Article

A Programmable Mechanical Maxwell's Demon

Zhiyue Lu [1,‡,*] and Christopher Jarzynski [2,3,4,‡,*]

1 James Franck Institute, University of Chicago, Illinois 60637
2 Institute for Physical Science and Technology, University of Maryland, College Park 20742
3 Department of Chemistry and Biochemistry, University of Maryland, College Park 20742
4 Department of Physics, University of Maryland, College Park 20742
* Correspondence: zhiyelu@gmail.com, cjarzyns@umd.edu
‡ These authors contributed equally to this work.

Version November 26, 2018 submitted to

Abstract: We introduce and investigate a simple and explicitly mechanical model of Maxwell's demon – a device that interacts with a memory register (a stream of bits), a thermal reservoir (an ideal gas) and a work reservoir (a mass that can be lifted or lowered). Our device is similar to one that we have briefly described elsewhere [1], but it has the additional feature that it can be programmed to recognize a chosen reference sequence, for instance, the binary representation of π. If the bits in the memory register match those of the reference sequence, then the device extracts heat from the thermal reservoir and converts it into work to lift a small mass. Conversely, the device can operate as a generalized Landauer's eraser (or copier), harnessing the energy of a dropping mass to write the chosen reference sequence onto the memory register, replacing whatever information may previously have been stored there. Our model can be interpreted either as a machine that autonomously performs a conversion between information and energy, or else as a feedback-controlled device that is operated by an external agent. We derive generalized second laws of thermodynamics for both pictures. We illustrate our model with numerical simulations, as well as analytical calculations in a particular, exactly solvable limit.

Keywords: Maxwell's demon; Shannon entropy; information engine; Landauer's principle; Szilard engine; second law of thermodynamics.

1. Introduction

The field of information thermodynamics traces its origins to a whimsical, 150-year-old thought experiment. In a letter to a friend [2], James Clerk Maxwell introduced a hypothetical "neat-fingered being", now universally known as *Maxwell's demon*, who brings about an apparent violation of the second law of thermodynamics, simply by observing the motions of gas molecules and manipulating a trapdoor to segregate faster from slower molecules. While Maxwell emphasized the role of the demon's intelligence, subsequent researchers – notably including Marian Smoluchowski [3] and Richard Feynman [4] – have considered whether a dumb device might be able to accomplish similar results, and if so, what the existence of such a device would imply about the status of the second law. In recent decades a consensus has formed around a perspective developed largely by Rolf Landauer, Oliver Penrose and Charles Bennett.[5–7] At the heart of this perspective is the notion that if Maxwell's demon were a purely physical machine, then the information it gathers must be stored in a physical memory register, commonly represented as a sequence of classical bits. The writing of this information increases the entropy of the bits, thereby (so the argument goes) compensating for the decrease of entropy that occurs elsewhere as the machine "violates" the second law. Bennett's analysis of chemical proofreading [8] provides an early model system illustrating this idea.

The past decade has seen renewed interest in this topic, motivated in part by its connections with fluctuation theorems and related advances in nonequilibrium statistical physics [9–11], as well as by improved experimental capabilities for manipulating small systems [12–23]. Progress has included

sharpened relationships between thermodynamic and information-theoretic quantities [24–32] as well as a variety of simple model systems more explicit than those explored in the past [1,18–21,33–46].

Two broad paradigms have emerged in these investigations – *autonomous* and *non-autonomous* demons. The non-autonomous paradigm echoes Maxwell's original idea: an external agent that is in a sense "outside of Physics" (the demon) performs feedback control on a material object (e.g. a trapdoor) to accomplish a task apparently prohibited by the second law. This task may be the creation of a temperature gradient as in Maxwell's scenario, or the conversion of heat into work as in many later models such as the Szilard engine[33]. The key idea is that the agent rectifies thermal fluctuations, using the information it gains by observing nanoscale motions. In this paradigm, the thermodynamic benefits delivered by the agent – such as work generated to lift a mass against gravity – are related to the amount of information it gathers about its surroundings.

By contrast, the autonomous paradigm is all-inclusive, in that the demon and, importantly, its memory are explicitly modeled as physical systems.[1,37,40,41,43–46] In this paradigm the goal is often to illustrate how a physical machine might actually accomplish results similar to those of Maxwell's imagined neat-fingered being, and to explore quantitatively how the thermodynamic benefits that the machine delivers are related to changes in the information content of its memory.

In the present paper we introduce and analyze a model of Maxwell's demon that can be interpreted within either the autonomous or the non-autonomous paradigm. Our model builds on one that we briefly described, with our colleague Dibyendu Mandal, in 2014.[1] Unlike earlier models involving systems making stochastic transitions among a discrete set of states[37,38,40–43,45,46], our model is entirely mechanistic – the demon and its memory consist of frictionless, moving components immersed in a dilute gas, evolving under Newtonian dynamics. Specifically, the demon is a rotational ring equipped with two blades and the memory is represented by a sequence of rotating paddles, as shown in Fig. 3 and discussed in greater detail in Sec. 2. We showed in Ref. [1] that if the system's memory is initialized in a "clean" state corresponding to the bit sequence "...00000...", then the mechanistic interplay between the ring, the paddles and the dilute gas produces rotational motion that lifts a small mass against gravity. In this mode of operation the entire contraption is an *information engine*, rectifying thermal fluctuations to convert heat into work – the fuel for this process is provided by the randomization of the memory, as the clean bit stream is converted to a "polluted" mixture of 0's and 1's. Conversely, if the memory begins in a random mixture "...01101...", then a large mass that drops with gravity can be harnessed to reset all the bits to 0's, illustrating Landauer's principle [5] that work is required to erase information.

In Ref. [1] a clean memory register was equated with the uniform bit sequence "...00000...". In principle, however, what matters is not uniformity but rather lack of randomness, as quantified by Shannon entropy. Let us use the term *generalized clean memory* to denote an arbitrary but *fully determined* bit sequence, for instance the binary representation of π. Since a fully determined sequence is entropically equivalent to the sequence "...00000...", a generalized clean memory should be able to serve as a thermodynamic resource to drive an information engine. This consideration motivates us to design a mechanical information engine that operates on a generalized clean memory. Our model is programmable, in the following sense: for any choice of pre-determined *reference sequence* – be it the binary representation of π, or the repeating sequence "...010101...", or for that matter the uniform sequence "...000000..." – we can program the system so that if the memory bits are initialized in this reference sequence then the machine operates as an information engine, lifting a small mass against gravity, as illustrated schematically in Fig. 1. Conversely, if the bits are initialized in a different sequence, then the energy from a falling large mass can be used to write the reference sequence onto the bits (rather than resetting them all to the 0 state as in Ref. [1]).

As we will describe in further detail, our system is programmed using a sequence of binary *programmable gates*. If these gates are fixed, prior to the start of the process, to match the chosen reference sequence, then the machine operates as an autonomous Maxwell demon. However, we can alternatively imagine that an external agent arranges the gates on the fly, one by one, using information

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 November 2018 doi:10.20944/preprints201811.0570.v1

Peer-reviewed version available at Entropy 2019, 21, 65; doi:10.3390/e21010065

based on real-time observations. By thus reinterpreting each programmable gate as a binary switch that is feedback controlled by the agent, our model becomes an illustration of a non-autonomous Maxwell demon. Analogous to the autonomous picture, where the system can operate in either an engine mode or an information copier mode, the agent-involved feedback control picture operates either as an engine or information recorder. (See Fig. 2.)

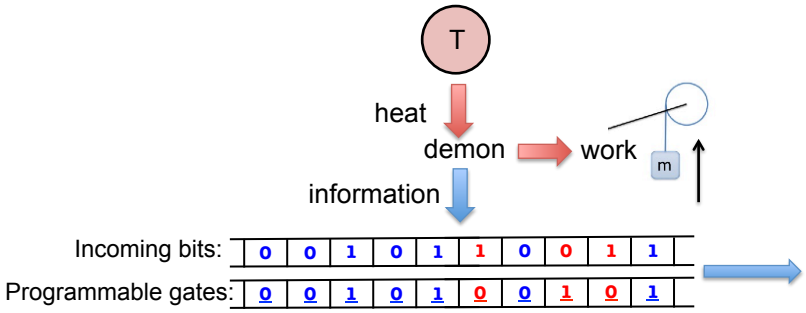
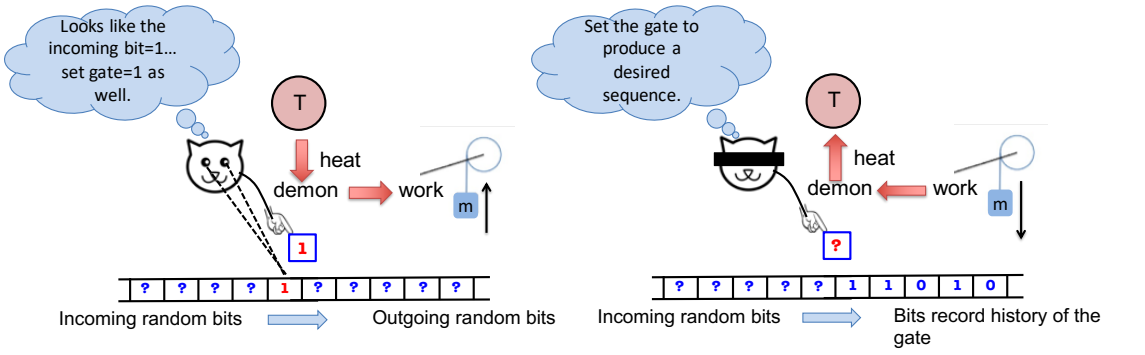


Figure 1. In our schematic conception of a programmable, autonomous Maxwell's demon, a fixed set of binary gates defines a *reference sequence*. As the demon interacts one bit at a time with an incoming sequence of memory bits, it is able to lift a small mass against gravity if the incoming bit sequence matches the reference sequence. As the demon writes information onto the memory bits, the outgoing sequence becomes less correlated with the reference sequence. Conversely, if the mass is large and falls against gravity, then this energy can be used to copy the reference sequence to the memory bits.



a. Feedback control engine powered by correct measurement b. Information eraser and recorder powered by mechanical work.

Figure 2. Alternatively, our model can illustrate a non-autonomous device operated via feedback control by an external agent. (a) In the engine mode, which resembles Szilard's thought experiment [33], the agent measures each incoming memory bit and switches a gate accordingly. When these measurements are accurate, the procedure induces a bias toward counter-clockwise rotation that can be harnessed to lift a small mass against gravity. (b) If the mass is large and falls against gravity, the energy that is released can be used to write a sequence chosen by the agent, onto the outgoing bit stream. In this mode the agent does not measure the incoming bits, but rather manipulates the gate to encode the desired sequence.

The paper is structured as follows. In Sec. 2 we describe the various components of our device, and we sketch how it can operate as an autonomous information engine. In Sec. 3 we describe in detail the three possible modes of operation of our autonomous device: as an engine, an eraser (or copier) and a "dud". In Sec. 4 we illustrate these modes of operation using numerical simulations, we solve explicitly for the behavior of the model in a particular "slow-moving" limit, and we consider its thermodynamic description, including its efficiency. In Sec. 5 we discuss how our model can be used to illustrate a non-autonomous machine, operated by an external agent using measurement and

feedback – as in Maxwell's original thought experiment – and we obtain a bound on the amount of work that this machine can deliver. We end in Sec. 6 with a brief summary and discussion.

