Nonequilibrium free energy, coarse-graining, and the Liouville equation

Brad Lee Holian
Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Harald A. Posch
Institute for Experimental Physics, University of Vienna, A-1090 Vienna, Austria

William G. Hoover
Department of Applied Sciences, University of California at Davis/Livermore, Livermore, California 94550
(Received 22 January 1990; revised manuscript received 2 April 1990)

The Helmholtz free energy is computed for an ensemble of initial conditions for a one-dimensional particle falling down a staircase potential, while in contact with a thermal reservoir. Initial conditions are chosen from the equilibrium canonical ensemble, with the gravitational field applied either as a step function (steady field) or a δ function (pulsed perturbation). The first case leads to a fractal steady-state distribution, while the second case leads to relaxation of a perturbed distribution back toward equilibrium. Coarse-graining is applied to the computation of the nonequilibrium entropy, with finer resolution in phase space accompanied by an increase in the number of trajectories. The limiting fine-grained (continuum) prediction of the Liouville equation is shown to be consistent with the numerical simulations for the steady state, but with incredibly slow (logarithmic) divergence appropriate to a lower-dimensional fractal distribution. On the other hand, simulations of the relaxation process show little or no sign of converging to the prediction obtained from the Liouville equation. Irreversible phase-space mixing of trajectories appears to be a necessary modification to the Liouville equation, if one wants to make predictions of numerical simulations in nonequilibrium statistical mechanics.

I. INTRODUCTION

The relationship between entropy, coarse-graining, and time-reversible atomistic dynamics has been a source of paradoxes since the earliest days of statistical mechanics.[1] Recent progress[2] has been made in understanding certain aspects of nonlinear response theory, the theoretical foundation for nonequilibrium statistical mechanics. Response theory assumes that the phase-space distribution function is continuous, as it is at equilibrium. If, in fact, the nonequilibrium distribution is *not* a smooth, continuous function, then severe difficulties arise in the mathematics of response theory.[2] One might suppose that the approach to equilibrium (such as in the relaxation following a pulsed perturbation) and the attainment of a nonequilibrium steady state (the response to steady external driving away from equilibrium) are simply inverse processes of each other. In this paper we will show that this common assumption is too simplistic, though both fundamental processes share one feature in their description, namely, Lyapunov instability, the tendency of nearby trajectories in phase space to diverge from each other exponentially with time.

In order to place the measurement of these two processes—relaxation and steady driving—on comparable footing, we need a characteristic function, which, when followed in time, gives an unambiguous signature of the approach to, or departure from equilibrium. Since an isothermal thermodynamic state is the most convenient (canonical) representation, we apply external forces to N molecules placed in a box of volume V in contact with a thermal reservoir at temperature T. The thermal bath guarantees that thermal equilibrium will occur when the system is pulsed, and the achievement of the steady state when the system is steadily driven, since work done on the system can be balanced by heat extracted from it. So that macroscopic irreversibility is not automatically guaranteed by using time-irreversible microscopic equations of motion (as in the case of stochastic forces for the coupling of the heat bath to the molecules), we will employ a deterministic, intrinsically time-reversible feedback method for thermostating the system.[3] The thermostat, as well as the external field, can be applied either homogeneously throughout the system to the molecules themselves (like a microwave oven), or heterogeneously to special molecules in boundary reservoirs, in which case the molecules in the sample obey the ordinary Newtonian (Hamiltonian) equations of motion. Without loss of generality, we will consider homogeneous thermostating in this paper.

In statistical mechanics, two limits are of central importance: The first is the infinite-ensemble, or continuum, limit (N_0 trajectories in $N_{\rm box}$ boxes in phase space, where both N_0 and $N_{\rm box}$ go to infinity, with their ratio held constant); the second is the thermodynamic limit (N particles in volume V, where both N and V go to infinity, with their ratio held constant). In this paper we do *not* emphasize the thermodynamic limit, which corresponds to an infinite-dimensional phase space. Rather, we choose to probe the basic mathematical structure of sta-

Lyapunov times (~ 50). We see a tendency that increases through the third snapshot and decreases somewhat in the fourth: The trajectories are flirting with the strange attractor of Fig. 2(b). This helps explain the apparent logarithmic ensemble-size dependence of the early-time width of the free-energy relaxation toward equilibrium. Ultimately, however, there is no question but that the ensemble relaxes back to equilibrium after 6–8 Lyapunov times, as represented by the coarse-grained free energy. We can only conclude that the Liouville continuity equation is inappropriate for statistical mechanics. The Boltzmann equation[10] (which introduces an explicitly irreversible term) and coarse-graining (which blurs some minor details of the distribution) appear to give qualitatively better descriptions of Nature. (Coarse-graining has its own pitfall of logarithmic divergence with ensemble size, however.)

Once again, it may be appropriate to add to the right-hand side of the Liouville equation an irreversible phase-space mixing term,[5] such as

$$\frac{\partial f}{\partial t}+\frac{\partial}{\partial \Gamma}\cdot(f\dot{\Gamma})=\lambda\nabla_{\zeta}^{2}(f-f_{0})\ . \tag{22}$$

Here, the coefficient $\lambda\sim\lambda_{\max}N_{0}^{-1/(2dN+1)}$ has units of the maximum Lyapunov exponent and becomes smaller as the phase-space resolution increases. The diffusion term

