Feynman's ratchet and pawl: An exactly solvable model

C. Jarzynski[1,*] and O. Mazonka[2,†]

[1]*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545*
[2]*Institute for Nuclear Studies, Świerk, Poland*
(Received 22 January 1999)

We introduce a simple, discrete model of Feynman's ratchet and pawl, operating between two heat reservoirs. We solve exactly for the steady-state directed motion and heat flows produced, first in the absence and then in the presence of an external load. We show that the model can act both as a heat engine and as a refrigerator. We finally investigate the behavior of the system near equilibrium, and use our model to confirm general predictions based on linear-response theory. [S1063-651X(99)07506-6]

PACS number(s): 05.40.−a

INTRODUCTION

Feynman's *ratchet and pawl* system [1] is a well-known (but not the earliest [2,3]) example of a proposed ''mechanical Maxwell's demon,'' a device whose purpose is to convert into useful work the thermal motions present in a heat reservoir. The idea is beautifully simple: set up a ratchet and pawl so that a wheel is allowed to turn only in one direction, then attach that wheel to a windmill whose vanes are surrounded by a gas at a finite temperature; see Fig. 46-1 of Ref. [1]. Every so often, an accumulation of collisions of the gas molecules against the vanes will cause the wheel to rotate by one notch in the allowed direction, but presumably never in the forbidden direction. Such rectification of thermal noise could be harnessed to perform useful work (such as lifting a flea against gravity), in direct violation of the second law of thermodynamics. Of course, in order for statistical fluctuations to cause rotation at a perceptible rate, the ratchet and pawl must be microscopic, and this points to the resolution of the paradox. If thermal motions of the gas molecules are sufficient to cause the wheel to rotate a notch, then the thermal motion of the pawl itself will occasionally cause it to disengage from the ratchet, at which point the wheel could move in the ''forbidden'' direction. Feynman compared the rates of the two processes—rotation in the allowed and the forbidden directions—and found them to be equal when the system is maintained at a single temperature. Thus no net rotation arises, and the second law is saved.

Since the failure of the ratchet and pawl system to perform work arises from thermal fluctuations of the pawl, a natural solution to the problem is to reduce these fluctuations by externally cooling the pawl to a temperature below that of the gas. In this case the device does indeed operate as designed, but this no longer constitutes a violation of the second law: the ratchet and pawl is now effectively a microscopic heat engine, capitalizing on a temperature difference to extract useful work from thermal motions.

While the ratchet and pawl was introduced in Feynman's *Lectures* primarily for pedagogical purposes, recent years have seen a renewed interest in this system [4–22], largely

due to the fact that analogous mechanisms have been proposed as simple models of motor proteins.

Our purpose in this paper is to introduce an exactly solvable model of Feynman's microscopic heat engine. This model is discrete rather than continuous, but it captures two essential features of the original example: (i) a periodic but asymmetric interaction potential between the ratchet and the pawl (corresponding to the sawtooth shape for the ratchet's teeth), and (ii) two ''modes'' of interaction (corresponding to the pawl being either engaged or disengaged from the ratchet). [Note that a particle (or, more generally, a reaction coordinate), evolving from one sufficienctly deep potential- (or free-) energy minimum to another, behaves much as if hopping from one site to another on a discrete lattice. See, for instance, Fig. 6 of Ref. [15].] These features are sufficient for the model to reproduce the behavior discussed by Feynman. Related discrete models of noise-induced transport have appeared in the literature [16–18]; however, in our model (as in Feynman's example) the transport is explicitly driven by a temperature difference between two reservoirs.

In Sec. I, we will introduce our model—a system in contact with two heat reservoirs—and consider it in the absence of an external load. We will solve for the net rate at which the device produces directed motion, in terms of the reservoir temperatures T_A and T_B [Eq. (14)]. We will also solve for the net rates of heat flow from reservoir A to the system, and from the system to reservoir B; see Fig. 3. For zero external load, these rates, $\dot{Q}_A$ and $\dot{Q}_B$, must be equal, as we will indeed find them to be. We will finally use these results to solve for the entropy production rate.

In Sec. II, we will allow the device to perform work against an external load f. We will again solve exactly for the directed motion and heat flows, in terms of $\mathbf{x} \equiv (f, T_A, T_B)$. In this situation, the heat flows $\dot{Q}_A$ and $\dot{Q}_B$ are not necessarily equal; the difference between them is the *power* $\dot{W}$, which the device delivers against the load f. When $\dot{W} > 0$, the system operates as a heat engine. Conversely, one can imagine that for some parameter values, the system will operate as a refrigerator, creating a net flow of heat from the colder to the hotter reservoir. In Sec. III we will use our analytical results to show that our model indeed exhibits both these behaviors (heat engine and refrigerator). Finally, in Sec. IV we will

---

*Electronic address: chrisj@lanl.gov
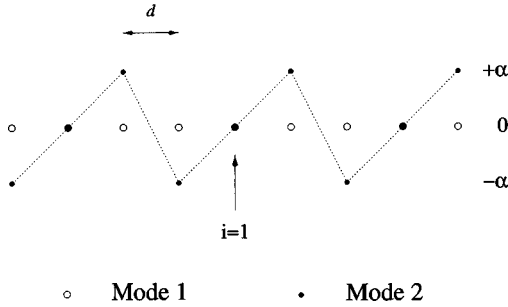†Electronic address: maz@iriss.ipj.gov.pl

FIG. 1. Potential energy $U_i^{(m)}$ is shown for both modes, $m = 1,2$, in the absence of an external load; the lattice spacing is d, and site 1 is labeled explicitly.

consider the near-equilibrium regime of small load and temperature difference. We will show that our model confirms general predictions based on linear-response theory.

I. ZERO EXTERNAL LOAD

Consider a particle that jumps between neighboring sites along a one-dimensional regular lattice, where d is the lattice spacing. We assume that the particle is coupled to a heat reservoir at temperature T_B, and that its jumps are thermal in nature. That is, the probability (per unit time) of making a jump to site $i+1$, starting from site i, is related to the probability rate of the inverse process by the usual detailed balance relation: $P_{i \to i+1}/P_{i+1 \to i} = \exp(-\Delta E/T_B)$, where $\Delta E = U_{i+1}^{(m)} - U_i^{(m)}$ is the instantaneous change in the particle's potential energy, associated with the jump from i to $i+1$. We use the notation $U_i^{(m)}$ to denote the potential energy of the particle at site i; the superscript m denotes the "mode" of the potential, to be explained momentarily. The integer index i runs from $-\infty$ to $+\infty$, and we are using units in which Boltzmann's constant $k_B = 1$.

Next, we assume that the potential energy function $U_i^{(m)}$ has two possible *modes*, $m = 1,2$, and that it changes stochastically between these two. In the first mode, the energy of each site is zero: $U_i^{(1)} = 0$. In the second mode, the energy is periodic: $U_i^{(2)} = \alpha \cdot [(i \bmod 3) - 1]$, where α is a positive constant with units of energy. As shown in Fig. 1, the second mode is a discrete version of an asymmetric sawtooth potential. We assume that the stochastic process governing the changes between modes is also a thermal process, driven by a heat reservoir at temperature T_A. Thus, the probability rate of a change to mode 2, starting from mode 1, relative to the probability rate of the reverse mode change, is given by the detailed balance factor $\exp(-\Delta E/T_A)$, where $\Delta E = U_i^{(2)} - U_i^{(1)}$ is now the (site-dependent) change in particle energy due to an instantaneous change from mode 1 to mode 2.

We now describe more precisely these two stochastic processes, governing the jumps of the particle, and the change between modes. We assume the processes are independent, and each is a Poisson process occurring at a rate Γ. That is, during every infinitesimal time interval δt, there is a probability $\Gamma \delta t$ that the particle will attempt a jump to a neighboring site. An "attempt" looks as follows. First, the particle decides (randomly, with equal probability) whether to try jumping to the left $(-d)$ or to the right $(+d)$. Then the Metropolis algorithm [23] is used to satisfy detailed balance:

if the value of ΔE associated with the jump is zero or negative, then the jump takes place; if $\Delta E > 0$, then the jump occurs with probability $\exp(-\Delta E/T_B)$. Similarly, during every infinitesimal time interval δt, there is a probability $\Gamma \delta t$ that the mode will attempt to change, and the attempt is accepted or rejected according to the Metropolis algorithm (at temperature T_A).

We have introduced three parameters, which we will view as being "internal" to the system: d, α, and Γ; these essentially set the relevant length (d), energy (α), and time (Γ^{-1}) scales. The two remaining parameters T_A and T_B, we will view as "external."

The analogy between our model and Feynman's ratchet and pawl runs as follows. First, the position of our particle corresponds to an angle variable θ specifying the orientation of the ratchet wheel. When the particle accomplishes a net displacement of three lattice sites, to the right or to the left, this is equivalent to the ratchet being displaced by one notch, or tooth; see Fig. 1. (Thus, the distance $3d$ in our model translates to an angular interval $\Delta \theta = 2\pi/N_{teeth}$, where N_{teeth} is the number of teeth along the perimeter of the ratchet wheel.) Since we want to keep track of the wheel over long intervals of time, possibly including many full rotations, θ varies from $-\infty$ to $+\infty$, rather than being a periodic variable from 0 to 2π.

