Mechanical models of Maxwell's demon with noninvariant phase volume

Kechen Zhang*
Department of Physics, Henan Medical University, Zhengzhou, Henan 450052, China

Kezhao Zhang†
Department of Physics, Peking University, Beijing 100871, China
(Received 9 April 1992)

This paper is concerned with the dynamical basis of Maxwell's demon within the framework of classical mechanics. We show that the operation of the demon, whose effect is equivalent to exerting a velocity-dependent force on the gas molecules, can be modeled as a suitable force field without disobeying any laws in classical mechanics. An essential requirement for the models is that the phase-space volume should be noninvariant during time evolution. The necessity of the requirement can be established under general conditions by showing that (1) a mechanical device is able to violate the second law of thermodynamics if and only if it can be used to generate and sustain a robust momentum flow inside an isolated system, and (2) no systems with invariant phase volume are able to support such a flow. The invariance of phase volume appears as an independent factor responsible for the validity of the second law of thermodynamics. When this requirement is removed, explicit mechanical models of Maxwell's demon can exist.

PACS number(s): 05.70.Ln, 05.40.+j, 05.90.+m, 03.20.+i

I. INTRODUCTION

In this paper, we treat Maxwell's demon [1,2] as a purely mechanical device, and seek to find the dynamical basis for the demon from the perspective of classical mechanics. Since the effect of Maxwell's demon is equivalent to exerting a velocity-dependent force on the gas molecules, we focus on what kind of force field is needed to implement this effect. The starting point is the mechanical system composed of N interacting point particles:

$$m_i\ddot{\mathbf{r}}_i=\mathbf{F}_i(\mathbf{r}_1,\ldots,\mathbf{r}_N,\dot{\mathbf{r}}_1,\ldots,\dot{\mathbf{r}}_N)\ , \tag{1}$$

where m_i and $\mathbf{r}_i$ are the mass and position of the ith particle. Within this framework, we will show that the effect of Maxwell's demon can be implemented by using suitable velocity-dependent force fields, and an essential requirement for the mechanical model of Maxwell's demon is that the phase-space volume of the system must be noninvariant, or *not* preserved during time evolution.

In the following sections, we will first construct a specific type of mechanical models of Maxwell's demon by using velocity-dependent force fields. These systems all have noninvariant phase volume. In order to examine whether the noninvariance is a general requirement, we will first formulate the second law of thermodynamics in an equivalent form which only involves nonvanishing and robust momentum flows. Because this formulation is readily expressed in terms of mechanics, we will be able to show rigorously that in systems with symmetric energy with respect to momentum reversal, violation of the second law requires noninvariance of phase volume.

It is interesting to note that this conclusion from our mechanical approach is consistent with that from information-theoretic or computational approaches [2–6], where it has been pointed out that the contraction of phase volume is needed for the operation of Maxwell's demon due to energy dissipation caused by information erasure [4].

It should be emphasized however that our models of Maxwell's demon perform fully automatically as "autonomous" mechanical systems without actually dissipating energy. In our approach, there is no explicit formulation of how information about the movement of the particles is gathered, processed, and used to control a door of some kind. Instead, we consider how the motion of the particles is affected by the velocity-dependent force fields such that the overall effects of Maxwell's demon are achieved. Because we always choose the force to be perpendicular to the velocity, keeping the kinetic energy of the particles unaffected, the mechanical demon does not need continuous input of free energy while it works.

II. MECHANICAL MODELS OF MAXWELL'S DEMON

A. A simple example

Consider a modified version of Maxwell's pressure demon (see Ref. [2], p. 6): Imagine a torus-shaped container is filled with gas and divided by a special "diaphragm" which allows gas molecules to pass only in one direction [Fig. 1(a)]. It is not difficult to see that a macroscopic momentum flow will arise and circulate clockwise indefinitely inside the torus regardless of the initial state of the system (with possible exceptions of measure zero). This flow emerges automatically as long as the total kinetic energy is nonzero, and is robust against any external disturbances. We call such sustaining and robust momentum flow *spontaneous.*

The operation of the diaphragm leads to violation of the second law of thermodynamics. For instance, we can

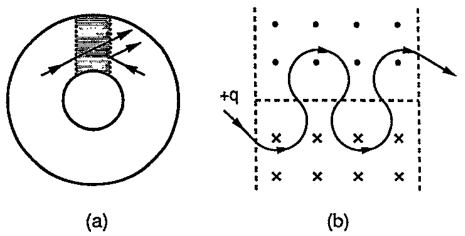

FIG. 1. (a) A Maxwell's demon as a diaphragm (shaded area) which allows gas molecules approaching from the left to pass through freely but reflects molecules approaching from the right. (b) Uniform magnetic fields with opposite directions in the two adjacent regions allow a positively charged particle to pass through along the midline via a piecewise circular path from left to right, but not reversely.

keep the torus in thermal contact with a single heat reservoir and extract work from the kinetic energy of the flow, say, by a small turbine. The joint operation of the diaphragm and the turbine will convert heat into work without other effects: This is a perpetual motion machine of the second kind.