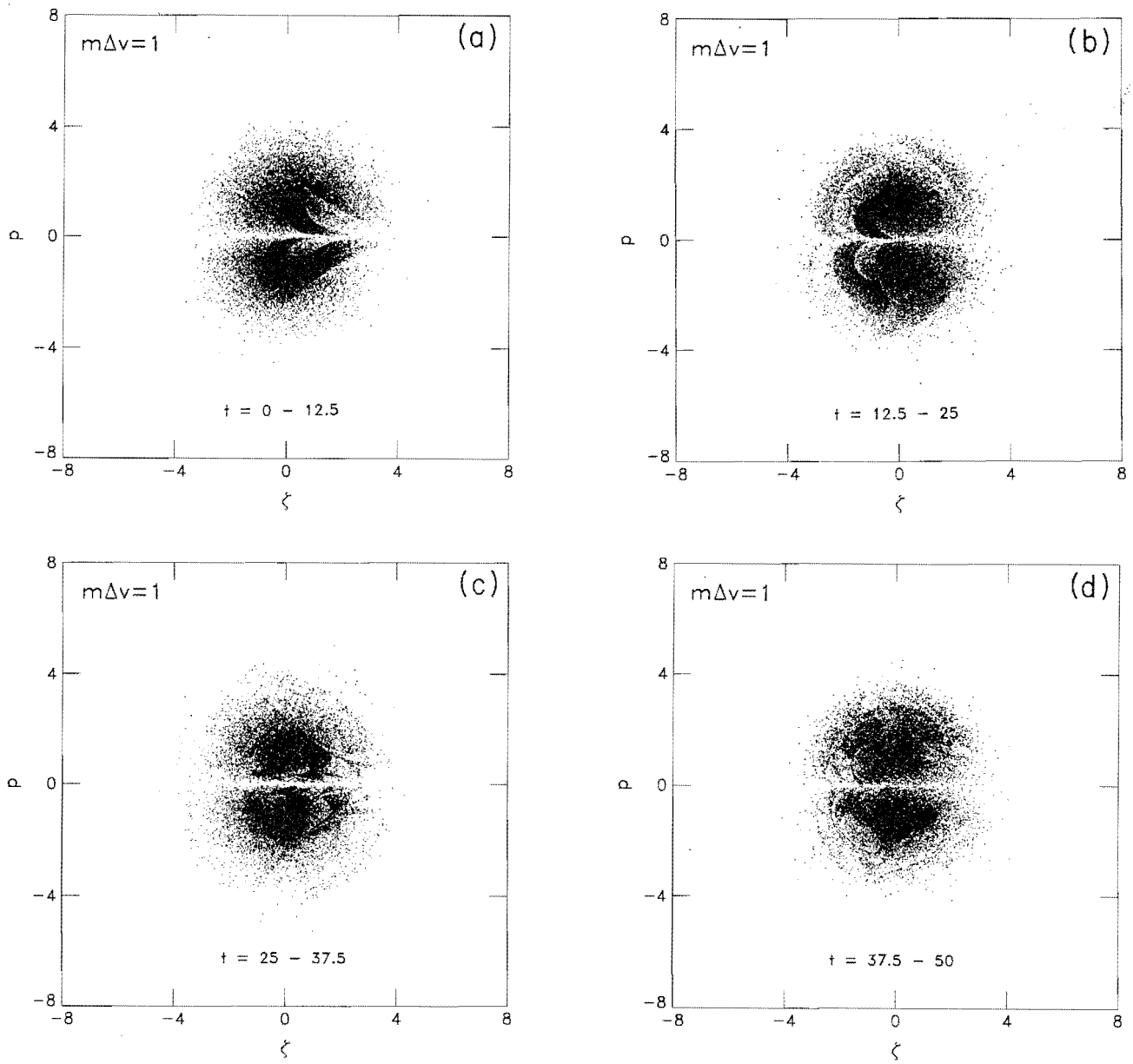

FIG. 10. Poincaré section for the pulse perturbation ($m\Delta v=1$), $N_0=10\,000$ initial conditions. Ranges of times after the δ-function pulse (spanning a total time of 50, or about two Lyapunov times) are shown in the four time-lapse accumulations of the transient distribution function, which is initially the equilibrium result of Fig. 2(a) shifted up by one unit along the momentum axis. Note the various features of the central part of the strange attractor [in Fig. 2(b)] that appear ephemerally, especially in the third snapshot; by about eight Lyapunov times, the distribution has once more settled down to the equilibrium one [see Fig. 2(a)].

in ζ mimics the phase-space mixing that occurs so dramatically in Fig. 10. When Eq. (22) is applied to the rate of change of mechanical observables (including the thermostating energy that is quadratic in ζ), the operations of time derivative d/dt and ensemble average $\langle\ \rangle$ commute. On the other hand, there arise additional terms for the entropy, or the free energy [Eq. (19)]:

$$\dot{A}=\langle \mathbf{J}\rangle\cdot\mathbf{X}-\lambda kT\int d\Gamma\frac{1}{f}[(\nabla_\zeta\Delta f)^2+dN\zeta\Delta f(\nabla_\zeta\Delta f)]\ , \tag{23}$$

where $\Delta f=f-f_0$. The first two terms are of order X^2 and the last is of order X^4; consequently, the correction to the Liouville-equation prediction is negative, at least for small fields. The last term can be of either sign and therefore contribute to the oscillatory behavior of ΔA [see Fig. 8(a)]. Qualitatively, these terms can account for the coarse-graining and phase-space mixing effects in both the steady- and pulsed-field results for the nonequilibrium free energy of the Galton staircase.

VI. CONCLUSIONS

For thermostated systems, the Helmholtz free energy is the appropriate characteristic function to show either the departure from, or return to, equilibrium. Unfortunately, predictions based on the Liouville continuity equation must be compared with finer and finer coarse-grained realizations of the engropy. For nonequilibrium steady states, this comparison is doomed because of the incredibly slow (logarithmic) divergence of entropy with ensemble size. It would seem intuitive that an ensemble of a few hundred experiments ought to suffice for a realization of the entropy, as it does for mechanical observables like the energy. However, the fractal nature of nonequilibrium steady-state distribution functions requires almost an infinite amount of information, or resolution, for the entropy.

It turns out that the really crucial test for the predictions using the Liouville equation is the relaxation to equilibrium from a pulsed external field, where the free energy is predicted to be a step function. On the contrary, numerical simulations show that relaxation to equilibrium occurs for the coarse-grained free energy after about eight Lyapunov times, regardless of ensemble size. Also, the distribution function recalls—fleetingly—the face of the strange attractor under steady driving, but finally it becomes once again a smooth, continuous function. Thus, any similarity between the processes of relaxation and the achievement of the nonequilibrium steady state is only transient. The notion that they are inverses of each other is an oversimplification.

Finally, we are forced to conclude that the mixing of trajectories in the process of equilibration is not described realistically by the Liouville equation. Phase-space mixing can be incorporated into a modified Liouville equation, which captures the essence of coarse-graining in an empirical way.

ACKNOWLEDGMENTS

We acknowledge very useful discussions with Kyozi Kawasaki, Eddie Cohen, Jerry Erpenbeck, Peter Milonni, Jim Hammerberg, Chris Patterson, Nicolas van Kampen, and David Chandler. Many of the focal points of this work originated in discussions with Ilya Prigogine.

[1]J. W. Gibbs, in *The Collected Works of J. Willard Gibbs* (Longmans, New York, 1931), Vol. 2.

[2]B. L. Holian, G. Ciccotti, W. G. Hoover, B. Moran, and H. A. Posch, Phys. Rev. A **39**, 5414 (1989); B. L. Holian, W. G. Hoover, and H. A. Posch, Phys. Rev. Lett. **59**, 10 (1987).

[3]S. Nosé, Mol. Phys. **52**, 255 (1984); J. Chem. Phys. **81**, 511 (1984); W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).

[4]B. J. Alder and T. Wainwright, in *Transport Processes in Statistical Mechanics*, edited by I. Prigogine (Interscience, New York, 1958).

[5]B. L. Holian, Phys. Rev. A **34**, 4238 (1986).

[6]M. C. Mackey, Rev. Mod. Phys. **61**, 981 (1989).

[7]See, for example, D. A. McQuarrie, *Statistical Mechanics* (Harper and Row, New York, 1976), pp. 35–40.

[8]W. G. Hoover, H. A. Posch, B. L. Holian, M. J. Gillan, M. Mareschal, and C. Massobrio, Molecular Simulations **1**, 79 (1987).

[9]H. A. Posch, W. G. Hoover, and B. L. Holian, Ber. Bunsenges. Phys. Chem. **94**, 250 (1990).

[10]L. Boltzmann, *Scientific Treatises of Ludwig Boltzmann*, edited by F. Hasenöhrl (Barth, Leipzig, 1909); see also *The Boltzmann Transport Equation*, edited by W. Thirring and E. G. D. Cohen (Springer, Vienna, 1973).

tistical mechanics by studying the limit of an infinite ensemble and its representation as a continuous distribution function in a finite-dimensional phase space.

Since the systems of interest are not thermally isolated, the initial conditions for the nonequilibrium experiments, performed under identical boundary conditions (external time-dependent driving), will be selected from an equilibrium canonical ensemble. If the system were thermally isolated, the characteristic function would be the entropy S, which would be a maximum at equilibrium. Over thirty years ago, Alder and Wainwright,[4] in a pioneering molecular-dynamics computer simulation, showed that the Boltzmann $\mathcal{H}$ function for a dilute gas, which is $-S/k$ (k is Boltzmann's constant), indeed decays to a minimum at equilibrium. However, for a finite, *thermostated* system, the Helmholtz free energy A is the appropriate characteristic function, and it reaches a minimum at equilibrium.

The Helmholtz free energy is given by

$$A(t)=E(t)-TS(t)=-kT\ln Z(t)\ , \tag{1}$$

where E is the energy of the system and Z is, by definition, the partition function. How do we measure such a thing for an ensemble of trajectories?

The energy E, like all mechanical observables, is most accurately and conveniently measured in the Lagrangian (Heisenberg), or co-moving, frame of reference in phase space. If all coordinates, momenta, and the thermostat heat-flow variable are collected into the vector Γ, then a trajectory in the multidimensional phase space is represented by $\Gamma_i(t)$, where i ranges from 1 to N_0, the ensemble size (number of different initial conditions). The energy function along a trajectory is $H_i(t)=H(\Gamma_i(t))$, and the ensemble average is

$$E(t)=\langle H(t)\rangle=\frac{1}{N_0}\sum_{i=1}^{N_0}H_i(t)$$
$$\rightarrow\int d\Gamma\, f_0(\Gamma)H(\Gamma(t))\ . \tag{2}$$

The continuum limit has been taken in order to obtain the latter expression,[2] in which the equilibrium (canonical) distribution function provides the weight of each trajectory and is given by

$$f_0(\Gamma)=\frac{1}{Z_0}\exp[-\beta H(\Gamma)]\ , \tag{3}$$

with $\beta=1/kT$ and the equilibrium partition function given by $Z_0=\int d\Gamma\exp(-\beta H)$.

