arXiv:2002.XXXXX

Variations on a Demonic Theme: Szilard's Other Engines

Kyle Ray* and James P. Crutchfield†
Complexity Sciences Center and Physics Department, University of California at Davis, One Shields Avenue, Davis, CA 95616

(Dated: March 21, 2020)

Szilard's now-famous single-molecule engine was only the first of three constructions he introduced in 1929 to resolve several paradoxes arising from Maxwell's demon. We analyze Szilard's remaining two demon models. We show that the second one, though a markedly different implementation employing a population of distinct molecular species and semi-permeable membranes, is informationally and thermodynamically equivalent to an ideal gas of the single-molecule engines. Since it is a gas of noninteracting particles one concludes, following Boyd and Crutchfield, that (i) it reduces to a chaotic dynamical system—called the Szilard Map, a composite of three piecewise linear maps that implement the thermodynamic transformations of measurement, control, and erasure; (ii) its transitory functioning as an engine that converts disorganized heat energy to work is governed by the Kolmogorov-Sinai entropy rate; (iii) the demon's minimum necessary "intelligence" for optimal functioning is given by the engine's statistical complexity, and (iv) it's functioning saturates thermodynamic bounds and so it is a minimal, optimal implementation. We show that Szilard's third model is rather different and addresses the fundamental issue, raised by the first two, of measurement in and by thermodynamic systems and entropy generation. Taken together, Szilard's suite of constructions lays out a range of possible realizations of Maxwellian demons that anticipated by almost two decades Shannon's and Wiener's concept of information as surprise and cybernetics's notion of functional information. This, in turn, gives new insight into engineering implementations of novel nanoscale information engines that leverage microscopic fluctuations and into the diversity of thermodynamic mechanisms and intrinsic computation harnessed in physical, molecular, biochemical, and biological systems.

PACS numbers: 05.45.-a 89.75.Kd 89.70.+c 05.45.Tp

Keywords: stochastic process, hidden Markov model, ϵ-machine, causal states, mutual information, information processing Second Law of Thermodynamics

I. INTRODUCTION

Since James Clerk Maxwell first proposed an intelligence that can violate the Second Law via accurate observations of individual molecules and precise control of boundary conditions, the idea has been revisited and challenged countless times. In his 1872 book on heat, Maxwell first formally introduced the seeming paradox: a "finite being" that could, in essence, capture individual thermal fluctuations to extract macroscopic amounts of work from a heat bath [1] in violation of the Second Law. Several years later, William Thomson (Lord Kelvin) dubbed these beings "Maxwell's Intelligent Demons" [2]. And so, the paradox of Maxwell's Demon was born: accurate observations and precise control can overcome the Second Law of Thermodynamics, rendering disordered heat energy into useful, ordered work. Over the following decades, many attempted resolutions [3] addressed purely mechanical limitations imposed by how a given demon acted on its observations to sort molecules.

Thomson makes this point quite explicitly in a lecture given before the Royal Institution in 1879, where he closes his abstract:

> The conception of the 'sorting demon' is merely mechanical, and is of great value in purely physical science. It was not invented to help us to deal with questions regarding the influence of life and of mind on the motions of matter, questions essentially beyond the range of mere dynamics. [4]

Thomson highlights two key distinctions made in early conceptions of the demon. First, the demon's primary task is to physically sort microscopic particles by their individual characteristics. Second, Maxwell's Demon (MD) cannot shed light on the influence of "mind" on the motion of matter. (This presumably addressed Maxwell's and others' repeated appeals to undefined notions such as "intelligent beings".)

Not until 1929, when Leo Szilard published his seminal work "On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings" [5], was a direct connection established between a thermodynamic cost and what Maxwell called intelligence—and what we now call "information".[1] In this, Szilard showed that

* kjray@ucdavis.edu
† chaos@ucdavis.edu

[1] Notably, Szilard discussed the manuscript's development with Albert Einstein [6].

both of Thomson's assertions could be relaxed. Notably, Szilard's constructions do not involve the direct manipulation of individual molecules, but always involve their observation (measurement) and control. The genius in this was to introduce a new, operational, and minimal definition of "mind" as storing information in physical states; thus, inextricably linking a demon with its physical instantiation. While Szilard acknowledged that the biological phenomena governing the working of a "finite being" were beyond the scope of physics, he delineated the minimal capabilities a mind needed to exhibit MD-like behavior and then created idealized machines with these abilities. Szilard's conclusion: if the Second Law is to hold, a physical memory's interaction with a thermodynamic system must entail entropy production.

Szilard is cited in much of the work on the subject since. A notable exception is Landauer's 1961 article [7] that, nonetheless, responds directly to Szilard's claim that measurement has an inherent entropy cost. Landauer argues that the thermodynamic cost arises instead from the demon's act to erase the measured information, necessary to "reset" itself to begin a new cycle afresh. From that point forward, and for the better part of a half century, that irreversible erasure of stored measurement information was the source of the compensating cost was taken as the resolution of Maxwell's paradox [8, 9]. Recent developments in information theory and nonequilibrium thermodynamics, though, allow for more precise accounting. The result of which is that, in the general case, measurement and erasure act as a conjugate pair—a decrease in the cost of one increases the cost of the other [10–14].

What this colorful history glosses over is that Szilard's 1929 work laid out three different constructions of thermodynamic machines. Taken together, they were his attempt to account more generally for how the flow of heat, work, and information (our modern word, not his[2]) drive each step of a thermodynamic process. In today's parlance we refer to these devices as *information engines* [14]. Since then, as history would have it, the descriptor "Szilard Engine" came to refer only to his first construction—the single-molecule engine. In light of recent experimental and theoretical developments allowing new treatments of information engines, it is pertinent to revisit Szilard's foundational work *en toto*. What additional insights can be gleaned from the other Szilard devices, if any? How do they compare to his first engine?

Below, we retrace Szilard's steps in constructing his second device and investigate his reasoning using more contemporary ideas and techniques for analyzing deterministic chaotic systems, information flow, and the energetics of nonequilibrium thermodynamic transformations. Once completed, we turn to his third construction that, as it turns out, is a novel view of the process of measurement itself, when two thermodynamic systems come into contact.

[2] "Information" appears only once and, then, in a narrative sense.

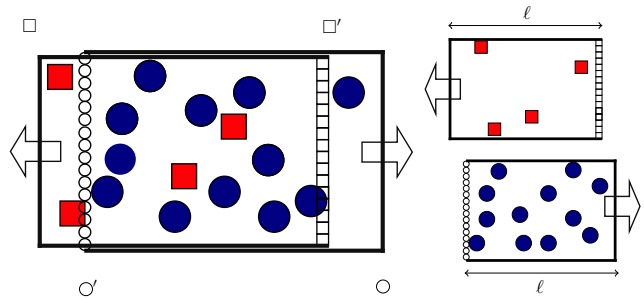

FIG. 1. Particle-type separation: (Left) Two constant-volume cylinders of length ℓ slide through each other, blue-circle particles ● are moved from the original left (L) volume to the right and red-square particles □ are unaffected. A membrane permeable to ○s (□s) is depicted as a line of squares (circles), as they are, in essence, walls for the □s (○s) only. (Right) Individual cylinders at end of the particle-separation process.

II. DEMON GAS: SZILARD'S SECOND ENGINE

Consider an ensemble of *demon-particle* molecules contained in a cylindrical tube in contact with a thermal reservoir at temperature T. Each demon-particle $i = 1, \ldots, N$ is defined by two variables: a particle-type variable $s_i \in \{\square, \bigcirc\}$ and a variable that relates to the demon's knowledge $y_i \in \{0, 1\}$ about the particle type. Demon i "knows" its molecule's type when y_i's value exactly correlates that of s_i. We refer to y_i as demon i's *memory*.

Particles spontaneously convert "monomolecularly"—Szilard's phrasing—from one type to the other at a given rate. This rate is chosen to maintain a particular desired equilibrium distribution $\rho_0(s)$ in which the probability of being one type is given by $\Pr(s_i = \square) = \delta$ and the other $\Pr(s_i = \bigcirc) = 1 - \delta$. Total particle number N is conserved. This equilibrium distribution of types can be enforced by there being an energy difference between the particle types or, perhaps, by spin statistics—as in the case of ortho- and para-hydrogen [15]. Thus, it is not necessary that the particle-type energies differ significantly. We assume that they do differ for the sake of generality, but not altogether for the sake of clarity. As such, we define the N-particle Hamiltonian:

$$H_0 = \epsilon_{\bigcirc} N_{\bigcirc} + \epsilon_{\square} N_{\square} + \sum_{i=1}^{N} \frac{p_i^2}{2m} \ , \qquad (1)$$

where $\epsilon_{\bigcirc}$ and $\epsilon_{\square}$ are the particle-type energies ($\Delta\epsilon = \epsilon_{\bigcirc} - \epsilon_{\square} > 0$), the particle numbers $N_{\bigcirc}$ and $N_{\square}$ sum to the total N, m is the particle mass, and p_i the ith particle's momentum.

The cylinder walls are impermeable to either type of particle, and there are four pistons inside. Two of these are also impermeable—paralleling Szilard, we also denote them ○ and □—and are initially set a distance ℓ apart. (The reason for reusing type labels as piston labels will become clear.) The other two pistons, ○′ or □′, are

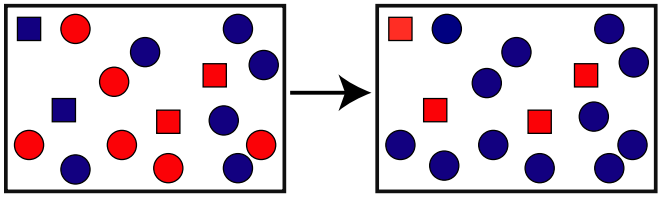

FIG. 2. Measurement in Szilard's second engine: Particle type variable $s \in \{\square, \bigcirc\}$ is depicted by shape and memory variable $y \in \{0, 1\}$ by color. (Left) initially uncorrelated demon-particle states—particle type is not correlated with memory (shape is not correlated with color). (Right) Configuration of the gas after measurement. Tracking from the left diagram to the right, the measurement process establishes a correlation between color (y) and shape (s): $\square \to$ red and $\bigcirc \to$ blue. There are only $\square$s and $\bullet$s.

permeable to only one of the two particle types. Each is set just inside of the impermeable pistons: $\bigcirc'$ being set next to $\square$ and $\square'$ next to $\bigcirc$. Refer to Fig. 1. These pistons are placed so they can slide one within the other, keeping the distance between $\bigcirc$ and $\bigcirc'$ and $\square$ and $\square'$ fixed at ℓ. We can think of this system as two overlapping cylinders of fixed length that can slide relative to each other; each having an impermeable wall ($\bigcirc$ or $\square$) at one end and a semi-permeable membrane ($\bigcirc'$ or $\square'$) at the other.