When a ratchet and pawl are "engaged"—that is, when the pawl actually presses against the teeth of the ratchet—then there exists an interaction energy, arising from the compression of the spring that holds the pawl against the ratchet, which has the form of a periodic sawtooth potential in the variable θ. In our model, the analogue of this interaction energy is the discrete sawtooth potential $U_i^{(2)}$; mode 2 thus corresponds to the situation in which the ratchet presses against the pawl. By contrast, mode 1 corresponds to the ratchet and pawl begin "disengaged," as will occur every so often as a result of thermal fluctuations of the pawl. Of course, in Feynman's system, the potential energy of the disengaged mode is always greater than that of the engaged mode (due to the spring compression needed to actually place the pawl out of reach of the ratchet's teeth), whereas in our model this is not the case: $U_i^{(1)} = 0$. This, however, does not change the problem in any qualitatively significant way.

As mentioned, the motion of the particle from site to site is a thermal process occurring at temperature T_B, while the stochastic "flashing" between modes occurs at temperature T_A. Thus, in the context of the analogy with the physical ratchet and pawl, T_B denotes the temperature of the gas surrounding the panes connected to the ratchet wheel, and T_A is the temperature at which the pawl is maintained.

To analyze the model, we first note that it maps nicely onto the problem of a spin-1/2 particle A coupled to a spin-1 particle B with the following energy function:

$$E(S_A, S_B) = \alpha \left(S_A + \frac{1}{2} \right) S_B, \tag{1}$$

where $S_A = \pm 1/2$ and $S_B = 0, \pm 1$. Here, spin A represents the mode (pawl), and spin B represents the particle (ratchet). When spin A is "down" $(S_A = -1/2)$, the energy of the system is independent of the state of spin B, as with mode 1; when spin A is "up," the energy is $-\alpha$, 0, or $+\alpha$, depend-
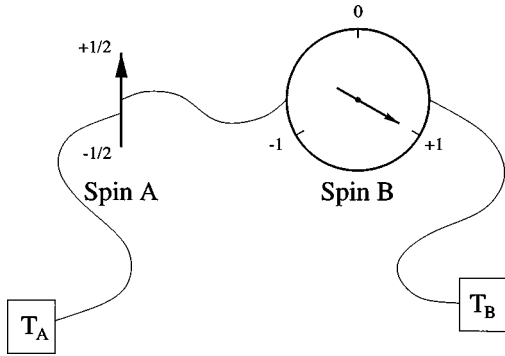
FIG. 2. Our system maps onto that of a spin-1/2 particle (A) coupled to a spin-1 particle (B). The former is depicted in the usual manner (as up or down), the latter by an arrow, which can point in one of three directions on the face of a clock. The thin lines denote the coupling between the two spins, as well as the coupling of each spin to a heat reservoir.

ing on the state of spin B, as with mode 2. Thus, changes in the state of spin A correspond to changes in the mode, whereas changes in the state of spin B correspond to the particle making a jump. If S_B changes from -1 to 0, or from 0 to $+1$, or from $+1$ to -1, then this amounts to the particle jumping to the right; the reverse processes (0 to -1, etc.) correspond to jumps to the left. We now couple spin A to a reservoir at temperature T_A, spin B to a reservoir at temperature T_B, with the stochastic dynamics as outlined above (two independent Poisson processes), and solve for the net drift of the particle.

This system can be visualized as shown in Fig. 2. Spin A flips between up and down; spin B performs sudden ''rotations'' by $\pm 120°$. A clockwise rotation corresponds to a jump to the right ($+d$) by the particle; counterclockwise ones translate into jumps to the left.

Our system has six possible states, listed in Table I. The dynamics of the particle is described by a Markov process, for which we can write a set of rate equations. Let $p_n(t)$ denote the probability that the system is found in state n at time t. Then

$$\frac{dp_n}{dt} = \Gamma \sum_{n' \neq n} [p_{n'}P(n' \to n) - p_n P(n \to n')], \quad n = 1, \ldots, 6.$$
(2)

Here, $\Gamma P(n \to n')$ is the probability rate at which the system, when its state is n, makes a transition to state n'. If the transition $n \to n'$ involves only a flip of spin A, then $P(n \to n')$ is the probability that such an attempt will be accepted, under the Metropolis rule. If the transition involves

TABLE I. The six states of the system, and their energies.

State (n)	S_A	S_B	$E(S_A, S_B)$
1	$-1/2$	-1	0
2	$-1/2$	0	0
3	$-1/2$	$+1$	0
4	$+1/2$	-1	$-\alpha$
5	$+1/2$	0	0
6	$+1/2$	$+1$	$+\alpha$

just a rotation of spin B (by $\pm 120°$), then $P(n \to n')$ is the probability of generating this move, given an attempt to change S_B. If the transition involves changes in the states of both spins, then $P(n \to n') = 0$. Examples of these rules are

$$P(1 \to 4) = 1, \quad P(1 \to 2) = \frac{1}{2}, \quad P(1 \to 5) = 0,$$
(3)

$$P(3 \to 6) = e^{-\alpha/T_A}, \quad P(4 \to 5) = \frac{1}{2} e^{-\alpha/T_B}, \quad P(4 \to 2) = 0.$$

The factors of one half come from the fact that, when the system attempts to rotate spin B, there are two possible states for which it can aim.

Our rate equations can be expressed as: $d\mathbf{p}/dt = \Gamma \mathcal{R} \mathbf{p}$, where $\mathbf{p} = (p_1, \ldots, p_6)^T$,

$$\mathcal{R} = \begin{pmatrix} -2 & \frac{1}{2} & \frac{1}{2} & \mu & 0 & 0 \\ \frac{1}{2} & -2 & \frac{1}{2} & 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & -1-\mu & 0 & 0 & 1 \\ 1 & 0 & 0 & -\mu - \frac{\nu + \nu^2}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 & \frac{\nu}{2} & -\frac{3+\nu}{2} & \frac{1}{2} \\ 0 & 0 & \mu & \frac{\nu^2}{2} & \frac{\nu}{2} & -2 \end{pmatrix},$$
(4)

and we have introduced the constants

$$\mu = e^{-\alpha/T_A} \quad \text{and} \quad \nu = e^{-\alpha/T_B}.$$
(5)

Note that μ and ν are monotonically increasing functions of T_A and T_B, and the temperature range $0 < T_A(T_B) < \infty$ translates to $0 < \mu$ (ν) < 1. We thus think of μ and ν as ''rescaled'' temperatures.

The long-time behavior of our system is governed by the steady-state distribution of probabilities $\bar{\mathbf{p}}$ which is the null eigenvector of $\mathcal{R}$ (i.e., $\mathcal{R}\bar{\mathbf{p}} = 0$). Determining $\bar{\mathbf{p}}$ is an exercise in Jordan elimination, and leads to the following result: $\bar{\mathbf{p}} = \mathbf{x}/N$, where

$$x_1 = 52\mu + 28\mu^2 + 12\nu + 19\nu^2 + 5\nu^3 + 21\mu\nu + 2\mu\nu^2 + 8\mu^2\nu,$$
(6)

$$x_2 = 36\mu + 16\mu^2 + 28\nu + 27\nu^2 + 5\nu^3 + 25\mu\nu + 8\mu\nu^2 + 2\mu^2\nu,$$
(7)

$$x_3 = 44\mu + 19\mu\nu + 20\nu + 49\nu^2 + 15\nu^3,$$
(8)

$$x_4 = 64 + 20\nu + 48\mu + 15\mu\nu,$$
(9)

$$x_5 = 24\mu + 40\nu + 20\nu^2 + 18\mu^2 + 30\mu\nu + 15\mu\nu^2,$$
(10)

$$x_6 = 22\mu^2 + 16\mu\nu + 44\mu\nu^2 + 14\mu^2\nu + 15\mu\nu^3 + 26\nu^2 + 10\nu^3,$$
(11)

and $N(\mu,\nu) = \Sigma_{i=1}^{6} x_i$ is a normalization factor.

When both temperatures go to zero, $\mu,\nu \to 0$, we get $\bar{\mathbf{p}}^T = (0,0,0,1,0,0)$. This makes sense: in that limit, the system freezes to the state of lowest energy, $n=4$.

We now address the question of net drift. We first define a net current, $J \equiv J_+ - J_-$, where J_+ is the rate at which spin B is observed to change from state 0 to state $+1$, and J_- is the rate of the reverse transitions, from $+1$ to 0. This can be interpreted by imagining an observer placed at two o'clock on the clock face depicting spin B in Fig. 2; J_+ (J_-) is then the rate at which the hand of the clock passes that observer in the clockwise (counterclockwise) direction; by "rate," we mean number of passes per unit time, averaged over an infinitely long interval of time. (Of course, we could just as well have placed our observer at six or ten o'clock; in the steady state, the current measured will be independent of where the observer is placed.) The current J represents the *net* average rate of clockwise revolutions of spin B. Since each revolution corresponds to 3 steps of the particle to the right, this translates into a particle drift:

$$v = 3\,dJ, \qquad (12)$$

where v denotes the (steady-state) average velocity of the particle.