The alternative way of looking at the ensemble average E is in the Eulerian (Schrödinger), or space-fixed, frame, where the N_0 trajectories are counted into $N_{\rm box}$ boxes fixed in phase space (box j is centered at Γ_j with volume $\Delta\Gamma_j$, and j ranges from 1 to $N_{\rm box}$). Then,

$$E(t)=\langle H(t)\rangle=\frac{1}{N_0}\sum_{j=1}^{N_{\rm box}}H_j n_j(t)$$
$$=\sum_{j=1}^{N_{\rm box}}H_j P_j(t)$$
$$\rightarrow\int d\Gamma\, H(\Gamma)f(\Gamma,t)\ . \tag{4}$$

Again, the continuum limit has been taken to get the latter expression.[2] The occupation number in box j is $n_j(t)$; the number of trajectories is

$$N_0=\sum_{j=1}^{N_{\rm box}}n_j(t) \tag{5}$$

and, therefore, the probability of occupation in box j is $P_j(t)=n_j(t)/N_0$, which is related to the distribution function f by $P_j(t)\simeq f(\Gamma_j,t)\Delta\Gamma_j$. (Note that both P and f are normalized to unity.) The number of boxes $N_{\rm box}$ is chosen so that the average occupation number in each box is $\langle n\rangle=N_0/N_{\rm box}$, a fixed ratio when we take the continuum limit. Because the boxes are coarse-grained in practical realizations, ensemble averages in the Eulerian picture are subject to more statistical noise than in the Lagrangian picture.

On the other hand, for the entropy, a measure of the phase-space probability density itself, there is no choice but to compute it in the Eulerian frame:

$$S(t)=\frac{k}{N_0}\ln\Omega(N_0,N_{\rm box})$$
$$=\frac{k}{N_0}\ln\left[N_0!\Big/\prod_{j=1}^{N_{\rm box}}n_j(t)!\right]$$
$$=\frac{k}{N_0}\left[N_0\ln N_0-\sum_{j=1}^{N_{\rm box}}n_j(t)\ln n_j(t)\right]$$
$$=-k\sum_{j=1}^{N_{\rm box}}P_j(t)\ln P_j(t)$$
$$\rightarrow -k\int d\Gamma\, f(\Gamma,t)\ln f(\Gamma,t)=-k\langle\ln f(t)\rangle\ . \tag{6}$$

The first line is Boltzmann's most famous equation. It states that the total entropy of the ensemble (N_0S) is Boltzmann's constant k times the logarithm of the number of ways $\Omega(N_0,N_{\rm box})$ that N_0 trajectories can be distributed among $N_{\rm box}$ boxes. Only the most probable configuration $\{n_j\}$ need be considered because of the huge factorials involved (even for only 100 trajectories in 100 boxes); therefore, Stirling's approximation to $\ln N_0!$ holds. The last line is obtained by taking the continuum limit, although we have arbitrarily thrown away a subtle, but well-known coarse-graining singularity, namely, $-k\ln\Delta\Gamma$. Usually, this singularity in the continuum expression for the entropy is dismissed as being trivial, since we say that we are only interested in entropy differences. Indeed, in this paper, we will report differences from equilibrium for each ensemble size.

While the entropy of a finite, thermostated system may oscillate rather wildly, the free energy either approaches (more or less monotonically) a minimum at equilibrium (A_0) or rises monotonically when the system is driven away from equilibrium:

$$\Delta A(t)=A(t)-A_0=-kT\ln\frac{Z(t)}{Z_0}$$

$$=kT\sum_{j=1}^{N_{\rm box}}P_j(t)\ln\frac{P_j(t)}{P_j^0}$$

$$\to kT\int d\Gamma\, f(\Gamma,t)\ln\frac{f(\Gamma,t)}{f_0(\Gamma)}\geq 0\ , \tag{7}$$

implying, as Gibbs showed,[1] that the accessible number of states at equilibrium is greater than under any other circumstance.

Earlier work identified the negative of Eq. (7) with T times "the nonequilibrium entropy, relative to the equilibrium distribution,"[5] which was shown in numerical realizations of thermal equilibration to rise unambiguously toward equilibrium. In a very interesting recent review article, which concentrates primarily on the closely related topic of irreversible dynamical systems, Mackey has independently made the same identification, except to call it the "conditional entropy."[6] From the preceding discussion, however, it is clear that we should more properly identify the characteristic function in Eq. (7) as the free-energy difference relative to the equilibrium value.

We should also note that a common way to derive[7] the equilibrium canonical distribution function f_0 is to maximize the entropy variationally with respect to f_0, subject to two constraints: (1) that the total energy of the ensemble be constant and (2) that f_0 be normalized. The latter constraint is eminently reasonable: It is nonsensical to imagine that trajectories can be either created or lost. The mathematical constraint on the total energy of the ensemble, however, has a serious physical flaw in its interpretation, namely, that the elements of the ensemble exchange energy among themselves and thereby interact with each other—a violation of our fundamental assumption of noninteracting trajectories. In fact, the correct physical picture of the thermal reservoir is that each member of the ensemble is submerged in it, and energy is exchanged between this heat bath and each element *independently*. Thus, the ensemble average can exhibit nonequilibrium fluctuations—even as we approach the continuum limit. It is only in the thermodynamic limit that fluctuations for each member of the ensemble disappear; only then is total energy conservation for the ensemble a good approximation. Since the formalism of statistical mechanics applies to systems with few degrees of freedom, we see clearly now that one must first take the continuum limit, and then the thermodynamic limit, if necessary. Taking the reverse order of limits can be very misleading.

The best way to think about the derivation of f_0 is to imagine minimizing the free energy variationally with respect to f_0, subject only to the constraint that f_0 be normalized. With this interpretation for a *thermostated* system with a finite number of particles, we see that, away from equilibrium, both the energy and entropy can fluctuate, so that the entropy is not necessarily a global maximum at equilibrium, as it is in isolated systems. Instead, the characteristic function that attains an unambiguous global minimum at equilibrium, except for random fluctuations, is the free energy.

II. ENSEMBLE DYNAMICS AND THE LIOUVILLE EQUATION

In the framework of the NVT or canonical ensemble, we now consider generalizable equations of motion for N particles in a d-dimensional box of volume V (periodic boundary conditions) in contact with a thermal reservoir at temperature T. In addition, we provide for the operation of an external driving force $\mathbf{X}(t)$. (For simplicity, we assume that both the thermostat and $\mathbf{X}$ act homogeneously throughout V.) We will be concerned with two cases: (1) step-function driving, $\mathbf{X}(t)=\mathbf{X}\Theta(t-t_0)$, leading to a nonequilibrium steady-state response; and (2) δ-function (pulsed perturbation) driving, $\mathbf{X}(t)=\Delta\mathbf{p}\,\delta(t-t_0)$, giving a response that relaxes back to equilibrium. For the example of mass transport, such as conductivity under an imposed electric or gravitational field, the equations of motion are

$$\dot\Gamma(\Gamma,t)=\begin{cases}\dot{\mathbf{q}}=\mathbf{p}/m\ , & \text{(8a)}\\ \dot{\mathbf{p}}=\mathbf{F}(\mathbf{q})-\nu\zeta\mathbf{p}+\mathbf{X}(t)\ , & \text{(8b)}\\ \dot\zeta=\nu(\mathbf{p}\cdot\mathbf{p}/dN\,mkT-1)\ . & \text{(8c)}\end{cases}$$

The phase space is $\Gamma=(\mathbf{q},\mathbf{p},\zeta)$, where $\mathbf{q}$ are the dN particle coordinates, $\mathbf{p}$ are the dN particle momenta, and ζ is the dimensionless heat-flow variable that describes the relative magnitude and direction of heat flow from the particles to the thermal reservoir (positive ζ means the particles lose kinetic energy to the bath and negative means the bath heats them up). Thus, the so-called "friction" coefficient $\nu\zeta$ can have either sign, in fact. The coupling of the heat bath to particles is governed by the rate-of-thermostating parameter ν: $\nu=0$ gives the usual Newtonian (Hamiltonian) equations of motion; in order for the thermostating to be efficient, it is best to choose ν to be on the order of the mean collision rate of the particles. The response of the particles to the thermostat is in the nature of integral feedback,[3] i.e., on the timescale of $1/\nu$. Furthermore, Eq. (8c) guarantees that the long-time average of the kinetic energy $K(\mathbf{p})=\mathbf{p}\cdot\mathbf{p}/2m$ will always be $\frac{1}{2}dNkT$, even at a nonequilibrium steady state. [Under the equilibrium thermostated equations of motion, the canonical distribution f_0 is a stationary solution to the Liouville continuity equation, to be presented later. Thus, for sufficiently mixing systems, this thermostat guarantees that the long-time average of an observable is equivalent to a canonical-ensemble average,[3] since any equilibrium trajectory will eventually visit every box in phase space, with probability $f_0(\Gamma)d\Gamma$.] From the total potential energy $\Phi(\mathbf{q})$, we obtain the internal forces $\mathbf{F}(\mathbf{q})=-\partial\Phi(\mathbf{q})/\partial\mathbf{q}$.