Szilard specifies a cyclic control protocol with three key transformations: *Measurement*: in which each particle's initial type is stored in its memory; *Control*: in which the system's thermodynamic resources are manipulated; and *Erasure*: in which the measurements are leveraged to return the overall system to its initial configuration. These steps generally outline the behavior of information engines as they leverage information resources to gain thermodynamic advantage; cf. Ref. [14].

The first step of the control cycle is *measurement*. Initially, the ensemble's particle-type distribution is given by $\rho_0(s)$ and the distribution $f(y)$ of the memory variable y is uncorrelated to particle type: $\Pr(s, y) = \rho_0(s) f(y)$. We choose the parameter γ to represent the initial distribution over the memory state of the particles, so that $f(y)$ is initially distributed as $\Pr(y_i = 0) = \gamma$ and $\Pr(y_i = 1) = 1 - \gamma$. During measurement, the current type s_i of each particle is imparted to its memory y_i such that each type $\square$ ($\bigcirc$) particle has its y variable set to 0 (1). Here, the distribution $f(y)$ changes so that the conditional distribution $f(y_i|s_i)$ is deterministic or, equivalently, the joint distribution over s and y is given by nonzero elements $\Pr(y_i = 0, s_i = \square) = \delta$ and $\Pr(y_i = 1, s_i = \bigcirc) = 1 - \delta$. See Fig. 2, where particle type is depicted via shape and particle memory via color.

Now, the cylinders slide relative to one another until the semi-permeable membrane ends come into contact. In doing so, the semipermeable membranes separate the particles by type. This is done without any input of work or heat since, from the perspective of each particle, its

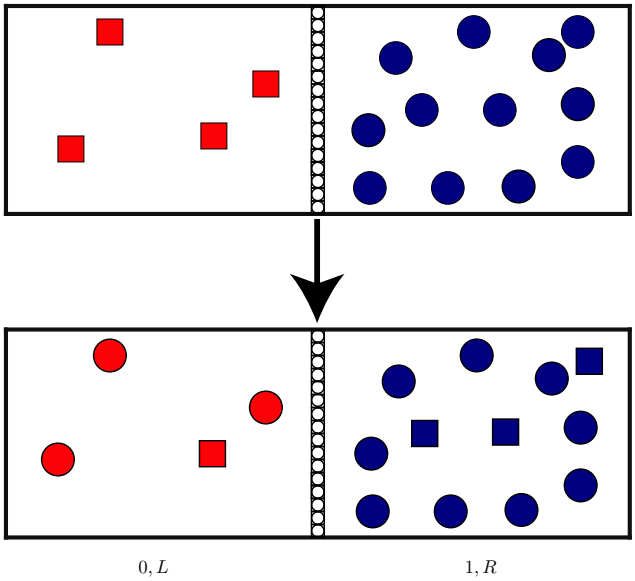

FIG. 3. Particle-type equilibration: (Top) Deterministic distributions $\Pr_L(s = \square) = \Pr_L(y = 0) = 1$ and $\Pr_R(s = \bigcirc) = \Pr_L(y = 1) = 1$ at the end of sliding-separation of Fig. 1. (Bottom) Distributions after a period of particle-type conversion. Particles are no longer separated by shape type $s \in \{\square, \bigcirc\}$, but still by memory state (color) $y \in \{0, 1\}$). That is, $\rho_L(s) = \rho_R(s) = \rho_0(s)$ but the memory state distribution in each compartment remains deterministic.

container is merely being translated; as demonstrated in Fig. 1. This transformation separates the particles into one of two compartments. Particles that were initially type $\square$ are all in the left volume (L) bounded by the pistons $\square$ and $\square'$; those that were type $\bigcirc$ have been shifted to the right container (R) that is bounded by the pistons $\bigcirc$ and $\bigcirc'$.

Time scales are important here. The separation must happen sufficiently slowly that the gas is always in equilibrium with respect to the container's spatial volume, but fast enough that no particles transition types during the process. This is not a generally prohibitive constraint, as we can assume the time-scale for a gas to fill its container uniformly is generally quick. After the separation, each particle type exists independently in a container of the same size as the initial container.

At this point, we introduce the subscript L or R, to denote whether we are discussing the distribution of particles in the left or the right compartment. Each compartment is no longer in equilibrium with respect to the type variable. See Fig. 3(Top). In principle, we can recover an equilibrium distribution with respect to H_0 individually within the containers by waiting for the system to re-thermalize with the heat bath or *control* the process with a protocol involving the input or output of work. See Fig. 3(Bottom). The latter is discussed in detail in Sec. III shortly.

At this point in the engine's operation, Szilard claims the

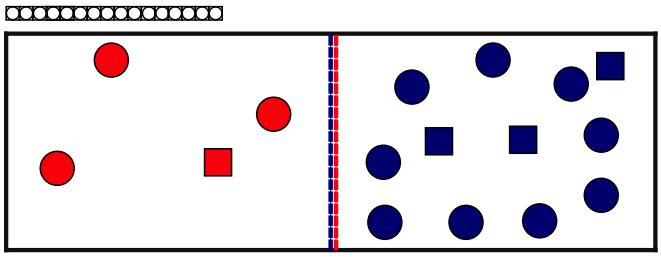

FIG. 4. First step to reintegrate molecules leading to the initial macrostate, replacing the type-semipermeable membranes with memory-state semipermeable membranes.

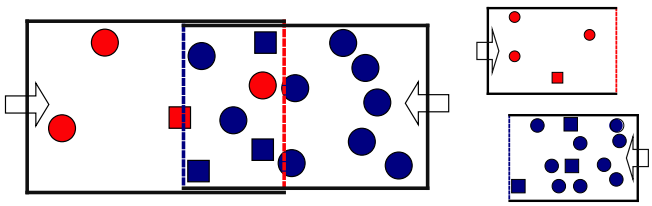

FIG. 5. Second step of reintegrating particles to return to the initial macrostate, leveraging the memory variable y with the newly inserted memory-state (color) semipermeable membranes.

"entropy has certainly increased". This is a familiar process and we expect it to increase the system entropy, since we increase the gas' effective volume. The entropy change ΔS from the initial macrostate to the final macrostate in which the particles have re-achieved equilibrium can be found by the Sakur-Tetrode equation (detailed in App. A), yielding:

$$\begin{aligned} \frac{\Delta S}{N} &= -k_{\mathrm{B}}\left(\delta \ln \delta + (1-\delta)\ln(1-\delta)\right) \\ &\equiv S(\delta)\ . \end{aligned} \tag{2}$$

The system's entropy increased, as Szilard claimed. If we re-establish the equilibrium distribution reversibly (through a *control* protocol) instead of spontaneously, then there must be a corresponding decrease $-S(\delta)$ in thermal reservoir entropy. Note that we cannot easily move the cylinders back into each other now, since there are particles of both types on each side of the semipermeable membranes.

Up to this point in the control protocol, the analysis only addressed the thermodynamics of particle type (shape)—variable s_i—not particle memory (color). That is, we have yet to use the particles' y_i variable. The engine now makes clever use of its memories (y_i) by exchanging the particle-type semi-permeable membranes with membranes that are semipermeable to memory states (color); see Fig. 4.

The system is then ready to operate the reverse strategy as when first type-separating the particles to bring them back into the same volume. See Fig. 5. Again, this is accomplished work-free, given that the process is done

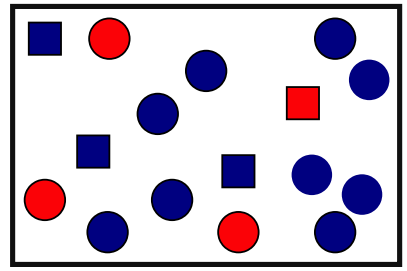

FIG. 6. Reintegration with sliding the memory-state (color) semipermeable membranes recovers the original distribution over particle type in the initial container.

on the proper time scale. Now that the particles are back within the original volume (L) again, they are no longer separated by color or shape. In this *erasure* process, we manage to bring the system back to its initial $\rho_0(s)$ macrostate, without interacting with the heat bath. The change in entropy for the system over the entire protocol cycle is, then, zero. The thermal reservoir, however, had a net decrease of entropy. At this point in the cycle, Szilard appeals to the validity of the Second Law, stating that [5]:

> If we do not wish to admit that the Second Law has been violated, we must conclude that ... the measurement of s by y, must be accompanied by a production of entropy.

Szilard's associating measurement with a change in thermodynamic entropy and giving the functional form Eq. (2) of the latter anticipates Shannon's communication theory and its measure of information [16] by nearly two decades.

The careful reader will notice several issues that require further investigation and definition. First, Szilard does not specify a mechanism that stores the memory variable y. Second, he does not investigate the work required to drive the reversible control transformation he postulates. Third, one notes that the final distribution over the memory variable y, while not correlated with type variable s at the cycle's end, is necessarily distributed so that $N\delta$ particles are in the $y = 0$ memory state and $N(1-\delta)$ particles are in the $y = 1$ memory state; that is, unless we include an additional erasure step that resets y to some arbitrary initial distribution. In addressing these (and related) concerns we shall see that, while the selection of the initial distribution $f(y)$ over memory variables y_i is arbitrary, the choice impacts the thermodynamic costs of measurement and erasure. First, we investigate the bounds on the work required to perform Szilard's reversible control transformation.

III. ENGINE VERSION 2.5

During the control step, each compartment begins in a nonequilibrium (completely deterministic) macrostate

$\rho_L(s)$ (or $\rho_R(s)$) (Fig. 3 (Top)) and ends in the canonical equilibrium macrostate ρ_0 (Fig. 3 (Bottom)). To understand the effects of this transformation, we appeal to recent developments in information theory and stochastic thermodynamics [17–19] that allow us to connect the Gibbs *statistical entropy*:

$$\begin{aligned} S(\rho) &= -k_{\mathrm{B}} \sum_{s\in\{\bigcirc,\square\}} \rho(s) \ln \rho(s) \\ &= k_{\mathrm{B}} \langle -\ln \rho \rangle_\rho \end{aligned}$$

to the energetics of the isothermal equilibration process.

The two compartments (L and R) interact separately with the heat bath, so we take the following process to be executed independently within each compartment. As such, we drop the L and R subscripts for clarity and take the final extensive quantities to be of the form $S(\rho) \equiv S(\rho_L) + S(\rho_R)$. Moreover, since the memory state remains deterministic within each compartment throughout the control process, the only relevant distribution is the marginal distribution—$\rho_L(s)$ or $\rho_R(s)$—over particle type.

