequences of having a deterministic transformation experience a perturbation from some outside source. Thus we are starting to consider open systems, the mathematical analog of the interaction between a system and a "heat bath." Discrete time systems are considered first in Sec. VIII.A, followed by continuous time systems with dynamics described by ordinary differential equations in Sec. VIII.B.

A. Open discrete time systems

Assume that, in general, a system evolves according to a given transformation $S(x_t)$. The qualifying phrase "in general" means that the transition $x_t \to S(x_t)$ occurs with probability $(1-\varepsilon)$. In addition, with probability ε, the value of x_{t+1} is uncertain. If $x_t=y$ is given, then, in this case, x_{t+1} may be considered as a random variable distributed with a density $K(x,y)$, which depends on y.

One interpretation of this process is that ε corresponds to the degree of coupling between the system under study and the heat bath. If this is the case, then the parameter ε can be thought of as a number related to the ratio of the fundamental frequency of operation of the basic deterministic system F_D to the frequency of the outside perturbation coming from the heat bath F_P. Thus when $F_D \ll F_P$, $\varepsilon \simeq 1$ and the system operates almost as a random walk, while with $F_D \gg F_P$, we have $\varepsilon \simeq 0$ and the system evolves almost completely deterministically. We will refer to the situation in which $0<\varepsilon<1$ as "loose coupling," while for $\varepsilon=1$, in which the influence of the external system is always experienced, we will speak of strong (or continuous) coupling. However, as will become clear in Sec. VIII.A.3, precisely the same formulation may be interpreted in a totally different fashion.

1. An operator equation

The first step in the study of this process is the derivation of an equation for the operator P^t that gives the prescription for passing from the density f_t of x_t to the density f_{t+1} of x_{t+1}. Assume that the dynamics of our system operate in a phase space X (with positive measure, of course), which is some measurable subset of R^d, and that the dynamics S are nonsingular.

If the density f_t of x_t is given, we wish to know the probability that some point x_{t+1} is in a subset A of the phase space. Clearly, x_{t+1} may be reached in only one of two ways: deterministically with probability $(1-\varepsilon)$ and stochastically with probability ε. In the deterministic case $x_{t+1}=S(x_t)$ and

$$\mathrm{Prob}_D(x_{t+1}\in A)=\mathrm{Prob}_D(S(x_t)\in A)\ . \qquad (8.1)$$

The index D denotes the deterministic situation. From the definition of the Frobenius-Perron operator, the density of $S(x_t)$ is $P_S f_t$ and, as a result,

$$\mathrm{Prob}_D(S(x_t)\in A)=\int_A P_S f_t(x)dx \ . \tag{8.2}$$

If the stochastic perturbation occurs, and if $y=x_t$, then

$$\mathrm{Prob}_P(x_{t+1}\in A|x_t=y)=\int_A K(x,y)dx \ .$$

(P denotes the perturbation.) Since x_t is a random variable with density f_t, it is also the case that

$$\mathrm{Prob}_P(x_{t+1}\in A)=\int_X \mathrm{Prob}_P(x_{t+1}\in A|x_t=y)f_t(y)dy \ .$$

Combining this relation with the previous one gives

$$\mathrm{Prob}_P(x_{t+1}\in A)=\int_A\left[\int_X K(x,y)f_t(y)dy\right]dx \ . \tag{8.3}$$

From Eqs. (8.1)–(8.3) it follows that

$$\mathrm{Prob}(x_{t+1}\in A)=(1-\varepsilon)\mathrm{Prob}_D(x_{t+1}\in A)+\varepsilon\mathrm{Prob}_P(x_{t+1}\in A)=\int_A\left[(1-\varepsilon)P_S f_t(x)+\varepsilon\int_X K(x,y)f_t(y)dy\right]dx \ .$$

Since A is an arbitrary subset of the phase space, the density f_{t+1} exists whenever f_t exists and is, in fact, given by

$$f_{t+1}(x)=(1-\varepsilon)P_S f_t(x)+\varepsilon\int_X K(x,y)f_t(y)dy \ .$$

Thus the full equation for the operator P describing the evolution of densities in this mixed system operating with both deterministic and perturbed elements is

$$Pf(x)=(1-\varepsilon)P_S f(x)+\varepsilon\int_X K(x,y)f(y)dy \ . \tag{8.4}$$

It is straightforward to show that the operator defined by (8.4) is a Markov operator.

In investigating the properties of the evolution of densities by the operator equation (8.4), and the consequent behavior of the entropy of these densities, some mild restrictions on both the transformation S and the kernel K are required. First, assume that the deterministic transformation S satisfies

$$|S(x)|\le a_0|x|+a_1 \tag{8.5}$$

throughout the phase space, where $a_0<1$ and a_1 are non-negative constants. Second, since for fixed y, $K(x,y)$ is a density, it clearly satisfies

$$K(x,y)\ge 0 \quad\text{and}\quad \int_X K(x,y)dx=1 \ .$$

These conditions in conjunction with the requirement that K be measurable means that K is a *stochastic kernel.* Furthermore we shall always assume that for every $\gamma>0$ there is a $\delta>0$ such that

$$\int_E K(x,y)dx\le\gamma \ ,$$

for every y in X and subset E of X with $\mu(E)\le\delta$, i.e., K is uniformly integrable in x. Finally, it will always be assumed that with $b_0<1$ and b_1 non-negative constants,

$$\int_X |x|K(x,y)dx\le b_0|y|+b_1 \ . \tag{8.6}$$

This condition is automatically satisfied if the phase space X is bounded; but if it is unbounded, then (8.6) prevents divergence of the trajectories to infinity.

2. Noise-induced effects in loosely coupled discrete time systems

We are now in a position to state our main results concerning the behavior of the entropy of a discrete time deterministic system coupled to a heat bath. The first of these guarantees the existence of at least one state of thermodynamic equilibrium and the evolution of the conditional entropy to a maximum, though not necessarily to zero, in the presence of noise. Thus the following result is equivalent to the second-order formulation of the second law of thermodynamics.