2. Programmable Maxwell's Demon

2.1. Components and basic design

As mentioned, the machine described in this paper is equipped with a binary reference sequence that can be preprogrammed to any desired pattern of 0's and 1's – for instance, the binary representation of π. This *reference sequence* is fixed, and is distinct from the sequence of *memory bits* that interact dynamically with the rotational ring. As we will argue, if the incoming memory bits match the reference sequence then the ring favors counter-clockwise (CCW) rotation that can be used to perform work against an external load.

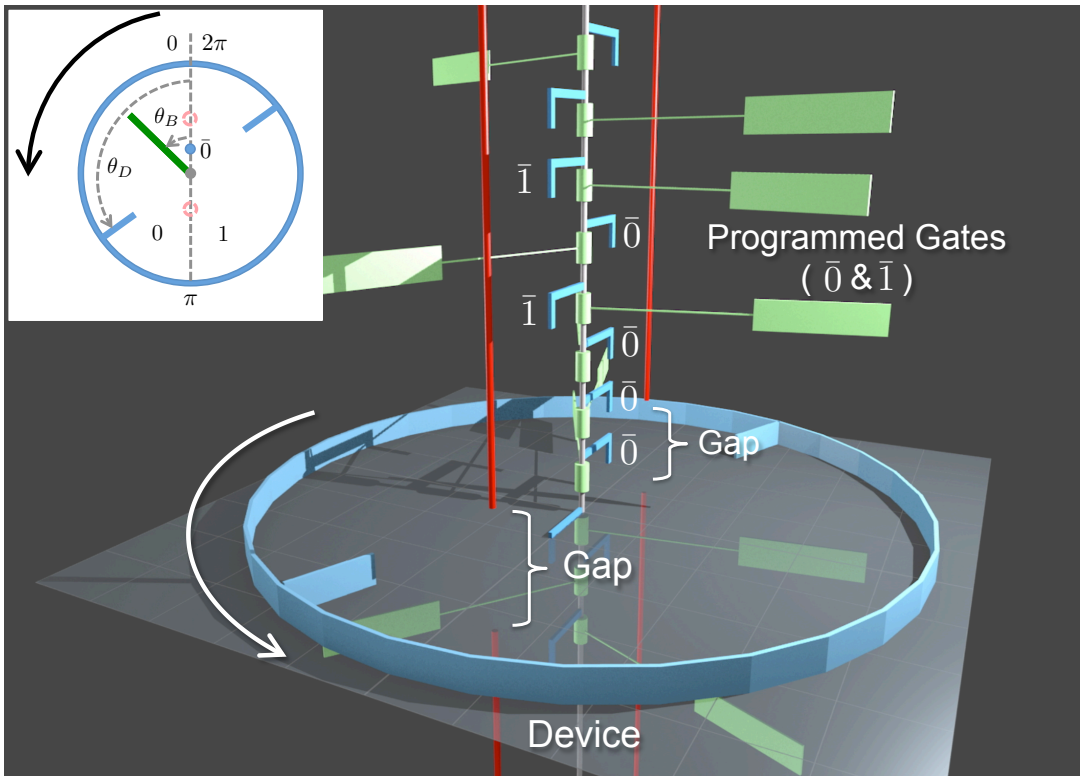


Figure 3. Snapshot of our programmable demon. A series of green paddles move down frictionlessly along the central axle. The paddles are separated by the red bars into binary states, left (0) and right (1) – see inset. Each bit passes by the rotational ring (the blue ring with two inward blades) for the same finite amount of time, during which it can change states. We claim that if the incoming bits $(000101\cdots)$ are in agreement with the programmed gates $(\bar{0}\bar{0}\bar{1}\bar{0}\bar{1}\cdots)$, then the ring favors CCW motion, which can be used to lift an external load. A top view of the system is shown in the inset. A video clip illustrating the dynamics of our demon is found at https://youtu.be/LkYljJ__-Cs

As illustrated in Figure 3, the entire machine consists of three components – a sequence of paddles acting as the bits of a memory register, a set of fixed gates that encode a preprogrammed binary reference sequence, and the demon that is realized by a rotational ring: a ring that interacts with the memory register via blade-paddle collision and can perform work via rotation against a constant external force. We now describe these components in detail, beginning with the paddles that constitute the bits of the memory register. These paddles rotate frictionlessly around a central axle. The orientation of a paddle is given by an angle θ_B. When $\theta_B \in (0, \pi)$ the paddle represents a bit in the 0 state, and when $\theta_B \in (\pi, 2\pi)$ it represents a bit in the 1 state. Two blocking bars (shown as vertical

red bars) located at angles 0 and π prevent each bit from spontaneously flipping between the 0 and 1 states. Each blocking bar contains a gap, as shown in Fig. 3. The central axle moves downward at a constant speed, carrying the bits and gates past the demon. The entire machine is immersed in an ideal gas in thermal equilibrium at temperature T. The gas particles collide elastically with the paddles, causing them to undergo Brownian-like rotation around the axle. For clarity the gas particles are not shown in the figure.

The preprogrammed reference sequence is encoded in a set of rigid gates attached to the central axle, which accompany the paddles as they move downward past the demon. These gates are shown as L-shaped blue bars extending perpendicularly from the axle. The orientation of a gate is fixed at either $\theta = 0$ (representing state $\bar{0}$) or $\theta = \pi$ (state $\bar{1}$). When a paddle and its gate arrive at the vertical location of the gaps on the red bars, the paddle is able to switch its state by passing through the gap that is not blocked by the gate. For example, if the gate is in state $\bar{0}$, the gap at $\theta = 0$ is blocked, and the bit paddle can switch its state by passing through the gap at $\theta = \pi$.

The rigid ring is equipped with two inward-pointing blades, attached at opposite locations. The ring rotates freely around the central axle but does not translate or wobble. The angular orientation of the ring is specified by θ_D; see inset of Figure 3. Like the paddles, the ring undergoes Brownian-like rotation due to elastic collisions between its blades and the gas particles. Additionally, the ring's blades can collide elastically with the paddles as they move past it. The ring is situated at the vertical height of the gaps in the blocking bars. The spacing between bits, the size of the gaps, and the vertical widths of the paddles and the ring's blades are set so that, at any time, there is exactly one paddle within the vertical range of the gap, and that paddle is simultaneously within the vertical collision range of the ring's blades. This paddle is called the *interacting bit*, and its gate is called the *engaged gate*. We will use the term *interaction interval* to denote the interval of time during which a given paddle acts as an interacting bit, and its gate acts as the engaged gate. The duration of the interaction interval, τ^{int}, is the same for each paddle and gate.

The life cycle of a given paddle (memory bit) then proceeds as follows. Prior to arriving at the vertical level of the ring, the orientation of the paddle, θ_B, performs Brownian-like motion but the binary state of the bit (0 or 1) is frozen due to the presence of the blocking bars. This binary state represent an *incoming* memory bit. Then, over the course of an interaction interval of duration τ^{int}, the paddle can switch between the 0 and 1 states, by passing through the gap that is not blocked by the reference gate; during this interval the paddle also interacts with the blades of the ring. Finally, after the interaction interval, as the paddle passes below the vertical level of the ring, the binary state of the paddle is once again frozen due to the blocking bars – at this point the paddle represents an *outgoing* memory bit.

2.2. Memory register – a sequence of bits

The binary state of an incoming memory bit (paddle), $b \in \{0, 1\}$, might or might not be the same as the binary state of the corresponding reference bit (gate), $g \in \{\bar{0}, \bar{1}\}$. We will characterize the cleanness of the incoming bit sequence $(\cdots b_{n-1}, b_n, b_{n+1} \cdots)$ by the degree to which it matches the fixed reference sequence $(\cdots g_{n-1}, g_n, g_{n+1} \cdots)$. If the binary state of each incoming memory bit matches that of the accompanying gate, i.e. if $b_n = g_n \, \forall n$, then the memory is considered to be perfectly clean. If the incoming sequence contains mismatches between memory and reference bits, then these mismatches are considered to be impurities that pollute the memory sequence.

Let $P_{\text{in}}(\text{same})$ denote the fraction of incoming bits that are correctly matched with their reference bits ($0\bar{0}$ or $1\bar{1}$), and $P_{\text{in}}(\text{diff})$ the fraction that are mismatched ($0\bar{1}$ or $1\bar{0}$). We assume that the probability of a mismatch is independent of the state of the reference bit, and the mismatches are statistically uncorrelated with one another. We quantify the cleanness of the incoming memory by the excess ratio of clean bits:

$$\delta = P_{\text{in}}(\text{same}) - \text{P}_{\text{in}}(\text{diff}) \in [-1, +1] \tag{1}$$

It is useful at this point to introduce a logical variable L that is the *Boolean equality* between the states of the bit and the gate: $L = B$ Exy G [47]. That is, the value of L is given by

$$\begin{aligned} l = \text{true} \equiv \text{same} \quad &\text{if} \quad b = g \\ l = \text{false} \equiv \text{diff} \quad &\text{if} \quad b \neq g \end{aligned} \tag{2}$$

Here and below, we use the capital letters B, G and L to refer to binary variables, and lower case b, g and l to denote the values of these variables. The sequences of incoming memory and reference bits together specify a sequence $(\cdots l_{n-1}, l_n, l_{n+1} \cdots)$, whose Shannon entropy, per bit, is given by

$$S_{\text{L,in}} = -\left[P_{\text{in}}(\text{same}) \log \text{P}_{\text{in}}(\text{same}) + \text{P}_{\text{in}}(\text{diff}) \log \text{P}_{\text{in}}(\text{diff})\right] \in [0, \log 2] \tag{3}$$

For the outgoing bits $(\cdots b'_{n-1}, b'_n, b'_{n+1} \cdots)$ we can similarly define $P_{\text{out}}(\text{same})$, $P_{\text{out}}(\text{diff})$ and

$$S_{\text{L,out}} = -\left[P_{\text{out}}(\text{same}) \log \text{P}_{\text{out}}(\text{same}) + \text{P}_{\text{out}}(\text{diff}) \log \text{P}_{\text{out}}(\text{diff})\right] \tag{4}$$

The difference $\Delta S_L = S_{\text{L,out}} - S_{\text{L,in}}$ quantifies the cleanness of the memory sequence, per bit, due to the interactions between the memory bits and ring. While the interaction between the memory bits and the demon might in principle induce correlations among the outgoing bits, in our analysis we will ignore these correlations.

2.3. Work reservoir – a mass that can be raised or lowered

In addition to the elements described above, an external load, Γ, acts on the ring in the clockwise (CW) direction. This load is generated by a mass that hangs from a string wrapped around the ring – the gravitational force on the mass produces a CW torque on the ring. If the ring rotates in the counter-clockwise (CCW) direction the mass is lifted upwards. This mass is not shown in Figure 3.