Assuming perfect control of the Hamiltonian at any point during the transformation allows us to design the most efficient protocol for the equilibration process. Consider the particle-cylinder system immediately after particle separation, in contact with a thermal bath at temperature T. Initially, the Hamiltonian is that given in Eq. (1). We break the process into two distinct steps, both steps are executed within each compartment as follows. First, we instantaneously shift the Hamiltonian from H_0 to $H_\rho = -k_{\mathrm{B}} T \ln \rho$. Tautologically, ρ is now the equilibrium distribution since by definition the canonical equilibrium probability distribution is $\rho = e^{-\beta H_\rho}$. Shifting the Hamiltonian requires a minimum amount $W_{\Delta H}$ of work given by the difference ΔH in the system's total energy under the two Hamiltonians:

$$W_{\Delta H} = \langle H_\rho \rangle_\rho - \langle H_0 \rangle_\rho \ .$$

Next, we quasistatically shift the Hamiltonian back to H_0, which keeps the system in equilibrium by definition.

The transformation is now complete—the Hamiltonian returned to H_0 and the system's macrostate is given by ρ_0. Energy conservation in the second step implies that thermal reservoir and system energies change according to the work W_{qs} invested in the transformation:

$$W_{\mathrm{qs}} = \Delta U_{\mathrm{res}} + \Delta U_{\mathrm{sys}} \ .$$

Assuming the reservoir maintains constant volume, we write the W_{qs} in terms of initial and final free energies:

$$W_{\mathrm{qs}} = F(\rho_0) - F(\rho) \ .$$

(Appendix B gives the details.) Then, the total work $W_{\mathrm{drive}} \equiv W_{\mathrm{qs}} + W_{\Delta E}$ to drive the two-step transformation is:

$$W_{\mathrm{drive}} = \langle H_0 \rangle_{\rho_0} - \langle H_0 \rangle_\rho + TS(\rho) - TS(\rho_0) \ .$$

Each term is readily interpreted in the present setting.

When considering the sum of both compartments—recall $\langle H_0 \rangle_\rho = \langle H_0 \rangle_{\rho_L} + \langle H_0 \rangle_{\rho_R}$—the energy expectation values for ρ and ρ_0 are the same. The average kinetic energy KE_{avg} will be the same since the whole system is thermalized to the same temperature, so we can neglect its contribution. For the initial nonequilibrium distribution ρ we have:

$$\langle H_0 \rangle_\rho = N\epsilon_{\bigcirc}\delta + N\epsilon_{\square}(1-\delta) \ .$$

And, under the ρ_0 distribution:

$$\begin{aligned} \langle H_0 \rangle_{\rho_0} =& N\delta(\epsilon_{\bigcirc}\delta + \epsilon_{\square}(1-\delta)) \\ &+ N(1-\delta)(\epsilon_{\bigcirc}\delta + \epsilon_{\square}(1-\delta)) \ . \end{aligned}$$

$\langle H_0 \rangle_{\rho_0}$ simplifies trivially to $\langle H_0 \rangle_\rho$. Together they make no contribution to W_{drive}. The $TS(\rho)$ term vanishes since the initial distribution of particle types within each compartment (L and R) is deterministic. The final term, the equilibrium state entropy, is $S(\rho_0) = NS(\delta)$. And so:

$$\begin{aligned} W_{drive} &= -TS(\rho_0) \\ &= -NTS(\delta) \ . \end{aligned}$$

It is now clear that the thermodynamic cycle is an engine. The driving work is negative, signifying that there is an opportunity to extract work from the heat bath. Once again, we are faced with the reality that the process of measurement must contain compensating thermodynamic costs or admit that Szilard's second engine is a type of perpetual motion machine.

IV. DEMON GAS AS A THERMODYNAMICAL SYSTEM

To investigate the cost of measurement thermodynamically, we must choose a specific implementation of the device. We start with a 3-dimensional unit cube in contact with a heat bath, inside are N particles moving about thermally. The previous section established that the work extracted by Szilard's engine is independent of the energy difference $\Delta\epsilon$. We are, then, free to set this difference to zero—yielding a box of particles that are all identical according to H_0. The particles need not interact with each other to perform any of the necessary operations, so we can choose them to be noninteracting. Thus, our system is an ideal gas of N identical particles.

The membranes separating the particles into the L and R compartments slide along the box's x axis. We take all particles to start in the region $x < \ell$, with an ideal barrier inserted along $x = \ell$ to keep the particles from moving thermally into the region $x > \ell$. In this way, we defined

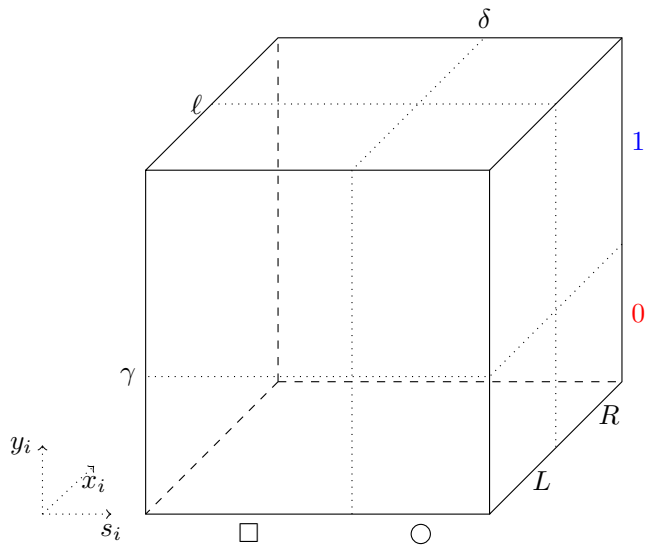

FIG. 7. Markov partition of a demon-particle's state space—3D unit box. The ith particle's position on the s axis corresponds to its particle type as $S_i = \square$ when $s_i < \delta$ (and $S_i = \square$ when $s_i > \delta$). The y-axis partition corresponds to the memory state as $Y_i = 0$ when $y_i < \gamma$ (and $Y_i = 1$ when $y_i > \gamma$). The depth dimension, parametrized by ℓ, corresponds similarly to particle position being in the left or right compartment. Note that ℓ must be equal to $\frac{1}{2}$ for the $L \leftrightarrow R$ transition to always be work free, as Szilard noted.

two compartments L and R corresponding respectively to $x_i < \ell$ and $x_i > \ell$.

We still need an operational definition of particle type which satisfies Szilard's requirements that there is a fixed particle-type equilibrium and that particles convert monomolecularly (N is constant) from one type to another. For our particle type, we choose the position of a particle along the s dimension. If the coordinate of a particle is $s_i < \delta$ or $s_i > \delta$ we consider it to be particle type $\square$ or $\bigcirc$, respectively. As the particles move about thermally, they cross back and forth across the line $s = \delta$ which exactly models Szilard's monomolecular conversion. Additionally, by choosing the parameter δ we are able to set our equilibrium distribution over particle type using the gas' tendency to quickly fill its container uniformly.

This choice for particle type also allows us to define semipermeable particle-type membranes as ideally impermeable membranes that cover only the region associated with the relevant particle type. It is well known that the physical position of particles stores information [7]. And so, we choose a particle's memory states to be stored in its y coordinate with 0 (1) corresponding to $y_i < \gamma$ ($y_i > \gamma$). See Figure 7 for the full partitioning of the demon-particle system.

We are now ready to (i) analyze this engine's thermodynamics, (ii) set up the symbolic dynamics for the gas particles, and (iii) analyze the engine's intrinsic computation.

A. Thermodynamics

The model introduced in Ref. [14] allows us to easily probe the thermodynamics of each step in the Szilard Engine V. 2.5 operation, as just described in Sec. II. Given this representation of Szilard's second engine, the overall thermodynamic cycle is the series of transformations shown in Fig. 8: *measure*, *control*, and *erase*. These operations are executed by inserting, sliding, and removing barriers.

The *measure* step, for example, involves three barriers. First, we insert a barrier along $s = \delta$. This is thermodynamically free, since the gas is identical on either side of the barrier. Next, we use a barrier perpendicular to the y axis that extends until $s < \delta$ to compress the particles that are in the $\square$ partition to fit entirely within the 0 partition. Similarly, we use a barrier perpendicular to the y axis that covers $s > \delta$ to compress the particles that are in the $\bigcirc$ partition to fit entirely within the 1 partition. This establishes the necessary correlation between type and memory state: all particles are either $\square$ or $\bullet$.

For the *control* step, the first operation separates particles by type into either the L or R partition. This involves translating the $\bullet$ particles to the R partition by inserting a barrier perpendicular to the x axis at $x = 0$ that covers from $s = \delta$ to $s = 1$. Then, along with the $s > \delta$ section of the initial barrier, this barrier translates the gas to the rear partition. This requires no interaction with the heat bath, since the volume of the $\bullet$ gas remains constant.

The second part of the control step expands along the particle-type dimension by allowing the two sections of the particle-type partition corresponding to $x < \ell$ and $x > \ell$ to slide independently of one another. The work W_{drive} the gas exerts on the barrier for an isothermal operation is calculated easily as $-\int P\mathrm{d}V$, with $P = Nk_{\text{B}}T/V$:

$$\begin{aligned} W_{\text{drive}} &= -\int_{\ell\delta\gamma}^{\ell\gamma} \frac{Nk_{\text{B}}T}{V}\mathrm{d}V - \int_{\ell(1-\delta)(1-\gamma)}^{\ell(1-\gamma)} \frac{Nk_{\text{B}}T}{V}\mathrm{d}V \\ &= Nk_{\text{B}}T\left(\ln\delta + \ln(1-\delta)\right) \\ &= -NTS(\delta)\ . \end{aligned}$$

This accords with the value calculated above. Thus, the model achieves the ideal efficiency bound.

