

Not All Fluctuations are Created Equal: Spontaneous Variations in Thermodynamic Function

James P. Crutchfield* and Cina Aghamohammadi†
*Complexity Sciences Center and Department of Physics,
University of California at Davis, One Shields Avenue, Davis, CA 95616*
(Dated: October 16, 2024)

We identify macroscopic functioning arising during a thermodynamic system’s typical and atypical behaviors, thereby describing system operations over the entire set of fluctuations. We show how to use the information processing second law to determine functionality for atypical realizations and how to calculate the probability of distinct modalities occurring via the large-deviation rate function, extended to include highly-correlated, memoryful environments and systems. Altogether, the results complete a theory of functional fluctuations for complex thermodynamic nanoscale systems operating over finite times. In addition to constructing the distribution of functional modalities, one immediate consequence is a cautionary lesson: ascribing a single, unique functional modality to a thermodynamic system, especially one on the nanoscale, can be misleading; likely masking an array of simultaneous, parallel thermodynamic transformations that together may also be functional. In this way, functional fluctuation theory alters how we conceive of the operation of biological cellular processes, the goals of engineering design, and the robustness of evolutionary adaptation.

PACS numbers: 05.70.Ln 89.70.-a 05.20.-y 05.45.-a

Keywords: large deviation theory, thermodynamic formalism, fluctuation spectrum, entropy rate, fluctuation relations, nonequilibrium steady state, Maxwell’s Demon, information ratchet, Information Processing Second Law of Thermodynamics

I. INTRODUCTION

Almost all processes—highly correlated, weakly correlated, or correlated not at all—exhibit statistical fluctuations. Often physical laws, such as the Second Law of Thermodynamics, address only typical realizations—those identified by Shannon’s asymptotic equipartition property [1] and that emerge in the thermodynamic limit of an infinite number of degrees of freedom and infinite time [2]. Indeed, our interpretations of the functioning of macroscopic thermodynamic cycles are so focused. What happens, though during atypical behaviors, during fluctuations?

The limitation to typical behaviors is particularly a concern when it comes to information processing in thermodynamic systems or in biological processes, since fluctuations translate into errors in performing designed computing tasks or in completing operations required for maintenance and survival, respectively. As a consequence, one realizes that the Information Processing Second Law (IPSL) only identifies thermodynamic functioning supported by a system’s typical realizations [3]. Now, since observing typical realizations is highly probable over long times and goes to probability one in the thermodynamic limit, a definition of system functional-

ity based on typicality is quite useful. However, this renders the IPSL substantially incomplete and practically inapplicable—ignoring fluctuations over finite times and in microscopic systems. This is unfortunate. For example, while a system’s typical realizations may operate as an engine—converting thermal fluctuations to useful work—even “nearby” fluctuations (atypical, but probable realizations) behave differently, as Landauer erasers—converting available stored energy to dissipate stored information. How do we account for functioning during fluctuations? And, over long times how, in fact, does a fluctuating system operate at all?

The following answers these questions by introducing constructive methods that identify thermodynamic functioning during any system fluctuation. It shows how to use the IPSL to determine functionality for atypical realizations and how to calculate the probability of distinct modalities occurring via the large-deviation rate function. The lesson is that, falling short of the thermodynamic limit, one cannot attribute a unique functional modality to a thermodynamic system.

To get started, the next section motivates our approach, reviewing its historical background and basic set-up. The development then reviews thermodynamic functioning in information engines and fluctuation theory proper, before bringing the two threads together to analyze functional fluctuations in a prototype information engine.

* chaos@ucdavis.edu

† caghamohammadi@ucdavis.edu

II. FROM SZILARD TO FUNCTIONAL INFORMATION ENGINES

Arguably, Szilard’s Engine [4] is the simplest thermodynamic device—a controller leverages knowledge of a single molecule’s position to extract work from a single thermal reservoir. As one of the few Maxwellian Demons [5] that can be completely analyzed [6], it exposes the balance between entropic costs dictated by the Second Law and thermodynamic functionality during the operation of an information-gathering physical system. The net work extracted exactly balances the entropic cost. As Szilard emphasized: while his single-molecule engine was not very functional, it was wholly consistent with the Second Law, only episodically extracting useful work from a thermal reservoir.

Presaging Shannon’s communication theory [7] by two decades, Szilard’s major contribution was to recognize the importance of the Demon’s information acquisition and storage in resolving Maxwell’s paradox [5]. The Demon’s informational manipulations had an irreducible entropic cost that balanced any gain in work. The role of information in physics [8] has been actively debated ever since, culminating in a recent spate of experimental tests of the physical limits of information processing [9–15] and the realization that the degree of the control system’s dynamical instability determines the rate of converting thermal energy to work [6].

Though many years ago, Maxwell [5] and then Szilard [4] were among the first to draw out the consequences of an “intelligent being” taking advantage of thermal fluctuations [16]. Szilard’s Engine, however, and ultimately Maxwell’s Demon are not very functional: Proper energy and entropy book-keeping during their operation shows their net operation is consistent with the Second Law. As much energy is dissipated by the Demon as it extracts from the heat bath [4]. There is no net thermodynamic benefit. Are there Demons that are functional?

Only rather recently was an exactly-solvable Maxwellian engine proposed that exhibited functionality, extracting net work each cycle by decreasing physical entropy at the expense of positive change in a reservoir’s Shannon information [17]. There, the Demon generated directed rotation leveraging the statistical bias in a memoryless information reservoir to compensate for the transfer of high-entropy energy in a thermal reservoir to low-entropy energy that performed the rotational work. Since then, an extensive suite of studies analyzed more complex *information engines* [3, 18–28]. Here, and in contrast several of these studies, we emphasize engines that leverage information reservoirs with large, unrestricted memories while interacting with complex, correlated environments.

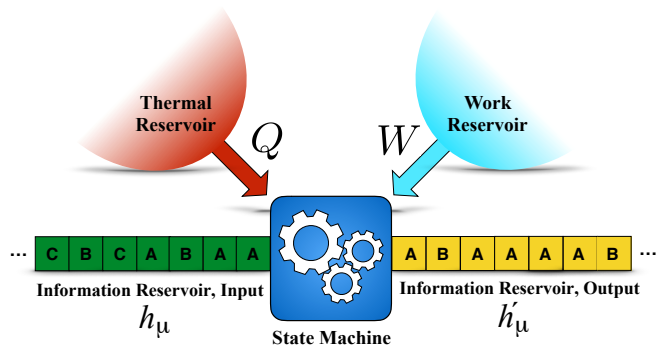


FIG. 1. Information engine: A thermodynamically-embedded state machine transforms symbols on the input tape with Shannon entropy rate h_μ to the output tape with Shannon entropy rate h'_μ . The input and output tapes comprise an information reservoir coupled, as are the thermal and work reservoirs, to the state machine. Tape symbols come from the same alphabet; e.g., as here, the set $\{A, B\}$. According to the Information Processing Second Law [3] by changing the Shannon entropies of the input and output symbol sequences the information engine functions to convert heat Q to work W or work to heat depending on the sign of the entropy change $h'_\mu - h_\mu$. Positive work and heat indicate energy flows into the Machine.

Figure 1 illustrates the general design for an information engine. The Demon, now denoted “State Machine”, is in contact with three reservoirs: thermal, work, and information. Each reservoir provides a distinct thermodynamic resource which the engine transforms. The thermal reservoir stores high-entropy energy; the work reservoir, low-entropy energy; and the information reservoir zero-energy Shannon information. The information reservoir consists of input and output tapes with cells storing discrete symbols.

The State Machine functions step by step. To process information on the tapes, it reads a symbol from an input cell and writes a symbol to an output tape cell and changes its internal state. The tapes then shift one cell presenting new input and output cells to the State Machine. In terms of the energetics, in the first step, a controller couples the symbol read from the input tape cell to the Machine. The controller may need positive or negative work from the work reservoir. The heat transfer is zero since, for our purposes here, we assume the process is relatively fast. In the second step, the state of the coupled cell-system transitions as a result of being in contact with thermal reservoir. Then the thermal reservoir induces a Markovian dynamics over the coupled cell-system joint states. This step is completely performed by the thermal reservoir and as a result there is heat transfer between the machine and thermal reservoir. The controller is absent and so the work done in this step is zero. In the third step, the controller decouples the output

state from machine state. Again, the work here can be nonzero, but the heat flow is zero.

There are three types of functioning. In the first, the state machine extracts heat from the thermal reservoir and performs work on work reservoir by producing output symbol sequences with higher entropy than the input sequences. In this case, we say the machine functions as an *engine*. In the second, machine decreases the output sequence entropy below that of the input by extracting work from work reservoir and dumping that energy to the thermal reservoir. In this way, the machine acts as an *information eraser*. Finally, the third (non)functionality occurs when the machine uses (wastes) work energy to randomize output. Since the randomization of the input can happen spontaneously without wasting work—similar to the engine mode—we say the machine functions as a *dud*; it is a wasteful randomizer.