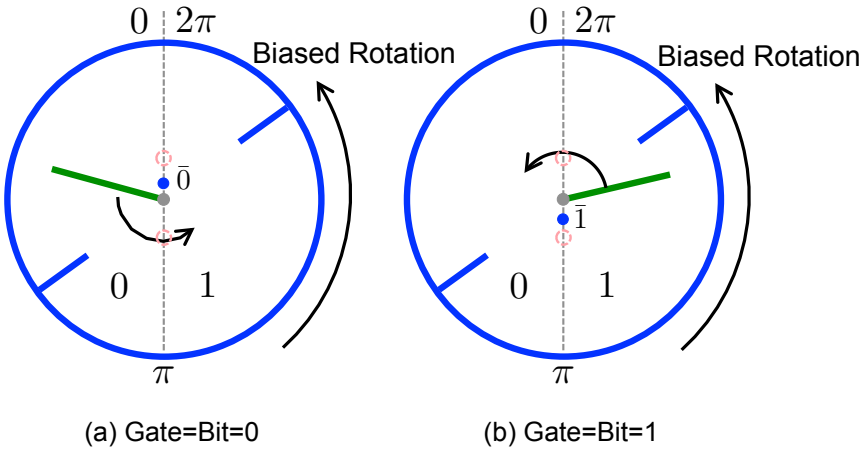


Figure 4. Engine Mode. The ring prefers CCW rotation when the bit starts with the state that is in agreement with its corresponding gate. The blue dots represent the programmed gates.

It is useful to understand the operation of our machine in the absence of this load, i.e. when $\Gamma = 0$. To that end, let us first assume that the incoming bit sequence is perfectly clean: the binary state of each memory bit matches that of its reference bit ($\delta = +1$). There are then two possible combinations for an incoming memory and reference bit, $(0\bar{0})$ and $(1\bar{1})$, illustrated in Figure 4. In the former case (Fig. 4a), the paddle is initially confined (by the blocking bars) within the angular range $\theta_B \in (0, \pi)$. During the interaction interval, this paddle has the opportunity to "expand" into the full circular range $(0, 2\pi)$, by swinging through the gap located at $\theta = \pi$. This opportunity produces a statistical bias that favors CCW rotation, which in turn induces a CCW rotational bias for the ring, due to the possibility of

collisions between the paddle and the ring's blades. For the incoming combination $(1\bar{1})$, the expansion of the memory bit during the interaction again interval favors CCW rotation as the reference gate blocks the gap at $\theta = \pi$ (Fig. 4b). In this manner, over the course of many interaction intervals the ring settles into a steady state in which the ring rotates systematically in the counterclockwise direction – the thermal fluctuations generated by collisions with the gas particles are rectified to produce directed rotation. In this steady state there is a continual exchange of energy (due to collisions) between the ring's blades and the gas, but this exchange does not lead to a net flow of energy in one direction.

By similar arguments, the maximally unclean situation ($\delta = -1$) produces an identical bias in the clockwise direction. More generally, each correctly matched pair of memory and reference bits generates a bias toward CCW rotation, while each mismatched pair generates a bias toward CW rotation. Hence, over many interaction intervals, our ring (in the absence of an external load) produces a rotational bias whose direction is CCW for $\delta > 0$ and CW for $\delta < 0$. The strength of the bias is quantified by $|\delta|$.

Let us now assume $\delta > 0$ and imagine that we add an external load, $\Gamma > 0$. If the load is sufficiently small then the bias generated by the ring will continue to produce CCW rotation (albeit at a lower rate than if the load were absent) thereby lifting the mass against gravity. In this situation the ring settles into a steady state in which energy is systematically withdrawn from the heat bath (gas) and delivered to perform mechanical work.

3. Operational Modes of the Programmable Demon

More generally, the behavior of our ring depends on four parameters that we consider to be tunable: the memory cleanness δ, the bath temperature T, the external load Γ, and the duration of the interaction interval τ^{int}. [1] Depending on the values of these four parameters, the machine operates in one of three different modes – as an information engine, an information eraser or a dud. In the limit $\tau^{\text{int}} \to \infty$ the model becomes analytically solvable (see Sec. 4.2), and its behavior is determined by the dimensionless parameters δ and $\beta\Gamma$, where $\beta = (k_B T)^{-1}$, as illustrated by the phase diagram shown in Figure 10. We now discuss each mode separately.

3.1. Engine mode

As mentioned in Sec. 2.3, for $\delta > 0$ and sufficiently small $\Gamma > 0$, the ring is able to convert energy drawn from the heat bath into work against the external load, thereby operating as an engine. In the limiting case $\delta = 1$, each incoming bit is matched perfectly to its reference bit, but this is no longer the case with the outgoing bits:

$$\delta' \equiv P_{\text{out}}(\text{same}) - P_{\text{out}}(\text{diff}) < 1 \tag{5}$$

More generally, when $\delta > 0$ and the ring operates in the engine mode we have

$$\delta > \delta' > 0 \tag{6}$$

as CCW rotation tends to generate mismatches between memory bits and reference bits. Eq. 6 indicates that there is greater uncertainty – less correlation with the reference bits – in the outgoing memory sequence than in the incoming sequence: $\Delta S_L > 0$. In effect, the decrease of thermodynamic entropy associated with the continual withdrawal of energy from the heat bath, is compensated by the increase of the Shannon entropy of the memory register. Our ring thus operates as an *information engine*, with a clean sequence of incoming bits serving as a thermodynamic resource that allows the system to convert heat from the bath into work against the load, without violating the second law of thermodynamics.

[1] All other parameters, such as the mass and density of gas particles, the length of the paddles, etc., are fixed in our model.

In the non-programmable engine of Ref. [1], an incoming sequence of 0's is converted into a mixture of 0's and 1's. It is natural to view this conversion as a process of *writing* information to the bit sequence. The outgoing pattern encodes information about the history of the ring, as outgoing 1's are correlated with CCW rotation during the corresponding interaction intervals. In the present model, by contrast, both the incoming and the outgoing sequences are mixtures of 0's and 1's. We can still view this as a process of writing information, provided this information is defined relative to the reference bits: a mismatch between an outgoing memory bit and its reference bit indicates a likelihood that CCW rotation occurred during that bit's interaction interval. Alternatively, for the present model we might view the incoming sequence as containing information (e.g. the binary digits of π), which is "digested" by the ring as it rectifies thermal fluctuations to generate work. Regardless of whether we interpret the ring as writing information onto a clean memory sequence, or digesting information contained in that sequence, the net result is the same: when the ring acts as an engine, the outgoing bit sequence is more disordered than the incoming one, $\Delta S_L > 0$.

3.2. Eraser mode

Now let us consider what happens when (1) the incoming bit sequence is maximally unclean ($\delta = 0$) i.e. the incoming bits are uncorrelated with the reference bits, and (2) a large mass produces a strong external load in the CW direction, $\Gamma > 0$. During a given interaction interval the mass drops as far as it can, producing CW rotation of the ring until the interacting paddle (bit) is pinched between one of the blades of the ring and the rigid engaging gate associated with that paddle, as illustrated in Figure 5. If the reference bit is in state $\bar{0}$, then the engaging gate is located at $\theta = 0$ and the paddle that encodes the memory bit is forced by the CW rotation into a state $0 < \theta_B \ll \pi$, corresponding to the binary state 0 (Figure 5a). Conversely, if the reference bit is in state $\bar{1}$, then the engaging gate is situated at $\theta = \pi$ and the paddle is forced into a state $\pi < \theta_B \ll 2\pi$, corresponding to the binary state 1 (Figure 5b). In either case, at the end of the interaction interval the memory bit matches the reference bit ($0\bar{0}$ or $1\bar{1}$).

In this mode of operation, the ring harnesses the gravitational energy of the falling mass to decrease the randomness in the bit sequence. Specifically, $\Delta S_L = -\log 2 < 0$, since the outgoing bits are perfectly matched to the reference bits; see Eqs. 3 and 4. This decrease in the Shannon entropy of the memory bit stream is compensated by an increase in the thermodynamic entropy of the heat bath, as the energy from the falling mass is ultimately dissipated into the bath.

The model developed in Ref. [1] displayed a similar mode of operation, with a falling mass converting an incoming sequence of 0's and 1's into an outgoing sequence of 0's. We referred to this mode as *Landauer's eraser*, as it illustrated Landauer's principle that heat must be dissipated in order to erase information. We will use the same terminology to refer to the mode of operation just described for the present model, although *Landauer's copier* might be more apt in this context, since the net effect is that the preprogrammed reference sequence is copied onto the memory bits.

3.3. Dud mode

It is useful to think of a clean memory ($\delta = 1$) as a thermodynamic resource, just as a mass that has been lifted against gravity is a thermodynamic resource. The engine and eraser modes represent an interplay between these two resources, in which one resource is depleted in order to increase the other. Thus in the engine mode, the cleanness of the memory bit stream is diminished in order to raise the mass against gravity, while in the eraser mode the gravitational potential energy of the mass is spent in order to obtain a clean memory. When the incoming bit stream is sufficiently clean and the external load (mass) is sufficiently small, the ring acts as an engine, whereas when the incoming bits are disordered and the mass is large, it acts as an eraser. For intermediate values of δ and Γ, the ring might act either in the engine mode or in the eraser mode, depending on the values of other parameters such as the interaction time τ^{int} and the temperature and density of the surrounding gas.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 November 2018 doi:10.20944/preprints201811.0570.v1

Peer-reviewed version available at Entropy 2019, 21, 65; doi:10.3390/e21010065

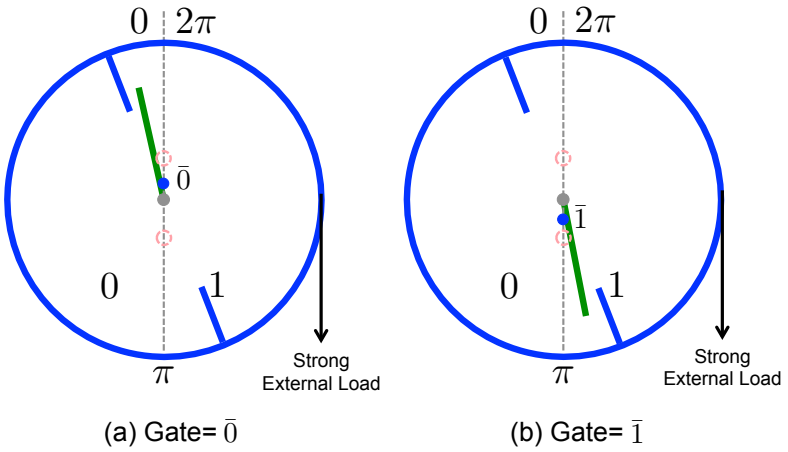


Figure 5. Eraser (Copier) Mode. Under a strong external load, CW rotation occurs until the bit becomes pinched between the engaging gate (shown as a blue dot situated on the gray dashed line) and a blade of the ring. The binary state of the memory bit then matches that of the reference bit.

There is also a third possibility: the mass drops while the disorder of the memory increases, $\Delta S_L > 0$. We call this the *dud* mode, since it represents a wasteful depletion of both thermodynamic resources. This mode arises either if the incoming memory sequence contains a surplus of mismatches over correct matches, $\delta < 0$, and the load $\Gamma > 0$ is not sufficiently strong to produce an even greater surplus of correct matches in the outgoing sequence – this is illustrated by the white area region appearing in the second quadrant in Fig. 10 – or if a surplus of correct matches in the incoming sequence is not sufficient to raise the mass against gravity, while simultaneously the load $\Gamma > 0$ is not sufficient to counter the tendency of the bits to randomize – this is illustrated by the narrow white tongue appearing in the first quadrant in Fig. 10.