We can also calculate the thermodynamic costs of the measurement and erasure transformations. In these, the gas' internal energy remains fixed and so $Q_{sys} = -W_{sys}$. To investigate the energy that is dissipated in the heat bath, we draw a relation between Q_{sys}, which is positive when heat flows into the system from the bath, and $Q_{diss} = -Q_{sys}$, which is positive when heat is being dissipated into the heat bath. For the measurement process,

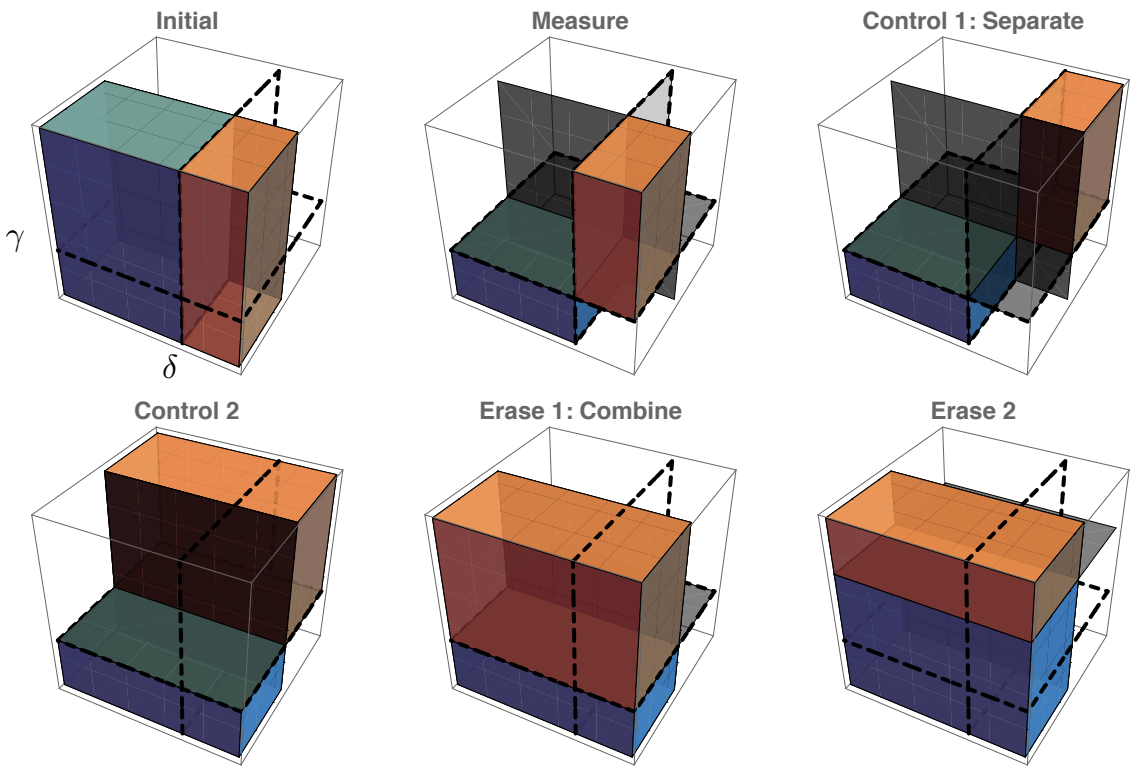

FIG. 8. Second Szilard engine's action on the unit-cube demon-particle state space decomposed into individual steps on an initially uniform distribution. Color illustrates which particles start as which type. The ideal barriers that are used to execute the protocol are depicted as dash-outlined gray planar partitions.

we have:

$$\begin{aligned} Q_M &= -\int_{\ell\delta}^{\ell\delta\gamma} \frac{N\delta k_{\mathrm{B}}T}{V}\mathrm{d}V - \int_{\ell(1-\delta)}^{\ell(1-\delta)(1-\gamma)} \frac{N(1-\delta)k_{\mathrm{B}}T}{V}\mathrm{d}V \\ &= Nk_{\mathrm{B}}T\left(-\delta\ln\gamma - (1-\delta)\ln(1-\gamma)\right) \\ &= Nk_{\mathrm{B}}T\left(\delta\ln\frac{1-\gamma}{\gamma} - \ln(1-\gamma)\right) \ . \end{aligned}$$

Figure 9 (Top) displays a contour plot of the measurement heat Q_M as a function of the partition parameters γ and δ. We see that the measurement thermodynamics strongly depends on these parameters and that heat will always be dissipated during measurement. To implement an efficient engine, then, we would select a set of parameters that minimizes the heat dissipated in measurement.

Measurement is only part of the overall engine cycle, though. There is also the erasure transformation. The first erasure step in Fig. 8 translates the particles back into the same $\{L, R\}$ partition; similar to the first *control* operation. The current model makes it abundantly clear that this not sufficient to return the gas to its initial state, though. The gas above and below the memory-state partition (inserted at the beginning of the *measurement* step) will not generally have the same pressure. We require an additional step to return the gas to it's initial maximum entropy state. Translating the boxes back to the L compartment does not require any thermodynamic input or output so this final step is the source of the thermodynamics of the erasure. The final step allows the gas to slide the partition that separates our memory states until the pressure on each side equalizes—until it rests at $y = \delta$. The barrier may then be removed at no cost or it may be left in the box and allowed to move freely along with the next cycles without affecting the thermodynamics. Calculating the energetic cost of this transformation is as simple as preceding, yielding:

$$Q_E = Nk_{\mathrm{B}}T\left((1-\delta)\ln\frac{1-\gamma}{1-\delta} + \delta\ln\frac{\gamma}{\delta}\right) \ .$$

It is not surprising that the entropy cost of erasure vanishes when $\delta = \gamma$, since then the barrier at γ is already in the equal pressure position before the final step. Figure 9(Bottom) shows that erasure does not incur a cost: instead, the erasure provides yet another opportunity to extract energy from the heat bath. This is as expected, as the erasure process always increases the entropy of the system. However, examining $Q_M + Q_E$ we see that choosing the parameters to maximize the energy extraction in erasure increases the cost of measurement commensurately. Suggestively, the total thermodynamic cost of measurement and erasure is algebraically independent of

the parameter γ:

$$\begin{aligned}\frac{Q_M + Q_E}{N k_{\mathrm{B}} T} &= \left((1-\delta)\ln\frac{1-\gamma}{1-\delta} + \delta \ln\frac{\gamma}{\delta}\right) \\ &\quad + \left(\delta \ln\frac{1-\gamma}{\gamma} - \ln(1-\gamma)\right) \\ &= -(1-\delta)\ln(1-\delta) - \delta\ln\delta\end{aligned}$$

Or:

$$\frac{Q_M + Q_E}{NT} = S(\delta)\ .$$

That is, the total combined cost of measurement and erasure depends only on δ, as in $NTS(\delta)$. This is exactly the energy necessary to compensate for the work extracted from the heat bath during control. Since the choice of γ does not affect the total work extracted from the heat bat, nor the total cost of the *measurement* and *erasure* processes together, one can set $\gamma = \delta$ so that erasure is cost neutral and all of the extracted work comes from the control process.

In this way, we need only consider the "cost" of measurement and the "revenue" from control. Of course, there is no net profit. Even in the most efficient system, the Second Law holds. And, this was one of Szilard's main points—the point that resolved Maxwell's paradox. By giving the demon (or control subsystem) a physical embodiment and properly accounting for its thermodynamics, there is no Maxwell demon paradox.

It is interesting to note that, just as in Ref. [14], the distinction between measurement and erasure turns out to be, in a sense, arbitrary. We may increase or decrease the cost of one, but we do so at the expense of the other. This harkens back to Szilard's original work, where he assigned entropy production to the measurement and then goes on to demonstrate with a specific measurement apparatus that the erasure step increases the entropy. (This apparatus is discussed in Sec. V below.) Szilard was not as much concerned about when the entropy was produced, as that the production had to be associated with the process of establishing and destroying correlation between particle type and the memory state.

This contrasts with the view advocated by Landauer and Bennett half a century after—the logical irreversibility of erasure solely determines thermodynamic costs [8, 9]. We now see, as others have recently emphasized [10–14], a more balanced view that there is a generalized principle bounding the total costs of measurement and erasure. Presumably, this is a constructive result that may lead to a design flexibility in future information engine implementations.

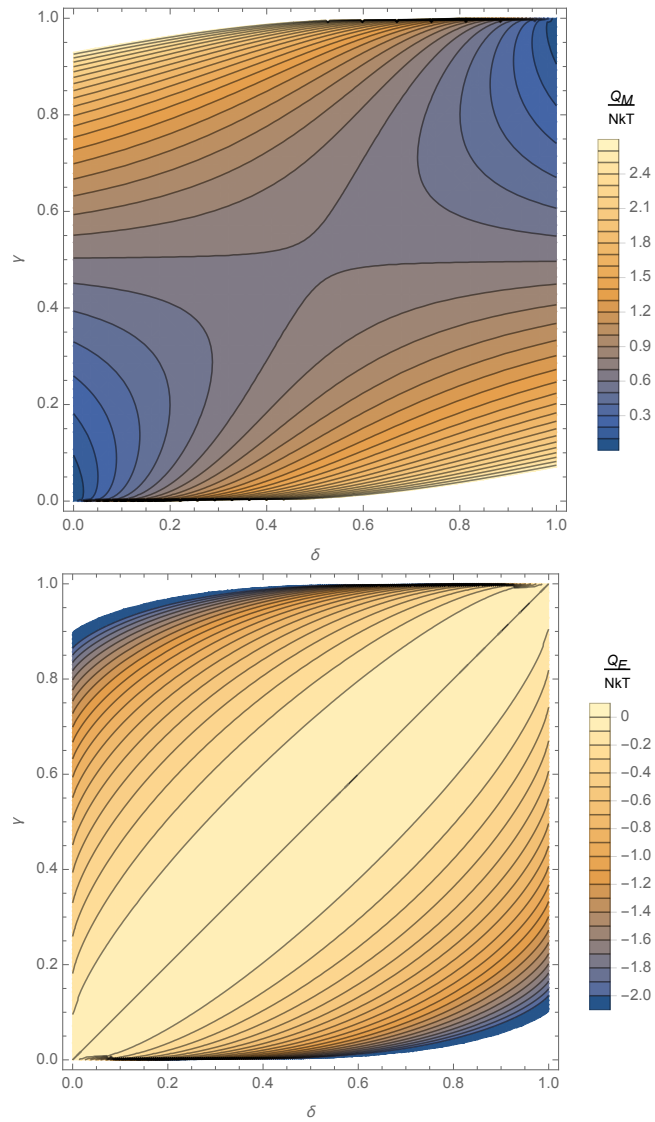

FIG. 9. Thermodynamic cost (heat dissipation) of measurement $Q_M/Nk_{\mathrm{B}}T$ (Top) and erasure $Q_E/Nk_{\mathrm{B}}T$ (Bottom) as a function of partition location parameters δ and γ.

B. Computational Mechanics of the Gas Symbolic Dynamics

The course-graining of the microstate-space's unit box, depicted in Fig. 7 is a Markov partition [20] of the microstate dynamics under the macroscopic thermodynamic transformations that make up the Szilard engine. This immediately suggests defining a vector of binary variables ($S_i \in \{\square, \bigcirc\}, Y_i \in \{0, 1\}, X_i \in \{L, R\}$), sequences of which exactly track of the engine's microscopic dynamical behavior.