Explicit expression for the quantities $J_\pm$ are given by

$$J_+ = \Gamma(\bar{p}_2 \mathcal{R}_{32} + \bar{p}_5 \mathcal{R}_{65}), \qquad (13a)$$

$$J_- = \Gamma(\bar{p}_3 \mathcal{R}_{23} + \bar{p}_6 \mathcal{R}_{56}). \qquad (13b)$$

(Since $\Gamma \mathcal{R}_{mn}$ is the transition rate to state m, *given* that the system is found in state n, $\bar{p}_n \Gamma \mathcal{R}_{mn}$ is the net rate at which transitions from n to m are observed to occur in the steady state.) Using our results for $\bar{\mathbf{p}}$, we get, after some algebra:

$$v(T_A, T_B) = -3\,d\frac{\Gamma}{N}(\mu - \nu)(1-\nu)(3\mu + 4). \qquad (14)$$

There are a number of things to note about this result. First, it implies that if $T_A > T_B$ (i.e., $\mu > \nu$), then there is a net flow of the particle to the left; if $T_B > T_A$, the particle drifts to the right. If $T_A = T_B$, then there is no drift, in agreement with Feynman's analysis (as well as the second law).

Next, notice that $v \to 0$ as $T_B \to \infty$ ($\nu \to 1$). In that limit, the change in energy arising from a jump to the left or to the right becomes negligible in comparison to the temperature of the reservoir, which drives those jumps; thus, from any lattice site, the particle is as likely to jump to the left as to the right, resulting in no net drift.

Finally, in the limit $T_A \to \infty$ ($\mu \to 1$), we get:

$$v = -21\,d\frac{\Gamma}{N}(1-\nu)^2, \quad T_A \to \infty. \qquad (15)$$

This is the limit in which changes between the modes occur independently of location of the particle: every attempt to change the mode is accepted. This is analogous to the situation studied by Astumian and Bier [5], where the "flashing" between the two modes of the potential is a Poisson process independent of the particle position.
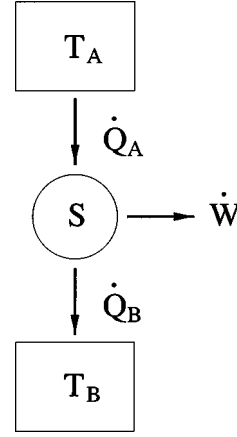

FIG. 3. Schematic representation of our system (S) in contact with two heat reservoirs (at temperatures T_A and T_B). As implied by the arrows, we define $\dot{Q}_A$ to be the net flow of heat from reservoir A to the system, and $\dot{Q}_B$ to be the heat flow from the system to reservoir B. Therefore, $\dot{W} = \dot{Q}_A - \dot{Q}_B$ is the power delivered as work against an external load. (In Sec. I, there is no such load; hence, $\dot{Q}_A = \dot{Q}_B$.)

We can also compute the average rates at which heat is transferred between the two reservoirs and our system. Whenever the system makes a transition from state 1 to state 4, or from state 6 to state 3, its energy drops by α; this energy is released into the reservoir at temperature T_A. Conversely, during the transitions $4 \to 1$ and $3 \to 6$, the system absorbs energy α from reservoir A. The net rate at which the system absorbs energy from reservoir A is then

$$\dot{Q}_A = \alpha\Gamma(\bar{p}_4 \mathcal{R}_{14} + \bar{p}_3 \mathcal{R}_{63} - \bar{p}_1 \mathcal{R}_{41} - \bar{p}_6 \mathcal{R}_{36}). \qquad (16)$$

We can similarly write down an expression for the rate at which heat flows from our system to reservoir B:

$$\dot{Q}_B = \alpha\Gamma(\bar{p}_5 \mathcal{R}_{45} + \bar{p}_6 \mathcal{R}_{56} + 2\bar{p}_6 \mathcal{R}_{46} - \bar{p}_4 \mathcal{R}_{54} - \bar{p}_5 \mathcal{R}_{65}$$
$$- 2\bar{p}_4 \mathcal{R}_{64}). \qquad (17)$$

Figure 3 illustrates the sign convention which we choose in defining $\dot{Q}_A$ and $\dot{Q}_B$. Plugging in the values for the components of $\bar{\mathbf{p}}$ and $\mathcal{R}$, we find that $\dot{Q}_A = \dot{Q}_B$, as we could have predicted, since in the steady state there is no net absorption of heat by the system, nor is any of the heat delivered as work against an external load. Thus, the particle drift is driven by a net passage of heat from A to B, by way of the system. The explicit expression for this heat flow is

$$\dot{Q}_{A \to B} = \dot{Q}_A = \dot{Q}_B = 3\frac{\alpha\Gamma}{N}(\mu - \nu)P(\mu,\nu), \qquad (18)$$

where $P(\mu,\nu) = 4 + 14\mu + 15\nu + 4\mu\nu + 5\nu^2 > 0$. The ratio $v/\dot{Q}_{A \to B}$ then gives us the average displacement of the particle, per unit of heat passed (via the system) from reservoir A to reservoir B:

$$\lim_{\tau \to \infty} \frac{\Delta x}{\Delta Q} = \frac{v}{\dot{Q}_{A \to B}} = -\frac{d}{\alpha}\frac{(1-\nu)(3\mu + 4)}{P(\mu,\nu)} < 0. \qquad (19)$$
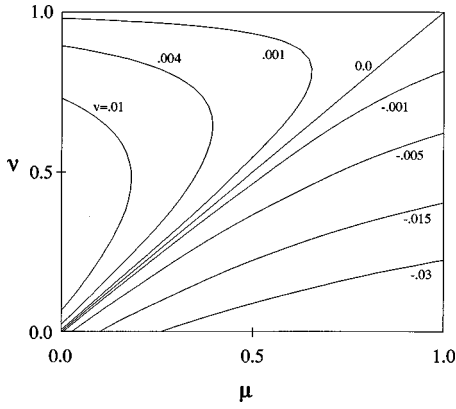
FIG. 4. Contour plot of the function $v(\mu,\nu)$, where $\mu = \exp(-\alpha/T_A)$ and $\nu = \exp(-\alpha/T_B)$ are the rescaled temperatures of the two reservoirs.

Here, Δx and ΔQ are the net particle displacement, and the net heat transferred from A to B, over a time interval τ. The factor $(\mu - \nu)$ in Eq. (18) guarantees that the direction of the heat flow is from the hotter to the cooler reservoir.

We can also compute the rate at which entropy is produced during this process. The rate of entropy production associated with the flow of heat from reservoir A to the system is $\dot{S}_A = -\dot{Q}_A/T_A$; and for reservoir B, $\dot{S}_B = \dot{Q}_B/T_B$. The net entropy production rate is thus

$$\dot{S} = \dot{S}_A + \dot{S}_B = \frac{T_A - T_B}{T_A T_B}\dot{Q}_{A\to B} \qquad (20)$$

$$= 3\frac{\Gamma}{N}\left(\ln\frac{\mu}{\nu}\right)(\mu - \nu)P(\mu,\nu) \geqslant 0. \qquad (21)$$

Figures 4 and 5 illustrate the results obtained in this section. Figure 4 is a contour plot of the drift v, as a function of the rescaled reservoir temperatures μ and ν. The contour $v = 0$ runs along the diagonal, $\mu = \nu$, as well as along ν
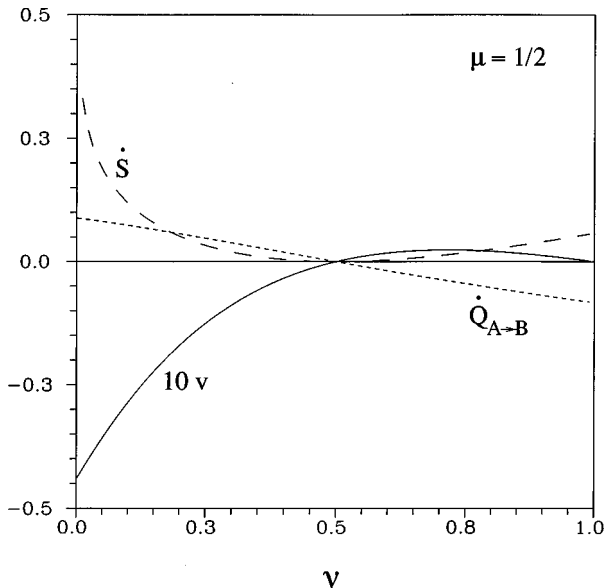

FIG. 5. Drift v (multiplied by 10), heat flow $\dot{Q}_{A\to B}$, and rate of entropy production $\dot{S}$, plotted as functions of ν, for fixed $\mu = 1/2$.
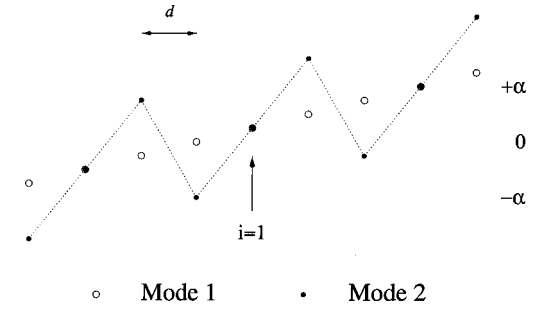

FIG. 6. Discrete, two-mode potential energy function $U_i^{(m)}$ is illustrated for the situation in which there is a nonzero external load. Here we have chosen $f = \alpha/4d$.