In the dud mode, the Shannon entropy of the memory sequence increases, $\Delta S_L > 0$, and the thermodynamic entropy of the surrounding gas increases, as it absorbs the energy of the falling mass.

4. Numerical and analytical results

4.1. Numerical simulations

We performed numerical simulations of our contraption immersed in a dilute gas, modeling the collisions between the gas particles and the paddles and blades as Poisson processes. The probability per unit time that a gas particle strikes a particular location of a given paddle or blade is determined by the temperature T and density of the gas, the angular velocity of the paddle or blade, and the radial location of the point of collision. During a given interaction interval we simulate the dynamics of the ring and the interacting bit as a sequence of events. Each event is a blade-paddle collision, a paddle-gate collision, or a collision of a gas particle with either the paddle or the blade. After each event the angular velocity of the blade and/or paddle is appropriately updated, and the next event is generated stochastically using the Gillespie algorithm [48]. At the end of the interaction interval the machine undergoes a bit renewal, in which the old interacting bit is replaced by a new one, whose angular location θ_B and velocity $\dot{\theta}_B$ are assigned randomly according to the values of δ and T.

Fig. 6 shows eleven angular trajectories of the angular rotation of the ring, $\theta_D(t)$, illustrating the engine mode and the dud mode. The simulations were performed at temperature $k_B T = 1$ and load $\Gamma = 0.05\, k_B T$, for eleven different values of the cleanness of the incoming memory bits, δ. Each simulation lasted for 2000 interaction intervals, representing 2000 incoming bits, with $\tau^{\text{int}} = 20$. The gates were prepared in the repeating binary sequence "...0101101011...". In agreement with the arguments of Sec. 3, when δ is close to 1, the ring undergoes systematic counterclockwise rotation

and the ring performs work against the external load, lifting the mass against gravity (engine mode). For less clean incoming sequences, with values $\delta \leq 0.6$, the ring can no longer overcome the external torque and rotates clockwise (dud mode).

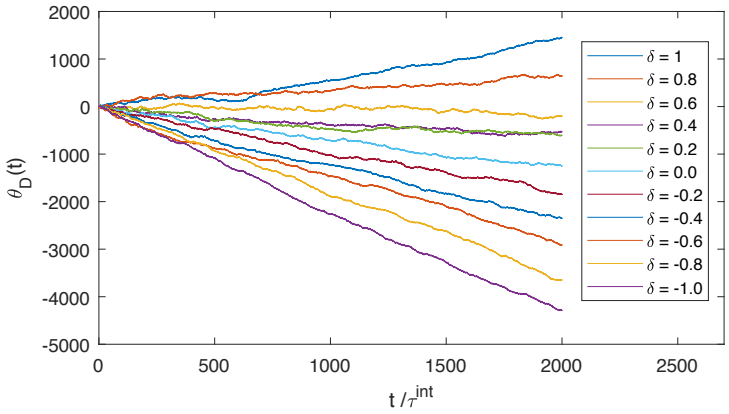


Figure 6. Trajectories of the ring's angular orientation for different values of δ at fixed load $\Gamma = 0.05\, k_B T$, with a bit renewal rate of 1 bit per 20 seconds ($\tau^{\rm int} = 20$). For $\delta = 1$ and $\delta = 0.8$ the ring performs work against the clockwise external torque, while for the other values of δ, the external load dominates and work is dissipated into the heat bath.

To illustrate the eraser mode, Fig 7 shows four trajectories simulated as in Fig. 6, except that we fix $\delta = 0.2$ and vary the external torque: $\Gamma[k_B T] = 0.1,\ 0.15,\ 0.2,\ 0.25$. As expected, the stronger the load, the faster the ring rotates in the CW direction, leading to more energy dissipated into the heat bath. We find that for $\Gamma \geq 0.15\, k_B T$, the outgoing sequence is cleaner than the incoming sequence of bits hence the ring functions as an eraser.

4.2. Analytical results for the slow-moving limit

Let us now consider the limit of long interaction time $\tau^{\rm int} \to \infty$. In this limit the behavior of the ring during one interaction interval becomes uncorrelated with its behavior in the next interval. The average work performed by the ring, W, and the Shannon entropy change of the memory tape, ΔS_L, can then be computed analytically and are given by Eqs. 7 and 11 below. We now sketch the approach that is taken to obtain these results, leaving the technical details to the Appendix.

Letting (θ_B, θ_D) denote the instantaneous configuration of the composite system – the interacting bit and the ring – we depict the relevant features of configuration space in Figure 8, with bold solid lines representing hard wall boundaries. Note that the boundary conditions depend on the state of the reference bit, $\bar{0}$ or $\bar{1}$, through the placement of the engaging gate at $\theta = 0$ or $\theta = \pi$.

During a given interaction interval, the ring and interacting bit undergo random collisions with the surrounding bath particles, while the external load imposes a potential energy contribution $\Gamma\theta_D$ that generates a CW torque on the ring. The ring and bit are confined within a single parallelogram-shaped cell in configuration space (see Fig. 8), and the composite system (θ_B, θ_D) has sufficient time to relax to equilibrium within this cell. Hence, if the composite system begins within a particular cell at the start of an interaction interval, then at the end of the interval its statistical state is given by a Boltzmann distribution restricted to that cell.

Let us suppose that during the initial interaction interval the composite system is found in one of the two shaded cells depicted in Fig. 8, depending on the state of the reference bit. Let $p^{\rm eq}_{\bar{0}}(\theta_B, \theta_D)$ and $p^{\rm eq}_{\bar{1}}(\theta_B, \theta_D)$ denote the equilibrium distributions restricted to these two cells. The correlations between θ_B and θ_D differ in these two distributions, but if we integrate either distribution over θ_B, then the resulting marginal equilibrium distributions for θ_D are identical: $p^{\rm eq}_D(\theta_D) = \int d\theta_B\, p^{\rm eq}_{\bar{0}} = \int d\theta_B\, p^{\rm eq}_{\bar{1}}$.

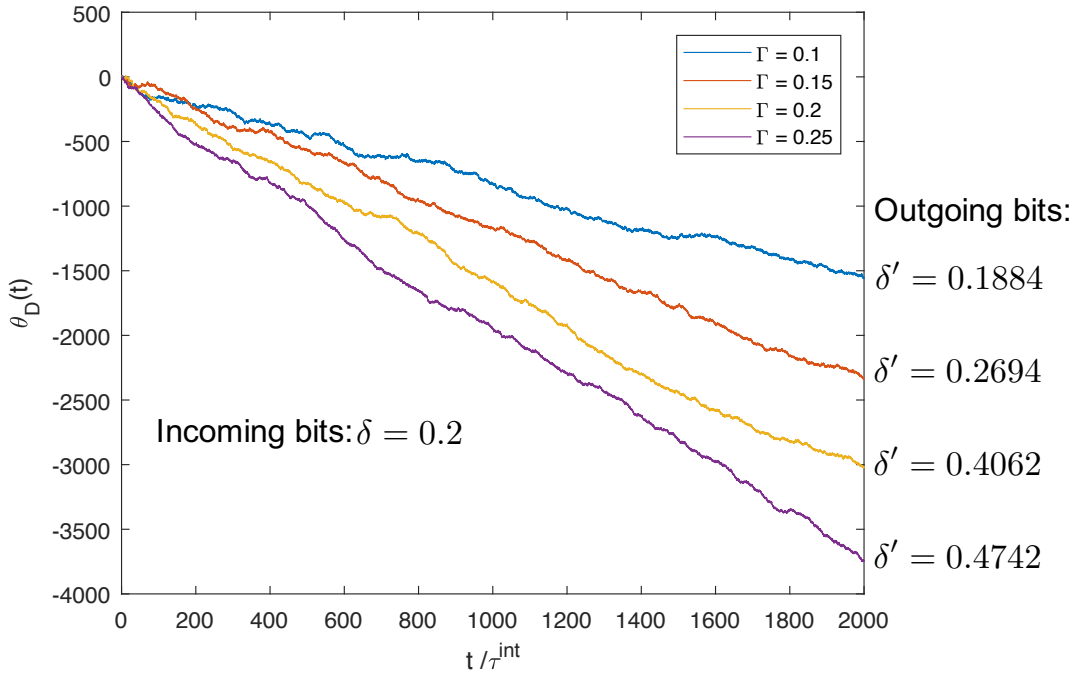
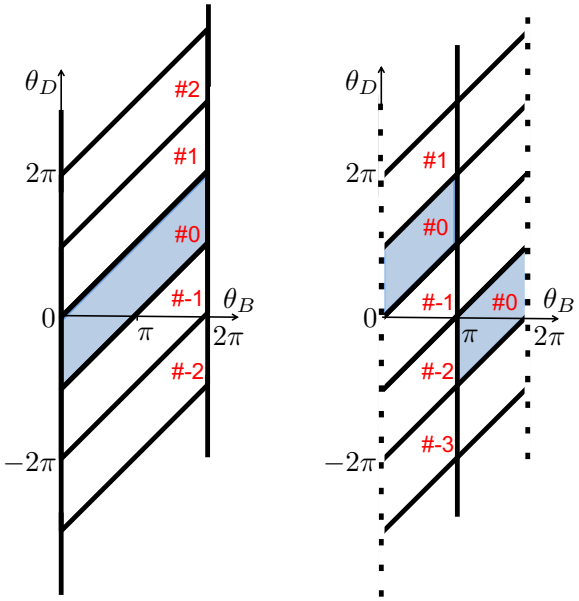


Figure 7. Trajectories of the ring's angular orientation for different values of CW external torque Γ, at fixed $\delta = 0.2$. For each trajectory, the ring rotates in the CW direction and thus the energy of the falling mass is dissipated into the heat bath. With increasing external torque ($\Gamma = 0.1k_BT$, $0.15k_BT$, $0.2k_BT$, $0.25k_BT$), the cleanness of the outgoing sequence of bits increases as well: $\delta' = 0.1884$, 0.2694, 0.4062, 0.4742. For $\Gamma \geq 0.15k_BT$ we obtain $\delta' > 0.2$ hence the ring acts as an eraser, removing randomness from the incoming sequence.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 November 2018 doi:10.20944/preprints201811.0570.v1

Peer-reviewed version available at Entropy 2019, 21, 65; doi:10.3390/e21010065



(a) Gate $= \bar{0}$ (b) Gate $= \bar{1}$

Figure 8. The configuration space of the ring and interacting bit. The tilted lines at $\theta_D - \theta_B = n\pi$ depict hard boundaries associated with a collision between the interacting bit paddle and either blade of the ring. The vertical solid lines correspond to the location of the engaging gate that blocks the paddle. This gate is located at $\theta = 0 = 2\pi$ when the reference bit is set to $\bar{0}$, panel (a), or at $\theta = \pi$ when the reference bit is set to $\bar{1}$, panel (b). The dashed lines in (b) represent periodic boundary conditions. The hard wall boundaries partition the configuration space into parallelogram-shaped *cells*, which are numbered as shown, with cell #0 shaded in each panel.