At each protocol step a compound symbol SYX is generated according to the particle's location in the state-space box. For example, a particle that ends a protocol step and generates the compound symbol $\square 0 L$ corresponds to

a particle that is currently type □, was particle type ○ when the most recent measurement was performed, and is in the Left compartment.

We must remind ourselves that the state space of this gas is large. There are N particles in three dimensions, so the full state space represents a $3N$ dimensional dynamical system. However, since the particles are noninteracting, Szilard's second engine is actually a collection (direct product) of N 3D particle state spaces. Applying computational mechanics' *predictive equivalence relation* collapses the $3N$-dimensional state-space to 3 dimensions of equivalent causal states [21]. Thus, we can use the symbolic dynamics of a single particle to find the engine's effective information processing behavior—and scale it to N particles with a pre-factor of N.

The problem simplifies even further since, having faithfully considered Szilard's initial problem statement, it is clear that the LR dimension of the state-space box is redundant in the current model. It was useful in Szilard's original construction to include a barrier that stops particles corresponding to the different memory states from intermixing. However, the current engine stores the memory and type states in positional coordinates, so the barrier used to compress the gas in the cycle's *measure* step already serves this purpose. Thus, we do not even need the full 3-dimensional state-space box to model the system's information and thermodynamic action.

Instead, we examine the action of Szilard's second engine on a 2-dimensional projection onto the sy plane of the box in Fig. 8. The resulting 2-dimensional map is nearly identical to the Szilard Map introduced in Ref. [14], constructed by considering Szilard's first or single-molecule engine. The only differences between these maps is a different initial state distribution. In fact, we could reconstruct the second Szilard engine to have the same initial state but, for the purpose of illustration, we will investigate the map under the current default memory state. At this point, though, one fully expects the results to agree with Ref. [14] in every fundamental sense.

We now track the probability density of a particle within the gas. Having abandoned tracking each particle's exact position within the box by using the course-graining into binary symbols, we now consider the actions of a deterministic map on the probability density as a whole. Each step in the process depicted in Fig. 8 compresses or expands the probability density along a particular dimension. The composite map that includes each step when $\delta = \gamma$ is given by:

$$\tau_{\text{Szilard}}(s, y) = \begin{cases} \left(\frac{s}{\delta}, y\delta\right) & s < \delta \\ \left(\frac{s-\delta}{1-\delta}, \delta + y(1-\delta)\right) & s > \delta \end{cases} .$$

Appendix D gives the maps for each individual step.

In accordance with the analysis in Ref. [14], if we choose to leave out the memory-state partitions that are added each cycle, we build up the same self-similar interleaving within the particle's state-space probability distribution as seen in the Baker's Map [22]. While the probability density is not uniform throughout each step of the map, we find that the distribution over the state space is uniform and constant for the composite map τ_{Szilard} above that includes each step in the protocol.

C. Information and Intelligence

We again apply computational mechanics' predictive equivalence relation—now not to the gas' microscopic state space but to the symbolic dynamics induced by Markov partition of Fig. 7. This leads directly to an ϵ-transducer [23] that captures the information processing embedded in the engine's operation. Figures 10 and 11 show the transducer for each dimension separately and Fig. 12 shows the ϵ-machine for the joint process. Then, retracing the steps in Ref. [14] establishes that Szilard's first and second engines are informationally and thermodynamically equivalent, though they arise from rather different implementations.

Composing Figs. 10 and 11 transducers with the period-3 input process—that specifying the measure-control-erase protocol, gives an ϵ-machine that generates the output process for particle type or for memory state. (In this case, this is trivially implemented by dropping the input symbols $\{M, C, E\}$ from the ϵ-transducer transitions.)

The two resulting ϵ-machines and that in Fig. 12 are counifilar [24]. The processes are not cryptic and this greatly simplifies calculating various informational properties. For example, the entropy rate of the joint system's machine (Fig. 12) is $\frac{1}{3}\,\mathrm{H}(\delta)$ per step, consistent with the analytical result for the Baker's Map from Pesin's theorem. However, there is a slight variation from Ref. [14]'s analysis of the statistical complexity C_μ—the information in an ϵ-machine's causal state distribution $\{\mathbf{S}\}$. It is immediately clear from Figs. 10 and 11 that C_μ^x and C_μ^y are equal. This was not the case in Ref. [14]. Thus, we see that the different choice of initial state symmetrizes the stored information with respect to particle type and memory state. The calculations for C_μ are straightforward, nonetheless:

$$\begin{aligned} C_\mu^x &= C_\mu^y \\ &= -\left[\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\left(\delta\log_2\frac{\delta}{3} + (1-\delta)\log_2\frac{1-\delta}{3}\right)\right] \\ &= \log_2 3 + \frac{2}{3}\,\mathrm{H}(\delta) \ . \end{aligned}$$

Similarly, we find that $C_\mu^{\text{joint}} = \frac{4}{3}\,\mathrm{H}(\delta) + \log_2 3$. These are quantitatively different from the results in Ref. [14]. However, that is the end of the differences. If we consider the relationship between the three, we recover that $C_\mu^{\text{joint}} = C_\mu^x + C_\mu^y - \log_2 3$. So, we see that the two engines have the same information related to synchronization of their two subsystems. For the original single-molecule engine these subsystems were demon memory and molecule

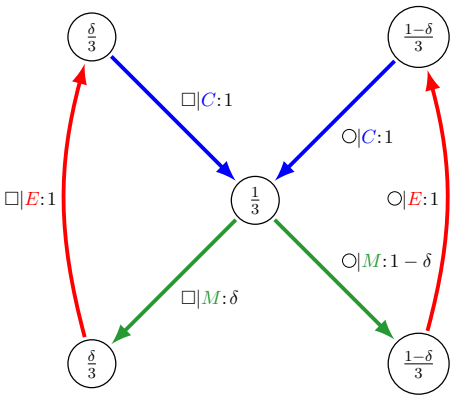

FIG. 10. ϵ-Transducer for the particle type (x) subsystem. Protocol steps are designated by color: (control, measure, erase) $\Leftrightarrow$ (blue, green, red). Numbers inside states correspond to the asymptotic state probability. The transition notation $s|d:p$ corresponds to emitting the symbol s with probability p given the driving symbol d.

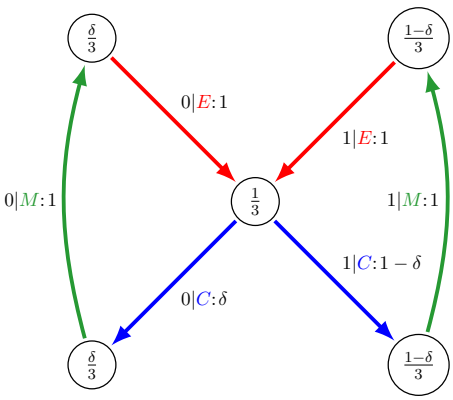

FIG. 11. ϵ-Transducer for the memory state (y) subsystem. Notation as in previous figure.

position; for engine Version 2.5, they are shape and color. In fact, we can explicitly check that all other informational measures agree with Ref. [14]. In doing so, we recover the asymptotic communication rate:

$$\lim_{L\to 0} \frac{\mathrm{I}[X_{0:L}; Y_{0:L}]}{L} = \frac{1}{3}\,\mathrm{H}(\delta) \ ,$$

the correlation rate:

$$\lim_{L\to 0} \frac{\mathrm{I}[S^x_{0:L}; S^y_{0:L}]}{L} = \frac{1}{3}\,\mathrm{H}(\delta) \ ,$$

and the dependence of the correlation during the protocol steps, which yields:

$$\mathrm{I}[X_0 : Y_0|M] = \mathrm{H}(\delta) \ .$$

The measurement step is where the single-symbol correlation is established. So, it stands to reason that the correlation dependence is found to be entirely in this step.

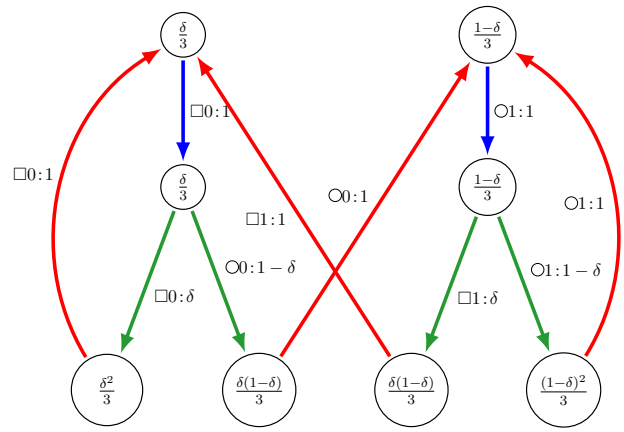

FIG. 12. ϵ-Machine for the joint demon-particle system: Protocol steps designated by color as in Figs. 10 and 11. Driving symbols are suppressed in the transition notation for clarity—$s:p$ corresponds to emitting the symbol s with probability p.

V. SZILARD'S THIRD ENGINE: MEASUREMENT AND STRONGLY COUPLED SYSTEMS

The Szilard Map stores its memory state in an additional spatial dimension. Section IV and Ref. [14] tease out the thermodynamic and information-processing consequences of this choice. The following introduces an alternative implementation of information storage—one introduced by Szilard himself.

After concluding that the measurement process in his engines must generate entropy, Szilard introduces a limit on the production of entropy from a binary measurement:

$$e^{-S_\square/k_\mathrm{B}T} + e^{-S_\bigcirc/k_\mathrm{B}T} \leq 1 \ ,$$

where $S_\square$ and $S_\bigcirc$ are the entropies that a protocol generates when taking the measurement value $\square$ or $\bigcirc$, respectively. Investigating this limit further, he adopts a specific mechanical system that performs the minimal measurement tasks that his engines require.

The essential tasks in measurement are as follows. First, establish a correlation between the instantaneous value of a fluctuating variable x and another variable y. Second, store that value in the "memory" of the second variable so that if x changes, y remains fixed. Finally, return to a default state so that the system is ready to preform another measurement.