Obviously even a small probability $\alpha \ll 1$ favoring the gas molecules to pass through the diaphragm in one direction is sufficient to generate the momentum flow. When probability α is very small, the maximum rate of entropy reduction by the operation of the diaphragm can be estimated as follows. Since the speed u of the nonvanishing macroscopic momentum flow along the longitude direction of the torus is also very small, the number of gas molecules passing through a unit area of the diaphragm during a unit time approximately equals

$$\alpha n \int_0^\infty (m/2\pi kT)^{1/2} \exp(-mv^2/2kT) v\, dv = \alpha n (kT/2\pi m)^{1/2} ,$$

where n is the number density of the gas molecules. On the other hand, this number flux should also equal the macroscopic speed u times the number density n. Thus

$$u \approx \alpha (kT/2\pi m)^{1/2} . \tag{2}$$

When the momentum flow is exploited to do work to the environment, e.g., via a turbine, the entropy of the system will be decreasing at the rate $dS/dt = -T^{-1} dW/dt$. Apparently, the rate for the work being extracted from the kinetic energy of the flow submits to $0 < dW/dt < nuA(mu^2/2)$, where A is the total area of the diaphragm. Using (2), we can get

$$-(k/4\pi)\alpha^2 nuA < \frac{dS}{dt} < 0 .$$

Note that in Fig. 1(a), the volume of the phase space is not preserved. In fact, the operation of the diaphragm implies contraction of the phase volume, because trajectories passing through the diaphragm may merge with those being reflected from the opposite side of the diaphragm. The noninvariance of the phase volume turns out to be of general importance. The dynamics of this system also happens to be irreversible, but it will become clear later that this feature is not essential.

It is important to realize that the effect of the diaphragm on the motion of the gas molecules is functionally equivalent to a velocity-dependent force field. Figure 1(b) is an example showing how a Lorentz force can serve as a field barrier which allows a charged particle to pass only in one direction. However, this is true only in the local region near the midline: If we also take into account the container walls at the periphery of the magnetic barrier (not illustrated in Fig. 1), it can be seen that the particle may leak through the barrier in the opposite direction via a sequence of collisions along the container walls. The net flow will be exactly zero no matter how the magnetic fields are arranged. This conclusion follows directly from our general theorem in Sec. III and the fact that the Lorentz force still preserves the phase volume.

B. "Implementation" of the demon using velocity-dependent force

Following the example of the pressure demon in Fig. 1(a), now we use a velocity-dependent force field to implement a barrier similar to the special diaphragm which allows particles to pass through preferably only in one direction. We consider the point-particle system (1), and choose the force to be perpendicular to the velocity, just like the Lorentz force, so that the total energy of the system is automatically conserved.

The concrete form of the force can be specified with much freedom without the danger of violating any laws in classical mechanics. We start with the simple case where the velocity $\mathbf{v}=(v_x,v_y,0)$ is restricted on the x-y plane. We modify the formulation of the Lorentz force and let the force be (in arbitrary units)

$$\mathbf{F}(\mathbf{v}) = (\mathbf{v}\times\hat{\mathbf{z}}) v \sin\theta , \tag{3}$$

where θ is the angle between $\mathbf{v}$ and the x axis, $v=|\mathbf{v}|$ is the speed, and $\hat{\mathbf{z}}$ is the unit vector of the z axis. Note that expression $(\mathbf{v}\times\hat{\mathbf{z}})$ is a Lorentz force, and the scalar factor $(v \sin\theta)$ only modifies the magnitude of the force. The direction of the force is reversed when the scalar factor is negative. Obviously $\mathbf{F}(\mathbf{v})$ is still perpendicular to $\mathbf{v}$. Since $\mathbf{F}(\mathbf{v})$ always points toward the positive x axis [Fig. 2(a)], we call $\hat{\mathbf{x}}$ the *attracting direction* of the force field. The symmetry of the force field with respect to the x axis is evident in Fig. 2. It is also clearly seen by rewriting Eq. (3) as

$$\mathbf{F}(\mathbf{v}) = \mathbf{v}\times(\hat{\mathbf{x}}\times\mathbf{v}) .$$

Evidently, this model can be immediately generalized into three dimensions by assuming the axial symmetry of $\mathbf{F}(\mathbf{v})$ around x.

Now consider the trajectory of a single particle of unit mass within a uniform force field specified by (3). The speed v of the particle is conserved in the motion. By $\dot{\mathbf{v}}=\mathbf{F}(\mathbf{v})$, we get