The distribution $p_D^{\rm eq}$ has support in the region $-\pi \leq \theta_D \leq 2\pi$. In the absence of an external load, both $p_{\bar{0}}^{\rm eq}$ and $p_{\bar{1}}^{\rm eq}$ are uniform distributions within the shaded regions, and $p_D^{\rm eq}(\theta_D)$ has the shape of an isosceles trapezoid. In the opposite limit of a strong external load $\Gamma \gg k_B T$, $p_D^{\rm eq}(\theta_D)$ is strongly concentrated near $\theta_D = -\pi$ (due to the Boltzmann factor $e^{-\beta\Gamma\theta_D}$), as the memory bit paddle becomes pinched between one of the ring's blades and the engaging gate.

At the start of the next interaction interval, the memory and reference bits are replaced, or *renewed*, by the arrival of a new paddle and engaging gate. The location of the engaging gate now reflects the new reference bit, $\bar{0}$ or $\bar{1}$. The state of the new memory bit, b, either matches or mismatches the reference bit, with a probability determined by the value of δ. We can treat the configuration of the incoming memory bit as a random, uniform sample either from the range $0 \leq \theta_B < \pi$ if $b = 0$, or from $\pi \leq \theta_B < 2\pi$ if $b = 1$. This renewal process instantaneously maps the final distribution of the composite system at the end of one interaction interval, into a new initial distribution at the beginning of the next interval, as the variable θ_B now refers to the new memory bit rather than the old one. This mapping depends on the state of the new bit, as illustrated in Figure 9. At the start of a new interaction interval, the bit and ring configurations, θ_B and θ_D, are uncorrelated.

If the machine (bit + ring) is found in cell #k during one interaction interval, and if the new, incoming memory and reference bits are correctly matched, then during the next interval it will be found in one of four possible cells, corresponding to a displacement $\Delta k = -1, 0, 1$ or 2, as illustrated in Figs. 8 and 9 for $k = 0$. The probability distribution for Δk is determined by considering how the equilibrium distribution restricted to the initial cell (#k) is redistributed by the mapping that occurs upon bit renewal. By similar arguments, if the incoming memory and reference bits are mismatched, then the displacement is $\Delta k = -2, -1, 0$ or 1.

The process then repeats itself over the next interaction interval: the probability distribution relaxes to equilibrium within each cell, and then renewal occurs when the new memory and reference

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 November 2018 doi:10.20944/preprints201811.0570.v1

Peer-reviewed version available at Entropy 2019, 21, 65; doi:10.3390/e21010065

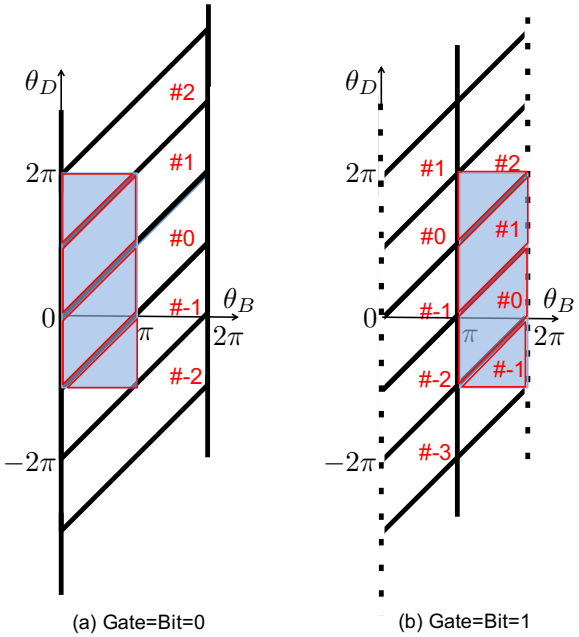


Figure 9. The shaded regions indicate the distribution of the composite system right after renewal, for the case when the memory bit is correctly matched with the reference bit. For purpose of illustration, we assume that just before the renewal the system was found in either one of the shaded cells shown in Fig. 8, both corresponding to #0. In (a), the new memory and reference bits are in the combined state $(0\bar{0})$, whereas (b) corresponds to the combination $(1\bar{1})$. The marginal probability distribution of the ring's angle, $p_D^{\rm eq}(\theta_D)$, is unaffected by the renewal mapping.

bits arrive. Thus, from one interaction interval to the next, we can treat the dynamics of the ring as a discrete time random walk along a lattice of cells, with each step Δk sampled randomly from a distribution that depends on whether the incoming memory and reference bits are matched or mismatched. The net result is that Δk can range from -2 to $+2$, with probabilities determined by the values of δ and Γ. On average, each positive (negative) unit increment in k corresponds to CCW (CW) rotation of the ring by half a circle.

Following the considerations discussed above, we have computed the probability distribution for Δk analytically, and from these results we have determined the average work performed by the ring, per interaction interval (see Appendix A for details):

$$W = \frac{\pi\beta\Gamma\delta - \pi\beta\Gamma\left[3\coth(\pi\beta\Gamma) + \operatorname{csch}(\pi\beta\Gamma)\right] + 4}{2\beta} \tag{7}$$

In the limit of a weak external load, Eq. 7 gives

$$W \approx \delta\pi\Gamma/2 \qquad (0 < \Gamma \ll k_B T) \tag{8}$$

and the ring acts as an engine when $\delta > 0$, in agreement with the discussion in Sec. 2.3. In the opposite limit of strong external load we get

$$W \approx (\delta - 3)\pi\Gamma/2 \qquad (\Gamma \gg k_B T) \tag{9}$$

hence $W < 0$, as expected. As a consistency check on Eq. 7, both of the limiting cases represented by Eqs. 8 and 9 can be verified by directly calculating the average displacement of θ_D per period, resulting from the renewal mapping illustrated in Figs. 8 and 9.

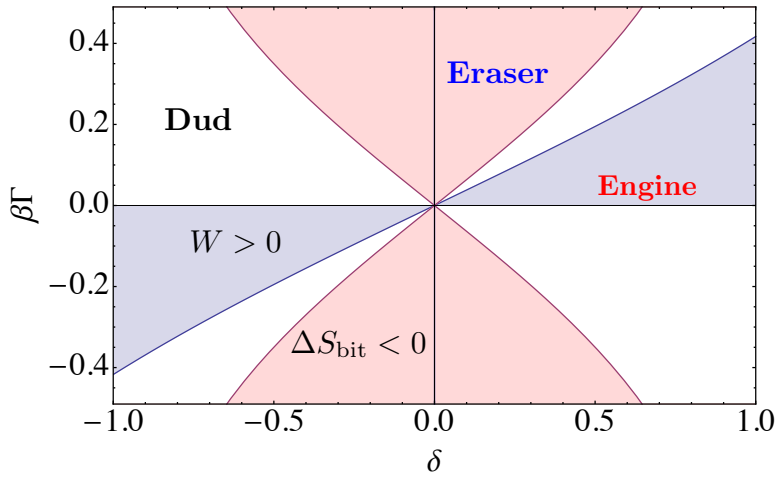


Figure 10. The phase diagram of the programmable Maxwell's demon in the limit $\tau^{\text{int}} \to \infty$. Here the behavior of the ring depends only on the sequence cleanness, δ, and the external torque scaled by bath temperature, $\beta\Gamma$. For finite values of τ^{int}, the behavior of the ring depends separately on three quantities β, Γ and δ.

Additionally, we can compute the fractions of bit-gate agreement and disagreement in the outgoing tape:

$$P_{\text{out}}(\text{same}) = \frac{e^{\Gamma\pi\beta}}{e^{\Gamma\pi\beta}+1} \quad , \quad P_{\text{out}}(\text{diff}) = \frac{1}{e^{\Gamma\pi\beta}+1}. \tag{10}$$

In the limit of a strong external load ($\Gamma \gg k_B T$) virtually all outgoing bits will be forced to match the reference bits, as each bit paddle becomes pinched between then ring's blade and the engaging gate; see Sec. 3.2. Per interaction period, the change of the Shannon entropy of the memory tape with respect to the gate is

$$\begin{aligned}\Delta S_L = &\frac{1-\delta}{2}\log\left(\frac{1-\delta}{2}\right) + \frac{1+\delta}{2}\log\left(\frac{1+\delta}{2}\right) \\ &- \frac{e^{\pi\beta\Gamma}}{e^{\pi\beta\Gamma}+1}\log\left(\frac{e^{\pi\beta\Gamma}}{e^{\pi\beta\Gamma}+1}\right) - \frac{1}{e^{\pi\beta\Gamma}+1}\log\left(\frac{1}{e^{\pi\beta\Gamma}+1}\right)\end{aligned} \tag{11}$$

where recall that the variable $L = B$ Exy G is the Boolean equality between the state of the bit and the state of the gate (see Sec. 2.2).

Combining Eqs. 7 and 11, we obtain (see Appendix B for details)

$$\Delta S_L - \frac{W}{k_B T} = D_{KL}\left[P_{in}|P_{out}\right] + \frac{\pi\beta\Gamma}{\tanh(\pi\beta\Gamma/2)} - 2 \tag{12}$$

where $D_{KL} \geq 0$ is the Kullback-Leibler divergence [49] between the incoming and outgoing bit distributions. Since $x/\tanh(x) > 1$ for all $x \neq 0$, Eq. 12 implies

$$k_B \Delta S_L - \frac{W}{T} \geq 0 \tag{13}$$

which is a strict inequality when $\Gamma \neq 0$. Because the work W is equal to the average energy extracted from the heat bath, per bit, the term $-W/T$ represents the net change in the thermodynamic entropy of bath. As a result, Eq. 13 can be viewed as a statement of the second law of thermodynamics: the sum of the entropy changes of the bit stream and heat bath must be non-negative. Notice that this interpretation relies on treating the information content of the bit stream (multiplied by k_B) as a genuine thermodynamic entropy, on par with the Clausius entropy.

Eq. 13 suggests natural definitions of the machine's thermodynamic efficiency in both the engine and the eraser mode. When the ring functions as an eraser, we have

$$W < k_B T \Delta S_L < 0 \tag{14}$$

and the efficiency is defined as

$$\eta_{\text{eraser}} = \frac{k_B T \Delta S_L}{W} < 1 \tag{15}$$

When the ring functions as an engine,

$$k_B T \Delta S_L > W > 0 \tag{16}$$

and the efficiency is defined as

$$\eta_{\text{engine}} = \frac{W}{k_B T \Delta S_L} < 1 \tag{17}$$

When the ring functions in the dud mode, $W < 0 < k_B T \Delta S_L$.

In Fig. 11 we plot the thermodynamic efficiency over the phase diagram of the machine. By definition $\eta > 0$ within the regions corresponding to the engine and eraser modes, but η drops to zero at the boundaries of these regions, where the ring becomes a dud. For example, a point on the boundary of the engine mode, with $\delta, \beta\Gamma > 0$, represents a stalled state. Here the ring generates just enough CCW torque to match the CW torque exerted by the external load (hence $W = 0$), nevertheless there is a positive rate of entropy generation in the bit stream ($\Delta S_L > 0$). If the load Γ is decreased by a small amount, then the ring will produce a slight CCW rotation, resulting in an engine with very low efficiency.