In this third construction of Szilard's, the variable to be measured x is the position of a pointer that moves back and forth according to a completely general protocol, either stochastic or deterministic. The variable y that stores the position is a function of the temperature of a body K that is mechanically connected to the end of the pointer. As this pointer moves back and forth, it brings K in contact with one of two intermediate temperature

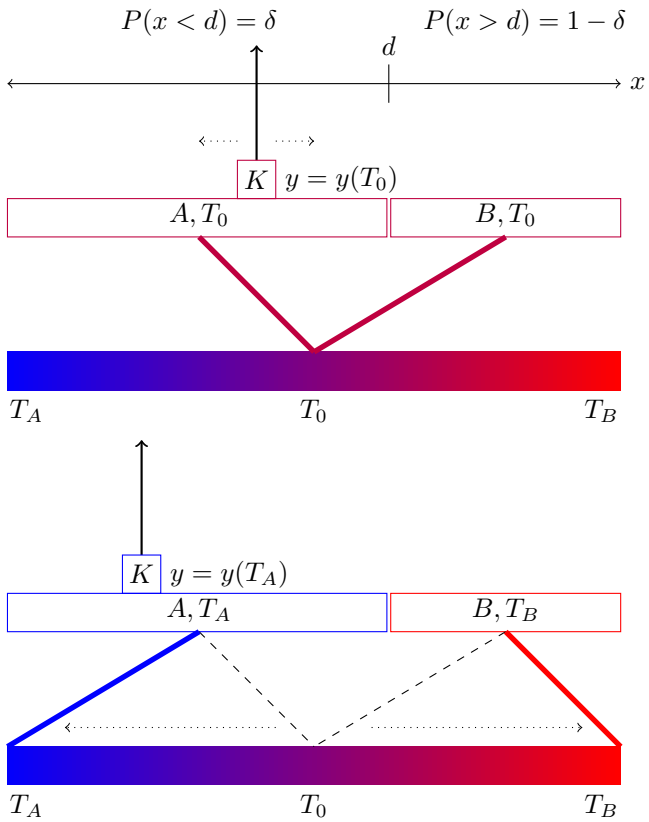

FIG. 13. (Top) Default state before measurement: Variable x tracks the position of the pointer and y is a function $y(T)$ of the temperature T of body K. (Bottom) Measuring position of the pointer by temperature of K. The pointer location at the time of measurement determines if K is cooled to T_A or heated to T_B. Consequently, y is set to either $y(T_A)$ or $y(T_B)$.

reservoirs, A or B. These reservoirs are connected by movable heat-conducting rods to a continuum of temperature reservoirs that span from a cold temperature T_A to a warm temperature $T_B > T_A$. Initially, both rods are connected to an intermediate temperature T_0; see Fig. 13.

By coarse-graining the position of the pointer into two regions (A, B) we are able to make a binary measurement. The measurement, which must happen during a timescale for which the pointer is stationary, involves moving the connecting rods through the continuum of heat reservoirs so that the intermediate reservoir $A(B)$ is cooled(heated) to $T_A(T_B)$. In this process K will either become heated or cooled depending on where the pointer was, see Fig. 13. This process can be done with arbitrarily small dissipation, if the process is done slowly enough such that the rod, the intermediate reservoirs, and K remain in thermal equilibrium at all times. We have now accomplished a binary measurement, where K is either at T_A or T_B, depending on the position of the pointer at the moment of measurement.

Next, the entire assembly of reservoirs is thermally isolated from the pointer and K so that, as the pointer continues to move, K maintains its temperature either at T_A or at T_B, even as the pointer leaves the interval it was in at the time of measurement. In this condition, the measurement value is stored in K's energy content. Now, to be ready to make another measurement, the system must return to its initial state. If one knew with certainty K's temperature, the system could be returned to the default state without entropy cost: Simply wait until the pointer is in the region that corresponds to K's temperature, bring the system back into contact with the reservoirs, and institute the measurement protocol in reverse. This is, of course, actually two different protocols—and requires knowledge (measurement) of the result of each measurement to decide which to implement on each cycle.

There is no single protocol that can blindly return the system to its original state without producing entropy. Anticipating Landauer's well-known argument for the bistable well [7] by more than three decades, Szilard notes that an increase in entropy "cannot possibly be avoided" because [5]:

> After the measurement we do not know ... whether [K] had been in connection with T_A or T_B in the end. Therefore neither do we know whether we should use intermediate temperatures between T_A and T_0 or T_0 and T_B.

We create, then, a single protocol that returns the system to its original state—the "erasure" process—and measure its total entropy generation. While the pointer is still uncoupled to the system, we return A and B to the equilibrium temperature T_0. Once again, this can be done reversibly on an appropriate timescale; see Fig. 14. Then, we bring K back into thermal equilibrium. This step cannot be done reversibly. This gives merit to the idea that erasure is the source of the entropic cost. Quantitative accounting for the entropy generation reveals additional insight.

All said, the body K undergoes a cyclic process, so the net change in the system entropy is zero. Thus, we consider the entropy change only in the reservoirs. If the pointer was at a location that caused K to cool (heat) to T_A (T_B), then the reservoir's entropy increases (decreases) during the measurement period by $\int \frac{dQ}{T}$. Similarly, when K is returned to T_0 the reservoir's entropy decreases (increases) by $\frac{\Delta E}{T_0}$. We see that, while only the erasure process causes the entropy of the universe to increase, both the measurement and erasure processes play a role in increasing and decreasing the reservoir entropy. Szilard was unconcerned with keeping measurement and erasure as two different actions since he already concluded that it was possible for either to produce or consume the reservoirs' entropic resources. This is an insight that only recently received renewed attention [10–14]. Furthermore, Szilard had also already concluded that the need to erase a binary random variable to a default state had unavoidable entropic costs.

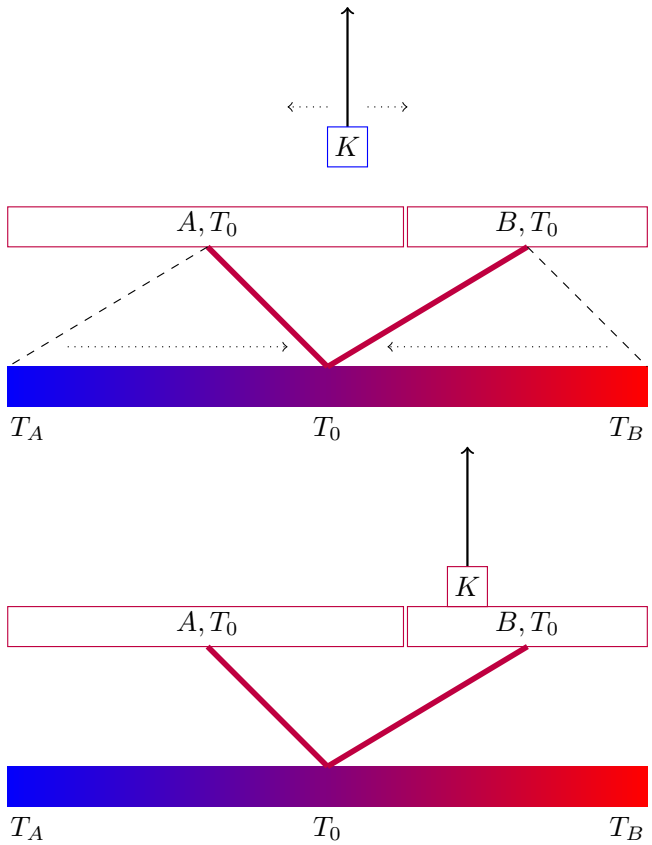

FIG. 14. (Top) While K stores the location of the pointer at the time of measurement, A and B are returned to T_0. (Bottom) K is returned to T_0 by thermal contact with A or B, incurring an unavoidable entropic cost.

To determine quantitatively the entropy gain from each measurement, Szilard adopts a 2-level system. The body K can be on one of two energy levels: a low energy state and a high energy state. Using standard canonical ensemble calculations, he shows that in the limiting process where the probability of the low (high) energy state at T_A (T_B) approaches unity the entropy generated by each process is:

$$\begin{aligned} S_A &= -k_{\mathrm{B}} \ln p \\ S_B &= -k_{\mathrm{B}} \ln q \ , \end{aligned}$$

where $p = p(T_0)$ and $q = q(T_0)$ are the probabilities that K is in the lower and upper energy state at temperature T_0, respectively. Szilard ended his analysis here. He does note that for this model:

$$e^{-S_A/k} + e^{-S_B/k} = 1$$

and that this represents the minimum amount of entropy generation necessary according to his limit:

$$e^{-S_A/k_{\mathrm{B}}T} + e^{-S_B/k_{\mathrm{B}}T} \leq 1 \ .$$

VI. SZILARD MEASUREMENT IN SZILARD ENGINES

To complete our analysis of Szilard's arguments, we couple the Szilard measurement device (SMD) above to one of his engines. For a simple physical picture, we specialize to the more familiar single-particle Szilard engine [5, 8, 14], where a classical particle in a box is used to extract work from a temperature bath by inserting a partition and allowing the particle to move the partition. In essence, this engine leverages the measurement of a thermal fluctuation to do work. The considerations above show that the multi-particle second engine has the same thermodynamic and information processing behavior as the more-oft-quoted single-particle engine, as treated in Ref. [14]. And so, we lose nothing by specializing to his first, simpler model.

Now, take the SMD's pointer to be mechanically connected to the particle inside the Szilard engine, so that the position of the pointer tracks the particle's thermal motion. The SMD is calibrated so that the particle on the left- (right-) hand side of the partition corresponds to the pointer being at $x < \delta$ ($x > \delta$). The SMD is thermally isolated from the rest of the engine, as having its own set of reservoirs is crucial to its operation. In this example, the body K plays the demon's role: the body changes length depending on its energy state, allowing K's state to select the engine's protocol; for example, by the position of a switch connected to K.

In this way, the entropy generated by one engine cycle is the sum of the entropy generated in the SMD's reservoirs and the particle-box system's reservoirs. The particle moves thermally through the entire box, so the probability that it falls in one or the other section of the box depends on the relative volume on either side of the inserted partition. During a cycle, the entropy generation in the system's reservoirs is proportional to either $\ln \delta$ or to $\ln(1-\delta)$ depending on if the particle starts on one or the other side of the partition, respectively. The mean entropy generated in the system reservoir over many cycles is then:

$$\begin{aligned} \langle \Delta S_{res_s} \rangle &\propto \delta \ln \delta + (1-\delta) \ln(1-\delta) \\ &= -H(\delta) \ . \end{aligned}$$

The mean entropy generation in the SMD's reservoirs over many cycles of the measurement process is:

$$\begin{aligned} \langle \Delta S_{res_m} \rangle &\propto \delta S_A + (1-\delta) S_B \\ &= -\delta \ln p - (1-\delta) \ln q \ . \end{aligned}$$

Adding the two contributions yields the average entropy generated in the universe per cycle:

$$\begin{aligned} \Delta S &\propto (1-\delta) \ln \frac{1-\delta}{1-p} + \delta \ln \frac{\delta}{p} \\ &= D_{KL}(\delta || \gamma) \ , \end{aligned}$$

where the relative information $D_{KL}(\cdot||\cdot)$ is positive for all values of $\delta \neq p$ and vanishes when $\delta = p$.