$$\dot{v}_x = v_y^2 , \tag{4}$$

$$\dot{v}_y = -v_x v_y . \tag{5}$$

To compute the trajectory, it is convenient to use θ as a

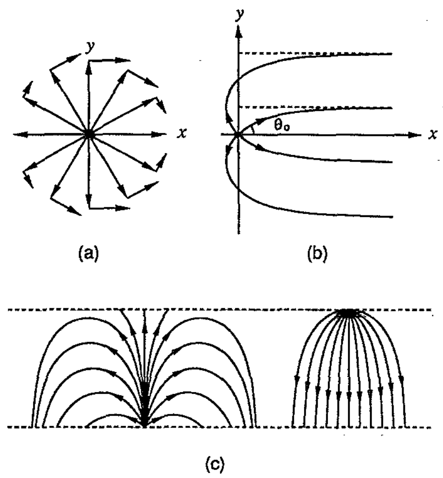

FIG. 2. (a) Dependence of the force given by (3) upon the direction of the velocity (represented by the radial arrows). (b) Trajectories starting from the origin approach the asymptote $y=\theta_0(-\pi<\theta_0<\pi)$. When $\theta_0>\pi/2$ or $<-\pi/2$, x attains its minimum $\ln|\sin\theta_0|$ at $y=\theta_0-\pi/2$ or $\theta_0+\pi/2$. (c) Uniform field between the dashed lines with the attracting direction pointing downward allows all particles approaching from above to pass through, but not the reverse, except the "leakage" caused by nearly perpendicular impinging.

parameter. From $v_x=v\cos\theta$, $v_y=v\sin\theta$, and Eqs. (4) and (5), we can obtain $\dot\theta=-v\sin\theta$. Thus, $dx/d\theta=v_x/\dot\theta=-\cot\theta$, and $dy/d\theta=v_y/\dot\theta=-1$. For the initial conditions $x=y=0$ and $\theta=\theta_0$, the solution is

$$x=\ln|\sin\theta_0|-\ln|\sin\theta| , \quad y=\theta_0-\theta .$$

The trajectory depends only on the initial direction θ_0 of the velocity, but not on the speed v. It always approaches the direction of $\hat{\mathbf{x}}$ [Fig. 2(b)].

It is interesting to note that similar equations were also derived in a Lorentz gas model with a time-varying thermostatting force [7]. It follows from (4) and (5) that $\partial\dot x/\partial x+\partial\dot y/\partial y+\partial\dot v_x/\partial v_x+\partial\dot v_y/\partial v_y=-v_x$. By Liouville's theorem, the phase volume $dx\,dy\,dv_x dv_y$ is expanding when $v_x<0$ and contracting when $v_x>0$. For trajectories going along the attracting direction, the phase volume is contracting (see also Ref. [7]).

Figure 2(c) illustrates how this force field is used to implement the special diaphragm of the pressure demon. Whereas a particle approaching the field barrier downward (in the attracting direction of the field) can always pass through, a particle approaching from below can pass the barrier directly only when $|\sin\theta_0|<\exp(-w)$, where w is the width of the field barrier and, as before, θ_0 is the angle between the approaching velocity and the attracting direction. This mechanism still works when we add in elastic collisions among the particles and between the particles and the container walls. In fact, within the force field the component of the total momentum along the attracting direction never decreases regardless of the interactions among the particles, because the interactions conserve the momentum while Eq. (4) ensures $\dot v_x\geq 0$ for each individual particle. In contrast to the Lorentz force mechanism in Fig. 1(b) where a particle may collide with a container wall again and again along the antiattracting direction, now a particle will leave the wall after a single collision to approach the attracting direction [8].

C. Modified models

In the model considered in the preceding section, there exists weak "leakage" of particles along the antiattracting direction due to nearly perpendicular impinging. Although the leakage can be made as small as desirable by increasing the width of the field barrier, we point out that it is possible to get rid of the leakage altogether by using force of the form

$$\mathbf{F}(\mathbf{v})=(\mathbf{v}\times\hat{\mathbf{z}})A(v)B(\theta) . \tag{6}$$

Consider the special case

$$\mathbf{F}(\mathbf{v})=(\mathbf{v}\times\hat{\mathbf{z}})v\tan(\theta/2) . \tag{7}$$

Now the situation is similar to that in (3) and $\hat{\mathbf{x}}$ is still the attracting direction, but in the antiattracting direction $(\theta\to\pm\pi)$, instead of approaching 0, the magnitude of the force now approaches infinity. This will prevent the leakage [9]. The trajectory now becomes

$$x=\cos\theta_0-\cos\theta+\ln(1-\cos\theta_0)-\ln(1-\cos\theta) ,$$

$$y=\theta_0-\theta+\sin\theta_0-\sin\theta .$$

As the initial angle $\theta_0\to\pm\pi$, the trajectory approaches a limit curve, whose minimum of x has the finite value

$$x_{\min}=\ln 2-1<0 \tag{8}$$

at $\theta=\pm\pi/2$. A field barrier thicker than $1-\ln 2$ will practically stop the leakage.

The original form of Maxwell's demon, the speed demon, can also be implemented by a force field of the type (6). It is helpful to notice that in a uniform force field of this type, the geometric shape of a trajectory is determined only by the function $B(\theta)$. Because the speed v is conserved, the function $A(v)$ serves only as a constant scaling factor for the equations of motion, and thus also as a scaling factor for the trajectory without altering its geometric shape. In the special case $A(v)=v$, such as in (3) and (7), the trajectory is independent of the speed v. We can use this special case as a reference. When $A(v)\neq v$, the trajectory is geometrically similar to the reference, but the size is scaled by a factor

$$v/A(v) .$$

As a consequence, if $v/A(v)$ is an increasing (decreasing) function of v, then the trajectory has a larger (smaller) size for higher (lower) speed of the particle.