III. ENVIRONMENT AND ENGINE REPRESENTATIONS

There are two technical points that need to be called out here. First, we imagine the engine interacts with a complex environment. This means that we allow the input sequence to be highly correlated with very long memory. Formally, the input sequence considered as a stochastic process is not necessarily Markovian. Denote the probability distribution over the input's bi-infinite random variable chain by $\mathbb{P}(\cdots X_{-1}X_0X_1\cdots)$, where X_t is the random variable at time t . Then, the input sequence's *Markov order* R is:

$$\mathbb{P}(X_t|\cdots X_{-1}X_0X_1\cdots X_t) = \mathbb{P}(X_t|X_{t-R}\cdots X_t) .$$

And so, by complex environment we mean that input sequences to the machine have large R —the environment remembers long histories. Second, even though the machine has finite number of states, we allow it to also have long memory. This simply means that, via its states, the machine can remember the last, perhaps large, number of inputs.

One concludes from the first point about complex environments that Markov chains are not powerful enough to represent correlated inputs, especially for the general case we analyze. We need a less restrictive representation and so use *hidden Markov models* (HMMs), which are known to be more powerful in the sense that, using only a finite number of internal states, they can represent infinite Markov-order processes. We use HMMs to represent the mechanisms generating both input sequences and output sequences.

A process \mathcal{P} 's HMM is given as a pair $\{\mathcal{S}, \{T^{(x)} : x \in$

$\mathcal{A}\}\}$. \mathcal{S} is HMM's hidden states. $T^{(x)}$ for any particular x is a substochastic matrix or state-to-state transition matrix for transitions that generate symbol x . \mathcal{A} is the alphabet of generated symbols.

Similarly, we conclude from the second point that more powerful machinery is needed to handle general stochastic mappings with long memory. We use stochastic finite-state *transducers* [29] as they are powerful enough to represent the mappings we use in the following. (Several of the technical contributions stem directly from showing how to work directly with these powerful representations.)

A transducer representation is a pair $\{\mathcal{S}, \{T^{(x \rightarrow y)} : x \in \mathcal{A}_x, y \in \mathcal{A}_y\}\}$. \mathcal{S} is transducer's states. $T^{(x \rightarrow y)}$ for any particular x and y is a substochastic matrix or state-to-state transition matrix for transitions that for input x generate symbol y . \mathcal{A}_x and \mathcal{A}_y are the alphabet for input and output symbols.

The following will demonstrate how these choices of representation greatly facilitate analyzing the dynamics and thermodynamics of information engines.

IV. THERMODYNAMIC FUNCTIONING: WHEN IS AN ENGINE A REFRIGERATOR?

Thermodynamic functionality is defined in terms of the recently introduced *information processing second law* (IPSL) [3] which bounds the thermodynamic resources required, such as work, to perform a certain amount of information processing:

$$\langle W \rangle \leq k_B T \ln 2 (h'_\mu - h_\mu) , \quad (1)$$

where k_B is Boltzmann's constant and T is the environment's temperature. The IPSL relates three macroscopic system measures: the input's Shannon entropy rate h_μ , the output's entropy rate h'_μ , and the the average work $\langle W \rangle$ done on the work reservoir per engine cycle:

$$\begin{aligned} h_\mu &= \lim_{\ell \rightarrow \infty} \frac{H[X_0, X_1, \cdots, X_{\ell-1}]}{\ell} , \\ h'_\mu &= \lim_{\ell \rightarrow \infty} \frac{H[X'_0, X'_1, \cdots, X'_{\ell-1}]}{\ell} , \text{ and} \\ \langle W \rangle &= \lim_{\ell \rightarrow \infty} \frac{1}{\ell} \sum_{w \in \mathcal{A}^\ell} \mathbb{P}(w) f(w) . \end{aligned} \quad (2)$$

Here, $H[\cdot]$ is the Shannon entropy of the specified random variables. $f(w)$ is defined as follows. Since the machine stochastically maps inputs to outputs a given input sequence w typically map to many distinct output sequences. Then $f(w)$ denotes the average work done by feeding word w to the machine, averaging over all the

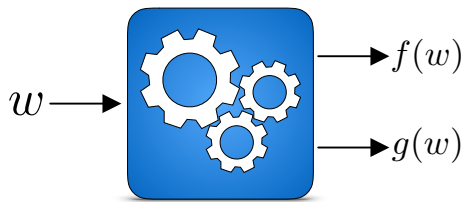


FIG. 2. Input-dependent work and information: Feeding in every single word w , on average the Machine generates work $f(w)$ and information $g(w)$.

possible mappings from w ; see Fig. 2.

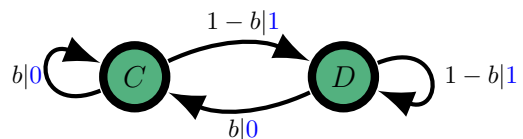
That is, thermodynamic functioning is determined by the signs of $\langle W \rangle$ and $h'_\mu - h_\mu$. Since there are two possible signs for each, there are four distinct cases. However, the IPLS forbids the case $\langle W \rangle > 0$ and $h'_\mu - h_\mu < 0$. And so, there are three thermodynamically functional modes: *Engine*, *Eraser*, and *Ineffective randomizer*; see Table I [3]. When operating as an Engine, the machine absorbs heat from thermal reservoir and converts it to work by mapping the input sequence to a higher entropy-rate output sequence. Thus, the net effect is to randomize the input. When operating as an eraser, the machine reduces the input entropy by consuming work from work reservoir and dumping it as high-entropy energy to the heat reservoir. In the third case, the machine does not function usefully at all. It is an ineffective randomizer, consuming work to randomize to input string. It wastes work, low-entropy energy.

Modality	Function	Net Work	Net Computation
Engine	Extracts high-entropy energy from the thermal reservoir, converts it into low-entropy work by randomizing output	$\langle W \rangle > 0$	$h'_\mu - h_\mu > 0$
Eraser	Uses low-entropy energy from work reservoir to reduce input randomness, exhausting high-entropy energy to thermal reservoir	$\langle W \rangle < 0$	$h'_\mu - h_\mu < 0$
Ineffective randomizer	Wastes stored work (low-entropy energy) to randomize output	$\langle W \rangle < 0$	$h'_\mu - h_\mu > 0$

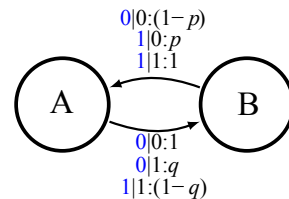
TABLE I. Thermodynamic functionings for information engines, as determined by the information processing second law of Eq. (1).

V. A FUNCTIONAL INFORMATION ENGINE

To ground these ideas, consider a prototype information engine—the *information ratchet* introduced in Ref. [3]. The engine, Fig. 3, specifies the distribution of inputs and the states and transition structure of the engine’s state machine. The inputs come from flipping a coin with bias b for Heads (‘0’). That is, the input is a memoryless, independent, and identically distributed (IID) stochastic process. Its generating mechanism is depicted as the hidden Markov model in Fig. 3(a) with two states, A and B . Together the current state and transition taken determine the statistics of the emitted symbol. Similarly, the engine’s mechanism is represented by the finite-state transducer in Fig. 3(b). Transducer transitions are labeled. For example, if the machine is in state B and the input is 0, then with probability p the output emitted is 1 and the machine state changes to A . This is shown by an edge labeled by $1|0 : p$ going from state A to B .



(a) Input Information Reservoir



(b) Input-Output Transducer

FIG. 3. (a) Hidden Markov model that generates a biased coin input string $x_t x_{t+1} \dots$ with bias $\Pr(X = 0) = b$. Edge labels $x : p$ indicate a state-to-state transition of probability p that emits symbol x . (b) The information engine’s internal mechanism is a transducer. Its edge labels $x|x' : p$ indicate a state-to-state transition of probability p taken on reading input symbol x that emits symbol x' . (Reprinted from Ref. [3] with permission.)

At this point only the engine’s information processing has been specified. To design a physical system that implements the transducer, we first define the energetics for inputs and for machine states and transitions:

$$\begin{aligned}
 E(0) &= E(1) = 0, \\
 E(A) &= 0, \text{ and} \\
 E(B) &= e_1,
 \end{aligned}$$

where e_1 is a parameter. Second, we define the energetics

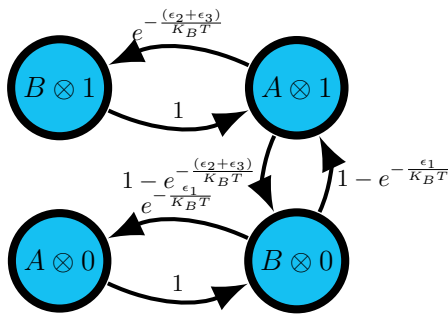


FIG. 4. Markovian detailed balance dynamics induced by contact with the thermal reservoir in the coupled system (input-symbol and machine state).

for joint symbol-states:

$$E(A \otimes 0) = 0, \quad E(B \otimes 0) = -\epsilon_1, \\ E(A \otimes 1) = -\epsilon_2, \quad E(B \otimes 1) = +\epsilon_3.$$

The energies ϵ_i are further constrained:

$$e^{(\epsilon_1 - \epsilon_2)/k_B T} = \frac{1 - e^{-(\epsilon_2 + \epsilon_3)/k_B T}}{1 - e^{-\epsilon_1/k_B T}}.$$

Third, we specify Markovian detailed-balanced dynamics over the coupled system (input + state machine) that is induced by thermal reservoir; see Fig. 4. To guarantee that this dynamic generates the same stochastic mapping as the transducer in Fig. 3(b), we must relate the energetics to stochastic-transition parameters p and q :

$$p = 1 - e^{-\epsilon_1/k_B T} \\ q = 1 - e^{-(\epsilon_2 + \epsilon_3)/k_B T}.$$

The average work done on the work reservoir is then:

$$\langle W \rangle = \frac{k_B T}{2} [(pb - q + qb) \ln(q/p) \\ + (1 - b)q \ln(1 - q) + pb \ln(1 - p)]. \quad (3)$$

See Ref. [3] for calculation details.