Theorem 8.1 (Lasota and Mackey, 1987). If S is a nonsingular transformation that satisfies (8.5) and K is a uniformly integrable stochastic kernel satisfying (8.6), then for $0<\varepsilon\le 1$ the operator P given by (8.4) is smoothing, and thus asymptotically periodic, whenever $a_0(1-\varepsilon)+b_0\varepsilon<1$.

Therefore, for any closed system whose dynamics evolve through the action of a nonsingular transformation S satisfying (8.5), placing it in contact with a second system whose effect on the first is a perturbation characterized by a kernel K satisfying (8.6) leads automatically to a situation in which the resulting open system is asymptotically periodic *regardless* of the nature of the original closed system S. Further, since this procedure induces asymptotic periodicity, we know that at least one state of thermodynamic equilibrium, characterized by a stationary density f_*, exists and that the conditional entropy $H(P^t f|f_*)$ is an increasing function with a limiting value given by $H_{\max}(f,f_*)$, as defined in Theorem 3.4.

Under certain circumstances involving loose coupling to a heat bath, there are even stronger results concerning the behavior of the entropy, corresponding to the third-order formulation of the second law of thermodynamics. One such case is as follows.

Assume that the value of the perturbation to the system S coming from the heat bath (when it occurs) at time $t+1$ is independent of the value of x_t. Then the stochastic kernel $K(x,y)$ is independent of y, and the kernel is simply $K(x,y)=g(x)$, where g is the density of the perturbations ξ_t. In this case, with the external perturbations independent of the state of the system S, the perturbations ξ_t could be interpreted as completely stochastic or as coming from another deterministic system. They could even be viewed as the trace of some deterministic system whose Frobenius-Perron operator has g as its unique stationary density. This is a slightly different situation from that explored in Sec. VII, where we considered the effect on the behavior of the entropy of only examining the trace of a system. Now we are considering the situation which could be interpreted as the perturbation of a system by the trace of another system.

In this case, the operator equation (8.4) takes the simple form

$$Pf(x)=(1-\varepsilon)P_S f(x)+\varepsilon g(x) , \tag{8.7}$$

and there are some surprising consequences of making the loose coupling independent of the state of the system S. Namely, the following.

Theorem 8.2 (Losta and Mackey, 1987). If P is the operator defined by (8.7), then $\{P^t f\}$ is f_* exact.

Thus, by the simple expedient of loosely coupling a system to a heat bath such that the system experiences perturbations that are independent of the state of the system, there will be a unique state of thermodynamic equilibrium, and the conditional entropy of the coupled system will globally converge to zero regardless of the nature of the original system S. In fact, it is even possible to write the unique stationary density f_* of the operator P in Eq. (8.7), characterizing the thermodynamic equilibrium, as

$$f_*=\sum_{k=0}^{\infty}(1-\varepsilon)^k P^k g .$$

To show that this is the unique stationary density, note that since $\|(1-\varepsilon)^k P^k g\| \le (1-\varepsilon)^k \|g\|$ by the contractive property of Markov operators, the series in f_* is absolutely convergent. Therefore substitution of the expression for f_* into Eq. (8.7) shows that $Pf_*=f_*$.

3. Strong coupling and additive noise

A much different interpretation of this perturbation at random times of a deterministically operating system is possible and related to strong coupling ($\varepsilon=1$) between a deterministic system and an external source of noise.

We consider the following. Take the quantities $\xi_0, \xi_1, \ldots,$ to be d-dimensional random vectors and let the phase space X be R^d. Then for a given $\{\xi_t\}$ and a dynamics S of two variables, $S{:}R^d \times R^d \to R$, we may assume that the system goes from $x_t=y$ to $x_{t+1}=S(y,\xi_t)$. Let $K(x,y)$ be the density of $S(y,\xi_t)$. Then the density will always exist if $S(y,z)$ as a function of z is nonsingular. If this is the case, then Eq. (8.4) with $\varepsilon=1$ describes the evolution of the densities corresponding to

$$x_{t+1}=S(x_t,\xi_t) . \tag{8.8}$$

These external perturbations, appearing through the $\{\xi_t\}$, could be entering the system in a variety of ways, for example, through fluctuations in coefficients or in a multiplicative fashion as

$$x_{t+1}=\xi_t S(x_t) .$$

To give a concrete and commonly considered example, suppose that the noise is additive, $S(y,z)=S(y)+z$, so (8.8) becomes

$$x_{t+1}=S(x_t)+\xi_t . \tag{8.9}$$

It is rather surprising that a dynamics of the form described by Eq. (8.9) may also appear as the consequence of taking a factor or trace (Lasota and Mackey, 1989). To illustrate this point we once again return to the baker transformation introduced in Eq. (5.2). As pointed out in Sec. VII.B, the x portion of the baker transformation has a trajectory determined by $x_{t+1}=T(x_t)$ where $T(x)=2x\,(\mathrm{mod}\,1)$. As a consequence, the baker transformation (5.2) can be rewritten in the equivalent form

$$x_{t+1}=T(x_t), \quad y_{t+1}=\tfrac{1}{2}y_t+\xi_t ,$$

where $\xi_t=\frac{1}{2}1_{[1/2,1]}(x_t)$. From classical results of Borel, the ξ_t defined in this way are independent random variables. As a consequence, if we take only the y factor of the baker transformation, we end up with a dynamics described by an equation that is simply a special case of (8.9). Thus taking a trace of a dynamical system in two different ways may yield results with quite different interpretations.

If the sequence $\{\xi_t\}$ of random variables in Eq. (8.9) has a common density g, then x_{t+1}, subject to the condition that $x_t=y$, has the conditional density $g(x-S(y))$; and since $K(x,y)=g(x-S(y))$, Eq. (8.4) becomes

$$Pf(x)=\int_{R^d} f(y)g(x-S(y))dy . \tag{8.10}$$

For the special case of additive noise, a comparison of Eqs. (8.10) and (8.4) shows that (8.10) can be derived independent of any assumption concerning the nonsingularity of S. Furthermore, in this case the condition given by Eq. (8.6) reduces to

$$m=\int_{R^d}|x|g(x)dx<\infty . \tag{8.11}$$

Thus we have an immediate corollary to Theorem 8.1 for systems with noise added, as in (8.9).