The total energy of the system of particles and thermostat is

$$H(\Gamma)=K(\mathbf{p})+\Phi(\mathbf{q})+\tfrac{1}{2}dNkT\zeta^2\ , \tag{9}$$

which includes a contribution from the thermostat of or-

der unity compared to the extensive (order N) quantities K and Φ, since $\langle \zeta^2 \rangle_0 = 1/dN$. Note also that the work done by the thermostat is quadratic in the heat-flow variable. The rate of change of H for the equations of motion, Eqs. (8), is therefore

$$\dot{H} = -dNkT\nu\zeta + \mathbf{J}\cdot\mathbf{X} , \tag{10}$$

where the flux is $\mathbf{J} = \mathbf{p}/m$ in response to the external force $\mathbf{X}$. Thus, the macroscopic equation of motion for E is

$$\dot{E} = \langle \dot{H} \rangle = \dot{Q} - \dot{W} , \tag{11}$$

where the rate of heat flow *into* the system is

$$\dot{Q} = -dNkT\nu\langle \zeta \rangle , \tag{12}$$

and the rate of work done *by* the system is

$$\dot{W} = -\langle \mathbf{J} \rangle \cdot \mathbf{X} . \tag{13}$$

For the cases we are interested in, both of these quantities are typically negative, i.e., work is done *on* the system and heat is extracted *from* it. For example, at the steady state, the flux is given by the transport coefficient α (conductivity in our mass-transport example), which is in general a positive-definite nonlinear function of $\mathbf{X}$, times the field $\mathbf{X}$: $\langle \mathbf{J} \rangle_{ss} = \underline{\alpha}(\mathbf{X})\cdot\mathbf{X}$, showing that $\dot{W}_{ss} = -\alpha X^2 < 0$. From the equations of motion, we see that $\nu\langle \zeta \rangle_{ss}$ is a positive frictional damping coefficient, so that

$$\dot{Q}_{ss} = -dNkT\nu\langle \zeta \rangle_{ss} < 0 .$$

Since $\dot{E}_{ss} = 0$, we expect to find in numerical simulations that $\dot{Q}_{ss} = \dot{W}_{ss}$. Equation (11) embodies the first law of thermodynamics.

The macroscopic equation of motion for the entropy of the ensemble necessarily involves not just a simple average over independent trajectories, but rather a probabilistic measure of the *density* of trajectories in all of phase space. For this, we need the Liouville continuity equation in its full generality to describe the flow of trajectories:

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial \boldsymbol{\Gamma}} \cdot (f\dot{\boldsymbol{\Gamma}}) = 0 . \tag{14}$$

The swarm of trajectories in the $(2dN+1)$-dimensional phase space is thus imagined to be a peculiar fluid of noninteracting "particles."[2] The Liouville equation can also be written in the form

$$\frac{df}{dt} + f\Lambda = 0 , \tag{15}$$

where the logarithmic expansion rate of local phase-space volume elements is

$$\Lambda = \frac{\partial}{\partial \boldsymbol{\Gamma}} \cdot \dot{\boldsymbol{\Gamma}}(\boldsymbol{\Gamma}, t) . \tag{16}$$

From the equations of motion [Eqs. (8)], we see that Λ has no explicit field dependence and therefore no explicit time dependence:

$$\Lambda = -\nu\zeta \frac{\partial}{\partial \mathbf{p}} \cdot \mathbf{p} = -dN\nu\zeta . \tag{17}$$

(This is the usual case for most nonequilibrium processes, though one could imagine bizarre equations of motion that would lead to more exotic expressions for Λ.) Applying the Liouville equation to the entropy $S = -k\langle \ln f \rangle$, we find that

$$\begin{aligned} T\dot{S} &= -kT\langle \dot{\ln} f \rangle \\ &= kT\langle \Lambda \rangle \\ &= -dNkT\nu\langle \zeta \rangle \\ &= \dot{Q} , \end{aligned} \tag{18}$$

where we rearrange Eq. (15) to read $\dot{\ln} f = \dot{f}/f = -\Lambda$. Equation (18) embodies the mathematical statement of the second law of thermodynamics, as predicted by the Liouville equation. This leads to a remarkably simple prediction for the behavior of the Helmholtz free energy:

$$\begin{aligned} \dot{A} &= \dot{E} - T\dot{S} \\ &= (\dot{Q} - \dot{W}) - \dot{Q} \\ &= -\dot{W} \\ &= \langle \mathbf{J} \rangle \cdot \mathbf{X} , \end{aligned} \tag{19}$$

or, since equilibrium is imagined to prevail for $-\infty < t < t_0$, so that $A(t_0) = A_0$, we have

$$\Delta A(t) = \int_{t_0}^{t} ds \langle \mathbf{J}(s) \rangle \cdot \mathbf{X}(s) . \tag{20}$$

The case of the pulsed perturbation, $\mathbf{X}(t) = \Delta\mathbf{p}\delta(t - t_0)$, deserves particular attention; there,

$$\Delta A(t) = \langle \mathbf{J}(t_0) \rangle \cdot \Delta\mathbf{p} = \langle \mathbf{p}(t_0) \rangle \cdot \Delta\mathbf{p}/m .$$

For t_{0-}, t_0, and t_{0+}, $\langle \mathbf{p}(t) \rangle$ is $\mathbf{0}$, $\frac{1}{2}\Delta\mathbf{p}$, and $\Delta\mathbf{p}$, respectively, so that $\Delta A(t > t_0) = |\Delta\mathbf{p}|^2/2m$. Since the distribution function $f(\boldsymbol{\Gamma}, t_{0+})$ is the equilibrium f_0 perturbed by translating the momentum origin, $f_0(\mathbf{q}, \mathbf{p} - \Delta\mathbf{p}, \zeta)$, the entropy is initially unaffected by the δ-function external force. For these thermostated systems, the surprising thing is that the Liouville equation for f predicts a step-function change in the free energy, with *no* subsequent relaxation back to equilibrium. For an isolated system, the equivalent conundrum (Gibb's paradox) is that the entropy never changes as the system relaxes.[1]

If we coarse-grain the available phase space into N_{box} discrete boxes ($\Delta\boldsymbol{\Gamma}$) and make the resolution finer and finer ($N_{box} \to \infty$), taking more and more initial conditions ($N_0 \to \infty$), with the ratio of trajectories to boxes N_0/N_{box} fixed, do we find that the free energy converges or diverges? And is this phase-space convergence (divergence) rapid or slow? What happens as the system approaches the steady state under step-function driving? Does the system, when perturbed by a δ-function pulse, ever equilibrate? In what sense, if any, are these two nonequilibrium processes equivalent? These questions are fundamental to the application of statistical mechanics to the real world.