The Shannon entropy rates of input and output sequences also can be calculated directly:

$$h_\mu = H(b) \\ \equiv -b \log_2 b - (1 - b) \log_2(1 - b) \\ h'_\mu = \frac{H(b(1 - p))}{2} + \frac{H((1 - b)(1 - q))}{2}. \quad (4)$$

Thus, the energies $\epsilon_{1,2,3}$ and control b are the only free parameters. They control the engine's behavior and, through the IPSL modalities in Table I, its functionality. Reference [3] gives a complete analysis of this information

engine's thermodynamic functioning.

Summarizing for general information engines, one specifies:

1. Input process as an HMM, and
2. Markovian detailed-balance dynamic over the coupled system of input and machine states as a finite-state transducer with consistent energy assignments.

This prepares us to analyze fluctuations in an information engine interacting with the complex environment specified by the input process.

VI. ENGINES IN FLUCTUATING ENVIRONMENTS: THE STRATEGY

Hidden in this and often unstated, but obvious once realized, Maxwellian Demons cannot operate unless there are statistical fluctuations. Szilard's Engine cleverly uses and skirts this issue since it contains only a single molecule whose behaviors, by definition, are nothing but fluctuations—single realizations. There is no large ensemble over which to average. The information gleaned by the engine's control system (Demon/Machine) is all about the “fluctuation” in the molecule's position. And, that information allows the engine to *temporarily* extract energy from a heat reservoir. In short, fluctuations are deeply implicated in the functioning of thermodynamic systems. The following isolates the underlying statistical mechanisms.

The distinct types of thermodynamic functioning—engine, eraser, or dud—are based on three average quantities: average work produced $\langle W \rangle$, the input sequences' Shannon entropy rate h_μ , and the output sequences' Shannon entropy rate h'_μ [3, 18–28, 30]. As a result, their definition's concern the thermodynamic limit of infinitely long sequences being fed into the machine. Of course, the situation is practically quite different: the engine works with and operates due to finite-length sequences.

To overcome this—and so to develop a theory of functional fluctuations—the following is burdened with precisely delineating the limitations inherent in the infinite-length definitions above. It shows that, for any finite length, the functionality definitions are limited to describing properties only of a unique subset of events—the so-called typical set of realizations as identified by the asymptotic equipartition property of information theory [1]. To do this, first we redefine the three quantities—work and entropy rates—as averages over all the possible input sequences of a given length. Second, we define three new unweighted-average quantities, but this time they are explicitly limited to typical realizations. Third,

we demonstrate that the differences between first three averages and second three can be made arbitrarily small. Since the second kind of averages are unweighted, the closeness result tells us that the average quantities are features of the typical set and not of any other subset of the input sequences. In point of fact, they do not describe atypical behaviors (statistical fluctuations) and so cannot be used to define thermodynamic functions arising from fluctuations.

One technical reason behind this result is that, for the three averages, the functions being averaged are linearly bounded from above by the input-sequence length. The conclusion is that the original quantities can give information only about system functionality for the specific subset of typical realizations. Of course, since observing realizations in this subset is highly probable for long sequences and has probability one in the thermodynamic limit of infinite length, the original functionality definition is quite useful. Our goal, though, is to show just how incomplete it is and in important ways that must be overcome to analyze fluctuations in functioning.

In short, the following consistently extends the original definitions to other realization subsets—the fluctuations or atypical sets. The net result is that the theory covers the set of any realization for any finite length. Given that, we introduce a method to calculate the new functionality for these different fluctuation subsets. This completes the picture of functional fluctuations for finite, but long lengths. We go on to find the large deviation rate for the new definition of functionality. An important contribution in this is that all of the results also apply to input sequences and machines with long memories, given that the latter are stochastic finite-state machines. This should be contrasted with developments, cited above, that assume memoryless or order-1 Markov systems. We return to discuss related work at the end, once the results are presented.

VII. FUNCTIONING SUPPORTED BY TYPICAL REALIZATIONS

A picture of a system’s behavioral fluctuations can be developed in terms of (and deviations from) asymptotic equipartition. Let’s review. Consider a given process \mathcal{P} and let \mathcal{A}^n denote the set of its possible length- ℓ realizations. Then, for an arbitrary $0 < \epsilon \ll 1$, the process’ *typical set* is:

$$A_\epsilon^\ell \equiv \{w : 2^{-\ell(h_\mu + \epsilon)} \leq \mathbb{P}(w) \leq 2^{-\ell(h_\mu - \epsilon)}, w \in \mathcal{A}^\ell\}. \quad (5)$$

This set consists of realizations whose probability scales with the process’ entropy rate [1, 31, 32]. Moreover, the

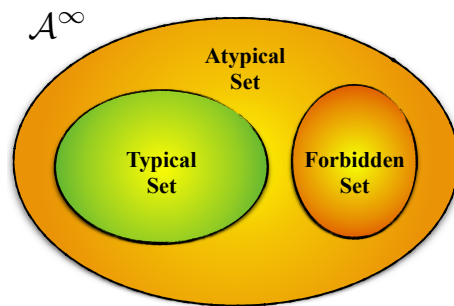


FIG. 5. For a given process, the space \mathcal{A}^∞ of all sequences is partitioned into forbidden sequences, sequences in the typical set, and sequences neither forbidden nor typical—the *atypical* or rare sequences.

Shannon-McMillan-Breiman theorem [7, 33, 34] gives the probability of observing one of these realizations. That is, for a given $\epsilon \ll 1$, sufficiently large ℓ^* , and $w \in \mathcal{A}^\ell$:

$$\mathbb{P}(w \in A_\epsilon^\ell) \geq 1 - \epsilon, \quad (6)$$

for all $\ell \geq \ell^*$. There are three lessons:

1. Asymptotic equipartition: Eq. (5) says that the probability of each sequence in the typical decays at approximately the same rate.
2. Typicality: Eq. (6) says that for large ℓ the probability of observing some typical realization goes to one. Overwhelmingly, they are what one observes.
3. Fluctuations: Conversely, the probability of observing realizations outside the typical set is close to zero. These are the sets of rare sequences or, what we call, *fluctuations*.

As a result, sequences generated by a stationary ergodic process fall into one of three partitions; as depicted in Fig. 5. The first contains sequences that are never generated; they fall in the *forbidden set*. The second is the typical set. And, the last contains sequences in a family of *atypical sets*—realizations that are rare to different degrees. Appendix A illustrates these for a Biased Coin Process.

What does this partitioning say about fluctuations in thermodynamic functioning? Recall the functionings identified by the IPSL, as laid out in Table I. That is, for a given input process, transducer, and temperature, thermodynamic functionality is controlled by three quantities: the average work $\langle W \rangle$ generated by the transducer when it operates on the input process, the Shannon entropy h_μ of the input process, and the Shannon entropy h'_μ of output process.

Appendix B proves that the difference between average work $\langle W(\ell) \rangle$ over all sequences and that $\langle W(\ell) \rangle_{TS}$

defined for typical set is small for sufficiently large ℓ . For all practical purposes, they are equal. This, together with recalling that $\langle W(\ell) \rangle_{TS}$ is an unweighted average of works $f(w)$ for $w \in \mathcal{A}_\ell^t$, provides an operational interpretation of works used in typical-set-defined functionality.

Similarly, App. C proves that the average generated information, when the transducer is fed the whole set, is essentially equal to the average information generated when the transducer is fed the typical set without probability weights.

From Eq. (5) it is also clear that Shannon entropy rate of the input process is also function of typical set. This demonstrates that all three quantities— $\langle W \rangle$, h_μ , and h'_μ —effectively measure properties of the typical set and not of other (atypical) partitions. Recalling that these three quantities also determine the thermodynamic via the IPSL functionality highlights that the previously-defined functionality is limited. Next, we remove this limitation, extending the thermodynamic functionality to the whole set of partitions.

VIII. FUNCTIONING OUTSIDE TYPICAL REALIZATIONS

The last section established that the average work $\langle W(\ell) \rangle$ and input and output entropy rates can be used, for $\ell \gg 1$, to identify system functionality for typical realizations. At last, “typical” has a precise operational meaning. Moreover, as $\ell \rightarrow \infty$ the fraction of information available about the functionality of realizations outside the typical set vanishes. Since the probability of observing realizations in the typical set at large ℓ approaches one, the definition of functionality based on $\langle W \rangle$ and the entropies is very useful.