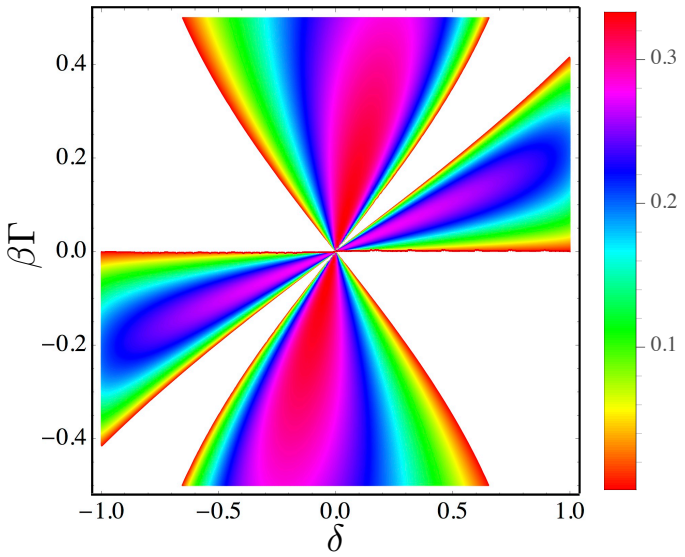


Figure 11. Efficiency plot of the programmable demon, obtained analytically in the limit $\tau^{\text{int}} \to \infty$. Since efficiency is defined only for the eraser and engine modes, the dud region is left blank.

4.3. Second law of thermodynamics in the slow moving limit

We have obtained Eq. 13 from our exact solution of the dynamics in the slow-moving limit, but the result has the character of a generalized, information-theoretic second law of thermodynamics (as already mentioned), and its validity may extend to finite values of τ^{int}. While it is difficult to establish this validity from first principles, we can make some progress by ignoring correlations (of any sort)

from one interval to the next, as we do in the following statistical treatment in which the variables B and G are treated as *information-bearing degrees of freedom* [50].

At the start of an interaction interval, let $P^{in}_{BG}(b,g)$ denote the joint probability to find the memory bit in state $b \in \{0,1\}$ and the reference gate in state $g \in \{\bar{0},\bar{1}\}$, and let $P^{in}_{B}(b)$ and $P^{in}_{G}(g)$ denote the corresponding marginal distributions. Let S_{BG}, S_B and S_G denote the Shannon entropies of these distributions.

Then

$$S^{in}_{BG} = S^{in}_{B} + S^{in}_{G} - I^{in}_{BG} \tag{18}$$

where

$$I^{in}_{BG} = \sum_{b,g} P^{in}_{BG} \log \frac{P^{in}_{BG}}{P^{in}_{B} P^{in}_{G}} \geq 0 \tag{19}$$

is the *mutual information* [51] between the bit and gate states. Defining similar quantities for the outgoing states, the net change in the combined entropy over one interaction interval is

$$\begin{aligned} \Delta S_{BG} &= \Delta S_B + \Delta S_G - \Delta I_{BG} \\ &= \Delta S_B - \Delta I_{BG} \end{aligned} \tag{20}$$

where $\Delta S_{BG} = S^{out}_{BG} - S^{in}_{BG}$, etc. Since the state of the gate remains fixed we have $\Delta S_G = 0$, whereas both S_B and I_{BG} typically change during the interaction interval.

We have used the variables B and G to specify the combined state of a memory and reference bit, but we could equally well specify this state using the variables L and G, leading to

$$\Delta S_{BG} = \Delta S_{LG} = \Delta S_L - \Delta I_{LG} \tag{21}$$

where ΔS_{LG}, ΔS_L and I_{LG} are defined as above, but with L in place of B.

The Hamiltonian analysis of Ref. [28] (see in particular Eq. 47 therein) suggests that the change in the Shannon entropy of the information-bearing degrees of freedom B and G obeys a generalized second law of thermodynamics: $W/k_BT \leq \Delta S_{BG}$. Combining with Eq. 21 gives us

$$\frac{W}{k_B T} \leq \Delta S_L - \Delta I_{LG} = \Delta S_L - I^{out}_{LG} \tag{22}$$

Here we have used our assumption that incoming mismatches are statistical uncorrelated with the state of the gate (Sec. 2.2) to set $I^{in}_{LG} = 0$. Since mutual information is non-negative, Eq. 22 immediately implies Eq. 13, but note that Eq. 22 provides a stronger bound than Eq. 13. In effect, if correlations develop between the reference gate G and the logical state L, then these correlations represent an "unused" information-thermodynamic resource. In the slow-moving limit, these correlations vanish since the demon and bit fully equilibrate, hence Eq. 22 reduces to Eq. 13 in that limit.

5. Our machine as a feedback controlled device

In previous sections we have presented our model as an autonomous system, whose various components (paddles, gas particles, etc.) evolve without external interference. With a slight modification our model can serve to illustrate a *non-autonomous* device: a machine that is manipulated via measurement and feedback. In this non-autonomous interpretation, the ring can again operate as an engine that lifts a mass against gravity, as we describe in Sec. 5.1. We then show how the inequality given by Eq. 22 for the autnomous case, can be translated into an inequality that applies to non-autonomous measurement and feedback, Eq. 26 below. Finally, in Sec. 5.3 we show how our model can be modified to act as a non-autonomous device that uses the energy of a dropping mass to write a desired target sequence to a stream of bits.

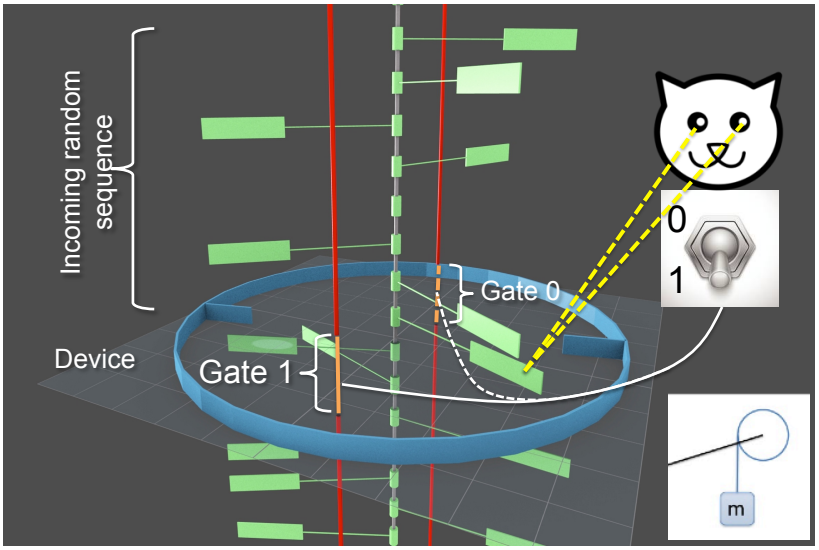


Figure 12. A non-autonomous version of our model. The snapshot is taken right at the time of bit renewal. The agent observes the state of the new bit (state 1) and simultaneously sets the gate $\bar{1}$ to be effective. Thus the new bit can switch state only through the unblocked gate $\bar{0}$. In this illustration, the agent's measurement is faithful and thus the ring is able to work in the engine mode.

5.1. Feedback controlled engine

Consider a setup that is essentially the same as that described in Sec. 2, but without the sequence of rigid reference gates (the blue L-shaped bars in Fig. 3). In their place is a single, switchable gate that can be set to block either one of the two gaps (in the red blocking bars) positioned at the vertical location of the ring. We will say that the gate is in the $\bar{0}$ state when it blocks the gap at $\theta = 0$, and in the $\bar{1}$ state when it blocks the gap at $\theta = \pi$; the latter case is depicted in Fig. 12.

Throughout this section we assume that the incoming bits arrive in a fully randomized sequence, with 0's and 1's distributed equally. We introduce an external agent who performs measurement and feedback on these bits; see Fig. 12. The agent observes each new bit as it arrives, and at the moment of bit renewal (when the incoming bit becomes the interacting bit) the agent sets the switchable gate accordingly: if it observes the incoming bit to be in state 0 (or 1), it sets the switchable gate to state $\bar{0}$ (or $\bar{1}$).

If the agent performs error-free measurements, faithfully identifying the state of each incoming bit, then from the perspective of the ring the situation is equivalent to the case $\delta = 1$ analyzed in Sec. 2. Namely, the blocked gate is matched with the state of the incoming bit so as to produce, during each interaction interval, a statistical bias in favor of CCW rotation. In the long run, this bias can cause a small mass to be lifted against gravity, systematically extracting energy from the heat bath and thereby reducing its entropy. Since (by assumption) the incoming bits arrive in a fully randomized sequence, the decrease in the entropy of the bath cannot be "paid for" by increasing the Shannon entropy of the bits. Rather, the model illustrates how an external agent, by performing measurement and feedback, can rectify fluctuations to produce an apparent violation of the second law of thermodynamics. Of course there is no real violation, as the physical nature of the agent is not being taken into account – like Maxwell, we have effectively inserted a "magical creature" into our model.

We further generalize this scenario to include the possibility of measurement errors. For each incoming bit, let ϵ denote the probability that the agent misidentifies the bit state and therefore blocks the "wrong" gate. This situation is equivalent to the one analyzed in Sec. 3.1, with $\delta = 1 - 2\epsilon$. For sufficiently small error rate ϵ and load Γ, the machine may still lift the mass against gravity, despite the measurement errors.

The non-autonomous model described in this section is similar to Maxwell's original thought experiment, and even more so to the Szilard engine [33], in which an agent determines whether a gas particle is within the left or right half of a box, then appropriately attaches a mass that can be lifted by the expansion of the single-particle gas. In our model, the "expansion" of a bit paddle from the half-circle to the full circle during each interaction interval plays the role of the expansion of the single-particle gas in the Szilard model. Note, however, that in the case of the Szilard engine the same gas particle is recycled from one iteration of the measurement-and-feedback process to the next, whereas our model uses a sequence of "gas particles" (incoming bits) that can act as a memory register. This allows our model to act not only as an engine but also as a device that writes information, as we discuss in Sec. 5.3.

5.2. *The second law of thermodynamics with feedback control*

We have noted the equivalence between the measurement-and-feedback scenario described in Sec. 5.1 (with error rate ϵ) and the autonomous engine of Sec. 3.1 (with $\delta = 1 - 2\epsilon$). Let us use this equivalence to obtain a second law inequality for the measurement-and-feedback process.