Corollary 8.3 (Lasota and Mackey, 1987). If S (nonsingular or not) is a transformation operating in the phase space R^d, satisfies inequality (8.5), and experiences an additive perturbation [as in Eq. (8.9)] with a finite first moment, then the sequence $\{P^t f\}$, where P is the Markov operator defined by Eq. (8.10), is asymptotically periodic for all densities f.

Hence for all situations in which perturbations are *added* to a transformation S, the effect is to induce asymptotically periodic behavior regardless of the nature of the original unperturbed dynamics S (remember that S may even be singular). Because of this, we also know that noise induces at least one state of thermodynamic equilibrium, whose stationary density is given by Eq. (3.16), and guarantees the smooth approach of the conditional entropy to a maximum (Theorem 3.4).

For some transformations, the induction of asymptotic periodicity by the addition of perturbations would not be at all surprising; e.g., the addition of a stochastic perturbation to a transformation with an exponentially stable periodic orbit gives asymptotic periodicity. However, the surprising content of Theorem 8.1 (and Corollary 8.3) is that even in a transformation that has aperiodic limiting behavior, the addition of noise will result in asymptotic periodicity.

This phenomenon is easy to illustrate by considering

$$x_{t+1}=S(x_t)\ (\mathrm{mod}\,1)\ , \tag{8.12}$$

where $S(x)=\alpha x+\lambda$, $0<\alpha<1$, and $0<\lambda<1$. This map is an example of a class of transformations considered by Keener (1980). From the results for general Keener transformations, there exists an uncountable set Λ such that for each λ in Λ the rotation number corresponding to the transformation (8.12) is irrational. For each such λ the sequence $\{x_t\}$ is not periodic and the invariant limiting set

$$\bigcap_{k=0}^{\infty} S^k([0,1]) \tag{8.13}$$

is a Cantor set. The proof of Keener's general result offers a constructive technique for numerically determining values of λ that approximate elements of the set Λ.

The transformation (8.12) satisfies the conditions of Corollary 8.3 and is, therefore, an ideal candidate to illustrate the induction of asymptotic periodicity by noise in a transformation whose limiting behavior is quite erratic in the absence of noise.

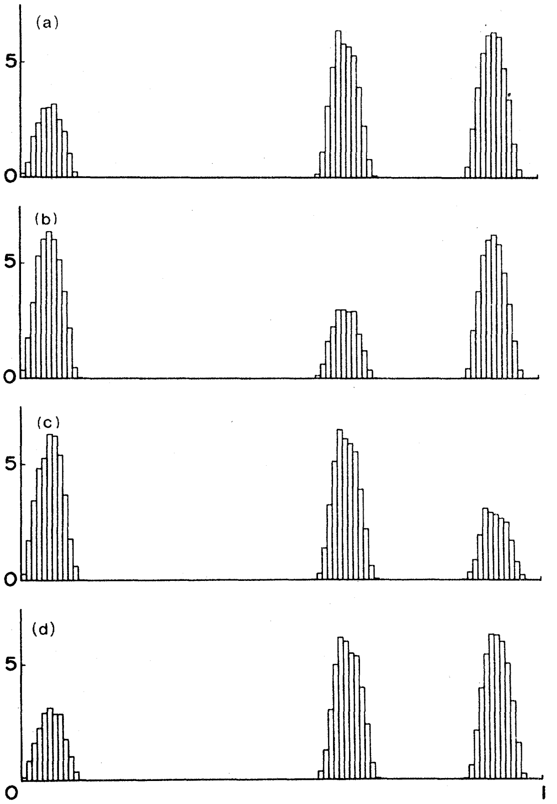

FIG. 8. Noise-induced asymptotic periodicity in the Keener map (8.12). See the text for parameter values and further discussion.

To be specific, pick $\alpha=\frac{1}{2}$ and use the results of Keener to show that $\lambda=\frac{17}{30}$ is close to a value in the set Λ for which the invariant limiting set (8.13) should be a Cantor set. Asymptotic periodicity is illustrated by studying

$$x_{t+1}=(\alpha x_t+\lambda+\xi_t)\ (\mathrm{mod}\,1)\ , \tag{8.14}$$

where the ξ_t are random numbers uniformly distributed on $[0,\theta]$. Figure 8 shows the effect of this stochastic perturbation for $\theta=\frac{1}{15}$ and these values of α and λ. Figures 8(a)–8(c) show the densities obtained after 10, 11, and 12 successive interactions of an initially uniform density $f(x)=1$ on [0,1]. The thirteenth density is identical with the tenth, the fourteenth with the eleventh, etc., thus indicating that the sequence of densities is asymptotically periodic with period 3. The bifurcation behavior of these densities as a function of the noise amplitude θ has been studied by Provatas and Mackey (1989).

4. f_* exactness induced by additive perturbations

The fact that additive noise in a system strongly coupled to its environment can induce asymptotic periodicity is important for the understanding of how the coupling of a system to a heat bath may cause the entropy to become an increasing function of time. However, under certain circumstances there are even stronger results concerning the effects of additive noise. Namely, additive noise may induce f_* exactness with a consequent increase in the conditional entropy to its maximal value of zero corresponding to the strongest (third-order) form of the second law of thermodynamics.

This is given in an extension of Corollary 8.3 that, in conjunction with Theorem 6.4, gives a sufficient condition for additive noise to cause the conditional entropy to approach zero. This result is as follows.