$= 1$ $(T_B \to \infty)$. The appearance of positive contours ($v > 0$) to the left of the diagonal, and negative ones to the right, illustrates the point that the drift is rightward when $T_B > T_A$ and leftward when $T_A > T_B$.

In Fig. 5 we have fixed the value of T_A by setting $\mu = 1/2$, and have plotted v, $\dot{Q}_{A\to B}$, and $\dot{S}$ as functions of ν. All three quantities hit zero at $\nu = 1/2$, where $T_A = T_B$: nothing interesting happens when the system is maintained at a single temperature. Note also that v and $\dot{Q}_{A\to B}$ are opposite in sign [in agreement with Eq. (19)], while $\dot{S}$ is always nonnegative (in agreement with the second law). Finally, note that $v \to 0$ as $\nu \to 1$ $(T_B \to \infty)$.

In plotting these two figures, we set all the internal parameters to unity: $\alpha = d = \Gamma = 1$.

II. NONZERO EXTERNAL LOAD

In this section we add an external load to our model. In Feynman's example, this load is a flea, attached by a thread to the ratchet wheel: when the wheel rotates in the appropriate direction, the creature is lifted against gravity. In our model, we add a slope to the discrete potential:

$$U_i^{(m)} \to U_i^{(m)} + ifd, \qquad (22)$$

where f is a real constant (and i is the lattice site); see Fig. 6. Effectively, f is a constant external force that pulls the particle leftward if $f > 0$, rightward if $f < 0$.

The presence of an external load allows the system to perform *work*. If, in the steady state achieved for fixed $\mathbf{x} = (f, T_A, T_B)$, the particle experiences a drift $v(\mathbf{x})$, then the *power* delivered against the external load is

$$\dot{W}(\mathbf{x}) = fv(\mathbf{x}). \qquad (23)$$

By conservation of energy, this must be balanced by heat lost by the reservoirs:

$$\dot{W} = \dot{Q}_A - \dot{Q}_B. \qquad (24)$$

Our approach to solving for the steady-state behavior is the same as in Sec. I, except that the presence of the term ifd in the potential changes the elements of $\mathcal{R}$, and, therefore, the steady-state probabilities $\bar{\mathbf{p}}$.

Because the acceptance probability of an attempted move in the Metropolis scheme has the form Prob $= \min\{1, \exp -\Delta E/T\}$, each matrix element $\mathcal{R}_{ij}$ is piecewise analytic in f. A little thought reveals that the f axis can be divided into four ranges of values, over each of which the elements of $\mathcal{R}$ can be written as analytic functions of α, d, f, T_A, and T_B. These ranges are

$$-\infty < f < -\alpha/d, \tag{25a}$$

$$-\alpha/d < f < 0, \tag{25b}$$

$$0 < f < 2\alpha/d, \tag{25c}$$

$$2\alpha/d < f < +\infty. \tag{25d}$$

In the two extreme ranges, Eqs. (25a) and (25d), stretching to $-\infty$ and $+\infty$, the slope is so steep that the potential energy function no longer has a sawtooth shape in mode 2. We will ignore these ranges and focus instead on the ones for which $U_i^{(2)}$ *does* look like a discrete sawtooth.

For range (25b), i.e., $-\alpha/d < f < 0$, the potential slopes downward with increasing i, and an explicit expression for $\mathcal{R}$ is

$$\mathcal{R}^b = \begin{vmatrix} -\dfrac{3}{2} - \dfrac{1}{2\sigma} & \dfrac{1}{2\sigma} & \dfrac{1}{2} & \mu & 0 & 0 \\[2mm] \dfrac{1}{2} & -\dfrac{3}{2} - \dfrac{1}{2\sigma} & \dfrac{1}{2\sigma} & 0 & 1 & 0 \\[2mm] \dfrac{1}{2\sigma} & \dfrac{1}{2} & -\dfrac{1}{2} - \mu - \dfrac{1}{2\sigma} & 0 & 0 & 1 \\[2mm] 1 & 0 & 0 & -\mu - \dfrac{\nu^2}{2\sigma} - \dfrac{\nu\sigma}{2} & \dfrac{1}{2} & \dfrac{1}{2} \\[2mm] 0 & 1 & 0 & \dfrac{\nu\sigma}{2} & -\dfrac{3}{2} - \dfrac{\nu\sigma}{2} & \dfrac{1}{2} \\[2mm] 0 & 0 & \mu & \dfrac{\nu^2}{2\sigma} & \dfrac{\nu\sigma}{2} & -2 \end{vmatrix}, \tag{26}$$

where

$$\sigma = e^{-fd/T_B}. \tag{27}$$

For range (25c), $0 \leq f < 2\alpha/d$, the potential slopes upward with i, and we have

$$\mathcal{R}^c = \begin{vmatrix} -\dfrac{3}{2} - \dfrac{\sigma}{2} & \dfrac{1}{2} & \dfrac{\sigma}{2} & \mu & 0 & 0 \\[2mm] \dfrac{\sigma}{2} & -\dfrac{3}{2} - \dfrac{\sigma}{2} & \dfrac{1}{2} & 0 & 1 & 0 \\[2mm] \dfrac{1}{2} & \dfrac{\sigma}{2} & -\dfrac{1}{2} - \mu - \dfrac{\sigma}{2} & 0 & 0 & 1 \\[2mm] 1 & 0 & 0 & -\mu - \dfrac{\nu^2}{2\sigma} - \dfrac{\nu\sigma}{2} & \dfrac{1}{2} & \dfrac{1}{2} \\[2mm] 0 & 1 & 0 & \dfrac{\nu\sigma}{2} & -\dfrac{3}{2} - \dfrac{\nu\sigma}{2} & \dfrac{1}{2} \\[2mm] 0 & 0 & \mu & \dfrac{\nu^2}{2\sigma} & \dfrac{\nu\sigma}{2} & -2 \end{vmatrix}. \tag{28}$$

Note that both $\mathcal{R}^b$ and $\mathcal{R}^c$ reduce to the matrix $\mathcal{R}$ of Sec. I, when $\sigma = 1$, i.e., $f = 0$.

As in Sec. I, the first order of business is to solve for $\bar{\mathbf{p}}$, using Jordan elimination, only now the process is considerably more tedious: terms do not cancel as nicely as when $f = 0$. The final results for the steady-state probability vectors $\bar{\mathbf{p}}^b$ and $\bar{\mathbf{p}}^c$ (corresponding to the two ranges of f values) are of the form

$$\bar{p}_n^j = \frac{P_n^j(\mu, \nu, \sigma)}{N^j(\mu, \nu, \sigma)}, \quad j = b, c \quad n = 1, \ldots, 6 \quad (29)$$

where the P_n^j's are finite polynomials in the variables μ, ν, and σ, and $N^j = \Sigma_{n=1}^6 P_n^j$ is a normalization factor. Explicit expression for the polynomials P_n^j are given in the Appendix.

We can now obtain $v(\mathbf{x})$, $\dot{Q}_A(\mathbf{x})$, and $\dot{Q}_B(\mathbf{x})$ from $\bar{\mathbf{p}}$, as in Sec. I. The results for the two ranges of f values, $j = b, c$, are

$$v^j(\mathbf{x}) = \frac{3}{2} d\, \Gamma \frac{X^j}{\sigma N^j}, \quad (30a)$$

$$\dot{Q}_A^j(\mathbf{x}) = \alpha \Gamma \frac{Y^j}{N^j}, \quad (30b)$$

$$\dot{Q}_B^j(\mathbf{x}) = \alpha \Gamma \frac{Y^j}{N^j} - \frac{3}{2} f d\, \Gamma \frac{X^j}{\sigma N^j}, \quad (30c)$$

where $X^j(\mu, \nu, \sigma)$ and $Y^j(\mu, \nu, \sigma)$ are polynomials for which explicit expressions are presented in the Appendix.

Note that the steady-state behavior described by Eq. (30) clearly satisfies energy conservation [see Eqs. (23) and (24)]:

$$\dot{Q}_A - \dot{Q}_B = fv. \quad (31)$$

Equation (30) is the central result of this paper (and reduces to the results of Sec. I when $f = 0$). In Sec. III, we use Eq. (30) to show that our model can act both as a heat engine and as a refrigerator. In Sec. IV, we consider the behavior of our system near equilibrium, and we use Eq. (30) to confirm predictions based on a general, linear-response analysis.