III. THE GALTON STAIRCASE

We choose as an example[8] of the equations of motion, Eqs. (8), a single particle on a one-dimensional (1D) "washboard" substrate (a sinusoidal potential) with spatial periodicity 2π:

$$\Phi(q)=1-\cos q \ , \tag{21}$$

where the unit of energy is such that the washboard well depth is 2, to be compared with $kT=1$ for the thermal reservoir. (For convenience, we choose the particle mass m to be 1 , so that the unit of time gives a harmonic vibrational period of 2π.) When an external gravitational field $X=mg$ is applied, we get the "Galton staircase" of Fig. 1. (The Galton Staircase is the 1D analog of the 2D Galton board, a triangular array of scattering pins. When balls are dropped down from the top of the Galton board, the binomial distribution is obtained in bins set up at the bottom. The 1D single-particle Galton Staircase can also be viewed as a two-particle periodic system interacting with the above potential. With one of the particles carrying charge $+e$ and applying an external electric field, the equations of motion for the relative coordinate and momentum can be interpreted as a single quasiparticle with the reduced mass; the center-of-mass motion then disappears from the problem.) We have chosen to study the 1D Galton staircase because it is an excellent prototype for a nonequilibrium statistical mechanical system: It is highly nonlinear, yet it has only three phase-space dimensions (coordinates q, momentum p, heat-flow variable ζ), the minimum possible dimensionality for exhibiting chaotic trajectories. It is therefore feasible in computer simulations to construct up to 100 bins in each dimension, a total of 10^6 bins; in comparison, consider a thermostated 3D fluid of 32 particles under periodic boundary conditions (a small molecular-dynamics system these days): Comparable resolution in phase space would require $100^{6N+1}=10^{386}$ boxes.

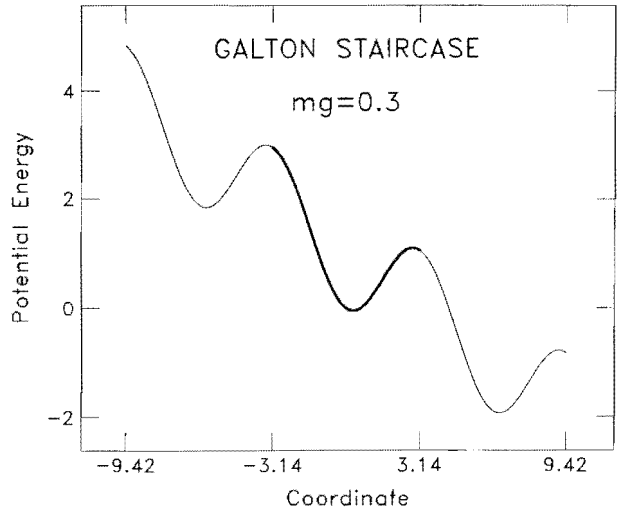

FIG. 1. Total potential energy (including gravitational) for the Galton staircase, $1-\cos q-0.3q$, as a function of coordinate q. the fundamental spatial period ($-\pi$ to π) is shown in heavy line. [In this and subsequent figures, the unit of energy is such that the equilibrium (zero-gravity) well depth of the potential is 2, as compared with the temperature of the thermal bath, $kT=1$. The particle mass $m=1$ and the unit of time is such that the harmonic vibrational period is 2π (the actual anharmonic period for $mg=0.3$ is 13.74; for $mg=0$, the period is 7.34). The thermostating rate, or coupling strength to the heat reservoir, is $\nu=0.316$.]

In computing entropy in the Eulerian picture for the Galton staircase, we divide phase space into boxes in the following way: The coordinate (modulo 2π, $-\pi<q<\pi$), the momentum ($-6<p<6$), and the heat-flow variable ($-6<\zeta<6$) are binned so that

$$N_{q\,\mathrm{box}}=N_{p\,\mathrm{box}}=N_{\zeta\,\mathrm{box}}=N_{\mathrm{box}}^{1/3} \ ,$$

with the number of boxes N_{box} as nearly as possible equal to the number of initial conditions N_0. (Since p and ζ are not periodic, but instead unbounded, the few trajectories that fall outside the designated boxes are lumped into the nearest ones.)

IV. STEP-FUNCTION EXTERNAL FIELD, $X(t)=mg\,\theta(t-t_0)$

When a particle on the Galton staircase is subjected to a steady gravitational field, it would accelerate indefinitely were it not for the restraint imposed by the thermostat. In fact, the thermostat guarantees that a steady-state downhill velocity is achieved. An interesting way to display the probability density (distribution function) for this fall down the staircase is to plot the Poincaré surface of section: Plot the point (ζ,p) whenever a trajectory in the ensemble crosses the $q=2\pi n$ plane ($n=0,\pm1,\pm2,\cdots$).

The equilibrium Poincaré section is shown in Fig. 2(a). The equilibrium trajectories describe a doughnutlike object, whose cross section reveals a dearth of probability for the particle to be frozen at the potential minimum ($q=2\pi n$) with zero momentum. In general, the thermostated equations of motion for many-body systems, Eqs. (8), have a strong instability for a special set of initial conditions, namely, $\mathbf{q}=$lattice sites (where $\mathbf{F}=\mathbf{0}$) and $\mathbf{p}=\mathbf{0}$ relative to the center of mass, giving $\zeta(t)=\zeta(0)-vt$. This instability leads to a noncanonical equilibrium distribution within a small coaxial volume in phase space in the neighborhood of the ζ axis. Outside the cylinder, the distribution approaches f_0; inside, it goes to zero. Except near the surface of this cylinder, the distribution satisfies the Liouville equation, just as the canonical f_0 does. This noncanonical artifact is, of course, more pronounced for low-dimensional thermostated systems, but even in the Galton staircase for $\nu=0.316$, the first three moments of the momentum are canonical and the fourth-order cumulant of the momentum is reduced only 2% from its canonical value of 3. Otherwise, the distribution is smooth and featureless.

On the other hand, the nonequilibrium steady state produces a fascinating multifractal picture in Fig. 2(b), resembling a triskelion. While the steady-state distribution is larger in extent than the equilibrium one, the actual volume is zero; it is a multifractal object, that is, at any given point its local dimensionality varies, but in general

is nonintegral and less than three, the dimensionality of f_0. Of course, at the steady state the distribution absolutely *cannot* have an overall dimensionality greater than the full phase space—that would be absurd. On the other hand, the distribution clearly does not shrink to a limit cycle of dimension 2, either. In fact, the dimensionality for $mg=0.3$ and $\nu=0.316$ has been measured[8,9] to be 2.47. (That is, the Poincaré section in Fig. 2(b) is ~1.5 dimensional, as measured by the average logarithm of the local area around a point, divided by the logarithm of the radial resolution, in the limit of zero radius.)

The dimensionality of the steady-state distribution is consistent with our arguments about Λ, the local logarithmic expansion rate of phase-space volume, and its relationship to the heat flow. That is, $\langle \Lambda \rangle_{ss} < 0$, so that the volume of the distribution must shrink to zero. The