As before, let $P^{in}_{BG}(b, g)$ denote the joint probability distribution describing the initial state of the bit and blocked gate – just after the agent has measured the bit and set the gate accordingly. During the interaction interval, $0 < t < \tau^{\text{int}}$, the machine operates autonomously, hence (see Sec. 4.3)

$$\frac{W}{k_B T} \leq \Delta S_{BG} = \Delta S_G + \Delta S_B - \Delta I_{BG} \tag{23}$$

Since the gate state G does not change during the interaction interval, $\Delta S_G = 0$. Also, since the fully randomized incoming bit stream contains equal populations of 0's and 1's, the same will be true (by symmetry) of the outgoing bit stream, hence $S^{in}_B = S^{out}_B = \log 2$, and $\Delta S_B = 0$. We thus get

$$\frac{W}{k_B T} \leq -\Delta I_{BG} = I^{in}_{BG} - I^{out}_{BG} \tag{24}$$

The initial mutual information is simply the information gained by the measurement process:

$$I^{in}_{BG} = I_{meas} = \log 2 + (1-\epsilon)\log(1-\epsilon) + \epsilon \log \epsilon \tag{25}$$

The final mutual information quantifies the degree to which B and G remain correlated at the end of the interval; we will refer to this value as the *residual* information: $I_{res} = I^{out}_{BG}$. We thus have

$$\frac{W}{k_B T} \leq I_{meas} - I_{res}, \tag{26}$$

i.e. the extracted work W is bounded by the amount of information gathered during the measurement, minus the amount "left over" at the end of the interval. Hence the gathered information is a thermodynamic resource, and the difference $I_{meas} - I_{res}$ represents the amount of that resource that is consumed, per interaction interval. Since $I_{res} \geq 0$, Eq. 26 immediately implies the weaker bound

$$\frac{W}{k_B T} \leq I_{meas}. \tag{27}$$

Eq. 27 was originally derived within the framework of stochastic thermodynamics by Sagawa and Ueda in Refs. [26,27], and Eq. 26 was subsequently obtained by the same authors in Refs. [29,52]. We also note that the net change in the mutual information between the bit and the gate, ΔI_{BG}, can be interpreted as the integrated *information flow*, within the bipartite approach developed by Horowitz and Esposito [30]. This information flow is negative (hence $I_{meas} - I_{res} > 0$), as information is consumed to extract energy to lift the mass.

5.3. Feedback controlled information recorder

In the eraser mode discussed in Sec. 3.2, our autonomous machine removes randomness from the incoming bit stream, replacing it with a preprogrammed sequence encoded in the reference gates. In the present context of an externally manipulated machine, let us imagine that the agent desires to write a particular target sequence, e.g. $011010\cdots$, to the bit stream. The agent does not perform measurements on the incoming bits, but as each bit arrives the agent sets the switchable gate to match the corresponding element of the target sequence. Then, as in Sec. 3.2, the CW torque produced by the gravitational pull of the mass produces a tendency to set the state of the interacting bit to match the desired target value, through the "pinching" mechanism illustrate in Fig. 5. The fidelity of the writing process increases with the torque Γ generated by the gravitational force on the mass, and the energy of the dropping mass is dissipated into the heat bath.

6. Concluding remarks

In this paper we have presented a model of a programmable, mechanical Maxwell's demon, that can be interpreted either as an autonomous device, as described in Sections 2 - 4, or as a non-autonomous device manipulated by external measurement and feedback control, as in Section 5. For these distinct interpretations, we have obtained distinct forms of the second law of thermodynamics, represented by Eqs. 22 and 26. While these results have been obtained within the specific context of our model, it would be useful to investigate whether they point to more general thermodynamic laws, in situations involving both autonomous and non-autonomous (i.e. feedback-controlled) devices. For instance, as indicated in Sec. 5.2, the inequalities given by Eqs. 26 and 27 have been obtained previously under assumptions of bipartite, Markovian dynamics [26,27,29,30,52]. By contrast, we have obtained these results within a Newtonian model of colliding particles and paddles, which suggests that they might be derived more generally within a classical, Hamiltonian framework.

Additionally, we have obtained analytical results for the work delivered by our device, Eq. 7, and the change in the Shannon entropy of the bits, Eq. 11, in the limit of a slowly-moving stream of bits, $\tau^{\rm int} \to \infty$. For finite $\tau^{\rm int}$, the interactions between the bits and the demon may induce statistical correlations among the outgoing bits. Such correlations, which could in principle act as a thermodynamic resource, have not been considered in our analysis. It would be interesting to investigate how these correlations might affect the inequalities that we have derived.

Acknowledgments: We gratefully acknowledge financial support from the U.S. Army Research Office under contract number W911NF-13-1-0390 (CJ). ZL acknowledge useful discussions with Zhixin Lu, Dibyendu Mandal, Sebastian Deffner, and David Sivak.

Appendix Work delivered per interaction interval.

To compute the mean work that our device delivers to life a mass against gravity, we must compute the mean angular displacement of the ring per interaction interval. This displacement can be determined by considering the transition from the end of an interaction interval (see Fig. 8) to the beginning of the next interaction interval (see Fig. 9). Recall that during a particular interaction interval, the composite system is confined within a single cell in (θ_B, θ_D)-space, as illustrated by the shaded region in Fig. 8(a). At the moment of bit renewal the value of θ_B changes suddenly, hence the system may find itself in a different cell immediately after bit renewal, as illustrated in Fig. 9(a). We characterize this transition by the change Δk in the cell index. At a coarse-grained level the evolution of the system from one interval to the next becomes a random walk of discrete jumps in k-space.

Here we compute the probability of the jumps conditioned on the agreement or disagreement between the state of the bit and its corresponding gate. At the moment of bit renewal, the state of the system can either remain in the same cell ($\Delta k = 0$), it can jump up or down by one cell ($\Delta k = \pm 1$), or up or down by two cells ($\Delta k = \pm 2$). Note however that the value $\Delta k = +2$ is possible only if the incoming bit matches the incoming gate (as in Fig. 9), and the value $\Delta k = -2$ can occur only if the bit and gate are mismatched. To illustrate how to compute the probabilities of these various events, let us imagine

that immediately after renewal, both the incoming bit and its reference gate are in the 0 state. Then the probability distribution for the system is partitioned among the four shaded regions appearing in Fig. 9(a), each of which corresponds to a particular value of Δk, and the distribution within each of these cells is inherited from the equilibrium distribution just prior to bit renewal (Fig. 8). By integrating over the distribution within each region, and then summing over all possible combinations of incoming bit and gate, we obtain the following results.

When the incoming bit agrees with its gate (e.g. bit=1 and gate=$\bar{1}$ or bit=0 and gate=$\bar{0}$), we have

$$P^{\text{same}}_{+2} \equiv P^{0,\bar{0}}_{+2} = P^{1,\bar{1}}_{+2} = \frac{\pi\beta\Gamma + e^{\pi\beta\Gamma}(\pi\beta\Gamma - 2) + 2}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{A1}$$

$$P^{\text{same}}_{+1} \equiv P^{0,\bar{0}}_{+1} = P^{1,\bar{1}}_{+1} = \frac{e^{\pi\beta\Gamma}\left[\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 3\right) + 2\right] - 2}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{A2}$$

$$P^{\text{same}}_{\text{stay}} \equiv P^{0,\bar{0}}_{\text{stay}} = P^{1,\bar{1}}_{\text{stay}} = \frac{e^{\pi\beta\Gamma}\left[\pi\beta\Gamma + e^{\pi\beta\Gamma}\left(-3\pi\beta\Gamma + 2e^{\pi\beta\Gamma} - 2\right)\right]}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{A3}$$

$$P^{\text{same}}_{-1} \equiv P^{0,\bar{0}}_{-1} = P^{1,\bar{1}}_{-1} = \frac{e^{2\pi\beta\Gamma}\left[\pi\beta\Gamma + e^{\pi\beta\Gamma}(\pi\beta\Gamma - 2) + 2\right]}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{A4}$$

$$P^{\text{same}}_{-2} \equiv P^{0,\bar{0}}_{-2} = P^{1,\bar{1}}_{-2} = 0 \tag{A5}$$

If the incoming bit mismatches its gate, we find

$$P^{\text{diff}}_{+2} \equiv P^{0,\bar{1}}_{+2} = P^{1,\bar{0}}_{+2} = 0 \tag{A6}$$

$$P^{\text{diff}}_{+1} \equiv P^{0,\bar{1}}_{+1} = P^{1,\bar{0}}_{+1} = \frac{\pi\beta\Gamma + e^{\pi\beta\Gamma}(\pi\beta\Gamma - 2) + 2}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{A7}$$

$$P^{\text{diff}}_{\text{stay}} \equiv P^{0,\bar{1}}_{\text{stay}} = P^{1,\bar{0}}_{\text{stay}} = \frac{e^{\pi\beta\Gamma}\left[\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 3\right) + 2\right] - 2}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{A8}$$

$$P^{\text{diff}}_{-1} \equiv P^{0,\bar{1}}_{-1} = P^{1,\bar{0}}_{-1} = \frac{e^{\pi\beta\Gamma}\left[\pi\beta\Gamma + e^{\pi\beta\Gamma}\left(-3\pi\beta\Gamma + 2e^{\pi\beta\Gamma} - 2\right)\right]}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{A9}$$

$$P^{\text{diff}}_{-2} \equiv P^{0,\bar{1}}_{-2} = P^{1,\bar{0}}_{-2} = \frac{e^{2\pi\beta\Gamma}\left[\pi\beta\Gamma + e^{\pi\beta\Gamma}(\pi\beta\Gamma - 2) + 2\right]}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{A10}$$

The net probability to take a step Δk at the moment of bit renewal is then given by

$$R_{\Delta k} = P_{\text{in}}(\text{same}) \cdot P^{\text{same}}_{\Delta k} + P_{\text{in}}(\text{diff}) \cdot P^{\text{diff}}_{\Delta k} = \frac{1+\delta}{2} P^{\text{same}}_{\Delta k} + \frac{1-\delta}{2} P^{\text{diff}}_{\Delta k} \tag{A11}$$

Over a long observation time, i.e. many interaction intervals, these discrete jumps in k-space produce average rotations by amounts $\Delta\theta_D = -2\pi,\ -\pi,\ 0,\ \pi,\ 2\pi$, per time interval τ^{bit}. The average work delivered per interval is simply the average rotation of the demon per interval multiplied, by the external torque:

$$\begin{aligned} W &= 2\pi\Gamma \cdot R_{+2} + \pi\Gamma \cdot R_{+1} - \pi\Gamma \cdot R_{-1} - 2\pi\Gamma \cdot R_{-2} \\ &= \frac{\pi\beta\Gamma\delta - \pi\beta\Gamma\left[3\coth(\pi\beta\Gamma) + \operatorname{csch}(\pi\beta\Gamma)\right] + 4}{2\beta} \end{aligned} \tag{A12}$$

Appendix Compatibility with the second law of thermodynamics.

Here we show that our analytical solution of the programmable demon in the slow-moving limit obeys the second law of thermodynamics.