III. SYSTEM AS HEAT ENGINE AND REFRIGERATOR

We can anticipate two different scenarios in which our system acts as a "useful" device: (i) $\dot{W} > 0$. In this case, the system is a *heat engine*, causing the particle to drift up the potential energy slope, with efficiency $\eta_{\text{eng}} = \dot{W}/\dot{Q}_>$, where $\dot{Q}_>$ is the rate of heat flow out of the hotter reservoir. (ii) $\dot{W} < 0$ and $\dot{Q}_< > 0$, where $\dot{Q}_<$ is the rate of heat flow out of the colder reservoir. Here the system is a *refrigerator*, with efficiency $\eta_{\text{ref}} = \dot{Q}_< / |\dot{W}|$. The particle drifts down the potential slope, and the resulting energy liberated allows for a net transfer of heat from the colder to the hotter reservoir, without violating the second law.

We will now use the results derived in Sec. II to show that our simple model indeed exhibits both these behaviors.

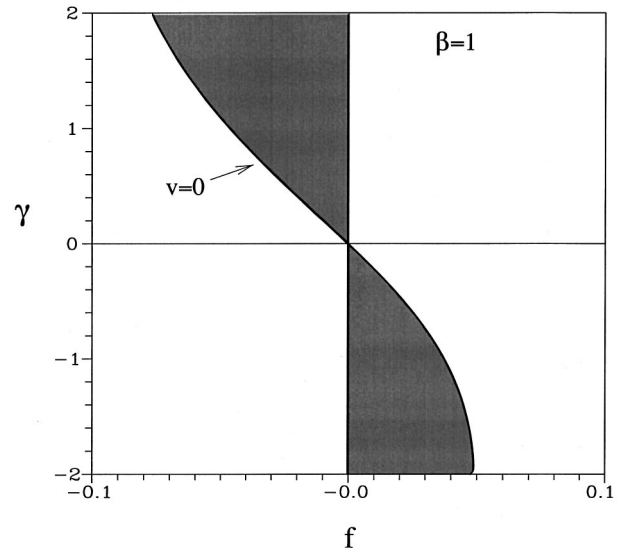The system is a heat engine when $v(f, T_A, T_B)$ and f are

FIG. 7. Contour $v(f, \gamma) = 0$ is shown, for fixed average inverse temperature $\beta = 1$. The shaded regions are those for which $fv > 0$, i.e., for which the system behaves as a heat engine.

of the same sign. Let us introduce the variables,

$$\beta = \frac{T_A + T_B}{2 T_A T_B}, \quad (32)$$

$$\gamma = \frac{T_A - T_B}{T_A T_B}, \quad (33)$$

and consider the behavior of our system in (f, γ) space, for a fixed value of β. (Note that β is the average inverse temperature of the reservoirs, and γ is the difference between inverse temperatures.) Because $\dot{Q}_A$, $\dot{Q}_B$, and v are mutually dependent [Eq. (31)], we can generically explore a measurable fraction of the space of steady-state behaviors by varying only two independent parameters, f and γ, while holding the third, β, fixed.

In Fig. 7 we plot the contour $v(f, \gamma) = 0$, having set $\beta = 1$, and $\alpha = d = \Gamma = 1$. (The range of γ values for this choice of β is $-2 < \gamma < 2$.) To the left of this contour, we have $v > 0$; to the right, $v < 0$. The shaded region thus represents the values of (f, γ) for which v and f are of the same sign, i.e., where the system behaves as a heat engine.

We can understand the placement of the shaded region as follows. Consider a point P on the positive γ axis: $f = 0$, $\gamma > 0$ (i.e., zero external load, $T_A > T_B$). From Sec. I, we know that the particle then drifts leftward, $v < 0$, although no work is performed, since $f = 0$. Let us now imagine tilting the potential slightly "downward" ($f < 0$). For a small enough tilt, we expect that the particle will continue to drift to the left, but now this drift is uphill, and, therefore, work is done against the external load: $\dot{W} > 0$. We conclude that points immediately to the left of the positive vertical axis will correspond to external parameters for which the system behaves as a heat engine, and Fig. 7 confirms this. (Similar reasoning applies for the negative γ axis, where the region $\dot{W} > 0$ appears to the right.) If we now continue to tilt the slope more and more downward (f increasingly negative), at fixed γ
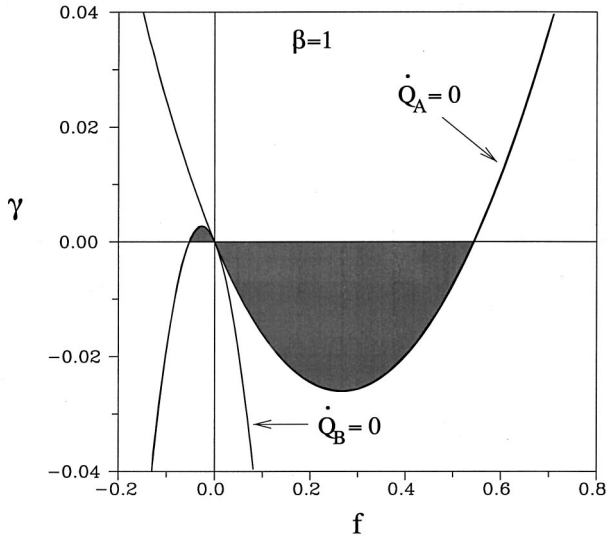
FIG. 8. The contours $\dot{Q}_A(f,\gamma)=0$ and $\dot{Q}_B(f,\gamma)=0$ are shown, for fixed $\beta=1$. In the shaded regions, there is a net flow of heat out of the colder reservoir; the system then acts as a refrigerator.

>0, we expect the leftward drift to become progressively slower, until for some tilt we get $v=0$. At this point the leftward ''thermal force'' exerted on the particle due to the temperature difference between the reservoirs, exactly balances the external load. This occurs at the boundary of the shaded region (the contour $v=0$): for more negative slopes, the particle slides down the slope, and the system no longer acts as a heat engine.

Our system is a refrigerator when $\dot{Q}_<>0$, where $\dot{Q}_<$ is the rate at which heat leaves the colder of the two reservoirs. In Fig. 8 we plot the contours $\dot{Q}_A(f,\gamma)=0$ and $\dot{Q}_B(f,\gamma)=0$, again for $\beta=1$. These two contours are tangent at the origin, and divide the plane of (f,γ) values as follows: $\dot{Q}_A$ is positive for points lying above the contour, and negative below. Similarly, $\dot{Q}_B>0$ ($\dot{Q}_B<0$) for points lying above (below) the contour $\dot{Q}_B=0$. Now, above the horizontal axis ($\gamma>0$) we have $T_A>T_B$, hence $\dot{Q}_<=-\dot{Q}_B$. The small shaded region in the second quadrant therefore represents values of (f,γ) for which reservoir B is the colder of the two reservoirs, *and* it is losing heat ($\dot{Q}_B<0$). Hence in this region our system acts as a refrigerator. Below the horizontal axis, $T_B>T_A$ and thus $\dot{Q}_<=\dot{Q}_A$. The larger shaded region in the fourth quadrant thus also represents refrigeration, only now reservoir A is being drained of heat.

We can understand the general shape of the shaded regions by assuming that, when the temperatures are equal ($\gamma=0$) and the slope is very small, the quantities v, $\dot{Q}_A$, and $\dot{Q}_B$ are linear functions of the slope:

$$v(f,0)=c_v f+O(f^2), \tag{34a}$$

$$\dot{Q}_A(f,0)=c_A f+O(f^2), \tag{34b}$$

$$\dot{Q}_B(f,0)=c_B f+O(f^2), \tag{34c}$$

with c_v, c_A, and $c_B\neq 0$. (Then $c_v<0$, since the particle cannot slide *up* the slope when the reservoir temperatures are

equal.) Energy conservation implies that the difference between $\dot{Q}_A$ and $\dot{Q}_B$ must be quadratic in f (since $\dot{Q}_A-\dot{Q}_B=\dot{W}=fv=c_v f^2$), hence $c_A=c_B$. Thus, for equal temperatures and a sufficiently small slope, one of the reservoirs will be losing heat and the other will be gaining it. If we now slightly lower the temperature of the reservoir that is losing heat, we have a refrigerator: heat flows out of the colder reservoir. In our model, we have $c\equiv c_A=c_B>0$, hence for points on the $\gamma=0$ axis immediately to the right of the origin, heat flows out of reservoir A and into reservoir B; immediately to the left of the origin the reverse holds true. This explains why, just to the right (left) of the origin, the shaded region corresponding to refrigeration hugs the horizontal axis from below (above). If we continue to the right along the line $\gamma=0$, increasing the value of f, the particle will drift ever more rapidly to the left as the slope becomes ever more steeply inclined. The potential energy lost as the particle slides down the incline gets dissipated into the reservoirs; for sufficiently large f, the rate of dissipation is great enough that both reservoirs become heated: $\dot{Q}_A<0$, $\dot{Q}_B>0$. This happens to the right of the point at which the contour $\dot{Q}_A=0$ crosses the horizontal axis with a positive slope; for values of f beyond this point the system can no longer operate as a refrigerator.