Theorem 8.4 (Lasota and Mackey, 1987). Assume that the transformation $S:R^d \rightarrow R^d$ and the density g of the perturbation satisfy inequalities (8.5) and (8.11). Further assume that there exists a point z_0 in the phase space R^d and a number

$$r_0>\frac{ma_0+a_1}{1-a_0}$$

such that

$$g(x)>0 \quad \text{for } |x-z_0|<r_0\ . \tag{8.15}$$

Then the Markov operator defined by Eq. (8.10) is f_* exact.

B. Open continuous time systems

1. An evolution equation

Section III.B considered the behavior of the entropy of densities evolving under the action of flows governed by the system of ordinary differential equations

$$x_i' = F_i(x), \quad i = 1, \ldots, d$$

operating in some region of R^d with initial conditions $x_i(0) = x_{i,0}$. Here we examine the behavior of the stochastically perturbed analog

$$x_i' = F_i(x) + \sum_{j=1}^{d} \sigma_{ij}(x)\xi_j, \quad i = 1, \ldots, d \tag{8.16}$$

with the same initial conditions, where $\sigma_{ij}(x)$ is the amplitude of the stochastic perturbation and $\xi_j = dw_j/dt$ is a "white-noise" term that is the derivative of a Wiener process.

Equation (8.16) is a stochastic differential equation. As in the case of a nonperturbed system of ordinary differential equations, if the functions $F_i(x)$ and $\sigma_{ij}(x)$ satisfy Lipschitz conditions

$$|F_i(x) - F_i(y)| \le L|x - y|, \quad x, y \in R^d, \quad i = 1, \ldots, d$$

and

$$|\sigma_{ij}(x) - \sigma_{ij}(y)| \le L|x - y|, \quad x, y \in R^d, \quad i, j = 1, \ldots, d ,$$

where $L > 0$, then Eq. (8.16) has a unique solution (Gikhman and Skorokhod, 1969).

Now we turn to a consideration of the equation that governs the evolution of the density function of the process $x(t)$ generated as the solution to the stochastic differential equation (8.16).

This density, denoted by $u(t,x)$, is defined by

$$\text{Prob}(x(t) \in B) = \int_B u(t,s)ds . \tag{8.17}$$

We set

$$a_{ij}(x) = \sum_{k=1}^{d} \sigma_{i,k}(x)\sigma_{j,k}(x) , \tag{8.18}$$

which is easily shown to be non-negative, and assume a stronger condition, known as the *uniform parabolicity condition*:

$$\sum_{i,j=1}^{d} a_{ij}(x)\lambda_i\lambda_j \ge \rho \sum_{i=1}^{d} \lambda_i^2 , \tag{8.19}$$

where $\rho > 0$. If the $a_{ij}(x)$ satisfy the uniform parabolicity condition (8.19) and if they and the $F_i(x)$ are continuous and C^3, then $u(t,x)$ exists and is differentiable.

The evolution equation for the density $u(t,x)$ is given by

$$\frac{\partial u}{\partial t} = -\sum_{i=1}^{d} \frac{\partial[F_i(x)u]}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2[a_{ij}(x)u]}{\partial x_i \partial x_j} . \tag{8.20}$$

This evolution equation is known variously either as the *Fokker-Planck equation* or the *forward Kolmogorov equation* and will be the foundation for our investigations into the effects of noise on the evolution of densities in systems whose dynamics are described by ordinary differential equations. The first term on the right-hand side is usually called a *drift term*, while the second is known as a *diffusion term*.

Equation (8.20) can also be written in the equivalent form

$$\frac{\partial u}{\partial t} = L^+ u , \tag{8.21}$$

where the operator L^+ is given by

$$L^+ = -\sum_{i=1}^{d} \frac{\partial}{\partial x_i} F_i(x) + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} a_{ij}(x) . \tag{8.22}$$

The *backward Kolmogorov equation* is

$$\frac{\partial u}{\partial t} = L^- u , \tag{8.23}$$

where the operator

$$L^- = -\sum_{i=1}^{d} F_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} \tag{8.24}$$

is adjoint to L^+. Remember that for any two L^1 functions u and v, operators L^+ and L^- that are adjoint to one another satisfy

$$\int_{R^d} v(L^+ u)dx = \int_{R^d} u(L^- v)dx .$$

2. Solutions of the Fokker-Planck equation

If the stochastic differential equation (8.16) has an initial condition $x(0)$ and an associated initial density g, then the solution $u(t,x)$ of the Fokker-Planck equation satisfies $u(0,x) = g(x)$. Further, if the solution of the Fokker-Planck equation (8.17) can be written in the form

$$u(t,x) = \int_{R^d} \Gamma(t,x,s)u(0,s)ds , \tag{8.25}$$

where the kernel Γ is independent of the initial density $u(0,x)$ and $\lim_{t \to 0} u(t,x) = u(0,x)$, then $u(t,x)$ is said to be a *generalized solution* of the Fokker-Planck equation. Under some standard regularity conditions on the coefficients of the Fokker-Planck equation the generalized solution is unique. Since u is a density, the generalized solution corresponds to the evolution of the system to a unique thermodynamic state.

From Eq. (8.25) for the generalized solution, a family $\{P^t\}_{t \ge 0}$ of integral operators can be defined by

$$P^0 f(x) = f(x), \quad P^t f(x) = \int_{R^d} \Gamma(t,x,s)f(s)ds . \tag{8.26}$$

If the generalized solution is unique, then the operator P is a Markov operator.

It is instructive to compare the dynamical equations for densities evolving under the action of continuous flows in the noise-free and noise-perturbed situations. For the noise-free closed system in Sec. III.A, the evolution of the Frobenius-Perron operator P^t is determined by the partial differential equation (3.2), the Liouville equation. When the very same system is subject to external noise from the environment, then the evolution of the Markov operator $P^t f$ is governed by the Fokker-Planck equation (8.20), which is just the same as the noise-free equation (3.2) with the addition of the diffusion term.