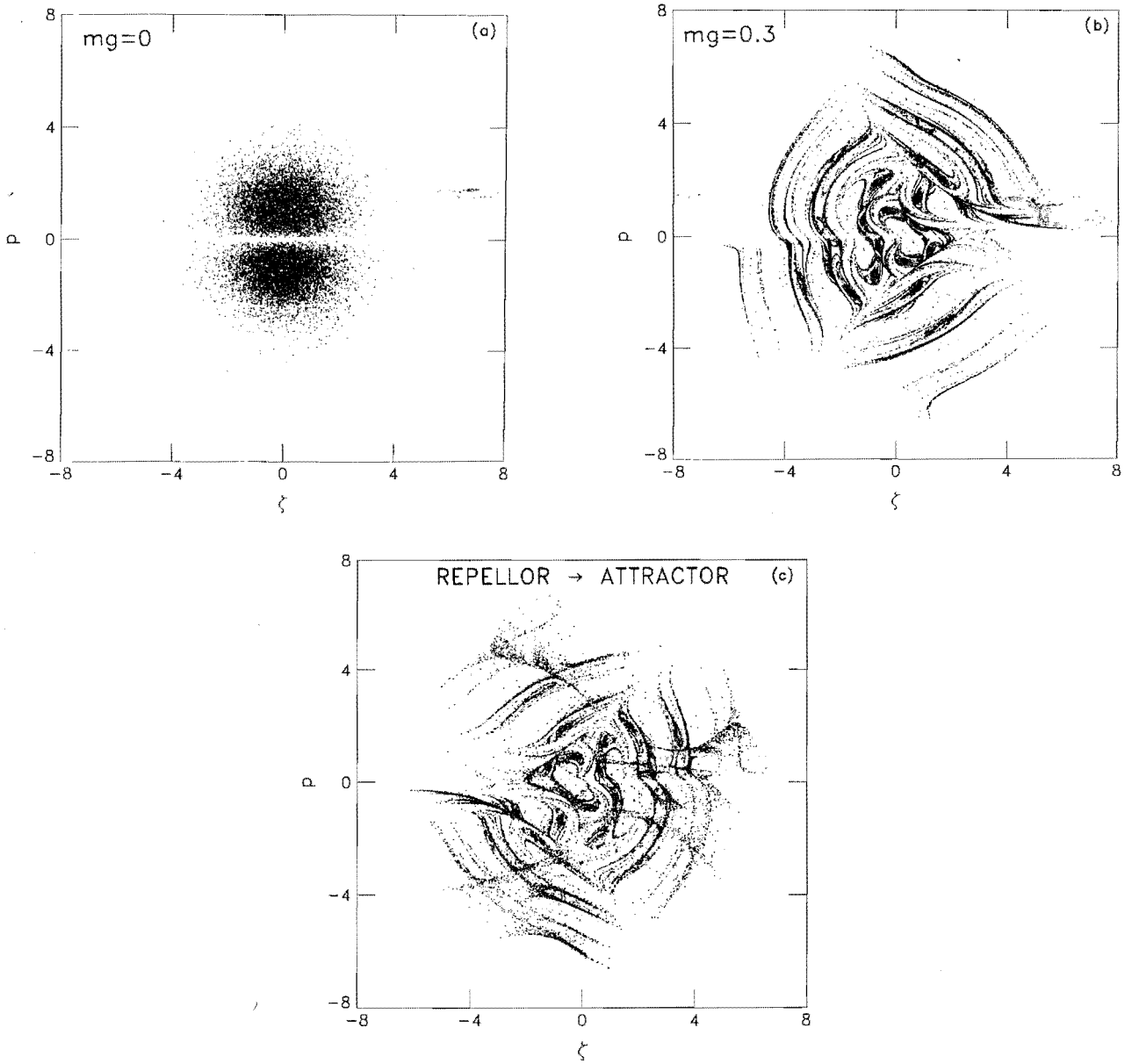

FIG. 2. Poincaré surface of section for the Galton staircase, where the momentum p is plotted vs the heat-flow coefficient ζ whenever the coordinate q reaches an integral multiple of 2π. (a) At equilibrium ($mg=0$), an ensemble of $N_0=10\,000$ initial conditions generates the Poincaré section at an average rate of 13.6 points per trajectory in a time of 50; note the smoothness of the equilibrium distribution. (b) At the steady state ($mg=0.3$), all trajectories collapse onto the multifractal strange attractor. After a time of ~2000 at the steady state, the Poincaré section is accumulated over a time of 50 at an average rate of 7.3 points per trajectory. The phase-space fractal dimensionality of the attractor is 2.47 (see Ref. 9). (c) By performing the time-reversal transformation ($q\rightarrow q$, $p\rightarrow -p$, $\zeta\rightarrow -\zeta$) on the $N_0=10\,000$ ensemble at a time of ~2000 after the establishment of the steady state, the strange repellor is obtained as a Poincaré section, which is accumulated for approximately one Lyapunov time (25). The repellor states are Lyapunov unstable, persisting for only about two Lyapunov times, during which the second law is violated (negative conductivity); here, after one Lyapunov time, faint wisps that destroy the inversion symmetry of the repellor can already be seen, as some trajectories begin to head again for the attractor—a process completed for virtually the entire ensemble after about four Lyapunov times.

Lyapunov exponents,[9] one for each dimension in phase space for this system, are $+0.0393$, 0, and -0.0842. The three cumulative sums represent exponential rates at which one-, two-, and three-dimensional elements in phase space grow with time; in particular, the sum of all the Lyapunov exponents is $\sum_i \lambda_i = \langle \Lambda \rangle_{ss}$. Thus, while the attractor to which the distribution collapses at the steady state has zero volume, nearby pairs of trajectories diverge exponentially with time from each other because $\lambda_{max} > 0$; hence the name "strange attractor."

In the pantheon of paradoxes in statistical mechanics, Loschmidt's paradox of macroscopic irreversibility arising from reversible microscopic equations of motion is represented by the so-called "strange repellor."[2] It is obtained by reversing velocities (and the heat-flow variable) after training the ensemble of trajectories onto the strange attractor—the time-reversal transformation, $t \to -t$, $q \to q$, $p \to -p$, $\zeta \to -\zeta$. The reversed trajectories obey the same equations of motion, but the miraculous thing is, the average flux is reversed—the particle falls *uphill* on average—and the transport coefficient (conductivity) is *negative*.

The resolution to this paradoxical violation of our intuition and experience (the second law of thermodynamics) is that the volume occupied by these strange repellor states is identically zero (like the attractor), so that the probability of actually observing such a violation of the second law is identically zero. Moreover, these states have the same Lyapunov exponents, except for sign: $+0.0842$, 0, and -0.0393. Thus, the repellor is absolutely unstable[9] on a time scale of $1/\lambda_{max} \sim 25$. The states that are approximately on the repellor soon find their way back to the attractor, as shown in Fig. 2(c). The wispy, non-repellor-like features on this Poincaré section collected within a time of 25 after the time-reversal transformation [which results in a 180° rotation of Fig. 2(b)] are the most unstable trajectories, which are already heading back to the attractor. By a time of 50, the average flux has gone from $-\langle J \rangle_{ss}$ to zero, and by a time 100, the flux has returned to $\langle J \rangle_{ss}$. This feature is insensitive to the time step for the central difference approximation to the equations of motion, as well as to the length of time spent at the steady state before the time reversal. (In contrast, after an impulse to an equilibrium ensemble, we have observed that reversibility depends weakly, i.e., logarithmically, on the time step. In principle, the reversibility time can be made arbitrarily long by using accurate and intrinsically time-reversible algorithms, such as central differences, and by using longer and longer computer word lengths. In practice, we have found this to be severely subject to the law of diminishing returns.) Generally, the entropy shows significantly more sensitivity to reversibility than mechanical observables such as the energy.

The steady-state plateau values of mechanical observables are attained on a time scale of 6–8 Lyapunov times ($\tau = 1/\lambda_{max} \sim 25$). For $X = mg = 0.3$ ($v = 0.316$), the ensemble average momentum, shown in Fig. 3, reaches a value of 0.149, so that the conductivity is $\alpha = \langle J \rangle_{ss}/X = 0.497$ and, thus, $\dot{W}_{ss} = -0.0447$. The heat-flow variable, shown in Fig. 4, reaches a steady-state

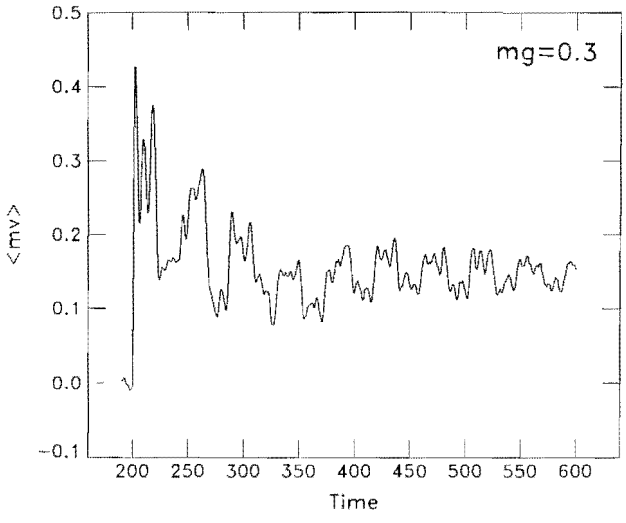

FIG. 3. Response of particle momentum $p = mv$ to a steady gravitational field ($mg = 0.3$) turned on at time $t_0 = 200$. The steady-state response (after 6–8 Lyapunov times) is $\langle mv \rangle_{ss} = 0.149$, so that the conductivity is 0.497. The average time for the particle to jump from one stair step to another is $42.17 = 2\pi/\langle v \rangle_{ss}$. The number of initial conditions (trajectories) is $N_0 = 25\,000$.

value of 0.142, so that $\dot{Q}_{ss} = -0.0449$. To well within the error bar of ± 0.0003, $\dot{Q}_{ss} = \dot{W}_{ss}$, thus verifying the first law of thermodynamics for the Galton staircase. The dependence of ensemble averages of mechanical observables upon ensemble size (we have studied $N_0 = 1600$, 4000, 10 000, 25 000, and 62 500) is completely negligible, while the root-mean-square magnitude of fluctuation about the mean shows the usual relative order $N_0^{-1/2}$ dependence.