However, one should not forget that this definition is limited, applying only to one particular subset of realizations. As a result, the associated definition of functionality gives an incomplete picture. How incomplete? Note that the size of the typical set grows like $2^{h_\mu \ell}$ and the size of the whole set, excluding forbidden realizations, grows as $2^{h \ell}$, where h is the input process’ *topological entropy* [35]. Generally, $h > h_\mu$ [36]. And so, the relative size of the typical set shrinks exponentially with ℓ as $2^{-(h-h_\mu)\ell}$, even though the probability of observing typical realizations converges to one. The lesson is that, at finite ℓ , only considering the typical set misses exponentially many— $2^{-(h-h_\mu)\ell}$ —possibly functional, observable realizations. With this as motivation, we are ready to define functionality for all realizations—typical and atypical—allowing one to describe “nearby” functionalities that arise during fluctuations out of the typical set. The goal is a complete the picture of functional fluctuations for finite, but long

realizations.

What engine functionalities do atypical realizations support? The very first step is to partition the set \mathcal{A}^ℓ of all possible realizations into the subsets of interest. How? We must find a suitable, physically-relevant parametrization of realization subsets. We call the collections a process’ *atypical sets*, using degrees of typicality as a parameter.

A key step in the last section was to realize that functionality is defined for unweighted sets of realizations. Recalling Eq. (5)’s definition of typical set, the normalized minus logarithm of probabilities—effectively a decay rate—of all the words in the typical set is sandwiched by small deviations ($\pm\epsilon$) from the Shannon entropy rate:

$$h_\mu - \epsilon \leq -\frac{1}{\ell} \log_2 \mathbb{P}(w) \leq h_\mu + \epsilon .$$

This is the main reason of why $\langle W \rangle_{TS}$ is approximately the unweighted average work and, consequently, why functionality is operationally defined for an unweighted set—the typical set. This provides an essential clue as to how to partition the set \mathcal{A}^ℓ of all possible realizations, at fixed length ℓ .

We collect all the realizations with the same probability in the same subset, labeling it with a decay rate denoted u :

$$\Lambda_{u,\ell} = \left\{ w : -\frac{\log_2 \mathbb{P}(w)}{\ell} = u, w \in \mathcal{A}^\ell \right\} . \quad (7)$$

Defining $\Lambda_u = \lim_{n \rightarrow \infty} \Lambda_{u,n}$, it is easy to show that $\Lambda_u \subset \mathcal{A}^\infty$ are disjoint and partition \mathcal{A}^∞ .

Technically, this definition for the (parametrized) subsets of interest is necessary to guarantee consistency with the previously-defined typical-set notion of functionality.

The parameter u , considered as a random variable, is sometimes called a *self process* [38]. Figure 6 depicts these subsets as “bubbles” of equal decay rate. Equation (5) says the typical set is that bubble with decay rate equal to the process’ Shannon entropy rate: $u = h_\mu$. All the other bubbles contain rare events, some rarer than others; in the sense that they exhibit faster or slower probability decay rates.

The previous section show that for $\ell \gg 1$ the averaging operator $\langle \cdot \rangle$ yields a statistic essentially about the typical set. Now, consider the situation in which we are interested in the functionality of another subset with decay rate $u \neq h_\mu$. How can we use the same operator to find functionality arising from this subset?

If someone presents us with another process \mathcal{P}^u whose typical set is Λ_u and we feed this new process into the system, instead of the original input process, then the operator can be used to identify the functionality of real-

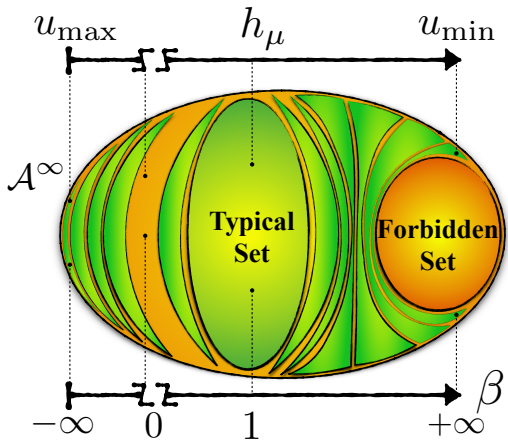


FIG. 6. Space \mathcal{A}^∞ of all sequences partitioned into subsets Λ_u —isoenergy-density or equal probability-decay-rate bubbles—in which all sequences in the same Λ_u have the same energy density u . The typical set is one such bubble with energy equal to Shannon entropy rate: $u = h_\mu$. Another important class is the forbidden set, in which all sequences do not occur. The forbidden set can also be interpreted as the subset of sequences with infinite positive energy. By applying the map \mathcal{B}_β to the process and changing β continuously from $-\infty$ to $+\infty$ (excluding $\beta = 0$) one can generate any atypical class of interest $\Lambda_{u,\ell}^\mathcal{P}$. $\beta \rightarrow -\infty$ corresponds to the most probable sequences with the largest energy density u_{\max} , $\beta = 1$ corresponds to the typical set and $\beta \rightarrow +\infty$ corresponds to the least probable sequences with the smallest energy density u_{\min} . (Reprinted with permission from Ref. [37].)

izations in Λ_u . Now, the question comes up as to whether this process exists at all and, if so, can we find it?

The answer to the first question is positive, since we made certain to define the atypical subsets in a way consistent with the definition of the typical set. And, by definition, all the sequences in the subset Λ_u have the same decay rate.

The answer to the second question is also positive. As argued earlier, we use hidden Markov models (HMMs) as our choice of process representation. Denote process \mathcal{P} 's HMM by $M(\mathcal{P}) = \{\mathcal{S}, \{T^{(x)} : x \in \mathcal{A}\}\}$. The question is now framed, What is $M(\mathcal{P}^u)$?

To answer, define a new process \mathcal{P}_β with HMM $M(\mathcal{P}_\beta) = \{\mathcal{S}, \{\mathbf{S}_\beta^{(x)}, x \in \mathcal{A}\}\}$. Notice both $M(\mathcal{P}_\beta)$ and $M(\mathcal{P})$ have the same states \mathcal{S} and same alphabet \mathcal{A} . The substochastic matrices of $M(\mathcal{P}_\beta)$ are related to the substochastic matrices of $M(\mathcal{P})$ via the following construction [37, 39]:

1. Pick a $\beta \in \mathbb{R}$.
2. For each $x \in \mathcal{A}$, construct a new matrix $\mathbf{T}_\beta^{(x)}$ for which $(\mathbf{T}_\beta^{(x)})_{ij} = (\mathbf{T}^{(x)})_{ij}^\beta$.
3. Form the matrix $\mathbf{T}_\beta = \sum_{x \in \mathcal{A}} T_\beta^{(x)}$.

4. Calculate \mathbf{T}_β 's maximum eigenvalue $\hat{\lambda}_\beta$ and corresponding right eigenvector $\hat{\mathbf{r}}_\beta$.
5. For each $x \in \mathcal{A}$, construct new matrices $\mathbf{S}_\beta^{(x)}$ for which:

$$(\mathbf{S}_\beta^{(x)})_{ij} = \frac{(\mathbf{T}_\beta^{(x)})_{ij} (\hat{\mathbf{r}}_\beta)_j}{\hat{\lambda}_\beta (\hat{\mathbf{r}}_\beta)_i}.$$

We defined the new process \mathcal{P}_β by constructing its HMM. We now use the latter to produce an atypical set of interest, say, $\Lambda_{u,\ell}$.

Theorem 1. *Within the new process \mathcal{P}_β , in the limit $\ell \rightarrow \infty$ the probability of generating realizations from the set $\Lambda_{u,\ell}$ converges to one:*

$$\lim_{\ell \rightarrow \infty} \Pr(\Lambda_{u,\ell}) = 1,$$

where the energy density is:

$$u = \beta^{-1} (h_\mu(\mathcal{P}_\beta) - \log_2 \hat{\lambda}_\beta). \quad (8)$$

Additionally, in the same limit the process \mathcal{P}_β assigns equal energy densities (probability decay rates) to all $w \in \Lambda_{u,\ell}$.

Proof. See Ref. [37].

In this way, for large ℓ the process \mathcal{P}_β typically generates realizations in the set $\Lambda_{u,\ell}$ and with the specified energy u . The process \mathcal{P}_β is variously called the *auxiliary*, *driven*, or *effective process* [40–42].

Using Eq. (8), one can show that for any u there exists a unique and distinct $\beta \in \mathbb{R}$ and, moreover, that u is a decreasing function of β . And so, we can equivalently denote the process \mathcal{P}_β by \mathcal{P}^u . More formally, every word in Λ_u with probability measure one is in the typical set of process \mathcal{P}_β . Thus, sweeping $\beta \in [-\infty, \infty]$ controls which subsets (atypical sets) outside the typical set we focus on. And, applying the operator $\langle \cdot \rangle$ determines the engine functionality for realizations in that subset, as we now show.

IX. FUNCTIONAL FLUCTUATIONS

Let's draw out the consequences and applications of this theory of functional fluctuations. First, we ground the results by identifying the range of functionality that arises as an information ratchet (introduced earlier) operates. Then, we turn to show how to calculate the probability of its fluctuating functionalities.