We can also understand why the two contours $\dot{Q}_A=0$ and $\dot{Q}_B=0$ ''kiss'' at the origin. The result $c_A=c_B$ means that $\dot{Q}_A$ and $\dot{Q}_B$ are (to leading order) equal along the horizontal axis $\gamma=0$, near the origin. However, they are also (exactly) equal along the vertical axis, since $\dot{W}=0$ when $f=0$. Thus, $\dot{Q}_A$ and $\dot{Q}_B$ are equal, to leading order in f and γ, for a small region around the origin: $\dot{Q}_A=\dot{Q}_B=cf+b\gamma$. This implies that their contours are both tangent to the line $\gamma=-cf/b$ at the origin.

Since we have expressions for $v(\mathbf{x})$, $\dot{Q}_A(\mathbf{x})$, and $\dot{Q}_B(\mathbf{x})$, we can compute the thermal efficiency η, when it acts as either a heat engine or a refrigerator. By the second law, these efficiencies must never exceed the Carnot efficiencies, η_{eng}^C and η_{ref}^C (which depend only on T_A and T_B). Ideally, we could use our exact results to find the maximum *relative efficiency* ($y=\eta/\eta^C$) that our model achieves, both as a heat engine and as a refrigerator. Unfortunately, the expressions for v, $\dot{Q}_A$, and $\dot{Q}_B$ are sufficiently complicated that we are not able to find these maxima analytically. However, at the end of Sec. IV, we will present analytical results for the maximum relative efficiencies achieved when the system operates *close to equilibrium*.

IV. LINEAR RESPONSE

When $f=0$ and $T_A=T_B\equiv\beta^{-1}$, our system is in equilibrium, and there results no average particle drift or heat flow. In Sec. III we briefly considered the behavior of our system *near* equilibrium [see, e.g., Eq. (34)]. We now consider this case in more detail. We will present a general analysis essentially the same as that of Jülicher, Ajdari, and Prost [13], and then show that the exact results obtained for our model confirm the predictions of this analysis.

For a sufficiently small load f and inverse temperature difference γ, and a fixed value of β (characterizing the in-

verse temperature of the equilibrium state around which we expand), we expect to be in the *linear response* regime: the particle drift and heat flows depend linearly on f and γ. Let us introduce the quantity

$$\Phi = \tfrac{1}{2}(\dot{Q}_A + \dot{Q}_B) \qquad (35)$$

roughly, a ''heat flux'' from A to B, and let us write, to leading order in f, γ:

$$\begin{pmatrix} v \\ \Phi \end{pmatrix} = \begin{pmatrix} \partial v/\partial f & \partial v/\partial \gamma \\ \partial \Phi/\partial f & \partial \Phi/\partial \gamma \end{pmatrix} \begin{pmatrix} f \\ \gamma \end{pmatrix}, \qquad (36)$$

with the derivatives of v and Φ evaluated at equilibrium ($f = \gamma = 0$). As per the arguments given at the end of the previous section, $\dot{Q}_A$ and $\dot{Q}_B$ are equal, to leading order in f and γ, near equilibrium.

The rate of entropy production is then

$$\dot{S} = -\frac{\dot{Q}_A}{T_A} + \frac{\dot{Q}_B}{T_B} = -\beta \dot{W} + \gamma \Phi = (f \quad \gamma) \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} f \\ \gamma \end{pmatrix}, \qquad (37)$$

where $M_{11} = -\beta \; \partial v/\partial f$, $M_{12} = -\beta \; \partial v/\partial \gamma$, $M_{21} = \partial \Phi/\partial f$, and $M_{22} = \partial \Phi/\partial \gamma$, evaluated at equilibrium. The second law implies that

$$\det \mathbf{M} \geqslant 0, \qquad (38)$$

whereas Onsager's reciprocity relation [24] predicts that $M_{12} = M_{21}$, or

$$-\beta \frac{\partial v}{\partial \gamma}\bigg|_{f=\gamma=0} = \frac{\partial \Phi}{\partial f}\bigg|_{f=\gamma=0}. \qquad (39)$$

Also, the diagonal elements of $\mathbf{M}$ must be positive: the particle must slide *down* the potential slope when the temperatures are equal ($M_{11} > 0$), and there must be a flow of heat from the hotter to the colder reservoir when the slope is zero ($M_{22} > 0$). Jülicher *et al.* [13] have obtained identical results for a molecular motor driven by a difference in chemical potential rather than temperature.

Using the exact results obtained in Sec. II, we differentiate v and Φ with respect to f and γ to evaluate the elements of the matrix $\mathbf{M}$ for our model:

$$\mathbf{M} = \begin{pmatrix} 3\beta^2 d^2 C(3+4\zeta) & \alpha\beta dC(1-\zeta) \\ \alpha\beta dC(1-\zeta) & \alpha^2 C[4+\zeta(29+\zeta)]/(4+3\zeta) \end{pmatrix}, \qquad (40)$$

where

$$\zeta = e^{-\alpha\beta}(0 < \zeta < 1), \quad C = \frac{3\zeta\Gamma}{(16+5\zeta)[1+\zeta(4+\zeta)]} > 0. \qquad (41)$$

[Recall that the expressions for $v(\mathbf{x})$, $\dot{Q}_A(\mathbf{x})$, and $\dot{Q}_B(\mathbf{x})$ differ according to the sign of f. We have verified that, regardless of whether we use the results valid for range (25b) ($f < 0$), or those for range (25c) ($f > 0$), we obtain the same results for the elements of $\mathbf{M}$.] The variable ζ is akin to μ
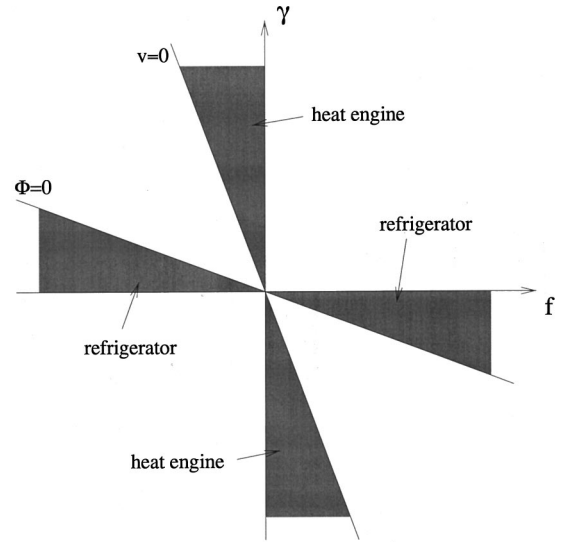

FIG. 9. General predictions based on linear response. The shaded regions adjacent to the vertical axis indicate the values of (f, γ) for which the system behaves as a heat engine; those adjacent to the horizontal axis indicate that the system is a refrigerator. The diagonal lines bounding these regions are the contours $v(f, \gamma) = 0$ and $\Phi(f, \gamma) = 0$. This figure essentially combines Figs. 7 and 8, for near-equilibrium values of $f, \gamma \approx 0$.

and ν: it is the ''rescaled temperature'' of the equilibrium state with respect to which the linear response behavior is defined. Then,

$$\det \mathbf{M} = \frac{9(\alpha\beta\zeta d\Gamma)^2(2+19\zeta+21\zeta^2)}{(4+3\zeta)(16+5\zeta)[1+\zeta(4+\zeta)]^2}, \qquad (42)$$

and we see by inspection that $M_{11}, M_{22} > 0$; that $M_{12} = M_{21}$, as mandated by Onsager reciprocity [Eq. (39)]; and that $\det \mathbf{M} > 0$, in agreement with the second law [Eq. (38)].

It is interesting to consider the operation of our system as a heat engine and refrigerator, in the linear-response regime. In this regime, the conditions for these two behaviors are $vf > 0$ for a heat engine (as before), and $\gamma\Phi < 0$ for a refrigerator (since $\Phi = \dot{Q}_A = \dot{Q}_B$, to leading order in f, γ). In Fig. 9, the shaded regions indicate values of (f, γ) for which the system acts as a heat engine or refrigerator (compare with Fig. 2 of Ref. [13]). The two diagonal lines, which form boundaries of these regions, are given by $v = 0$ and $\Phi = 0$. The slopes of these lines are

$$\lambda_{v=0} = -\frac{M_{11}}{M_{12}}, \quad \lambda_{\Phi=0} = -\frac{M_{21}}{M_{22}}. \qquad (43)$$

The second law, by requiring that $\det \mathbf{M} \geqslant 0$, guarantees that these shaded regions do not overlap: $|\lambda_{v=0}| \geqslant |\lambda_{\Phi=0}|$.