Figure 3 shows how this property is exploited to implement the speed demon. Now Eq. (7) is the reference case with $A(v)=v$. In this case, a particle can penetrate a barrier up to a distance of $1-\ln 2$ [perpendicular-impinging limit, see expression (8)]. For an arbitrary function $A(v)$, this distance becomes $(1-\ln 2)v/A(v)$ due to the scaling property. Therefore, in Fig. 3 a parti-

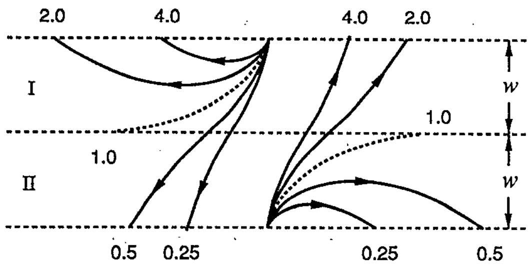

FIG. 3. A barrier composed of two uniform force fields of type (6) allows a particle to pass downward only if its speed $v<1$, and upward only if $v>1$. In region I: $A(v)=v^2$, and the attracting direction of the field is upward. In region II: $A(v)=1$, and the attracting direction is downward. In both regions, $B(\theta)=\tan(\theta/2)$. The trajectories shown are the limit curves for perpendicular impinging along the antiattracting directions. Numbers refer to speeds. $w=1-\ln 2$.

cle may pass through barrier II where $A(v)=1$ only if its speed $v>w/(1-\ln 2)$. Once it has passed barrier II, it is along the attracting direction of barrier I and thus has no problem going out. Similarly, a particle may pass through barrier I followed by barrier II only if its speed $v<(1-\ln 2)/w$. The overall effect is to favor high speed particles to pass upward and low speed particles to pass downward.

Notice that broken time-reversal symmetry is not necessary for the operation of Maxwell's demon, although in Fig. 1(a) the operation of the diaphragm is irreversible. The example of Fig. 2 has reversible dynamics. In general, the system with velocity-dependent force of the general form (6) is time reversible if and only if $\mathbf{F}(-\mathbf{v})=\mathbf{F}(\mathbf{v})$, which is equivalent to

$$B(\theta+\pi)=-B(\theta)\ ,$$

where the function $B(\theta)$ is of period 2π. As for the invariance of phase volume, it follows from Liouville's theorem that for a single particle in a field (6), the phase volume $dx\,dy\,dv_x dv_y$ is preserved if and only if $B(\theta)$ is a constant.

III. IS NONINVARIANCE OF PHASE VOLUME NECESSARY FOR MAXWELL'S DEMON?

A. Spontaneous momentum flow (SMF)

The mechanical models of Maxwell's demon considered in the previous sections all have noninvariant phase volume. Is this noninvariance an accidental feature or a general requirement?

One closely related result comes from the computational approach: It was shown that erasure of information is needed to prepare the memory of Maxwell's demon for the next round of computation [4]. The erasure of information implies energy dissipation, which in turn implies phase volume contraction. However, this mechanism of phase volume contraction is different from that in our mechanical models, where the energy is always conserved. Another difference is that in our models the demon is implemented simply as velocity-dependent force fields, which do not have internal states for information storage and processing. It seems rather natural to regard these models as purely mechanical devices.

For a rigorous mechanical approach to the problem of noninvariance of phase volume, we will elaborate the concept of spontaneous momentum flow first introduced in the model of Fig. 1. In general, by *spontaneous momentum flow* (SMF), we mean a sustaining and robust momentum flow inside an isolated mechanical system. Robustness is an essential feature of SMF. A purely sustaining momentum flow needs not be a spontaneous one. For instance, when a hard-sphere gas is confined by a spherical container with a perfectly smooth and elastic surface, a momentum flow is also expected to sustain if the total angular momentum

$$\mathbf{L}=\sum_i \mathbf{p}_i\times\mathbf{r}_i\ ,$$

an invariant of this system, is initially nonzero. Unlike the torus system of Fig. 1(a), this flow cannot survive external disturbances. It is unable to restore itself once it is destroyed.

It will be shown later that the second law of thermodynamics holds if and only if SMF does *not* exist. The concept of SMF is an imaginary one since no known physical systems exhibit it. Nevertheless, SMF will prove to be a very helpful theoretical tool. On the one hand, the second law can be formulated equivalently in terms of SMF. On the other hand, the descriptive definition of SMF is readily reformulated in rigorous mechanical terms.