For convenience, define $x \equiv \delta$ and $y \equiv (e^{\lambda}-1)/(e^{\lambda}+1)$, where $\lambda \equiv \pi\beta\Gamma$. Then Eq. 11 becomes

$$\Delta S_L = \frac{1-x}{2}\ln\frac{1-x}{2} + \frac{1+x}{2}\ln\frac{1+x}{2} - \left(\frac{1-y}{2}\ln\frac{1-y}{2} + \frac{1+y}{2}\ln\frac{1+y}{2}\right) \tag{A13}$$

and the dimensionless work done per bit (Eq. 7) is

$$\beta W = \frac{\lambda x}{2} - \frac{\lambda}{2}\left[3\coth(\lambda) + \operatorname{csch}(\lambda)\right] + 2 \tag{A14}$$

$$= \frac{x}{2}\ln\frac{1+y}{1-y} + 2 - \frac{1}{2}\left(\frac{2}{y} + y\right)\ln\frac{1+y}{1-y} \tag{A15}$$

Taking the difference, we get

$$\Delta S_L - \beta W = -2 + \frac{1}{y}\ln\frac{1+y}{1-y} + \frac{x}{2}\ln\frac{(1+x)(1-y)}{(1-x)(1+y)} + \frac{1}{2}\ln\frac{(1+x)(1-x)}{(1-y)(1+y)} \tag{A16}$$

$$= \frac{x+1}{2}\ln\frac{(1+x)/2}{(1+y)/2} + \frac{1-x}{2}\ln\frac{(1-x)/2}{(1-y)/2} + \frac{1}{y}\ln\frac{1+y}{1-y} - 2 \tag{A17}$$

where $x \in [-1,+1]$ and $y \in (-1,+1)$. Note also that $P_{in}(\text{same/diff}) = (1 \pm x)/2$ (see Eq. 1) and $P_{out}(\text{same/diff}) = (1 \pm y)/2$ (see Eq. 10), which allows us to rewrite the above expression as

$$\Delta S_L - \beta W = D_{KL}\left[P_{in}|P_{out}\right] + \frac{1}{y}\ln\frac{1+y}{1-y} - 2 \tag{A18}$$

where $D_{KL} \geq 0$ is the Kullback-Leibler divergence [49] between the incoming and outgoing bit distributions.

Next, we show that

$$\frac{1}{y}\ln\frac{1+y}{1-y} - 2 \geq 0 \tag{A19}$$

by expanding the logarithm as an infinite series:

$$\frac{1}{y}\ln\frac{1+y}{1-y} - 2 = \frac{1}{y}\left(\sum_{n=0}^{\infty}\frac{(-1)^n y^{n+1}}{n+1} + \sum_{n=0}^{\infty}\frac{y^{n+1}}{n+1}\right) - 2 \tag{A20}$$

$$= \frac{1}{y}\sum_{n=0}^{\infty}\frac{2y^{2n+1}}{2n+1} - 2 \tag{A21}$$

$$= \sum_{n=1}^{\infty}\frac{2y^{2n}}{2n+1} \geq 0 \tag{A22}$$

where the equality is achieved when $y = 0$. Alternatively, we can rewrite the left side in terms of λ:

$$\frac{1}{y}\ln\frac{1+y}{1-y} - 2 = \frac{\lambda}{\tanh(\lambda/2)} - 2 = 2\left(\frac{\lambda/2}{\tanh(\lambda/2)} - 1\right) \geq 0 \tag{A23}$$

where the last inequality follows since $|a| \geq |\tanh a|$ for any real a.

We thus confirm that in the slow-moving limit, our Maxwell's demon satisfies $\Delta S_L - \beta W \geq 0$, where the equality is achieved only when the external force is absent and the incoming sequence is totally random.

1. Lu, Z.; Mandal, D.; Jarzynski, C. Engineering Maxwell's demon. *Phys. Today* **2014**, *67*, 60–61.
2. Maxwell, J.C. *The Scientific Letters and Papers of James Clerk Maxwell*; Vol. 2: 1862-1873, Cambridge University Press: Cambridge, 1995.
3. Smoluchowski, M. Experimentell nachweisbare, der üblichen Thermodynamik widersprechende Molekularphänomene [Experimentally verifiable phenomena of molecules contradicting usual thermodynamics]. *Physikalische Zeitschrift (German)* **1912**, *13*, 1069–1080.
4. Feynman, R.P.; Leighton, R.B.; Sands, M. *The Feynman Lectures on Physics*; Addison-Wesley: Reading, MA, 1966.
5. Landauer, R. Irreversibility and Heat Generation in the Computing Process. *IBM J. Res. Dev.* **1961**, *5*, 183–191.
6. Penrose, O. *Foundations of statistical mechanics*; Vol. 42, Pergamon Press: Oxford, 1970.
7. Bennett, C.H. Logical Reversibility of Computation. *IBM J. Res. Dev.* **1973**, *17*, 525–532.
8. Bennett, C.H. Dissipation-error tradeoff on proofreading. *BioSystems* **1979**, *11*, 85–91.
9. Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
10. Crooks, G.E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* **1999**, *60*, 2721.
11. Seifert, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics* **2012**, *75*, 126001.
12. Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S.B.; Tinoco Jr, I.; Bustamante, C. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* **2005**, *437*, 231.
13. Kay, E.R.; Leigh, D.A.; Zerbetto, F. Synthetic molecular motors and mechanical machines. *Angewandte Chemie International Edition* **2007**, *46*, 72–191.
14. Kim, K.H.; Qian, H. Fluctuation theorems for a molecular refrigerator. *Physical Review E* **2007**, *75*, 022102.
15. Gavrilov, M.; Chétrite, R.; Bechhoefer, J. Direct measurement of weakly nonequilibrium system entropy is consistent with Gibbs–Shannon form. *Proceedings of the National Academy of Sciences* **2017**, *114*, 11097–11102.
16. Serreli, V.; Lee, C.F.; Kay, E.R.; Leigh, D.A. A molecular information ratchet. *Nature* **2007**, *445*, 523.
17. Bérut, A.; Arakelyan, A.; Petrosyan, A.; Ciliberto, S.; Dillenschneider, R.; Lutz, E. Experimental verification of Landauer's principle linking information and thermodynamics. *Nature* **2012**, *483*, 187.
18. Koski, J.V.; Maisi, V.F.; Pekola, J.P.; Averin, D.V. Experimental realization of a Szilard engine with a single electron. *Proceedings of the National Academy of Sciences* **2014**, *111*, 13786–13789.
19. Koski, J.V.; Kutvonen, A.; Khaymovich, I.M.; Ala-Nissila, T.; Pekola, J.P. On-Chip Maxwell's Demon as an Information-Powered Refrigerator. *Phys. Rev. Lett.* **2015**, *115*, 260602.
20. Camati, P.A.; Peterson, J.P.; Batalhao, T.B.; Micadei, K.; Souza, A.M.; Sarthour, R.S.; Oliveira, I.S.; Serra, R.M. Experimental rectification of entropy production by Maxwell's demon in a quantum system. *Physical review letters* **2016**, *117*, 240502.
21. Chida, K.; Desai, S.; Nishiguchi, K.; Fujiwara, A. Power generator driven by Maxwell's demon. *Nature communications* **2017**, *8*, 15310.
22. Elouard, C.; Herrera-Martí, D.; Huard, B.; Auffèves, A. Extracting Work from Quantum Measurement in Maxwell's Demon Engines. *Physical Review Letters* **2017**, *118*, 260603.
23. Cottet, N.; Jezouin, S.; Bretheau, L.; Campagne-Ibarcq, P.; Ficheux, Q.; Anders, J.; Auffèves, A.; Azouit, R.; Rouchon, P.; Huard, B. Observing a quantum Maxwell demon at work. *Proceedings of the National Academy of Sciences* **2017**, *114*, 7561–7564.
24. Zurek, W.H. Thermodynamic cost of computation, algorithmic complexity and the information metric. *Nature* **1989**, *341*, 119.
25. Leff, H.; Rex, A.F. *Maxwell's Demon 2 Entropy, Classical and Quantum Information, Computing*; CRC Press, 2002.
26. Sagawa, T.; Ueda, M. Generalized Jarzynski equality under nonequilibrium feedback control. *Physical review letters* **2010**, *104*, 090602.
27. Sagawa, T.; Ueda, M. Nonequilibrium thermodynamics of feedback control. *Phys. Rev. E* **2012**, *85*, 021104.
28. Deffner, S.; Jarzynski, C. Information Processing and the Second Law of Thermodynamics: An Inclusive, Hamiltonian Approach. *Phys. Rev. X* **2013**, *3*, 041003.

29. Sagawa, T.; Ueda, M. Role of mutual information in entropy production under information exchanges. *New Journal of Physics* **2013**, *15*, 125012.
30. Horowitz, J.M.; Esposito, M. Thermodynamics with continuous information flow. *Physical Review X* **2014**, *4*, 031015.
31. Barato, A.; Seifert, U. Unifying three perspectives on information processing in stochastic thermodynamics. *Physical review letters* **2014**, *112*, 090601.
32. Parrondo, J.M.R.; Horowitz, J.M.; Sagawa, T. Thermodynamics of information. *Nat. Phys.* **2015**, *11*, nphys3230.
33. Szilard, L. über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Z. Physik* **1929**, *53*, 840–856.
34. Cao, F.J.; Feito, M.; Touchette, H. Information and flux in a feedback controlled Brownian ratchet. *Physica A: Statistical Mechanics and its Applications* **2009**, *388*, 113–119.
35. Vaikuntanathan, S.; Jarzynski, C. Modeling Maxwell's demon with a microcanonical Szilard engine. *Physical Review E* **2011**, *83*, 061120.
36. Kim, S.W.; Sagawa, T.; De Liberato, S.; Ueda, M. Quantum Szilard engine. *Phys. Rev. Lett.* **2011**, *106*, 070401.
37. Mandal, D.; Jarzynski, C. Work and information processing in a solvable model of Maxwell's demon. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 11641–11645.
38. Tu, Y. The nonequilibrium mechanism for ultrasensitivity in a biological switch: Sensing by Maxwell's demons. *Proceedings of the National Academy of Sciences* **2008**, *105*, 11737–11741.
39. Mehta, P.; Schwab, D.J. Energetic costs of cellular computation. *Proceedings of the National Academy of Sciences* **2012**.
40. Mandal, D.; Quan, H.T.; Jarzynski, C. Maxwell's refrigerator: an exactly solvable model. *Phys. Rev. Lett.* **2013**, *111*, 030602.
41. Deffner, S. Information-driven current in a quantum Maxwell demon. *Physical Review E* **2013**, *88*, 062128.
42. Strasberg, P.; Schaller, G.; Brandes, T.; Esposito, M. Thermodynamics of a physical model implementing a Maxwell demon. *Physical review letters* **2013**, *110*, 040601.
43. Barato, A.C.; Seifert, U. An autonomous and reversible Maxwell's demon. *EPL (Europhysics Letters)* **2013**, *101*, 60001.
44. Chapman, A.; Miyake, A. How an autonomous quantum Maxwell demon can harness correlated information. *Physical Review E* **2015**, *92*, 062125.
45. Boyd, A.B.; Mandal, D.; Crutchfield, J.P. Identifying functional thermodynamics in autonomous Maxwellian ratchets. *New Journal of Physics* **2016**, *18*, 023049.
46. Ouldridge, T.E.; Govern, C.C.; ten Wolde, P.R. Thermodynamics of computational copying in biochemical systems. *Physical Review X* **2017**, *7*, 021004.
47. Givant, S.; Halmos, P. *Introduction to Boolean algebras*; Springer Science & Business Media, 2008.
48. Gillespie, D.T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **1977**, *81*, 2340–2361.
49. Kullback, S.; Leibler, R.A. On information and sufficiency. *The annals of mathematical statistics* **1951**, *22*, 79–86.
50. Bennett, C.H. Notes on Landauer's Principle, Reversible Computation, and Maxwell's Demon. *Stud. Hist. Phil. Mod. Phys.* **2003**, *34*, 501.
51. Cover, T.M.; Thomas, J.A. *Elements of information theory*; John Wiley & Sons, 2012.
52. Sagawa, T.; Ueda, M. Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics. *Physical review letters* **2012**, *109*, 180602.