The *efficiency* of our system, when operating as a heat engine, is given by

$$\eta_{\text{eng}} = \frac{\dot{W}}{|\Phi|} \qquad (44)$$

(again using $\Phi = \dot{Q}_A = \dot{Q}_B$ to leading order). The *Carnot efficiency* defined for the temperatures T_A and T_B is

$$\eta_{\text{eng}}^C = \frac{|T_A - T_B|}{T_>} = \frac{|\gamma|}{\beta} + O(\gamma^2). \tag{45}$$

Then we can get an explicit expression for the relative efficiency ($y = \eta/\eta^C$) of our system, in the linear-response regime:

$$y_{\text{eng}} \equiv \frac{\eta_{\text{eng}}}{\eta_{\text{eng}}^C} = -\frac{1}{\lambda} \frac{M_{11} + M_{12}\lambda}{M_{21} + M_{22}\lambda}, \tag{46}$$

where $\lambda = \gamma/f$. The relative efficiency is the same for all points along any straight line through the origin, and Eq. (46) gives that relative efficiency as a function of the slope λ of the line. A similar analysis holds for the case of refrigeration:

$$y_{\text{ref}} \equiv \frac{\eta_{\text{ref}}}{\eta_{\text{ref}}^C} = -\lambda \frac{M_{21} + M_{22}\lambda}{M_{11} + M_{12}\lambda}. \tag{47}$$

Note that y_{ref} happens to be the inverse of y_{eng}, although the two expressions are valid for different ranges of λ values, corresponding to the shaded regions in Fig. 9.

The results of the previous two paragraphs were derived with the implicit assumption that $M_{12}, M_{21} > 0$. This happens to be true for our model, but, in general, these elements can be either positive or negative (or zero), so long as they are equal. If M_{12} and M_{21} were negative, then the shaded regions would occur in the the first and third quadrants of the (f, γ) plane, and the negative signs would not appear in Eqs. (46) and (47).

The above results imply that, near equilibrium, a microscopic device operating between two reservoirs either can act *both* as a heat engine and as a refrigerator (if $M_{12} = M_{21} \neq 0$), or will exhibit neither behavior (if $M_{12} = M_{21} = 0$). For instance, in a microscopic ratchet-and-pawl device, if the sawteeth on the wheel have a symmetric shape, the system cannot behave as a heat engine: $M_{12} = 0$, as is obvious by symmetry. What is not so immediately obvious, but follows from the condition $M_{12} = M_{21}$, is that it is equally impossible for a system with symmetric teeth to operate as a refrigerator, in the linear-response regime.

Finally, for a given inverse temperature β of the equilibrium state, we can solve for the *maximal* relative efficiency achievable in the linear-response regime, by maximizing y_{eng} and y_{ref} with respect to λ. It turns out that the maximal efficiencies are equal in the two cases, and depend only on a single parameter r characterizing the behavior of the system near equilibrium:

$$y^{\max} \equiv y_{\text{eng}}^{\max} = y_{\text{ref}}^{\max} = \frac{r}{(1 + \sqrt{1-r})^2},$$

$$r(\beta) = \frac{M_{12}M_{21}}{M_{11}M_{22}} \quad (0 \leq r \leq 1). \tag{48}$$

The value of $y^{\max}$ increases monotonically as r goes from 0 to 1.

From Eq. (40), we find that our model gives

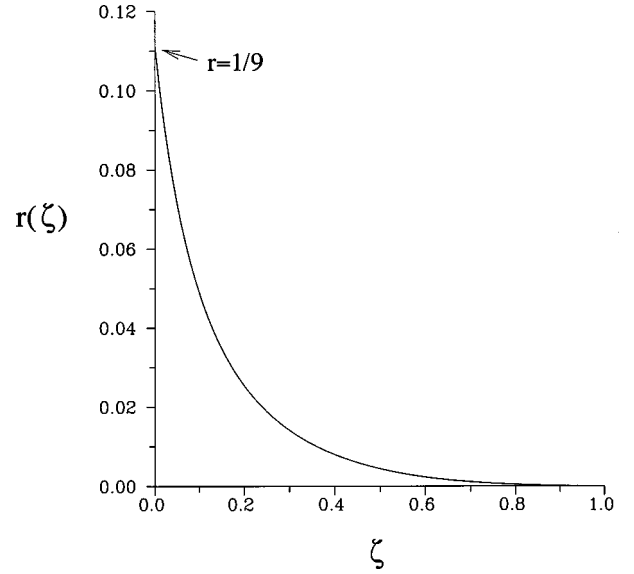$$r = \frac{(1-\zeta)^2(4 + 3\zeta)}{3(3 + 4\zeta)[4 + \zeta(29 + 9\zeta)]}, \tag{49}$$

FIG. 10. Quantity $r(\zeta)$ is plotted for the entire range of values of the (rescaled) equilibrium state temperature ($0 < \zeta < 1$).

which depends only on the rescaled temperature $\zeta = \exp(-\alpha\beta)$ of the equilibrium state. In Fig. 10 we plot $r(\zeta)$. We see that r approaches a limiting value of 1/9 as the equilibrium temperature $T = \beta^{-1}$ goes to zero (i.e., $\zeta \to 0$), and decreases to zero as $T \to \infty$ (i.e., $\zeta \to 1$). For the limiting value $r = 1/9$, our model gives a maximal relative efficiency,

$$y^{\max}(T \to 0) = \frac{1}{17 + 12\sqrt{2}} \approx 0.0294. \tag{50}$$

This is the best relative efficiency, which our system can achieve near equilibrium.

We have carried out a cursory numerical search—10^8 points sampled randomly in (f, μ, ν) space—and have found, *away from equilibrium*, relative efficiencies as high as $y \approx 0.0432$ (heat engine) and $y \approx 0.0647$ (refrigerator). These are greater than the near-equilibrium value quoted in Eq. (50), but still far short of unity. This suggests that the efficiency of our model is always considerably lower than the corresponding Carnot efficiency. Such a conclusion is in agreement with Parrondo and Espanol [19], and Sekimoto [20], who have argued that Feynman's analysis on the point of efficiency—in which he concluded that the ratchet and pawl would operate at Carnot efficiency—was in error. We note also that Hondou and Takagi [21], as well as Magnasco and Stolovitzky [22], have shown that, within a Langevin model for two degrees of freedom (e.g., the ratchet and pawl) coupled to two different heat reservoirs, Carnot efficiency cannot be achieved.

V. SUMMARY

Our aim in this paper has been to introduce a discrete model of Feynman's ratchet and pawl, and to solve exactly for the behavior of that model as a function of external load (f) and reservoir temperatures (T_A, T_B). The central result, Eqs. (30), gives the average directed motion (v) and heat flows ($\dot{Q}_A, \dot{Q}_B$), in the steady state. We have shown that our

model can act both as a heat engine and as a refrigerator, and we have investigated its behavior in the near-equilibrium, linear-response regime.

ACKNOWLEDGMENTS

APPENDIX

Here we present explicit expressions for the polynomials P_n^j, X^j, and Y^j appearing in Sec. II. The P_n^j's are obtained by Jordan elimination, performed on the matrices $\mathcal{R}^j$, $j=b,c$, and X_j and Y_j then follow from Eqs. (12), (13), (16), (17), (29), and (30).

$$P_1^b = \nu^3\sigma(1+\sigma+3\sigma^2) + \nu\sigma^2\{4+3\mu+(4+\mu)(1+2\mu)\sigma+[4+3\mu(3+2\mu)]\sigma^2\} + 4\mu\sigma[3+\sigma(5+3\mu+5\sigma+4\mu\sigma)]$$
$$+ \nu^2\{4+\sigma[4+2\mu+\sigma(6+\sigma+\sigma^2+3\sigma^3)]\},$$

$$P_2^b = \nu\sigma^2\{4+3\mu+(12+7\mu)\sigma+[12+\mu(15+2\mu)]\sigma^2\} + \nu^2[4+2(6+\mu)\sigma+6(1+\mu)\sigma^2+\sigma^3+3\sigma^4+\sigma^5]$$
$$+ \nu^3\sigma[1+\sigma(3+\sigma)]+4\mu\sigma\{3+\sigma[3+\mu+3(1+\mu)\sigma]\},$$

$$P_3^b = 4\mu\sigma[3+\sigma(5+3\sigma)]+\nu^3\sigma[1+\sigma(5+9\sigma)]+\nu\sigma^2\{4(1+\sigma+3\sigma^2)+\mu[3+\sigma(9+7\sigma)]\}$$
$$+ \nu^2\{4+\sigma[12+\sigma(18+\sigma+5\sigma^2+9\sigma^3)]\},$$

$$P_4^b = \sigma\{16+4(6+\nu)\sigma+6(4+\nu)\sigma^2+10\nu\sigma^3+\mu[4+\sigma(20+\nu+24\sigma+5\nu\sigma+9\nu\sigma^2)]\},$$