For an exact formulation of SMF, consider the mechanical system composed of N interacting point particles. Let V be a fixed spatial region. The total momentum inside V at a given time is

$$\mathbf{J}_V(\zeta)=\sum_{i=1}^{N}\mathbf{p}_i\chi_V(\mathbf{r}_i)\ , \tag{9}$$

where $\mathbf{r}_i$ and $\mathbf{p}_i=m_i\dot{\mathbf{r}}_i$ are the position and momentum of the ith particle at that time, and $\chi_V(=0,1)$ is the indicator function of region V. Obviously the total momentum $\mathbf{J}_V(\zeta)$ is a function of the state of the system $\zeta\equiv(q_1,\ldots,q_s,p_1,\ldots,p_s)$ with $s=3N$.

We say that there is a spontaneous momentum flow in the system if the long-term average

$$\bar{\mathbf{J}}_V=\lim_{\tau\to\infty}\frac{1}{\tau}\int_0^{\tau}\mathbf{J}_V(T_t\zeta)dt \tag{10}$$

is the same for all initial states ζ from the same energy surface Σ (with a possible exception of the measure zero), and is nonzero for some spatial region V. Here $T_t\zeta$ denotes that state of the system at time t starting from the initial state $\zeta\in\Sigma$.

In this formulation, the nonvanishing long-term average captures the sustaining feature of SMF. The independence of the average upon the initial state implies the robustness of SMF, because the same flow should appear whenever the system attains the same energy. It is not always necessary to strictly require that the long-term average be the same for all initial states from the whole energy surface Σ. Depending on the nature of the external disturbances and the structure of the system, the collection of all permissible initial states in the presence of disturbances may form only a subset of Σ. Although the requirement for the independence all over Σ seems appropriate for the examples considered in this paper, it is understood that a generalized formulation of SMF is possible.

The momentum flux at a single point can be defined as the limit of $\mathbf{j} \equiv \lim \bar{\mathbf{J}}_V / V$ as V shrinks to zero around that point, if the limit exists. For the simple example considered in Sec. II A [Fig. 1(a)], we can estimate that $|\mathbf{j}| \approx nmu$ with u being given by (2).

A requirement on the rate of convergence should also be included in the exact formulation of SMF. Consider the difference defined by

$$D_V(\tau,\zeta) \equiv \left| \frac{1}{\tau} \int_0^\tau \mathbf{J}_V(T_t\zeta)dt - \bar{\mathbf{J}}_V \right| .$$

By (10) we know that $D_V(\tau,\zeta) \to 0$ as $\tau \to \infty$ for each fixed initial state $\zeta \in \Sigma$. But there is no restriction on how the rate of convergence varies with different initial states. Roughly speaking, it is desirable to ensure that a stable momentum flow can be established in a practically finite time for all initial states. It is too strong to require that the convergence be uniform across the energy surface Σ, namely, there exists a function $\varepsilon(\tau)$ such that $\varepsilon(\tau) \to 0$ monotonically as $\tau \to \infty$, and

$$D_V(\tau,\zeta) < \varepsilon(\tau) \tag{11}$$

always holds for all $\zeta \in \Sigma$. For our future consideration, we need only require

$$\lim_{\tau \to \infty} \int_\Sigma D_V(\tau,\zeta) d\mu(\zeta) = 0 , \tag{12}$$

which includes the uniform convergence (11) as a special case. Here μ is the measure on the energy surface and $\mu(\Sigma)$ is finite. We conjecture that our exact formulation of SMF with the weak requirement (12) applies to the models considered in the previous sections. Furthermore, if SMF is physically meaningful, an upper bound for the averaged difference $\int_\Sigma D_V(\tau,\zeta)d\mu(\zeta)$ should exist as a function of the time span τ, and this function should approach zero reasonably fast as $\tau \to \infty$.

B. Formulation of the second law by SMF

From the specific examples in the previous sections, we have already seen that Maxwell's demon can be used to generate SMF. We will show that there is a general relationship between SMF and possible violations of the second law: (i) the existence of a perpetual motion machine of the second kind implies the existence of SMF; and (ii) the existence of SMF implies the existence of a perpetual motion machine. Due to the descriptive nature of the original formulations of the second law, we can only use the descriptive definition of the SMF in the following consideration.

Part (i). Because different formulations of the second law are equivalent, let us stick to the Kelvin-Planck statement and suppose there exists a perpetual motion machine which can extract heat from a single reservoir (say, a water tank) and convert it into work without other effects. To show that statement (i) is true, we need only to show that we can use this machine to generate SMF. First, the work produced by this machine can always be applied to start and maintain a momentum flow, for instance, a circular water current in the tank, with the aid of some stirring apparatus [Fig. 4(a)]. The heat generated by the stirring goes back to the water tank (the heat reservoir). The large system composed of the perpetual motion machine, the water tank, and the stirring apparatus is itself an isolated system where (at least in the tank) a momentum (water) flow will persist. Because the perpetual motion machine is expected to perform robustly, this flow is also robust against external disturbances. Since this flow is both sustaining and robust, it is therefore a SMF.

Part (ii). To show that statement (ii) is true, let us suppose there exists an isolated system exhibiting SMF. We take it for granted that we can always extract work from the kinetic energy of the flow, say, by a small turbine [Fig. 4(b)]. To regain the kinetic energy of the flow, just keep this system in thermal contact with a single heat reservoir. Since the spontaneous flow in the system is expected to persist regardless of the disturbances induced by the turbine and contact with the reservoir, the joint operation of the system and the turbine will convert heat first into momentum flow and then into work without other effects. This violates the Kelvin-Planck statement of the second law.