3. The behavior of entropy

In Sec. IV.C we showed directly that the entropy of a closed continuous time system with dynamics described by ordinary differential equations was absolutely constant and equal to the entropy of the initial density with which the system was prepared. As pointed out, this result is not surprising, since dynamics described by ordinary differential equations are reversible; Theorem 3.2 shows that the entropy is constant for all reversible systems.

With the material of the previous sections we are now in a position to examine the effects of perturbations on the entropy of these continuous time reversible systems.

In presenting our first result we introduce the concept of a *Liapunov function*, by which we mean any function V with the following properties.

(1) $V(x)\geq 0$ for all x.

(2) $\lim_{|x|\to\infty} V(x)=\infty$.

(3) $(\partial V/\partial x_i)$ and $(\partial^2 V/\partial x_i x_j)$ are continuous for $i,j=1,\ldots,d$.

(4) $V(x)\leq Ke^{Mx}$ and $|\partial V(x)/\partial x_i \leq Ke^{M|x|}$, where K and M are positive constants.

The existence of a Liapunov function V that satisfies the differential inequality

$$\sum_{i,j=1}^{d} a_{ij}(x)\frac{\partial^2 V}{\partial x_i \partial x_j}+\sum_{i=1}^{d} F_i(x)\frac{\partial V}{\partial x_i}\leq -\alpha V(x)+\beta\ , \tag{8.27}$$

where α and β are positive constants, is involved with a sufficient condition stating that the evolution of densities by the Fokker-Planck equation is described by an f_* exact Markov operator. Specifically we have third-order form of the second law.

Theorem 8.5 (Lasota and Mackey, 1985). Assume that there is a unique generalized solution of the Fokker-Planck equation and that there exists a Liapunov function V satisfying (8.27). Then the Markov operator P^t defined by the generalized solution of the Fokker-Planck equation and given in (8.26) is f_* exact, and the conditional entropy $H_c(P^t f|f_*)$ approaches its maximal value of zero as $t\to\infty$.

It is a rather simple demonstration that the unique stationary density f_* defined by

$$\lim_{t\to\infty} P^t f(x)=f_*(x)\ ,$$

the existence of which is guaranteed by this theorem, is, under the conditions of the theorem, given by the (unique) solution of the elliptic equation

$$-\sum_{i=1}^{d}\frac{\partial[F_i(x)u]}{\partial x_i}+\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2[a_{ij}(x)u]}{\partial x_i\partial x_j}=0\ . \tag{8.28}$$

As a simple example of the application of Theorem 8.5, consider the two-dimensional oscillator system

$$\frac{dr}{dt}=r(c-r^2),\quad \frac{d\theta}{dt}=2\pi\ , \tag{8.29}$$

in (r,θ) space. The system (8.29) is an example of a system with a supercritical Hopf bifurcation. For $c<0$ the origin $r_*=0$ is the globally stable steady state, while for $c>0$ all solutions are attracted to the limit cycle defined by $r=\sqrt{c}$. Following Mackey *et al.* (1989) we consider the effects of noise in the analogous one-dimensional system

$$\frac{dx}{dt}=x(c-x^2)\ , \tag{8.30}$$

obtained by ignoring the angular coordinate θ in Eqs. (8.29). For Eq. (8.30), when $c<0$ all solutions are attracted to the single steady state $x_*=0$. Further, when $c>0$ the steady state $x_*=0$ is unstable and $x(t)\to\sqrt{c}$, if $x(0)=x_0>0$, while $x(t)\to -\sqrt{c}$ for $x_0<0$.

In the presence of additive noise, the stochastic differential equation corresponding to Eq. (8.30) is

$$\frac{dx}{dt}=x(c-x^2)+\sigma\xi\ ; \tag{8.31}$$

so there is always a positive probability that $x(t)$ may take on negative values, starting from a positive position, and vice versa. Therefore it is natural to consider this problem for $-\infty<x<\infty$.

The stationary solution of the Fokker-Planck equation corresponding to (8.31) is

$$f_*(x)=K_1 e^{\beta x^2(2c-x^2)/4c}\ , \tag{8.32}$$

where $\beta=2c/\sigma^2$. The normalization constant K_1 always exists and thus the $f_*(x)$ defined by (8.32) is a stationary density. To use Theorem 8.5, let $V(x)=x^2$ so V is a Liapunov function. Inequality (8.27) becomes

$$2\sigma^2+(2c+\alpha)x^2-2x^4\leq\beta\ .$$

This is clearly satisfied for arbitrary fixed $\alpha>0$ and sufficiently large $\beta>0$, thus proving the f_* exactness of the generalized solution of the Fokker-Planck equation corresponding to Eq. (8.31) for additive noise applied to the system (8.30). As a consequence of this we also know that the effect of the additive noise has been to induce the conditional entropy to approach its maximal value of zero.

In Fig. 9(a) we graph the location of the maxima of the stationary density given in Eq. (8.32) as a function of the parameter c. As might be expected on intuitive grounds, for $c<0$ the stationary density $f_*(x)$ has a single maximum centered at $x=0$, the location of the globally stable steady state of the unperturbed system (8.30). Once $c>0$, the stationary density $f_*(x)$ shows two maxima centered at $x=\pm\sqrt{c}$, the locally stable steady states of Eq. (8.30), and a local minimum at the unstable steady state $x=0$. Thus, if we were to calculate various system averages and plot them as a function of the parameter c, a qualitatively similar pattern would emerge. If, for example, the parameter were to be viewed as related to sys-