$$P_5^b = \nu^2[4+\mu+5(2+\mu)\sigma+3(2+3\mu)\sigma^2]+2\mu\sigma[4+\mu+(4+3\mu)\sigma+(4+5\mu)\sigma^2]$$
$$+ 2\nu\sigma^2[4+\mu+(8+5\mu)\sigma+(8+9\mu)\sigma^2],$$

$$P_6^b = 2\mu\nu\sigma^2[2+\mu+(2+3\mu)\sigma+(4+3\mu)\sigma^2]+2\mu^2\sigma[3+\sigma(5+3\sigma)]+\nu^3\sigma\{2+4\sigma(1+\sigma)+\mu[1+\sigma(5+9\sigma)]\}$$
$$+ \nu^2\{4+6\sigma+2\sigma^2(3+\sigma[1+2\sigma(1+\sigma)])+\mu\{3+\sigma[11+\sigma(15+\sigma+5\sigma^2+9\sigma^3)]\}\}). \tag{A1}$$

$$X^b = \nu^3\sigma(-1-2\sigma+3\sigma^3)+\nu\sigma^2\{-4-3\mu-2\mu\sigma+4\mu\sigma^2+3[4+\mu(5+2\mu)]\sigma^3\}+\nu^2[-4-3(4+\mu)\sigma-3(4+3\mu)\sigma^2$$
$$- (1+9\mu)\sigma^3+(2+\mu)\sigma^4+(6+5\mu)\sigma^5+(13+9\mu)\sigma^6]+2\mu\sigma(-6+\sigma\{-4+6\sigma^2+3\mu[-1+(-1+\sigma)\sigma]\}),$$

$$Y^b = -\nu^3\sigma[3+\sigma(5+7\sigma)]+\nu\sigma^2\{-2\mu^2(1+\sigma)(1+2\sigma)+\mu[-1+(3-5\sigma)\sigma]+4(1+\sigma+\sigma^2)\}-2\mu\sigma\{2(1+\sigma+\sigma^2)$$
$$+ \mu[5+\sigma(9+7\sigma)]\}+\nu^2[8+\mu(-1+\sigma-3\sigma^2)+\sigma(10+\sigma\{12+\sigma[3+\sigma(5+7\sigma)]\})]. \tag{A2}$$

$$P_1^c = \nu^3\sigma[1+\sigma(3+\sigma)]+4\mu\sigma[5+\mu(5+2\sigma)+\sigma(5+3\sigma)]+\nu\sigma^2\{2\mu^2(3+\sigma)+4(1+\sigma+\sigma^2)+3\mu[3+\sigma(3+\sigma)]\}$$
$$+ \nu^2\{4+2\mu+\sigma[6+\sigma(4+\sigma+3\sigma^2+\sigma^3)]\},$$

$$P_2^c = \nu^3\sigma(3+\sigma+\sigma^2)+4\mu\sigma[\mu+3\mu\sigma+3(1+\sigma+\sigma^2)]+\nu\sigma^2\{2\mu^2\sigma+4[3+\sigma(3+\sigma)]+\mu[15+\sigma(7+3\sigma)]\}$$
$$+ \nu^2(12+2\mu(3+\sigma)+\sigma\{6+\sigma[4+\sigma(3+\sigma+\sigma^2)]\}),$$

$$P_3^c = 4\mu\sigma[5+3\sigma(1+\sigma)]+\nu^3\sigma[9+\sigma(5+\sigma)]+\nu\sigma^2\{4[1+\sigma(3+\sigma)]+\mu[9+\sigma(7+3\sigma)]\}$$
$$+ \nu^2[16+\sigma(14+\sigma\{4+\sigma[9+\sigma(5+\sigma)]\})],$$

$$P_4^c = \sigma[24+2\sigma(3+2\sigma)(4+\nu+\nu\sigma)+\mu(28+\sigma\{4(4+\sigma)+\nu[9+\sigma(5+\sigma)]\})],$$

$$P_5^c = 2\mu\sigma\{4(1+\sigma+\sigma^2)+\mu[3+\sigma(5+\sigma)]\}+\nu^2\{10+6\sigma+4\sigma^2+\mu[9+\sigma(5+\sigma)]\}+2\nu\sigma^2\{4[2+\sigma(2+\sigma)]$$
$$+\mu[9+\sigma(5+\sigma)]\},$$

$$P_6^c = 2\mu^2\sigma[5+3\sigma(1+\sigma)]+2\mu\nu\sigma^2\{2(1+\sigma)^2+\mu[3+\sigma(3+\sigma)]\}+\nu^3\sigma\{2[2+\sigma(2+\sigma)]+\mu[9+\sigma(5+\sigma)]\}$$
$$+\nu^2\{2[3+\sigma(3+\sigma\{2+\sigma[2+\sigma(2+\sigma)]\})]+\mu[15+\sigma(11+\sigma\{3+\sigma[9+\sigma(5+\sigma)]\})]\}. \tag{A3}$$

$$X^c = \sigma(\nu\sigma^2[-4-5\mu+4\mu(2+\mu)\sigma+2(2+\mu)^2\sigma^2+(4+3\mu)\sigma^3]+\nu^3(-3\sigma+2\sigma^3+\sigma^4)$$
$$+2\mu\sigma\{-10+6\sigma^3+\mu[-5+\sigma(-1+3\sigma)]\}+\nu^2\{-22+\mu[-15+\sigma(-5+\sigma\{-1+\sigma[9+\sigma(5+\sigma)]\})]$$
$$+\sigma[-8+\sigma(-2+\sigma\{7+\sigma[10+\sigma(6+\sigma)]\})]\}),$$

$$Y^c = -\{\nu^3\sigma[5+\sigma(7+3\sigma)]+\nu\sigma^2\{4(1+\sigma+\sigma^2)-2\mu^2[3+\sigma(2+\sigma)]-\mu[-3+\sigma(5+\sigma)]\}-2\mu\sigma\{2(1+\sigma+\sigma^2)$$
$$+\mu[9+\sigma(7+5\sigma)]\}+\nu^2[10-\mu[-1+\sigma(3+\sigma)]+\sigma(12+\sigma\{8+\sigma[5+\sigma(7+3\sigma)]\})]\}. \tag{A4}$$

---

[1] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading, MA, 1966), Vol. I, Chap. 46.

[2] M. Smoluchowski, Phys. Z. **13**, 1069 (1912); in *Vortage uber die Kinetische Theorie der Materie und der Elektrizitat*, edited by M. Planck (Teubner und Leipzig, Berlin, 1914), pp. 89–121. For a numerical study of ''Smoluchowski's trapdoor,'' see Ref. [3] below.

[3] P. A. Skordos and W. H. Zurek, Am. J. Phys. **60**, 867 (1992).

[4] M. O. Magnasco, Phys. Rev. Lett. **71**, 1477 (1993).

[5] R. D. Astumian and M. Bier, Phys. Rev. Lett. **72**, 1766 (1994).

[6] J. Prost, J.-F. Chauwin, L. Peliti, and A. Ajdari, Phys. Rev. Lett. **72**, 2652 (1994).

[7] J.-F. Chauwin, A. Ajdari, and J. Prost, Europhys. Lett. **27**, 421 (1994).

[8] J. Rousselet, L. Salome, A. Ajdari, and J. Prost, Nature (London) **370**, 446 (1994).

[9] L. P. Faucheux, L. S. Bourdieu, P. D. Kaplan, and A. J. Libchaber, Phys. Rev. Lett. **74**, 1504 (1995).

[10] C. R. Doering, Nuovo Cimento D **17**, 685 (1995).

[11] R. D. Astumian, Science **276**, 917 (1997).

[12] M. Bier, Contemp. Phys. **38**, 371 (1997).

[13] F. Jülicher, A. Ajdari, and J. Prost, Rev. Mod. Phys. **69**, 1269 (1997), Sec. II.

[14] P. Hänggi and R. Bartussek, in *Nonlinear Physics of Complex Systems: Current Status and Future Trends*, edited by J. Parisi, S. C. Müller, and W. Zimmermann (Springer-Verlag, Berlin, 1996), p. 294.

[15] R. D. Astumian and M. Bier, Biophys. J. **70**, 637 (1996).

[16] L. Schimansky-Geier, M. Kschischo, and T. Fricke, Phys. Rev. Lett. **79**, 3335 (1997).

[17] A. B. Kolomeisky and B. Widom, J. Stat. Phys. **93**, 633 (1998).

[18] H. Ambaye and K. W. Kehr, e-print cond-mat/9901034.

[19] J. M. R. Parrondo and P. Espanol, Am. J. Phys. **64**, 1125 (1996).

[20] K. Sekimoto, J. Phys. Soc. Jpn. **66**, 1234 (1997).

[21] T. Hondou and R. Takagi, J. Phys. Soc. Jpn. **67**, 2974 (1998).

[22] M. O. Magnasco and G. Stolovitzky, J. Stat. Phys. **93**, 615 (1998).

[23] N. Metropolis *et al.*, J. Chem. Phys. **21**, 1087 (1953).

[24] L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).