We have therefore shown that the Kelvin-Planck statement of the second law is equivalent to the statement *in any isolated system, no spontaneous momentum flow exists.* Because SMF can be formulated rigorously in terms of mechanics, we will be able to consider the dynamical conditions required for violating the second law via SMF by any mechanical devices, which include Maxwell's demon as a special case.

C. A nonexistence result

Our theorem about the nonexistence of spontaneous momentum flow is as follows. The point-particle system (1) cannot exhibit SMF if (i) its energy function E is symmetric under momentum reversal, namely,

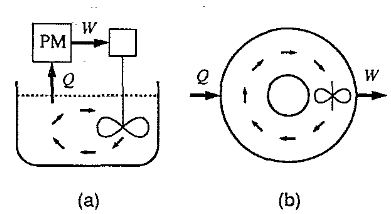

FIG. 4. Schematic diagrams showing that (a) a perpetual mobile (PM) can be used to generate spontaneous flow; and conversely, (b) a system exhibiting spontaneous flow can be used to build a perpetual mobile. See text.

$$E(\tilde{\zeta})=E(\zeta)\ , \tag{13}$$

where $\tilde{\zeta}=(q_1,\ldots,q_s,-p_1,\ldots,-p_s)$ denotes the state obtained by inverting the momenta in state $\zeta=(q_1,\ldots,q_s,p_1,\ldots,p_s)$; and (ii) the phase volume $d\Omega=dq_1\cdots dq_s dp_1\cdots dp_s$ is invariant during time evolution, and the total phase volume is finite for finite energy.

The energy function E need not coincide with a Hamiltonian function of the system. But in the case where the energy function is also a Hamiltonian, Eq. (13) is the condition for the time reversibility of the system [10]. Note that energy function of the general form

$$E(\zeta)=\sum_{i=1}^{s} p_i^2/2m_i+\Phi(q_1,\ldots,q_s)$$

has momentum-reversal symmetry. The corresponding system need not be conservative. For example, velocity-dependent forces, such as the Lorentz force, can also be included as long as they do not alter the energy. In fact, all the demon models considered in Sec. II have an energy function of this form.

From condition (ii), we know that the microcanonical measure μ on the energy surface given by

$$d\mu=d\Sigma/|\nabla E| =d\Sigma\left\{\sum_i\left[\left[\frac{\partial E}{\partial q_i}\right]^2+\left[\frac{\partial E}{\partial p_i}\right]^2\right]\right\}^{-1/2},$$

with $d\Sigma$ being the area element on the energy surface Σ, is finite and invariant during time evolution. Furthermore, it is easily seen that measure μ is symmetric with respect to momentum reversal in the sense that for all measurable $A\subset\Sigma$

$$\mu(\tilde{A})=\mu(A)\ , \tag{14}$$

where $\tilde{A}=\{\tilde{\zeta}:\zeta\in A\}$.

It should be mentioned that the mere existence of a symmetric and invariant measure on the energy surface is sufficient for proving the theorem. No other details about the measure are necessary. Condition (ii) can therefore be generalized as follows.

(ii′) There exists a finite measure μ on the energy surface Σ such that μ is invariant during time evolution and is symmetric with respect to momentum reversal [Eq. (14)].

To prove the theorem, suppose there exists a system which exhibits spontaneous flow while having an invariant, symmetric, and finite measure μ on Σ. According to our rigorous formulation, we have

$$\bar{\mathbf{J}}_V=\lim_{\tau\to\infty}\frac{1}{\tau}\int_0^\tau \mathbf{J}_V(T_t\zeta)dt\neq 0 \tag{15}$$

for some V, and $\bar{\mathbf{J}}_V$ is the same for all initial states $\zeta\in\Sigma$. Here T_t is the group of transformations on the energy surface Σ with $T_aT_b=T_{a+b}$ for all real numbers a and b. Moreover, the difference

$$D_V(\tau,\zeta)=\left|\frac{1}{\tau}\int_0^\tau \mathbf{J}_V(T_t\zeta)dt-\bar{\mathbf{J}}_V\right|$$

submits to

$$\lim_{\tau\to\infty}\int_\Sigma D_V(\tau,\zeta)d\mu(\zeta)=0\ .$$

Thus, given an arbitrary $\varepsilon>0$, we can choose τ large enough such that

$$\int_\Sigma D_V(\tau,\zeta)d\mu(\zeta)<\varepsilon\ . \tag{16}$$

On the other hand,

$$\begin{aligned}\int_\Sigma D_V(\tau,\zeta)d\mu(\zeta) &\geq \left|\int_\Sigma\left[\frac{1}{\tau}\int_0^\tau \mathbf{J}_V(T_t\zeta)dt-\bar{\mathbf{J}}_V\right]d\mu(\zeta)\right| \\ &=\left|\frac{1}{\tau}\int_\Sigma\int_0^\tau \mathbf{J}_V(T_t\zeta)dt\,d\mu(\zeta)-\bar{\mathbf{J}}_V\right|\ ,\end{aligned} \tag{17}$$