A. An Information Ratchet Fluctuates

Recall the information ratchet introduced in Sec. IV, but now set its Markov dynamic parameters $p = 0.2$ and $q = 0.6$ and put it in contact with an information reservoir that generates IID symbol sequences with bias $b = 0.9$. Operating the input reservoir for a sufficiently long time, with high probability we observe a sequence that has nearly 90% 0s in it. Using Eqs. (3) and (4), we see positive work $\langle W \rangle > 0$ and positive entropy production $h'_\mu - h_\mu > 0$. Then, according to the IPSL functionalities in Table I, the ratchet typically operates as an engine.

What thermodynamic functionalities occur when the input fluctuates outside the typical set? Sweeping β controls which subsets outside the typical set are expressed and, consequently, which fluctuation subsets are accessible. Recall that the input process is specified by the unifilar HMM in Fig. 3(a). For this input, as a result of the ratchet design, $M(\mathcal{P}_\beta)$ is the same as $M(\mathcal{P})$, except that b is shifted to $\hat{b} = b^\beta / (b^\beta + (1-b)^\beta)$. Different sequence-probability decay rates u are calculated from Eq. (8). Then, feeding the new process to the ratchet, $\langle W \rangle$ is calculated from Eq. (3), again by changing b to \hat{b} . Denote this work quantity $\langle W \rangle(u)$. Figure 7 showing the dissipated work $\langle W \rangle(u)$ and the difference between output and input Shannon entropy rate versus fluctuation decay rate u . There are several observations to make, before associating thermodynamic function.

First, let's locate the input typical set. It occurs at a u such that $\beta = 1$. The figure identifies it with vertical line, so labeled.

Second, the input process' ground states occur as $\beta \rightarrow \infty$ since u is decreasing function of β . As a consequence of Eq. (7) this subset corresponds to the sequence with the highest probability. In this case, this is the all-0s sequence with $u_{\min} = -\log_2(b) \simeq 0.152$. The other extreme is at u_{\max} , corresponding to the lowest-probability, allowed sequence. This is the all-1s sequence with $u_{\max} = -\log_2(1-b) \simeq 3.32$. Note that there is only a single sequence associated with u_{\max} and only one with u_{\min} .

Third, to complete the task of identifying function, we must determine the average work $\langle W \rangle$ as a function of energy u . From the figure, we see that the dissipated work $\langle W \rangle$ is linear in the decay rate u . Appendix D derives this and also shows that the maximum work, over all subsets—all β or all allowed decay rates u —is independent of the input process bias. This is perhaps puzzling as bias clearly controls the ratchet's thermodynamic behavior. Thus, assuming an IID input, the maximum work is a property of the ratchet itself and not the input—the maximum work playing a role rather analogous to how

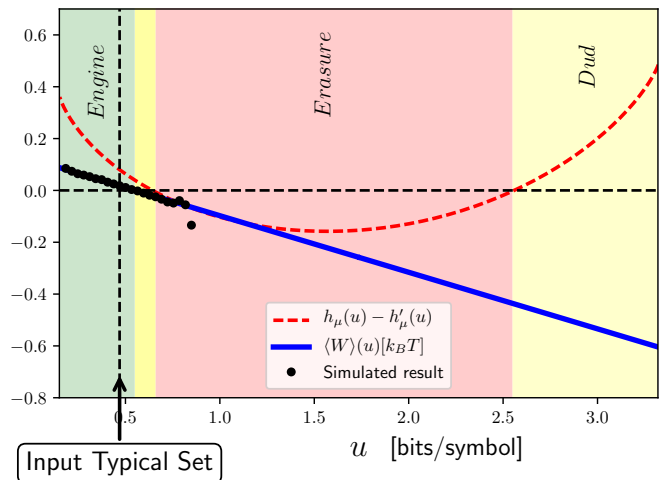


FIG. 7. Average work $\langle W \rangle(u)$ (blue line) and difference $h'_\mu - h_\mu$ between output and input Shannon entropy rate, respectively, (red dashed line) versus decay rate u for different atypical sets (fluctuations). In this, information transducer with parameters $p = 0.2$ and $q = 0.6$ is driven by an IID input source with bias $b = 0.9$. The Table I has been used to identify functionality of different fluctuations subsets: Engine (green), Eraser (red), and Dud (yellow, two regions).

Shannon's channel capacity is a channel property.

To better understand how the ratchet operates thermodynamically, consider the ground state of the input process; which as just noted has only a single member, the all-0 sequence with zero entropy rate $h_\mu = 0$. If we feed this sequence into the ratchet, the ratchet adds stochasticity which appears in the output sequence. The first 0 fed to the ratchet leads to a 0 on the output. For the next 0 fed-in, with probability p the ratchet outputs 1 and with probability $1-p$ it outputs 0. The entropy rate of output sequence then is $h'_\mu = \frac{1}{2} H(p) \simeq 0.36$.

To generate this sequence we simply use the ϵ -machine in Fig. 3 with $b = 1$. With this biased process as input, using Eq. (3) we find $\langle W \rangle(u_{\min}) \simeq 0.0875 > 0$. Table I then tells us that if we feed the ground state of the input process to the ratchet, it functions as an engine. At the other extreme U_{\max} , the only fluctuation subset member is the all-1s sequence with $h_\mu = 0$. Again, the ratchet adds stochasticity and the output has $h'_\mu = \frac{1}{2} H(q) \simeq 0.485$. To generate this input sequence we simply use the ϵ -machine in Fig. 3 with $b = 0$. With this process as an input, we use Eq. (3) again and find negative work $\langle W \rangle(U_{\max}) \simeq -0.6$. Table I now tells us that feeding in this extreme sequence (input fluctuation) the ratchet functions as a Dud.

Overall Table I allows one to identify the regimes of u associated with distinct thermodynamic functionality. These are indicated in Fig. 7 with the green region corresponding to Engine functioning, red to Eraser

functioning, and yellow to Dud. We conclude that the ratchet’s thermodynamic functioning depends substantially on fluctuations and so will itself fluctuate over time. In particular, Engine functionality occurs only at relatively low input fluctuation energies, seen on Fig. 7’s left side, and encompasses the typical set, as a consequence of our design. Rather nearby the Engine regime, though, is a narrow one of no functioning at all—a Dud. In fact, though the ratchet was designed as an Engine, we see that over most of the range of fluctuations, with the given parameter setting, the ratchet operates as an Eraser.

B. Probable Functional Fluctuations

In this way, we see that typical-set functionality can be extended to all input realizations—that is, to all fluctuation subsets. The results give insight into the variability of thermodynamic function and a direct sense of its robustness or lack thereof. Now, we answer two questions that are particularly pertinent in the present setting of events (sequences) whose probabilities decay exponentially fast and so may be practically never observed. How probable are fluctuations in thermodynamic functioning? And, the related question, how probable are each of the fluctuation subsets? Exploring one example, we will show that the functional fluctuations are, in fact, quite observable not only with short sequences, perhaps expectedly, but also over relatively long sequences, such as $\ell = 100$.

The second question calls for determining $\mathbb{P}(w \in \Lambda_{u,\ell})$. However, in the large- ℓ limit this quantity vanishes. So, it is rather more natural to ask *how* it converges to zero. Since we are considering ergodic stationary processes, we can apply the large deviation principle: the probability of every subset $\Lambda_{u,\ell}$ vanishes exponentially with ℓ . However, each subset $\Lambda_{u,\ell}$ has different exponent which is the subset’s *large deviation rate* [38]:

$$I(u) = \lim_{\ell \rightarrow \infty} -\frac{1}{\ell} \log_2 \mathbb{P}(w \in \Lambda_{u,\ell}) .$$

Since all these w have the same probability decay rate u , $\mathbb{P}(w)$ decomposes to two components. The first gives the number $|\Lambda_{u,\ell}|$ of sequences in the subset and the second, the probability $2^{-\ell u}$ of individual sequences. That is:

$$\begin{aligned} I(u) &= \lim_{\ell \rightarrow \infty} -\frac{1}{\ell} \log_2 \mathbb{P}(w \in \Lambda_{u,\ell}) \\ &= \lim_{\ell \rightarrow \infty} -\frac{1}{\ell} \log_2 (|\Lambda_{u,\ell}| 2^{-\ell u}) \\ &= u - \lim_{\ell \rightarrow \infty} \frac{1}{\ell} \log_2 (|\Lambda_{u,\ell}|) . \end{aligned}$$