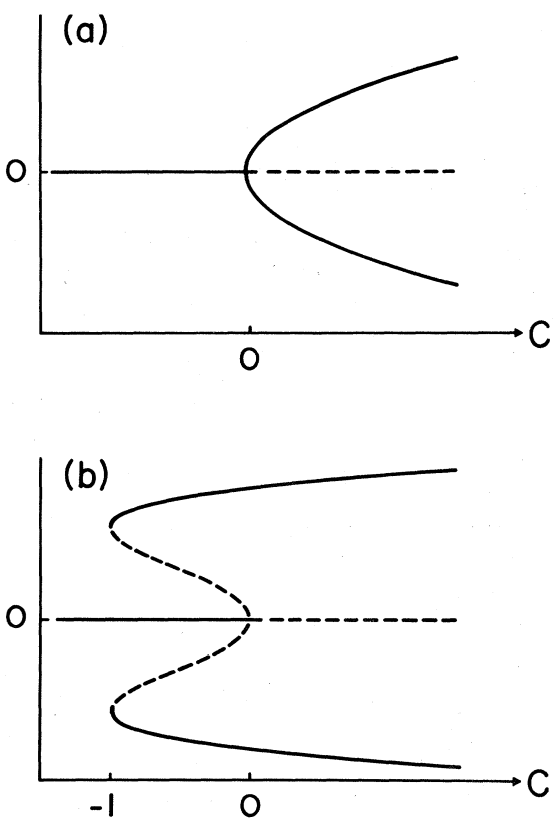

FIG. 9. Locations of the maxima (solid lines) and minima (dashed lines) in the stationary densities of the noise-induced f_* exact systems (8.31) and (8.38). The behavior in (a) as a function of the parameter c is like that seen in second-order phase-transition curves, while that of (b) is like that observed in first-order phase transitions. See the text for more detail.

tem temperature or pressure, the resulting graph has the characteristics of a second-order phase transition (Horsthemke and Lefever, 1984). It is important to realize that this appearance is due exclusively to the presence of the additive noise in the system.

The addition of noise to continuous time systems may result in different types of behavior similar to those found in first-order phase transitions, as illustrated by considering a second system

$$\frac{dr}{dt}=r(c+2r^2-r^4),\quad \frac{d\theta}{dt}=2\pi ,$$

which has a subcritical Hopf bifurcation at $c=-1$. As in the previous example we treat the effects of noise in the one-dimensional system

$$\frac{dx}{dt}=x(c+2x^2-x^4) . \tag{8.33}$$

The solutions of Eq. (8.33) have the following behavior. For $c<-1$ all solutions $x(t)\to 0$ regardless of the initial condition x_0. However, for $-1<c<0$ there is a tristability in that

$$x(t)\to\begin{cases}-\sqrt{1+(1+c)^{1/2}}, & \text{for } x_0<-x_*^+ ,\\ 0, \text{ for } -x_*^-<x_0<x_*^- , & \\ \sqrt{1+(1+c)^{1/2}}, & \text{for } x_*^+<x_0 ,\end{cases}$$

where $x_*^+=\sqrt{1+(1+c)^{1/2}}$ and $x_*^-=\sqrt{1-(1+c)^{1/2}}$. For $c>0$, the steady state $x_*=0$ becomes unstable, and this tristable behavior gives way to a bistability such that

$$x(t)\to\begin{cases}-\sqrt{1+(1+c)^{1/2}}, & \text{for } x_0<0 ,\\ \sqrt{1+(1+c)^{1/2}}, & \text{for } x_0>0 .\end{cases}$$

The stochastic differential equation corresponding to Eq. (8.33) is

$$\frac{dx}{dt}=x(c+2x^2-x^4)+\sigma\xi . \tag{8.34}$$

As before, we consider this problem for $-\infty<x<\infty$. The stationary solution of the Fokker-Planck equation corresponding to the stochastic differential equation (8.34) is

$$f_*(x)=K_2e^{\beta x^2(3c+3x^2-x^4)/6c} , \tag{8.35}$$

where β is as before, and again K_2 always exists. Therefore $f_*(x)$ is a stationary density. Furthermore, use of the same Liapunov function allows us to apply Theorem 8.5 to prove the convergence of the conditional entropy to zero.

The differences between these two systems immediately become apparent in Fig. 9(b), which again shows the location of the maxima of the stationary density (8.35). For $c<-1$, the stationary density $f_*(x)$ has a single maximum located at $x=0$, the globally stable steady state of the unperturbed system (8.33). For $-1<c,<0$, where the tristable behavior occurs, the stationary densities still have an absolute maximum at $x=0$, but they also display maxima at $x=\pm\sqrt{1+(1+c)^{1/2}}$ that become progressively more prominent as c increases. Finally, for $c>0$ the stationary density has absolute maxima located at $x=\pm\sqrt{1+(1+c)^{1/2}}$ and a local miminum at $x=0$. Translated into the language of a phase diagram, this behavior is precisely what is found in first-order phase transitions.

A second theorem giving a strong (third-order) version of the second law may also be established. This is given in the following.

Theorem 8.6. Assume that there is a unique generalized solution (8.25) to the Fokker-Planck equation (8.20), and that there is a unique stationary density

$$\lim_{t\to+\infty}u(t,x)=\lim_{t\to+\infty}P^tf(x)=f_*(x) .$$

Then

$$\lim_{t\to+\infty}H_c(P^tf|f_*)=0 .$$

Proof. From Eqs. (3.13) and (8.21) and from the adjointness of L^+ and L^-, we may write

$$\frac{dH_c}{dt} = -\int_X u \left[L^- \log\left[\frac{u}{f_*}\right]\right] dx \ . \qquad (8.36)$$

However, it is easy to show that

$$\begin{aligned} L^- \log\left[\frac{u}{f_*}\right] &= \frac{f_*}{u} L^- \left[\frac{u}{f_*}\right] \\ &\quad - \frac{1}{2}\left[\frac{f_*}{u}\right]^2 \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial}{\partial x_i}\left[\frac{u}{f_*}\right] \frac{\partial}{\partial x_j}\left[\frac{u}{f_*}\right] . \end{aligned}$$

Again using the adjointness of L^+ and L^-, and the fact that since f_* is a stationary density it satisfies Eq. (8.28), or $L^+ f_* \equiv 0$, we finally have

$$\begin{aligned} \frac{dH_c}{dt} = \frac{1}{2}\int_X \left[\frac{f_*^2}{u}\right] \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial}{\partial x_i}\left[\frac{u}{f_*}\right] \\ \times \frac{\partial}{\partial x_j}\left[\frac{u}{f_*}\right] dx \ . \qquad (8.37) \end{aligned}$$

Since the a_{ij} satisfy the uniform parabolicity condition (8.19), we have

$$\frac{dH_c}{dt} \geq 0 \ ,$$

with the equality holding if and only if $u = f_*$.