where, as usual, $\mu(\Sigma)=1$ has been assumed for simplicity. Changing the order of the integrations, we first consider

$$\begin{aligned}\int_\Sigma \mathbf{J}_V(T_t\zeta)d\mu(\zeta)&=\int_{T_{-t}\Sigma}\mathbf{J}_V(\zeta)d\mu(T_{-t}\zeta)\\ &=\int_\Sigma \mathbf{J}_V(\zeta)d\mu(\zeta)\ .\end{aligned} \tag{18}$$

Here the first equality is generally true, and the second equality holds because T_t preserves the measure μ. It follows from definition (9) that

$$\mathbf{J}_V(\tilde{\zeta})=-\mathbf{J}_V(\zeta)\ . \tag{19}$$

From Eq. (13) we know that $\zeta\in\Sigma$ implies $\tilde{\zeta}\in\Sigma$, and vice versa. The symmetry of the energy surface and symmetry of the measure μ [Eq. (14)] together with Eq. (19) imply that the last integral in Eqs. (18) cancels out and equals zero. Therefore the right-hand side of expression (17) equals $|\bar{\mathbf{J}}_V|$. Combining (17) with (16) yields

$$\varepsilon>\int_\Sigma D_V(\tau,\zeta)d\mu(\zeta)\geq|\bar{\mathbf{J}}_V|\ .$$

The arbitrariness of ε contradicts the assumption $\bar{\mathbf{J}}_V\neq 0$. Hence $\bar{\mathbf{J}}_V\equiv 0$ for all spatial regions V. This completes the proof.

It is important to note that our theorem does not rely on any assumption of the ergodicity of the system. In the following section, we will actually need to apply the theorem to nonergodic systems. If the system is ergodic, however, the theorem still holds, and the proof becomes simpler. One can then replace the time average in (15) by the average on the energy surface and thus directly goes to the last step of (18).

D. Remarks and conclusions

The purpose of introducing spontaneous momentum flow is to transform the original Kelvin-Planck statement concerning a perpetual motion machine of the second kind into an equivalent formulation which involves only an isolated mechanical system. Although the perpetual mobile itself is expected to be an open system which converts heat into work, it can be incorporated into a large system which is isolated and exhibits SMF [cf. Fig. 4(a)]. We have shown that the existence of the perpetual mobile implies the existence of SMF, and vice versa.

Because an exact formulation of SMF can be expressed solely in terms of mechanics, we are able to prove the nonexistence of SMF under the assumption of symmetric energy with respect to momentum reversal. The conclusion is that any point-particle systems with invariant phase volume cannot exhibit SMF. As a consequence, such systems cannot serve as Maxwell's demon (see below). If we want to model Maxwell's demon as mechanical systems with symmetric energy, as in the models in Sec. II, then the noninvariance of phase volume is a necessary requirement. For comparison, a system based on Lorentz force and potential force, as in Fig. 1(b), cannot serve as Maxwell's demon, because phase volume is preserved by its dynamics:

$$m\dot{\mathbf{v}}=q\mathbf{v}\times\mathbf{B}(\mathbf{r})-q\nabla\phi(\mathbf{r}) .$$

To show why it follows from the nonexistence theorem that Maxwell's demon itself must have noninvariance of phase volume, suppose we *can* implement Maxwell's demon as a mechanical system with invariant phase volume. This will lead to contradiction. Obviously we can use the demon to construct a perpetual motion machine of the second kind by adding necessary apparatus which, as usual, are themselves conservative mechanical systems with invariant phase volume. Then, as in Fig. 4(a), we can add in more conservative mechanical parts to the perpetual mobile, like the water tank and stirring apparatus, to get a large system exhibiting SMF. Since the phase volume of each subsystem is preserved, the phase volume of the large composite system must also be preserved. The existence of SMF in this system contradicts our theorem.

IV. DISCUSSION

The theorem about the nonexistence of spontaneous momentum flow stresses the strong connection between invariance of phase volume and the second law of thermodynamics. Not surprisingly, the results are consistent with the conventional approach of equilibrium statistical mechanics, where the preservation of the phase volume by Liouville's theorem is usually considered as the starting point for its basic postulate on equal *a priori* probability.

The aforementioned connection, however, is not always obvious. For example, it has been suggested that some integrable or nearly integrable systems might be used to violate the second law of thermodynamics, because according to the Kolmogorov-Arnold-Moser (KAM) theorem those systems can survive the presence of weak perturbations without going into chaos [11]. Our nonexistence theorem practically rules out this possibility, because the KAM theorem [12] concerns perturbation on Hamiltonian systems, where phase volume is always preserved.