The size of subsets also grows exponentially with ℓ ,

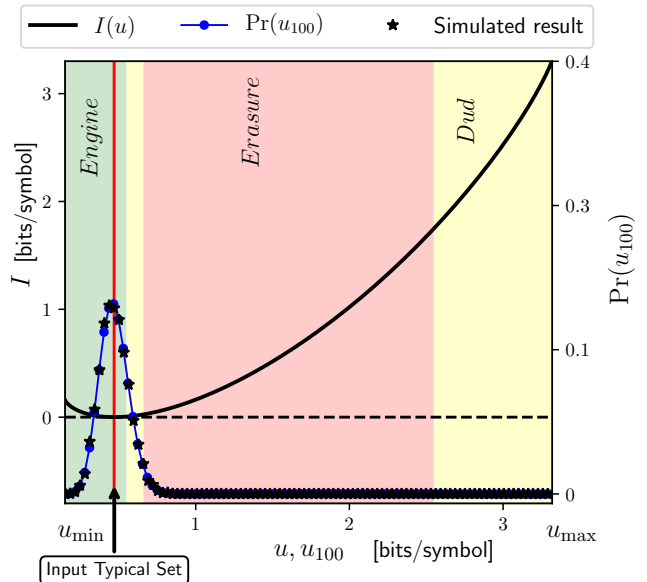


FIG. 8. Probability of fluctuations in thermodynamic functioning: Large-deviation rate function $I(u)$ (solid black line) and the theoretically-predicted probability $\Pr(u_{100})$ of fluctuation subsets for length $\ell = 100$ input realizations (dotted-solid blue line). Star tokens denote estimates from numerical simulation which validate the analytical results due to their close fit.

each subset with different exponent. To monitor this, we define new function:

$$S(u) = \lim_{\ell \rightarrow \infty} \frac{1}{\ell} \log_2 |\Lambda_{u,\ell}| .$$

Previously, we showed that $S(u) = h_\mu(\mathcal{P}_\beta)$, where $h_\mu(\mathcal{P}_\beta)$ is \mathcal{P}_β Shannon entropy and $u = \beta^{-1}(h_\mu(\mathcal{P}_\beta) - \log_2 \hat{\lambda}_\beta)$ from Eq. (8) [37]. These results allow one to calculate $I(u)$ for any subset using the expressions:

$$\begin{aligned} I(u) &= (\beta^{-1} - 1) h_\mu(\mathcal{P}_\beta) - \beta^{-1} \log_2 \hat{\lambda}_\beta \text{ and} \\ u &= \beta^{-1} (h_\mu(\mathcal{P}_\beta) - \log_2 \hat{\lambda}_\beta) . \end{aligned}$$

Figure 8 plots $I(u)$ for our example information ratchets. As with the previous figure, when realizations from the typical set are fed in, the transducer functions as an Engine. We now see that the typical set has zero large deviation rate. That is, in the limit of infinite length the probability of observing realizations in the typical set goes to one. In terms of thermodynamic functioning, the transducer operates as an Engine over long times with probability one. Complementarily, in the infinite length limit, the probability of the other “fluctuation” subsets vanishes.

In reality, though, one only observes finite-length sequences. And so, the operant question here is, are func-

tional fluctuations observable at finite lengths? As we alluded earlier, the expectation is that short sequences should enhance their observation.

Consider the input process in Fig. 3(a) and assume the input’s realization length is $\ell = 100$. We have 2^{100} distinct input sequences that are partitioned into 101 fluctuation subsets with different energy densities—subsets of sequences with ℓ 0s and $100 - \ell$ 1s for $\ell = 0, 1, \dots, 100$. Let’s calculate the probability of each of these fluctuations subsets occurring analytically. The probability of each versus its energy is shown in Fig. 8 as the blue dotted line. To distinguish it from the energy density of fluctuation subsets at infinite length we label the energy density of each of these sets with u_{100} , the index 100 reminds us that we are examining input sequences of length $\ell = 100$. There are 101 blue points on the figure, each representing one of the fluctuation subsets. (Most are obscured by other tokens, though.) If we feed the first 13 of the 101 fluctuation subsets (the first 13 blue points in the left of the figure) to the transducer, it functions as an Engine. Summing the probabilities of these Engine subsets, we see that the transducer functions as an Engine 80% of the time. Which is quite probable, even though it operates on sequences of length 100 that are individually highly improbable.

To verify the analytical results we also performed extensive numerical simulations that drove the ratchet with a sequence of length $\ell = 10^6$. We divided the input sequence into time intervals of length 100 and estimated the generated work and other observables, such as energy, during each interval. Star tokens in Fig. 7 show the estimated average work in each interval with decay rate u versus the decay rate itself. The numerical estimates agree closely with the analytical result. Figure 8 also shows the probabilities of each of these atypical subsets estimated from the simulations, which also validates the analytical results.

Let’s return to the remaining question: How probable are fluctuations in thermodynamic functioning? The answer is given by the large deviation rate for $\langle W \rangle(u)$? Since $\langle W \rangle$ is a function of u one can use the contraction principle [38] and relate the large deviation rate of $\langle W \rangle(u)$ in terms of large deviation rate of u via:

$$\tilde{I}(y = \langle W \rangle(u)) = \min_{u: y = \langle W \rangle(u)} I(u) .$$

Since $\langle W \rangle(u)$ is a one-to-one function the minimization above may be removed.

X. DISCUSSION

A. Related work

The new results here on memoryful information engines are also complementary to previous studies of fluctuations in the efficiency of a nanoscale heat engine [43–45], a particular form of information engine.

B. Relation to Fluctuation Theorems

To head-off confusion, and anticipate a key theme, note that “statistical fluctuation” above differs importantly from the sense used to describe variations in mesoscopic quantities when controlling small-scale thermodynamic systems. This latter sense is found in the recently-famous fluctuation theorem for the probability of positive and negative entropy production ΔS during macroscopic thermodynamic manipulations [46–52]:

$$\frac{\Pr(\Delta S)}{\Pr(-\Delta S)} = e^{\Delta S} .$$

Both kinds of fluctuation are ubiquitous, often dominating equilibrium finite-size systems and finite and infinite nonequilibrium steady-state systems. Differences acknowledged, there are important connections between statistical fluctuations in microstates observed in steady state and fluctuations in thermodynamic variables encountered during general control: For one, they are deeply implicated in expressed thermodynamic function. Is a system operating as an engine—converting thermal fluctuations to useful work—or as an eraser—depleting energy reservoirs to reduce entropy—or not functioning at all?

XI. CONCLUSION

We synthesized statistical fluctuations—as entailed in Shannon’s Asymptotic Equipartition Property [1] and large deviation theory [38, 53, 54]—and functional thermodynamics—as determined using the new informational Second Law [3]—to predict spontaneous variations in thermodynamic functioning. In short, there is simultaneous, inherently parallel, thermodynamic processing that is functionally distinct and possibly in competition. This strongly suggests that, even when in a nonequilibrium steady state, a single nanoscale device or biomolecule can be both an engine and an eraser. And, we showed that these functional fluctuations need not be rare. This complements similar previous results on fluctuations in small-scale engine efficiency [44, 55, 56]. The

conclusion is that functional fluctuations should be readily observable and the prediction experimentally testable.

A main point motivating this effort was to call into question the widespread habit of ascribing a single functionality to a given system and, once that veil has lifted, to appreciate the broad consequences. To drive them home, since biomolecular systems are rather like the information ratchet here, they should exhibit measurably different thermodynamic functions as they behave. If this prediction holds, then the biological world is vastly richer than we thought and it will demand of us a greatly refined vocabulary and greatly improved theoretical and experimental tools to adequately probe and analyze this new modality of parallel functioning.

That said, thoroughness forces us to return to our earlier caveat (Sec. IX) concerning not conflating various “temperatures”. If we give the input information reservoir and the output information reservoir physical implementations, then the fluctuation indices U_{in} and U_{out} take on thermal physical meaning and so can be related to the ratchet’s thermodynamic temperature T . Doing so, however, would take us too far afield here, but it will be necessary for a complete understanding.

Looking forward, there are many challenges. First, note that technically speaking we introduced a fluctuation theory for memoryful stochastic transducers, but by way of the example of Ref. [3]’s information ratchet. A thoroughgoing development must be carried out in much more generality using the tools of Refs. [29, 57], if we are to fully understand the functionality of thermodynamic processes that transform inputs to outputs, environmental stimulus to environmental action.

Second, the role of Jarzynski-Crooks theory for fluctuations in thermodynamic observables needs to be made explicit and directly related to statistical fluctuations, in the sense emphasized here. One reason is that their theory bears directly on controlling thermodynamic systems and the resulting macroscopic fluctuations. To draw the parallel more closely, following the fluctuation theory for transitions between nonequilibrium steady states [58], we could drive the ratchet parameters p and q and input bias b between different functional regimes and monitor the entropy production fluctuations to test how the theory fares for memoryful processes. In any case, efficacy in control will also be modulated by statistical fluctuations.

Not surprisingly, there is much to do. Let’s turn to a larger motivation and perhaps larger consequences to motivate future efforts.

As just noted, fluctuations are key to nanoscale physics and molecular biology. We showed that fluctuations are deeply implicated both in identifying thermodynamic function and in the very operation of small-scale systems. In fact, fluctuations are critical to life—its proper and ro-

bust functioning. The perspective arising from parallel thermodynamic function is that, rather than fluctuations standing in contradiction to life processes, potentially corrupting them, there may be a positive role for fluctuations and parallel thermodynamic functioning. Once that is acknowledged it is a short step to realize that biological evolution may have already harnessed them to good thermodynamic effect. Manifestations are clearly worth looking for.