It is interesting that if we write $a_{ij}(x) = \sigma^2 h(x)$ where $0 \leq h(x) \leq 1$, then Eq. (8.37) can be written as

$$\frac{dH_c}{dt} = \sigma^2 K \ ,$$

where $K \geq 0$. This clearly shows how the rate of convergence of the conditional entropy to zero is dependent on the amplitude σ of the noise.

IX. SUMMARY

Here we have identified thermodynamic states with densities and shown how the evolution of these states may be studied by an examination of the evolution of densities under the action of integral operators. Stationary densities of these operators correspond to states of thermodynamic equilibrium. Invoking a maximum entropy principle, we have demonstrated that the entropy will have an absolute maximum if and only if a thermodynamic state is characterized by a generalization of the density of the microcanonical or canonical ensemble. Ergodicity of the system dynamics has been shown to be both necessary and sufficient for the existence of a unique state of thermodynamic equilibrium. Alternately, the dynamic property of asymptotic periodicity may imply that there are multiple states of relative thermodynamic equilibrium.

With respect to the question of convergence, asymptotic periodicity has been shown to be sufficient for the convergence of the entropy to a relative maximum that will, in general, depend on the way in which the system is prepared. The much stronger property of exactness is both necessary and sufficient for the global convergence of the entropy to its maximum value.

Given the dual observations that exactness is a property that only irreversible systems may have and that all of the laws of physics are framed in terms of reversible system dynamics, a clear problem exists with current attempts to formulate a satisfactory statistical mechanics. The first possibility is that the laws of physics are incorrectly formulated, which is not considered further here. Three alternate possibilities are considered.

In the first we show that if the system dynamics are at least mixing, then nontrivial coarse graining of the phase space will lead to a situation in which the entropy will evolve to its maximum value, but it will do so irrespective of the direction of time. In the second, closely related to the first, we consider the possibility that we have access to only a restricted number of the dynamical variables (trivial coarse graining, or taking the trace of a larger dynamics) and show that under this circumstance the entropy may appear to evolve to a relative or absolute maximum. In the third and last possibility we examine the effects of interactions with a heat bath. Under a variety of circumstances this may lead either to asymptotic periodicity and a consequent increase of the entropy to a relative maximum, or to exactness and the global increase of the entropy to an absolute maximum.

ACKNOWLEDGMENTS

I would like to thank my colleagues M. Bunge, M. Grant, A. Longtin, J. Milton, and Zeng Wan-Zhen (McGill University), A. Lasota and R. Rudnicki (Silesian University, Poland), U. an der Heiden (Witten, F.R.G.), and O. Rössler (Tübingen, F.R.G.) for their critical comments on this work. This work was initiated at the Universities of Oxford and Bremen, 1986–1987, and I would like to thank Professor J. D. Murray (Oxford) and Professor H. Schwegler (Bremen) for their hospitality during this period. I am especially indebted to Professor Schwegler for his continued interest and support in reading and commenting on almost every aspect of this work. This research was supported by the NSERC (Canada) through Grant No. A-0091 and the SERC (Great Britain) through Grant No. GR/D/13573.

REFERENCES

Abraham, R., and J. E. Marsden, 1978, *Foundations of Mechanics* (Benjamin Cummings, Reading, MA).

an der Heiden, U., 1985, in *Delay Equations, Approximation and Application,* edited by G. Meinardus and G. Nurnberger (Birkhauser, Basel), p. 147.

an der Heiden, U., and M. C. Mackey, 1982, J. Math. Biol. **16**, 75.

Anosov, D. V., 1963, Sov. Math. Dokl. **4**, 1153.

Anosov, D. V., 1967, Proc. Steklov Inst. Math. **90**, 1.

Arnold, V. I., and A. Avez, 1968, *Ergodic Problems of Classical Mechanics* (Benjamim, New York).

Bergmann, P. G., 1967, in *Delaware Seminar in the Foundations of Physics,* edited by M. Bunge (Springer, New York), p. 1.

Boltzmann, L., 1909, *Wissenschaftliche Abhandlungen,* edited by Fritz Hasenohrl (Barth, Leipzig).

Bratteli, O., and D. Robinson, 1987, *Operator Algebras and Quantum Statistical Mechanics,* Vol. 2 (Springer, New York).

Clausius, R., 1879, *The Mechanical Theory of Heat* (MacMillan, London).

Davies, P. C. W., 1974, *The Physics of Time Asymmetry* (University of California, Berkeley, Los Angeles).

Denbigh, K. G., and J. S. Denbigh, 1985, *Entropy in Relation to Incomplete Knowledge* (Cambridge University, Cambridge).

Dorfle, M., 1985, J. Stat. Phys. **40**, 93.

Dugas, R., 1959, *La Theorie Physique au sens de Boltzmann* (Griffon, Neuchatel, Switzerland).

Ehrenfest, P., and T. Ehrenfest, 1959, *The Conceptual Foundations of the Statistical Approach in Mechanics,* a translation by M. J. Moravcsik of their 1912 article in *Encyklopadie der Mathematischen Wissenschaften,* IV.2.II.6, published by B. G. Teubner, Leipzig (Cornell University, Ithaca).

Farquhar, I. E., 1964, *Ergodic Theory in Statistical Mechanics* (Wiley, London).

Fer, F., 1977, *L'irreversibilite, Fondement de la Stabilite du Monde* (Gauthier-Villars, Paris).

Gal-Or, B., 1974, in *Modern Developments in Thermodynamics,* edited by B. Gal-Or (Wiley, New York), p. 3.

Gal-Or, B., 1981, *Cosmology, Physics, and Philosophy* (Springer, New York).