We see that, once again, there is no Maxwell demon paradox, as the total entropy generation is positive. In the case that $p = \delta$, the mean entropy produced in the SMD reservoirs during the measurement cycle is exactly enough to compensate the decrease of entropy in the system's reservoirs during the work extraction. This encoding of the information yields the most efficient cycle, in accordance with the analysis in the previous sections and Ref. [14].

However, there is no physical requirement that $p = \delta$. The inner workings of body K need not match where the partition between the physical regions A and B lies. It is tempting to make a direct comparison between p and the parameter γ in Ref. [14]'s Szilard Map, but there is a distinct difference between the two. Under the action of the Szilard Map, as introduced in Ref. [14], there is inherent interaction between the parameters δ and γ that manifests itself in the density of the ideal gas that serves as the engine's "working fluid". When one part of the gas is compressed into a particular memory partition, the size of that partition determines the cost of the next step. If the partition is small, it is "more difficult" to squeeze the same number of particles in. Consequently, both δ and γ appear in the cost of both measurement and erasure. This coupling between the two dimensions becomes relevant when we take into account the total cost of measurement and erasure, finding that γ drops out of the consideration. The result is an engine that is ideally efficient for every parameter setting.

When coupling the SMD to the first engine, the importance in the inherent interaction of δ and γ in the Szilard Map becomes even more apparent. Unlike γ, the engine is no longer ideally efficient for any choice of parameter p. Instead, we must choose the distribution of K's energy states at the equilibrium temperature to have the same distribution as the particles position states. If not, the engine suffers additional dissipation from a mismatch of our system and our measurement device. (Reference [25] recently considered the energy costs of such mismatches.)

Looking across the sweep of progress since his original results, we now see that Szilard's construction is a concrete example of Ashby's *Principle of Requisite Variety* [26]: the variety of actions available to a control system must match the variety of perturbations it is able to compensate. Specifically, Szilard recognized that a minimal control system for a binary measurement must have two states. (Yet again, Szilard predated the cybernetics era by several decades.)

However, we also see a stricter requirement used to avoid unnecessary dissipation. The actual distribution over the controller's internal states must be the same as the system's. This also touches on the general arguments put forth in Ref. [27] that consider the efficiency of a thermodynamically embedded *information ratchet* which interacts with an information reservoir to extract work.

Finally, one sees a clear parallel between the information-theoretic concept of optimal encoding [17] in which minimizing the memory needed to store a particular message, the highest probability events are given the shortest codewords. In short, it should not be surprising that it is optimal to match our controller to the system. However, it is gratifying to see such a clear and straightforward example—an example unfortunately ignored by Szilard's future colleagues.

The SMD model of measurement also provides a clear physical picture of adding memory to a Maxwell Demon engine. If we imagine that the SMD has two bodies K_1 and K_2 that store information, the single-particle Szilard engine can operate for two cycles without having to go through an erasure process. Instead of erasing the first body at the end of the first cycle, the SMD moves on to operate on K_2—leaving K_1 in whatever final state the first cycle determined. This avoids increasing the universe entropy while extracting work from the Szilard engine heat bath. This violation of the Second Law is only transient, though. To preform a third cycle, the SMD must erase K_1 or K_2 to store the next measurement. At the point immediately before erasure in each cycle, the joint system must pay the entropic cost for 2 fewer measurements than it has made.

It is easy to see how this construction generalizes to larger physical memories consisting of N memory "bits" $K_1, K_2, \ldots, K_N$. With an N-bit memory, the joint system of the Szilard engine and the SMD can continue to extract work from the engine's reservoirs for N cycles before having to finally pay the cost of its first measurement. From that point forward, though, every new measurement must be associated with an erasure of a previous one. This restores the Second Law with respect to the erased measurement. Each measurement is eventually paid for and, as the number of cycles grows large, the transient leverage from having a large memory becomes less and less noticeable.

VII. CONCLUSION

Since Szilard's day in 1929, the once-abstract conception of a molecular-scale "neat fingered and very observant" being [1] that interacts with heat and information reservoirs has only become more tenable, as modern computing emerged and micromanipulators were invented and then miniaturized through nanofabrication techniques. Thus, understanding the workings of information engines—microscopic machines interacting with such reservoirs—is now highly relevant, especially compared to the days when Maxwell first offered up the idea as a pedagogical absurdity. This is evinced by, if nothing else, a constant and increasing stream of recent efforts that take Szilard's original single-molecule engine as a jumping off point to investigate how measurement, information, thermodynamics, and energy interact with one another in support functional behaviors [28–33].

Szilard's early models grounded Maxwell's demon in physical embeddings. Since their introduction they provided the bedrock for much debate and occasional insights over the past century, largely through his first, brilliantly-simple single-particle engine. Here, we found that his second (multiparticle) engine, though more obtuse in construction, captures all of the same interesting consequences suggested by the first. Additionally, it maps exactly on to the first engine's operation by setting the demon memory state to another positional coordinate as in Ref. [14]. In several important ways, though, his second engine is more physical and plausible. And so, the multiparticle-membrane engine is more robust to criticism arising from concerns about applying classical statistics to the behavior of the first engine's single particle. Thus, the second engine's relationship to its single-particle sibling supports the physicality of the limits on information costs as developed in Refs. [8, 9, 19, 27, 34].

Here, we re-emphasized the connection between Maxwell's demon and deterministic chaos. This has been discussed in several settings [14, 18, 34, 35]. Reference [14], for example, noted that the degree of chaos—the Kolmogorov-Sinai entropy of the engine's equivalent chaotic dynamical system—is the rate at which the engine (transiently) extracts disorganized thermal energy transforming it into work. In this way, the dynamical-systems connection provides a powerful approach to precisely accounting for the simultaneous flow of energy and information. Indeed, the connection is fated. A device that exhibits demon-like behavior is a machine that takes microscopic thermal fluctuations and amplifies them to macroscopic effect. A deterministic chaotic system is one in which microscopic variations in initial conditions yield macroscopic variation in its trajectory. With this in mind and reflecting on the century-plus history of clever constructions of Maxwellian demons, the current constructions appear especially suited to deepening our understanding of the relationship between energy and information.

VIII. LOOKING AHEAD

Beyond the conceptual insights that arise from Szilard's various engines, we can even be somewhat literal-minded. Szilard's first engine has been the inspiration for a diverse set of models and experimental realizations [36–40]. Recent developments in nanofabrication suggest attempting to realize Szilard's multi-particle engine, as well. For example, the graphene membrane fabrication techniques discussed in Ref. [41] can provide macroscale membranes with tunable pore size, pore density, and mechanical strength that are well suited to molecular gas separation. With the right gas ensemble, it is possible these are good candidates for the semi-permeable membranes required by Szilard's second engine. When coupled with modern nanomechanical device design, a tantalizing engineering challenge to implement Szilard's second engine presents itself.

Szilard's engines are simple enough to be readily analyzed, as we showed, with all hitherto relevant thermodynamic and information calculations analytically solvable. This thorough-going look at Szilard's original constructions gave reassuring results—results consistent with the fundamentals of both information theory and thermodynamics. Curiously, they are also consistent with very recent developments in nonequilbrium thermodynamics and fluctuation theory.

In particular, our investigation raised new questions. For one, Szilard's inequality $e^{-S_1/k} + e^{-S_2/k} \leq 1$ is more akin to very modern fluctuation theorems [18, 42–44] than to the fluctuation theories of his contemporaries. Is this another realm in which Szilard was prescient? This would not be surprising given that Szilard anticipated Shannon's information theory by two decades, Wiener's cybernetics, and the rise and fall of Landauer's Principle by half a century. We speculate that Szilard's constructions can again provide a simple platform—one giving a new view of detailed fluctuation theorems in action.

ACKNOWLEDGMENTS

We thank Alec Boyd for helpful discussions and the Telluride Science Research Center for their hospitality during visits. This material is based upon work supported by, or in part by, FQXi Grants FQXi-RFP-1609 and FQXi-RFP-IPW-1902, the John Templeton Foundation grant 52095, and U.S. Army Research Laboratory and the U. S. Army Research Office under contracts W911NF-13-1-0390 and W911NF-18-1-0028.

[1] J. C. Maxwell. *Theory of Heat*. Longmans, Green and Co., London, United Kingdom, ninth edition, 1888. 1, 13

[2] W. Thomson. Kinetic theory of the dissipation of energy. *Nature*, 9:441 EP –, 04 1874. 1

[3] H. Leff and A. Rex. *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*. Taylor and Francis, New York, 2002. 1

[4] W. Thomson. The sorting demon of Maxwell. In *Roy. Soc. Proc.*, volume 9, pages 113–114, 1879. 1

[5] L. Szilard. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Z. Phys.*, 53:840–856, 1929. 1, 4, 11, 12

[6] W. Lanouette and B. Szilard. *Genius in the Shadows: A Biography of Leo Szilard, The Man Behind The Bomb*. Skyhorse Publishing, New York, New York, 2013. 1

[7] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.*, 5(3):183–191, July 1961. 2, 6, 11

[8] C. H. Bennett. Thermodynamics of computation—A review. *Intl. J. Theo. Phys.*, 21:905, 1982. 2, 8, 12, 14

[9] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Develop.*, 5(3):183–191, 1961. 2, 8, 14

[10] K. Shizume. Heat generation required by information erasure. *Phys. Rev. E*, 52(4):3495–3499, 1995. 2, 8, 11

[11] F. N. Fahn. Maxwell's demon and the entropy cost of information. *Found. Physics*, 26:71–93, 1996.

[12] M. M. Barkeshli. Dissipationless information erasure and Landauer's principle. *arXiv:0504323*.

[13] T. Sagawa. Thermodynamics of information processing in small systems. *Prog. Theo. Phys.*, 127(1):1–56, 2012.