Another aspect of the connection between the invariance of phase volume and the second law is the existence of mechanical models of Maxwell's demon when phase volume is *not* preserved. Although our models of Maxwell's demon do not disobey any laws in classical mechanics and the dynamics of the model in Sec. II B bears formal resemblance to that of the Lorentz gas model with isokinetic thermostatting force [7], they are apparently not constructed as real physical systems which violate the second law. Nonetheless, the models are helpful to illustrate the fact that the validity of the second law relies on the dynamics of the underlying mechanical systems. This validity cannot be justified by the laws of classical mechanics alone. Invariance of phase volume appears as an additional factor which is responsible for this validity. In the hypothetical case where phase volume is not preserved, possible violation of the second law still cannot be ruled out theoretically. Looking from another angle, if the second law is taken as a fundamental assumption, then the invariance of phase volume may be considered as a constraint imposed by the second law on the allowable dynamics of the mechanical systems.

Finally, we will provide a rather intuitive argument for the general relationship between SMF and phase volume contraction. The basic idea is that the states of the system corresponding to SMF, if it exists, are expected to form an attractor in phase space due to the robustness of SMF. For a more formal consideration, suppose there is an isolated system exhibiting spontaneous flow. Let Σ be the energy surface and $F\subset\Sigma$ be the collection of all states that specify an instantaneous spatial distribution of momenta close to that of the long-term average limit. Although the exact meaning of "close" is undefined here, we expect

$$\mu(F)\ll 1$$

because most states on the energy surface Σ *should correspond* to thermodynamic equilibrium rather than to a flow state. Here $\mu(\Sigma)=1$ is assumed. But F is also expected to be an "attractor" on Σ, namely, starting from an arbitrary initial state $\zeta\in\Sigma$, after a long time t there should be a large probability $p(\approx 1)$ for the system to be in a flow state, or $T_t\zeta\in F$. The stability of the probability demands that

$$\lim_{t\to\infty}\mu(G\cap T_{-t}F)/\mu(G)=p$$

for all measurable $G\subseteq\Sigma$ [13]. We need only consider the simple situation $G=\Sigma$ and get

$$\lim_{t\to\infty}\mu(T_{-t}F)=p\approx 1\gg\mu(F) .$$

This requires $\mu(T_{-t}F)\gg\mu(F)$ for large t. Hence the phase volume must be contracting somewhere surrounding F. In other words, we have shown that the existence of SMF implies contraction of phase volume, which conclusion is consistent both with our results in the previous sections and with the results of the computational approach [4].

*Present address: Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92093.

†Present address: Department of Physics, University of California, San Diego, La Jolla, CA 92093.

[1] J. C. Maxwell, *Theory of Heat*, 6th ed. (Appleton, New York, 1880), pp. 328–329.

[2] *Maxwell's Demon: Entropy, Information, Computing*, edited by H. S. Leff and A. F. Rex (Princeton University Press, Princeton, 1990).

[3] E. Fredkin and T. Toffoli, Int. J. Theor. Phys. **21**, 219 (1982).

[4] R. Landauer, Phys. Rev. Lett. **53**, 1205 (1984); Nature (London) **335**, 779 (1988); Phys. Today **44** (5), 23 (1991).

[5] W. H. Zurek, Phys. Rev. Lett. **53**, 391 (1984); Nature (London) **341**, 119 (1989); Phys. Rev. A **40**, 4731 (1989).

[6] C. M. Caves, Phys. Rev. Lett. **64**, 2111 (1990); C. M. Caves, W. G. Unruh and W. H. Zurek, *ibid.* **65**, 1387 (1990).

[7] B. Moran, W. G. Hoover, and S. Bestiale, J. Stat. Phys. **48**, 709 (1987); W. G. Hoover, *ibid.* **42**, 587 (1986); see also W. G. Hoover, *Computational Statistical Mechanics* (Elsevier, Amsterdam, 1991).

[8] When the width w of the field barrier is very small, it is not difficult to estimate the probability $\alpha(\ll 1)$ favoring the particles to pass through the barrier along the preferred direction (cf. Sec. II A). By assuming a uniform Maxwell distribution of velocity at both sides of the barrier and ignoring the collisions between the particles, we can evaluate the probability as $\alpha \approx 1-\exp(-2w)$ with $w \ll 1$.

[9] The approach to infinity is not crucial. An alternative simple choice would be $\mathbf{F}(\mathbf{v})=(\mathbf{v}\times\hat{\mathbf{z}})v\,\mathrm{sgn}(\sin\theta)$, where $\mathrm{sgn}(x)=1$ or -1 when $x>0$ or <0, and $\mathrm{sgn}(0)=0$.

[10] R. L. Devaney, Trans. Am. Math. Soc. **218**, 89 (1976).

[11] J. Ford, in *Directions in Chaos*, edited by Hao Bai-Lin (World Scientific, Singapore, 1987), Vol. 1, pp. 1–16.

[12] J. Moser, *Stable and Random Motions in Dynamical Systems* (Princeton University Press, Princeton, 1973); V. I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd ed. (Springer-Verlag, New York, 1989), pp. 399–415.

[13] K. Zhang, Phys. Rev. A. **41**, 1893 (1990); **42**, 1012 (1990).