The steady-state plateau value of the entropy, on the other hand, is much more slowly achieved, taking 5–10 times longer than mechanical observables. In Fig. 5 we show the nonequilibrium Helmholtz free energy,

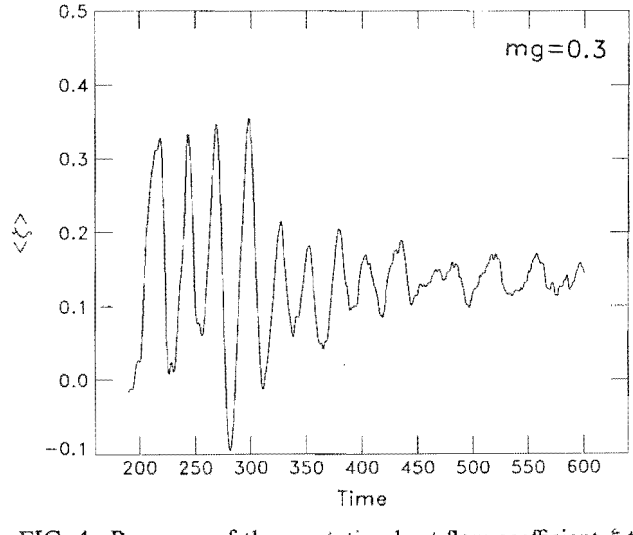

FIG. 4. Response of thermostating heat-flow coefficient ζ to a steady field ($mg = 0.3$) turned on at $t_0 = 200$. The steady-state value is $\langle \zeta \rangle_{ss} = 0.142$, for ensemble size $N_0 = 25\,000$.

$A = E - TS$, for the $mg = 0.3$ Galton staircase. At early times, the Liouville prediction of Eq. (20) for ΔA is a quadratic upturn with time, since the response of J is increasing linearly with time. At late times, when J has reached a steady-state value, ΔA grows linearly with time. It is clear from Fig. 5(a) that these numerical results agree well with the early-time prediction, for times substantially less than a Lyapunov time. However, in Fig. 5(b), the plateau values for ensemble sizes differing by a factor of 2.5 are evenly spaced, suggesting a logarithmic dependence on N_0. Thus, while the computer simulations appear to be reaching for the Liouville prediction of divergence with increasing ensemble size, the approach

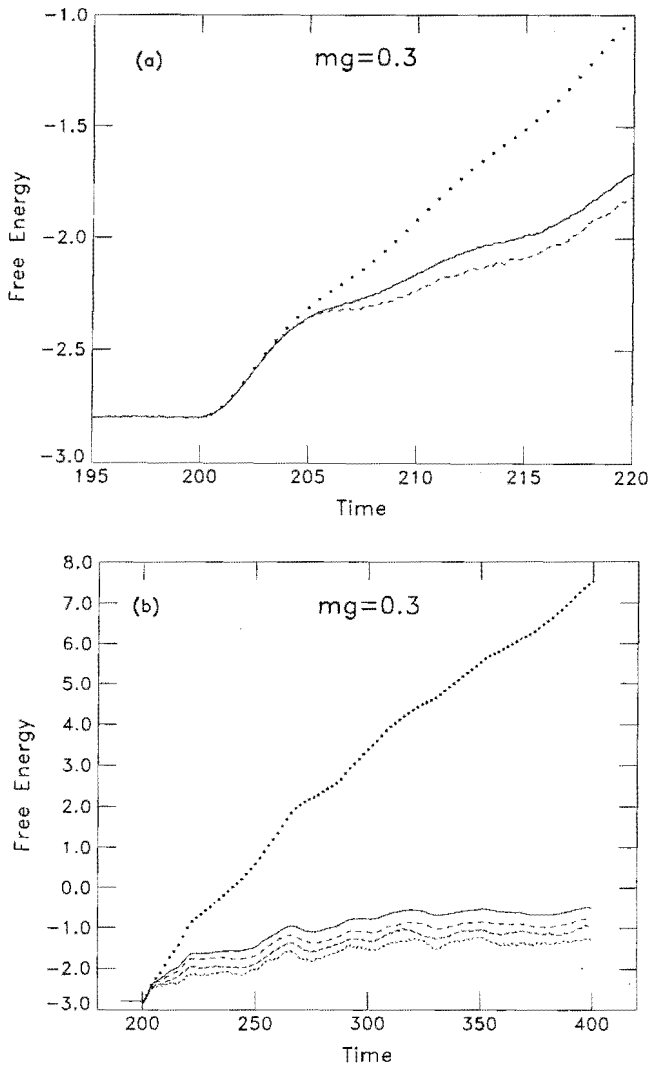

FIG. 5. Nonequilibrium free energy $A = E - TS$ for the Galton staircase under a steady gravitational field ($mg = 0.3$) turned on at time $t_0 = 200$. (a) The early quadratic behavior with time, in close agreement with the Liouville prediction (dots), i.e., time integral of the flux times the field, is shown for ensemble sizes $N_0 = 25\,000$ (solid line) and 10 000 (dashes). (b) At longer times, the free energy for various ensemble sizes separate logarithmically: $N_0 = 25\,000$ (solid line), 10 000 (long dashes), 4000 (medium dashes), 1600 (short dashes), and Liouville equation prediction (dots); asymptotes ($\sim -2.3 + 0.5 \log_{10} N_0$) are fully achieved after times $t \sim t_0 + 1500$.

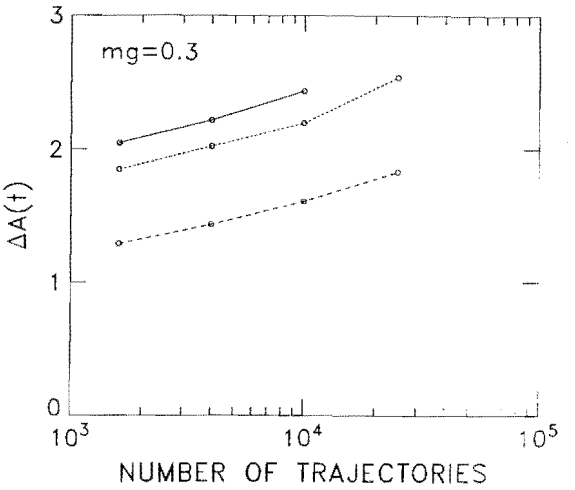

FIG. 6. Free-energy rise $\Delta A(t) = A(t) - A(t_0)$ as a function of ensemble size N_0 for various times, $t - t_0$: 65 (dashed line), 400 (dotted line), 2000 (essentially infinite time, solid line); note that the rise is clearly logarithmic with N_0.

is pitifully slow. This signature of the singular fractal distribution is shown in Fig. 6, confirming that the resolution in the free energy depends logarithmically on ensemble size, even as early as 2–3 Lyapunov times. The case of step-function driving shows that the mathematics of nonequilibrium statistical mechanics, as represented by the prediction of the Liouville equation, is utterly impractical in application, though not necessarily wrong. For example, we estimate that in order to achieve agreement with Eq. (20) for about eight Lyapunov times, we would need to simulate 10^{15} trajectories on the computer.