It now seems highly likely that fluctuations engender more than mere health and homeostasis. It is a commonplace that biological evolution is nothing, if not opportunistic. If so, then it would evolve cellular biological thermodynamic processes that actively leverage fluctuations. Mirroring Maxwell’s Demon’s need for fluctuations to operate, biological evolution itself advances only when there are fluctuations. For example, biomolecular mutation processes engender a distribution of phenotypes and fitnesses; fodder for driving selection and so evolutionary innovation. This, then, is *Darwin’s Demon*—a mechanism that ratchets in favorable fluctuations for positive thermodynamic and then positive survival benefit. The generality of results and methods here give new insight into thermodynamic functioning in the presence of fluctuations that should apply at many different scales of life, including its emergence and evolution.

ACKNOWLEDGMENTS

We thank Fabio Anza, Alec Boyd, John Mahoney, Dibyendu Mandal, Sarah Marzen, Paul Riechers, Gregory Wimsatt for helpful discussions. JPC thanks the Santa Fe Institute and the Telluride Science Research Center for their hospitality during visits. This material is based upon work supported by, or in part by, FQXi Grant number FQXi-RFP-1609, 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-13-1-0340 and grant W911NF-18-1-0028.

AUTHOR CONTRIBUTIONS STATEMENT

CA and JPC conceived of the project, developed the theory, and wrote the manuscript. CA performed the calculations. JPC supervised the project.

**COMPETING FINANCIAL INTERESTS
STATEMENT**

The authors declare that they have no competing financial interests.

Appendix A: Typical Set for a Biased Coin

What is A_ϵ^n for a biased coin with bias b ? The typical set is defined by;

$$A_\epsilon^n = \{w \in \mathcal{A}^n : 2^{-n(h_\mu + \epsilon)} \leq \Pr(w) \leq 2^{-n(h_\mu - \epsilon)}\} .$$

The probability of a biased coin generating a particular sequence w with k heads is $b^k(1-b)^{(n-k)}$. And so, for w to be in the typical set we must have:

$$\ell b - \frac{n\epsilon}{\log \frac{b}{1-b}} \leq k \leq nb + \frac{n\epsilon}{\log \frac{b}{1-b}} .$$

Since k is an integer:

$$\left\lceil nb - \frac{n\epsilon}{\log \frac{b}{1-b}} \right\rceil \leq k \leq \left\lfloor nb + \frac{n\epsilon}{\log \frac{b}{1-b}} \right\rfloor .$$

For example, in the case where $n = 1000$, $b = 0.6$, and $\epsilon = 0.01$, we have:

$$582 \leq k \leq 617 .$$

This means that those length $n = 1000$ sequences with 582 to 617 Heads are in the typical set.

Appendix B: Work Bounds

Average work for finite length ℓ is:

$$\langle W(\ell) \rangle = \frac{1}{\ell} \sum_{w \in \mathcal{A}^\ell} \mathbb{P}(w) f(w) . \quad (\text{B1})$$

Recall that $f(w)$ is the average work generated by the transducer when fed the word w ; see Fig. 2.

Now, let us say we are only interested in the engine's functionality when operating on sequences in a particular partition—those in the typical set. To determine the functionality, we first define a new probability distribution for the typical set:

$$\widetilde{\mathbb{P}}(w) = \begin{cases} \mathbb{P}(w) / \sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) & w \in \mathcal{A}_\epsilon^\ell \\ 0 & w \notin \mathcal{A}_\epsilon^\ell \end{cases} .$$

Using it, we define a new average work for finite ℓ se-

quences in the typical set:

$$\langle W(\ell) \rangle_{TS} = \frac{1}{\ell} \sum_{w \in \mathcal{A}_\epsilon^\ell} \widetilde{\mathbb{P}}(w) f(w) .$$

There are two important observations. First, this statistic gives no information about works generated by sequences outside of $\mathcal{A}_\epsilon^\ell$, since the probability distribution vanishes there. Second, for every pair of typical sequences $w_1, w_2 \in \mathcal{A}_\epsilon^\ell$:

$$\widetilde{\mathbb{P}}(w_1) \simeq \widetilde{\mathbb{P}}(w_2) ,$$

since Eq. (5) bounds the sequences' probabilities: $2^{-\ell(h_\mu + \epsilon)} \leq \Pr(w) \leq 2^{-\ell(h_\mu - \epsilon)}$. Effectively, $\langle W(\ell) \rangle_{TS}$ is an unweighted average over $f(w)$ s for the words in typical set.

Now, consider Eq. (B1) and decompose its righthand side into two parts: the share of typical sequences and the share of atypical sequences:

$$\sum_{w \in \mathcal{A}^\ell} \mathbb{P}(w) f(w) = \sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) f(w) + \sum_{w \notin \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) f(w) . \quad (\text{B2})$$

The second term is bounded from above:

$$\begin{aligned} \sum_{w \notin \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) f(w) &\leq \left(\sum_{w \notin \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) \right) \max\{f(w) : w \notin \mathcal{A}_\epsilon^\ell\} \\ &\leq \epsilon \times \max\{f(w) : w \notin \mathcal{A}_\epsilon^\ell\} . \end{aligned}$$

$f(w)$, the work generated by any length- ℓ sequence $w \in \mathcal{A}^\ell$, is also bounded:

$$\ell \alpha_{\min} \leq f(w) \leq \ell \alpha_{\max} ,$$

where α_{\min} and α_{\max} are the minimum and maximum one-shot works, respectively. Here, $\alpha_{\max} > 0$ and $\alpha_{\min} < 0$, which are due to the finiteness of energy for the machine states and coupled input-symbol-states. As a result, we have:

$$\sum_{w \notin \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) f(w) \leq \ell \epsilon \alpha_{\max} .$$

Similarly, it has a lower bound:

$$\sum_{w \notin \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) f(w) \geq \ell \epsilon \alpha_{\min} .$$

Now, we turn to decompose Eq. (B2)'s first term into

two parts:

$$\begin{aligned} \sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) f(w) &= \sum_{w \in \mathcal{A}_\epsilon^\ell} \frac{\mathbb{P}(w)}{\sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)} f(w) \\ &+ \sum_{w \in \mathcal{A}_\epsilon^\ell} \left(\mathbb{P}(w) - \frac{\mathbb{P}(w)}{\sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)} \right) f(w). \end{aligned} \quad (\text{B3})$$

The second term in Eq. (B3) can be written as:

$$\begin{aligned} \sum_{w \in \mathcal{A}_\epsilon^\ell} \left(\mathbb{P}(w) - \frac{\mathbb{P}(w)}{\sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)} \right) f(w) &= \\ \left(1 - \frac{1}{\sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)} \right) \sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) f(w), \end{aligned}$$

To go further one must note that the coefficient of the sum on the righthand side is negative. As a result:

$$\begin{aligned} \sum_{w \in \mathcal{A}_\epsilon^\ell} \left(\mathbb{P}(w) - \frac{\mathbb{P}(w)}{\sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)} \right) f(w) &\leq \left(1 - \frac{1}{\sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)} \right) \min\{f(w) : w \notin \mathcal{A}_\epsilon^\ell\} \sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) \\ &\leq \left(1 - \frac{1}{\sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)} \right) \ell \alpha_{\min} \sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) \\ &\leq \left(1 - \frac{1}{1 - \epsilon} \right) \ell \alpha_{\min}. \end{aligned}$$

Giving an upper bound on the second term in Eq. (B3).

Similarly, one can give it a lower bound:

$$\sum_{w \in \mathcal{A}_\epsilon^\ell} \left(\mathbb{P}(w) - \frac{\mathbb{P}(w)}{\sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)} \right) f(w) \geq \left(1 - \frac{1}{1 - \epsilon} \right) \ell \alpha_{\max}.$$

Using Eqs. (B2) and (B3) and these upper and lower bounds, we have:

$$\delta_{\min} \leq \langle W(n) \rangle - \langle W(n) \rangle_{TS} \leq \delta_{\max},$$

where:

$$\begin{aligned} \delta_{\min} &= \epsilon \left(\alpha_{\min} - \frac{\alpha_{\max}}{1 - \epsilon} \right) \text{ and} \\ \delta_{\max} &= \epsilon \left(\alpha_{\max} - \frac{\alpha_{\min}}{1 - \epsilon} \right). \end{aligned}$$

Recalling that $\alpha_{\max} > 0$ and $\alpha_{\min} < 0$, we have $\delta_{\max} > 0$ and $\delta_{\min} < 0$.

Thus, the difference between average work $\langle W(\ell) \rangle$ over all sequences and that $\langle W(\ell) \rangle_{TS}$ defined for typical set is small for sufficiently large ℓ . For all practical purposes, they are equal. This, together with recalling that $\langle W(\ell) \rangle_{TS}$ is an unweighted average of works $f(w)$ for $w \in \mathcal{A}_\epsilon^\ell$, provides an operational interpretation of the previously-defined functionality.