Gibbs, J. W., 1902, *Elementary Principles in Statistical Mechanics* (Yale University, New Haven, CT).

Gikhman, I. I., and A. V. Skorokhod, 1969, *Introduction to the Theory of Random Processes* (Saunders, Philadelphia).

Goldstein, S., B. Misra, and M. Courbage, 1981, J. Stat. Phys. **25**, 111.

Goldstein, S., and O. Penrose, 1981, J. Stat. Phys. **24**, 325.

Grad, H., 1952, J. Phys. Chem. **56**, 1039.

Grandy, W. T., 1988, *Foundations of Statistical Mechanics,* Vols. I and II (Reidel, Dordrecht).

Hopf, E., 1932, Proc. Natl. Acad. Sci. U.S.A. **18**, 333.

Horsthemke, W., and R. Lefever, 1984, *Noise Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology* (Springer, Berlin).

Jaynes, E. T., 1957, Phys. Rev. **106**, 620.

Katz, A., 1967, *Principles of Statistical Mechanics* (Freeman, San Francisco).

Keener, J. P., 1980, Trans. Am. Math. Soc. **261**, 589.

Kestin, J., and J. R. Dorfman, 1971, *A Course in Statistical Thermodynamics* (Academic, New York).

Khinchin, A. I., 1949, *Mathematical Foundations of Statistical Mechanics* (Dover, New York).

Komornik, J., and A. Lasota, 1987, Bull. Pol. Acad. Sci. (Math.) **35**, 321.

Koopman, B. O., 1931, Proc. Natl. Acad. Sci. U.S.A. **17**, 315.

Krylov, N. S., 1979, *Works on the Foundations of Statistical Physics* (Princeton University, Princeton, NJ).

Kurth, R., 1960, *Axiomatics of Classical Statistical Mechanics* (Pergamon, Oxford).

Lasota, A., T. Y. Li, and J. A. Yorke, 1984, Trans. Am. Math. Soc. **286**, 751.

Lasota, A., and M. C. Mackey, 1985, *Probabilistic Properties of Deterministic Systems* (Cambridge University, Cambridge, England).

Lasota, A., and M. C. Mackey, 1987, Physica D **28**, 143.

Lasota, A., and M. C. Mackey, 1989, J. Math. Anal. Appl. **138**, 232.

Lebowitz, J. L., 1973, in *Nonlinear Problems in the Physical Sciences and Biology,* edited by I. Stakgold, D. D. Joseph, and D. H. Sattinger, Lecture Notes in Mathematics, Vol. 322 (Springer, Berlin), p. 193.

Lebowitz, J. L., and O. Penrose, 1973, Phys. Today **36**, (2) 23.

Lewis, R. M., 1967, J. Math. Phys. **8**, 1448.

Lorenz, E. N., 1980, Ann. N.Y. Acad. Sci. **357**, 282.

Loschmidt, J., 1876, Wien. Ber. **73**, 128.

Ma, S.-K., 1985, *Statistical Mechanics* (World Scientific, Philadelphia, Singapore).

Mackey, G. W., 1974, Adv. Math. **12**, 178.

Mackey, M. C., and L. Glass, 1977, Science **197**, 287.

Mackey, M. C., A. Longtin, and A. Lasota, 1989, J. Stat. Phys. (in press).

Misra, B., and I. Prigogine, 1981, in *Long Time Prediction in Dynamics,* edited by C. W. Horton, L. E. Reichl, and V. G. Szebehely (Wiley, New York), p. 21.

Misra, B., I. Prigogine, and M. Courbage, 1979, Physica A **98**, 1.

Pathria, R. K., 1972, *Statistical Mechanics* (Pergamon, Oxford).

Penrose, O., 1970, *Foundations of Statistical Mechanics* (Pergamon, Oxford).

Penrose, O., 1979, Rep. Prog. Phys. **42**, 1937.

Penrose, R., 1979, in *General Relativity,* edited by S. W. Hawking and W. Israel (Cambridge University, Cambridge, England), p. 581.

Petersen, K., 1983, *Ergodic Theory* (Cambridge University, Cambridge, England).

Prigogine, I., 1980, *From Being to Becoming: Time and Complexity in the Physical Sciences* (Freeman, San Francisco).

Provatas, N., and M. C. Mackey, 1989, "Bifurcation structure of noise induced asymptotic periodicity," Center for Nonlinear Dynamics, McGill University preprint.

Reichl, L. E., 1980, *A Modern Course in Statistical Physics* (University of Texas, Austin).

Rochlin, V. A., 1964, Am. Math. Soc. Transl. (2) **39**, 1.

Ruelle, D., 1969, *Statistical Mechanics* (Benjamin, Reading, MA).

Scalapino, D. J., 1961, "Irreversible statistical mechanics and the principle of maximum entropy," Ph.D. dissertation (Stanford University).

Sinai, Y., 1963, Sov. Math. Dokl. **4**, 1818.

Sinai, Y., 1970, Russ. Math. Surv. **25**, 137.

Skagerstam, B. S., 1974, Z. Naturforsch. **29a**, 1239.

Spohn, H., 1978, J. Math. Phys. **19**, 1227.

Thirring, W., 1983, *Quantum Mechanics of Large Systems* (Springer, New York).

Tolman, R. C., 1938, *The Principles of Statistical Mechanics* (Oxford University, Oxford).

Truesdell, C., 1960, in *Ergodic Theories,* edited by P. Caldirola (Academic, New York), p. 21.

Voigt, J., 1981, Commun. Math. Phys. **81**, 31.

Von Neumann, J., 1932, Proc. Natl. Acad. Sci. U.S.A. **18**, 70.

Walters, P., 1982, *An Introduction to Ergodic Theory,* Graduate Texts in Mathematics, Vol. 79 (Springer, New York).

Wehrl, A., 1978, Rev. Mod. Phys. **50**, 221.

Zermelo, E., 1896, Ann. Phys. (Leipzig) **57**, 485.