[14] A. B. Boyd and J. P. Crutchfield. Maxwell demon dynamics: Deterministic chaos, the Szilard map, and the intelligence of thermodynamic systems. *Phys. Rev. Lett.*, 116:190601, 2016. 2, 3, 6, 8, 9, 10, 11, 12, 13, 14

[15] R. K. Pathria and P. D. Beale. *Statistical Mechanics*. Butterwork-Heinemann, Oxford, United Kingdom, second edition, 1996. 2

[16] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948. 4

[17] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012. 5, 13

[18] U. Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports Prog. Phys.*, 75(12):126001, 2012. 14

[19] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa. Thermodynamics of information. *Nature Physics*, 11(2):131, 2015. 5, 14

[20] A. Lasota and M. C. Mackey. *Probabilistic Properties of Deterministic Systems*. Cambridge University press, Cambridge, United Kingdom, 1985. 8

[21] J. P. Crutchfield. Between order and chaos. *Nature Physics*, 8(January):17–24, 2012. 9

[22] S. H. Strogatz. *Nonlinear Dynamics and Chaos: with applications to physics, biology, chemistry, and engineering*. Addison-Wesley, Reading, Massachusetts, 1994. 9

[23] N. Barnett and J. P. Crutchfield. Computational mechanics of input-output processes: Structured transformations and the ϵ-transducer. *J. Stat. Phys.*, 161(2):404–451, 2015. 9

[24] C. J. Ellison, J. R. Mahoney, and J. P. Crutchfield. Prediction, retrodiction, and the amount of information stored in the present. *J. Stat. Phys.*, 136(6):1005–1034, 2009. 9

[25] A. Kolchinsky and D. H Wolpert. Dependence of dissipation on the initial distribution over states. *J. Stat. Mech.: Th. Expt.*, 2017(8):083202, 2017. 13

[26] W. Ross Ashby. *An Introduction to Cybernetics*. John Wiley and Sons, New York, second edition, 1960. 13

[27] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Leveraging environmental correlations: The thermodynamics of requisite variety. *J. Stat. Phys.*, 167(6):1555–1585, 2016. 13, 14

[28] S. Still. Thermodynamic cost and benefit of memory. *Phys. Rev. Let.*, 124(5):050601, 2020. 13

[29] J. Bengtsson, M. N. Tengstrand, A. Wacker, P. Samuelsson, M. Ueda, H. Linke, and S. M. Reimann. Quantum Szilard engine with attractively interacting bosons. *Phys. Rev. Let.*, 120(10):100601, 2018.

[30] M. H. Mohammady and J. Anders. A quantum Szilard engine without heat from a thermal reservoir. *New J. Physics*, 19(11):113026, 2017.

[31] S. Vaikuntanathan and C. Christopher. Modeling Maxwell's demon with a microcanonical Szilard engine. *Phys. Rev. E*, 83(6):061120, 2011.

[32] L. B. Kish and C. G. Granqvist. Energy requirement of control: Comments on Szilard's engine and Maxwell's demon. *Europhys. Lett.*, 98(6):68001, 2012.

[33] W. H. Zurek. Eliminating ensembles from equilibrium statistical physics: Maxwell's demon, Szilard's engine, and thermodynamics via entanglement. *Phys. Rep.*, 755:1–21, 2018. 13

[34] G. M. Zaslavsky. From Hamiltonian chaos to Maxwell's demon. *Chaos*, 5:653–661, 1995. 14

[35] J. M. R. Parrondo. The Szilard engine revisited: Entropy, macroscopic randomness, and symmetry breaking phase transitions. *Chaos: Interdisc. J. Nonlin. Sci.*, 11(3):725–733, 2001. 14

[36] T. Admon, S. Rahav, and Y. Roichman. Experimental realization of an information machine with tunable temporal correlations. *Phys. Rev. Let.*, 121(18):180601, 2018. 14

[37] J. V. Koski, V. F. Maisi, J. P. Pekola, and D. V. Averin. Experimental realization of a Szilard engine with a single electron. *Proc. Natl. Acad. Sci. USA*, 111(38):13786–13789, 2014.

[38] K. V. Koski, V. F. Maisi, T. Sagawa, and J. P. Pekola. Experimental observation of the role of mutual information in the nonequilibrium dynamics of a Maxwell demon. *Phys. Rev. Let.*, 113(3):030601, 2014.

[39] J. V. Koski and J.P. Pekola. Maxwell's demons realized in electronic circuits. *Compt. Rend. Phys.*, 17(10):1130–1138, 2016.

[40] L. B. Kish and C.-G. Granqvist. Electrical Maxwell demon and Szilard engine utilizing Johnson noise, measurement, logic and control. *PloS One*, 7(10), 2012. 14

[41] K. Choi, A. Droudian, R. W. Wyss, K.-P. Schlichting, and H. G. Park. Multifunctional wafer-scale graphene membranes for fast ultrafiltration and high permeation gas separation. *Sci. Adv.*, 4(11):eaau0476, 2018. 14

[42] T. Sagawa and M. Ueda. Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics. *Phys. Rev. Lett.*, 109(18):180602, 2012. 14

[43] C. Jarzynski. Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale. *Annu. Rev. Condens. Matter Phys.*, 2(1):329–351, 2011.

[44] G. E. Crooks and S. E. Still. Marginal and conditional second laws of thermodynamics. *arXiv:1611.04628*. 14

Appendix A: Entropy Change

Our goal is to determine ΔS in Szilard's second engine. The Sakur-Tetrode equation, the starting point, is:

$$S = Nk_{\mathrm{B}} \ln \frac{V}{N} + \frac{3}{2} Nk_{\mathrm{B}} \ln \frac{4\pi m U}{3h^2 N} + \frac{5}{2} Nk_{\mathrm{B}} \ .$$

Terms that remain constant throughout an engine cycle can be neglected for our purposes. Cursory inspection

reveals that ΔS will be determined by, at most:

$$Nk_{\mathrm{B}} \ln \frac{V}{N} + \frac{3}{2} Nk_{\mathrm{B}} \ln \frac{U}{N} \ .$$

In our case, the energy density term also drops out, since both the initial and final macrostates reach the equilibrium distribution $N_{\square} = \delta N$ and $N_{\bigcirc} = (1-\delta)N$. And so:

$$\frac{U}{N} = \delta \epsilon_A + (1-\delta)\epsilon_B + \mathrm{KE}_{\mathrm{avg}} \ ,$$

for any number of particles, where $\mathrm{KE}_{\mathrm{avg}}$ is the average kinetic energy. Finally, since each container has the same volume, the volume term does not contribute. Calculating the resulting entropy change is straightforward, but requires attention. After some algebra, we have:

$$\begin{aligned} \frac{\Delta S}{k_{\mathrm{B}}} &= -\delta \ln N_{\square} - (1-\delta) \ln N_{\bigcirc} + \ln N \\ &= -\left(\delta \ln \delta + 1 - \delta \ln(1-\delta)\right) \\ &= S(\delta) \ . \end{aligned}$$

Appendix B: Free Energies

A similar need arises for obtaining the free energy in Szilard's second engine. We starting observing that:

$$W_{\mathrm{qs}} = \Delta U_{\mathrm{res}} + \Delta U_{\mathrm{sys}} \ .$$

If reservoir volume remains constant, we note that $Q_{\mathrm{res}} = \Delta U_{\mathrm{res}}$. Then, using the expression above, we find $\Delta S_{\mathrm{res}} = Q_{\mathrm{res}}/T$ to be given by:

$$T\Delta S_{\mathrm{res}} = W_{\mathrm{qs}} - \Delta U_{\mathrm{sys}} \ .$$

Thus, the total entropy change, including both the system and the reservoir must then be:

$$T\Delta S_{\mathrm{res}} + T\Delta S_{\mathrm{sys}} = W_{\mathrm{qs}} + T\Delta S_{\mathrm{sys}} - \Delta U_{\mathrm{sys}} \ .$$

The difference of $T\Delta S_{\mathrm{sys}} - \Delta U_{\mathrm{sys}}$ is nothing more than the change in the system's free energy $-\Delta F_{\mathrm{sys}}$. Following Szilard's statement, we choose a reversible process ($\Delta S_{\mathrm{res}} + \Delta S_{\mathrm{sys}} = 0$). This allows us to find the work to drive the quasistatic step by:

$$W_{\mathrm{qs}} = F(\rho_0) - F(\rho) \ .$$

Adding W_{qs} and $W_{\Delta H}$ yields the the total driving work for both steps of the coming to equilibrium process:

$$\begin{aligned} W_{drive} &= \langle H_\rho \rangle_\rho - \langle H_0 \rangle_\rho + \left(\langle H_0 \rangle_{\rho_0} - TS(\rho_0)\right) \\ &\quad - \left(\langle H_\rho \rangle_\rho - TS(\rho)\right) \\ &= \langle H_0 \rangle_{\rho_0} - \langle H_0 \rangle_\rho + TS(\rho) - TS(\rho_0) \ . \end{aligned}$$

Appendix C: An Erasure Alternative

Consider a different choice for the final erasure step in the Szilard Map.

We could simply remove the partition and allow the gas to spontaneously re-equilibrate. This gives the same relationship in terms of entropy, but we do not get the advantage of a clear way to extract work that can be harnessed by the entropy increase.

Specifically, the change in entropy when mixing two identical gasses at different densities depends only on that part of the entropy given by $\ln \frac{V}{N}$. Initially, there are two separate gasses with the relevant entropy components:

$$S_A + S_B = Nk_{\mathrm{B}}\delta \ln \frac{\ell\gamma}{\delta N} + Nk_{\mathrm{B}}(1-\delta) \ln \frac{\ell(1-\gamma)}{(1-\delta)N} \ .$$

In the final state, we have a single gas:

$$S_F = Nk_{\mathrm{B}} \ln \frac{\ell}{N} \ .$$

The difference gives the entropy change:

$$\begin{aligned} \frac{\Delta S}{Nk_{\mathrm{B}}} &= \ln \frac{1}{N} - \delta \ln \frac{\gamma}{\delta N} - (1-\delta) \ln \frac{1-\gamma}{(1-\delta)N} \\ &= -\left((1-\delta) \ln \frac{1-\gamma}{1-\delta} + \delta \ln \frac{\gamma}{\delta}\right) \ . \end{aligned}$$

Appendix D: Szilard Engine Maps

The three discrete-time maps of the unit-square state-space that correspond to *measurement*, *control*, and *erasure* are, respectively:

$$\tau_{\mathrm{M}}(s,y) = \begin{cases} (s, y\gamma) & s < \delta \\ (s, y(1-\gamma)+\gamma) & s > \delta \end{cases} \ ,$$

$$\tau_{\mathrm{C}}(s,y) = \begin{cases} \left(\frac{s}{\delta}, y\right) & y < \gamma \\ \left(\frac{s-\delta}{1-\delta}, y\right) & y > \gamma \end{cases} \ ,$$

and:

$$\tau_{\mathrm{E}}(s,y) = \begin{cases} \left(s, \frac{y}{\gamma}\delta\right) & y < \gamma \\ \left(s, \frac{(y-\gamma)(1-\delta)}{1-\gamma} + \delta\right) & y > \gamma \end{cases} \ .$$

Taken together, and specializing to the case where $\delta = \gamma$, we have the composite map:

$$\tau_{\mathrm{Szilard}}(s,y) = \begin{cases} \left(\frac{s}{\delta}, y\delta\right) & s < \delta \\ \left(\frac{s-\delta}{1-\delta}, \delta + y(1-\delta)\right) & s > \delta \end{cases} \ .$$