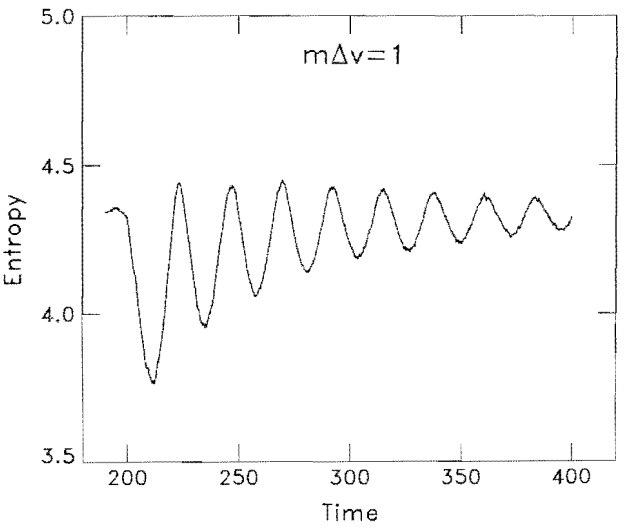

FIG. 7. Entropy S of the thermostated response to a pulsed field ($m\Delta v = 1$) imposed at time $t_0 = 200$, for an ensemble of $N_0 = 25\,000$ initial conditions at equilibrium (equilibrated from $t = 0$ to t_0); the particle relaxes toward equilibrium on a time scale of 6–8 Lyapunov times (150–200). Note that, unlike the entropy of an *isolated* system, which rises monotonically toward a maximum at equilibrium, the thermostated entropy oscillates noticeably (even overshooting its final equilibrium value), with a period of 22.9, about one Lyapunov time.

V. δ-FUNCTION PULSED PERTURBATION, $X(t)=\Delta p\,\delta(t-t_0)$

When a particle is given an impulse on an otherwise horizontal (equilibrium) Galton staircase, the distribution function is perturbed by translating the momentum origin of f_0 by Δp. As we pointed out earlier, the entropy is therefore continuous across the δ function. Soon, however, the entropy drops, as shown in Fig. 7. It recovers its equilibrium value in a decidedly nonmonotonic, strongly oscillatory way. The oscillations correspond to a mechanical bouncing down the staircase, with a period of about one Lyapunov time. With $\Delta p=1$ and $kT=1$, the average kinetic energy is half the potential barrier, so that there are many trajectories in the ensemble where the particles can fall a long way before coming to rest after the initial "kick."

The free energy is indeed a much smoother characteristic function for this thermostated system: in Fig. 8(a), ΔA looks very much like the relaxation toward equilibrium of the Boltzmann $\mathcal{H}$ function (negative entropy) for an *isolated* system. The general shape of ΔA, especially at long times, is Lorentzian, $[1+(t/\tau)^2]^{-1}$, regardless of ensemble size [Fig. 8(b)]. These numerical results differ dramatically from the step-function shape predicted by Eq. (20) from the Liouville equation, particularly after 6–8 Lyapunov times. At early times, i.e., less than two Lyapunov times, the width of the Lorentzian appears to grow logarithmically with ensemble size, very much like the steady-field case (see Fig. 9).

The early-time relaxation process can be illuminated by generating a series of Poincaré "time-lapse exposures" following the pulse perturbation: Figure 10 shows a series of four such exposures, spanning a total of two

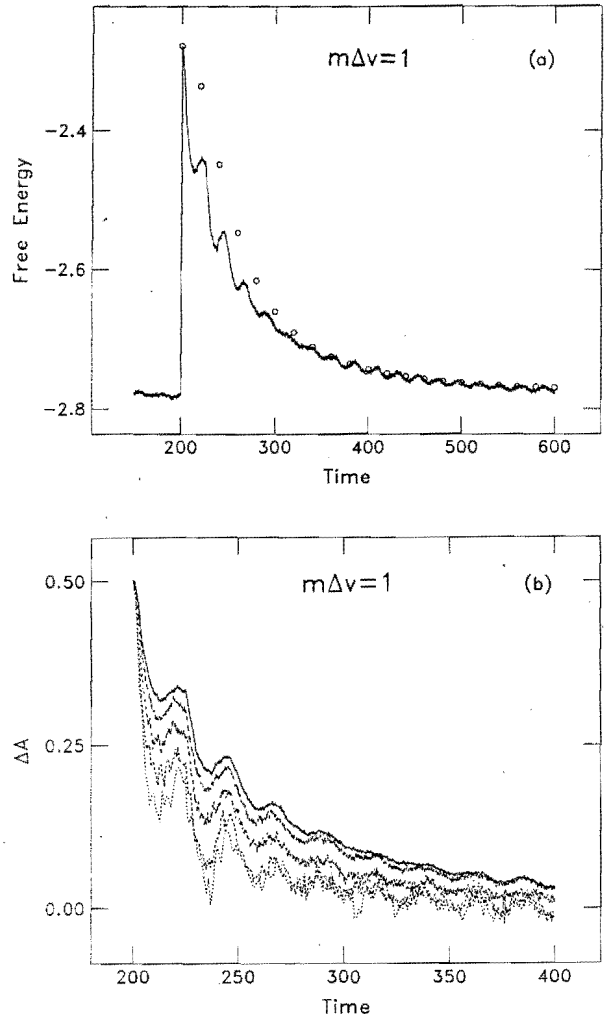

FIG. 8. Relaxation of the free energy with time toward equilibrium, following a δ-function external field ($m\Delta v=1$) imposed at time $t_0=200$. (a) At times $t-t_0$ greater than ~150, the response (solid line, $N_0=62\,500$) behaves like a Lorentzian (the circles are for a Lorentzian $\tau=56$). (b) The ensemble-size dependence is represented by the solid line ($N_0=62\,500$) and increasingly shorter dashes (25 000, 10 000, 4000), down to the dotted line (1600). At early times, the width of ΔA increases weakly with N_0.

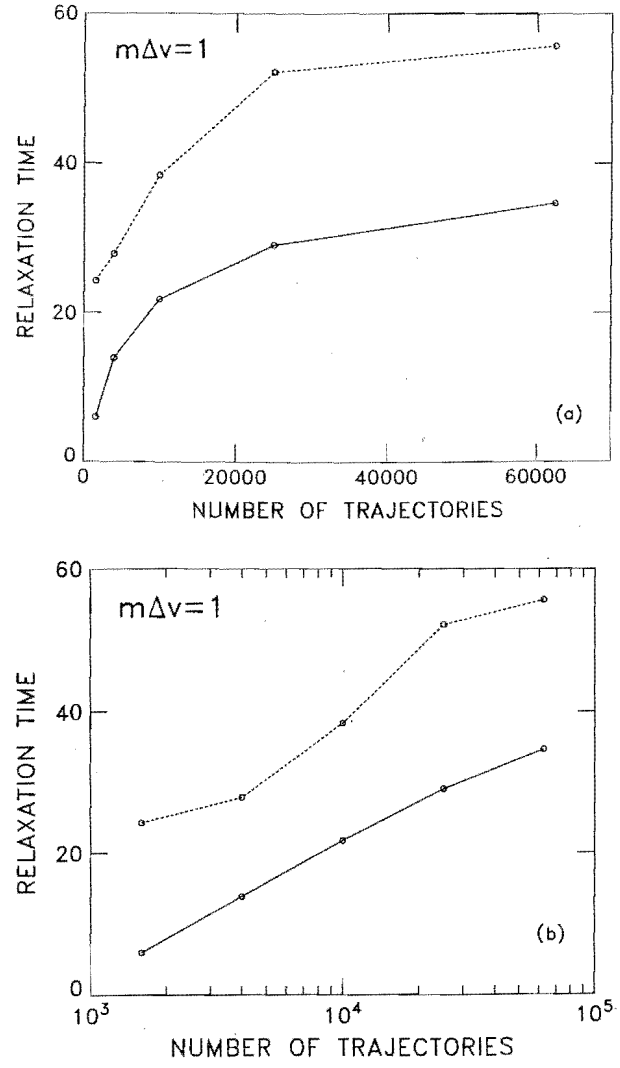

FIG. 9. Relaxation time for the free energy after a δ-function pulse ($m\Delta v=1$), as a function of ensemble size. Full width at half maximum is shown as a solid line connecting data points; the Lorentzian fit at 3τ is shown as a dashed curve. (a) Linear-linear (τ vs N_0). (b) Log-linear (τ vs $\log_{10}N_0$). Note that the earlier-time fit is more nearly logarithmic with N_0, while the later-time fit converges with N_0.