Appendix C: Information Bounds

The Shannon entropy rate of the output process for finite length ℓ is:

$$\Delta H'(\ell) = \frac{1}{\ell} \sum_{w' \in \mathcal{A}^\ell} \mathbb{P}'(w') \log_2 \mathbb{P}'(w').$$

Here, $\mathbb{P}(\cdot)$ refers to probability of output sequences under the process generated by the transducer and w' is an output sequence. Rewriting the sum in the form:

$$\sum_{w' \in \mathcal{A}^\ell} \mathbb{P}'(w') \log_2 \mathbb{P}'(w') = \sum_{w, w' \in \mathcal{A}^\ell} \mathbb{P}(w) \mathbb{P}(w'|w) \log_2 \mathbb{P}'(w'),$$

where $\mathbb{P}(w'|w)$ is the conditional probability of the transducer generating output sequence w' when reading input w . Now, defining:

$$g(w) = \sum_{w' \in \mathcal{A}^\ell} -\mathbb{P}(w'|w) \log_2 \mathbb{P}'(w'),$$

one writes the Shannon entropy rate in a form paralleling Eq. (B1):

$$\Delta H'(\ell) = \frac{1}{\ell} \sum_{w \in \mathcal{A}^\ell} \mathbb{P}(w) g(w), \quad (\text{C1})$$

where $g(w)$ is the average information generated by the word w when passing through transducer; see Fig. 2.

We can also monitor the information generated by feeding in only the typical set with:

$$\Delta H'_{TS}(\ell) = \frac{1}{\ell} \sum_{w \in \mathcal{A}_\epsilon^\ell} \widetilde{\mathbb{P}(w)} g(w).$$

Similar to analyzing the generated works, one decomposes the sum in Eq. (C1) into two parts:

$$\sum_{w \in \mathcal{A}^\ell} \mathbb{P}(w) g(w) = \sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) g(w) + \sum_{w \notin \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) g(w). \quad (\text{C2})$$

The second term in Eq. (C2) is bounded above:

$$\begin{aligned} \sum_{w \notin \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)g(w) &\leq \left(\sum_{w \notin \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) \right) \max\{g(w) : w \notin \mathcal{A}_\epsilon^\ell\} \\ &\leq \epsilon \times \max\{g(w) : w \notin \mathcal{A}_\epsilon^\ell\} \end{aligned}$$

From the definition, one sees that there are upper bounds on $g(w) < \ell$, where the bound can only be reached when the input is a Fair Coin Process and the transducer maps every word fairly to all possible output sequences. This means the second term in Eq. (C2) is bounded from above:

$$\sum_{w \notin \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)g(w) \leq \epsilon \ell .$$

We can similarly analyze the first term in Eq. (C2):

$$\begin{aligned} \sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)g(w) &= \sum_{w \in \mathcal{A}_\epsilon^\ell} \widetilde{\mathbb{P}}(w)g(w) \\ &\quad + \sum_{w \in \mathcal{A}_\epsilon^\ell} (\mathbb{P}(w) - \widetilde{\mathbb{P}}(w))g(w) . \end{aligned}$$

The second term here is negative and is bounded from below:

$$\begin{aligned} \sum_{w \in \mathcal{A}_\epsilon^\ell} (\mathbb{P}(w) - \widetilde{\mathbb{P}}(w))g(w) &= \\ &\left(1 - \frac{1}{\sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)} \right) \sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)g(w) \\ &\geq \left(1 - \frac{1}{1 - \epsilon} \right) \sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w)g(w) \\ &\geq \left(1 - \frac{1}{1 - \epsilon} \right) \max\{g(w) : w \notin \mathcal{A}_\epsilon^\ell\} \sum_{w \in \mathcal{A}_\epsilon^\ell} \mathbb{P}(w) \\ &\geq \left(\frac{-\epsilon}{1 - \epsilon} \right) \max\{g(w) : w \notin \mathcal{A}_\epsilon^\ell\} \\ &\geq \ell \left(\frac{-\epsilon}{1 - \epsilon} \right) . \end{aligned}$$

Combining these two bounds:

$$\frac{-\epsilon}{1 - \epsilon} \leq \Delta H'(\ell) - \Delta H'_{TS}(\ell) \leq \epsilon ,$$

one concludes the average generated information, when the transducer is fed the whole set, is essentially equal to the average information generated when the transducer is fed the typical set without probability weights.

Appendix D: Work is a Linear Function of Decay Rate

First let us calculate u as a function of β . Recall that they related via $u = \beta^{-1}(h_\mu(\mathcal{P}_\beta) - \log_2 \hat{\lambda}_\beta)$.

Using step 2 and 3 for the HMM model shown in Figure 3 we have

$$\mathbf{T}_\beta = \begin{bmatrix} b^\beta & (1-b)^\beta \\ b^\beta & (1-b)^\beta \end{bmatrix} .$$

Calculating the maximal eigenvalue $\hat{\lambda}_\beta$, we find:

$$\log \lambda_\beta = \log_2(b^\beta + (1-b)^\beta) .$$

The Shannon entropy of process \mathcal{P}_β , $h_\mu(\mathcal{P}_\beta)$ is equal to the Shannon entropy of biased coin with bias $\hat{b} = b^\beta / (b^\beta + (1-b)^\beta)$:

$$\begin{aligned} h_\mu(\mathcal{P}_\beta) &= - \left(\frac{b^\beta}{b^\beta + (1-b)^\beta} \log_2 \frac{b^\beta}{b^\beta + (1-b)^\beta} \right. \\ &\quad \left. + \frac{(1-b)^\beta}{b^\beta + (1-b)^\beta} \log_2 \frac{(1-b)^\beta}{b^\beta + (1-b)^\beta} \right) . \end{aligned}$$

It is straightforward, now, to calculate u from these:

$$\begin{aligned} u &= \frac{-b^\beta}{b^\beta + (1-b)^\beta} \log_2(b) \\ &\quad + \frac{-(1-b)^\beta}{b^\beta + (1-b)^\beta} \log_2(1-b) . \end{aligned}$$

In the next step we need to calculate the work $\langle W \rangle$ from Equation 3 for the input process \mathcal{P}_β by replacing b with \hat{b} :

$$\langle W \rangle(\beta) = \frac{k_B T}{2} \left(-q \log(q/p) + q \log(1-q) + \frac{cb^\beta}{b^\beta + (1-b)^\beta} \right) ,$$

where $c = (p+q) \log(q/p) + p \log(1-p) - q \log(1-q)$. Now it is easy to see that

$$\begin{aligned} \langle W \rangle(u) &= \frac{k_B T}{2} \left(-q \log \left(\frac{q}{p} \right) + q \log(1-q) \right. \\ &\quad \left. + c \frac{u + \log(1-b)}{\log(1-b) - \log(b)} \right) . \end{aligned}$$

It is easy to see that

$$\begin{aligned} W_{\max} &= \max_\beta \langle W \rangle \\ &= \begin{cases} W_{\max}^- \equiv k_B T (-q \log(p/q) - q \log(1-q)) & c < 0 \\ W_{\max}^+ \equiv k_B T (p \log(p/q) - p \log(1-p)) & c \geq 0 \end{cases} , \end{aligned}$$

which is in both cases independent of the bias for the

input process b .

-
- [1] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, second edition, 2006.
- [2] H. B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. Wiley, New York, second edition, 1985a.
- [3] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Identifying functional thermodynamics in autonomous Maxwellian ratchets. *New J. Physics*, 18:023049, 2016.
- [4] L. Szilard. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Z. Phys.*, 53:840–856, 1929.
- [5] J. C. Maxwell. *Theory of Heat*. Longmans, Green and Co., London, United Kingdom, ninth edition, 1888.
- [6] 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.
- [7] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948.
- [8] L. Brillouin. *Science and Information Theory*. Academic Press, New York, second edition, 1962.
- [9] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano. Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nat. Physics*, 6:988–992, 2010.
- [10] B. Lambson, D. Carlton, and J. Bokor. Exploring the thermodynamic limits of computation in integrated systems: Magnetic memory, nanomagnetic logic, and the Landauer limit. *Phys. Rev. Lett.*, 107:010604, 2011.
- [11] A. Berut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz. Experimental verification of Landauer’s principle linking information and thermodynamics. *Nature*, 483:187, 2012.
- [12] Y. Jun, M. Gavrilov, and J. Bechhoefer. High-precision test of Landauer’s principle in a feedback trap. *Phys. Rev. Lett.*, 113:190601, 2014.
- [13] M. Madami, M. d’YAquino, G. Gubbiotti, S. Tacchi, C. Serpico, and G. Carlotti. Micromagnetic study of minimum-energy dissipation during Landauer erasure of either isolated or coupled nanomagnetic switches. *Phys. Rev. B*, 90:104405, 2014.
- [14] J. P. Pekola. Towards quantum thermodynamics in electronic circuits. *Nat. Physics*, 11:118–123, 2015.
- [15] J. V. Koski, A. Kutvonen, I. M. Khaymovich, T. Ala-Nissila, and J. P. Pekola. On-chip Maxwell’s demon as an information-powered refrigerator. *Phys. Rev. Lett.*, 115:260602, 2015.
- [16] W. Thomson. Kinetic theory of the dissipation of energy. *Nature*, 9:441, 1874.
- [17] D. Mandal and C. Jarzynski. Work and information processing in a solvable model of Maxwell’s demon. *Proc. Natl. Acad. Sci. USA*, 109(29):11641–11645, 2012.
- [18] D. Mandal, H. T. Quan, and C. Jarzynski. Maxwell’s refrigerator: An exactly solvable model. *Phys. Rev. Lett.*, 111:030602, 2013.
- [19] P. Strasberg, G. Schaller, T. Brandes, and M. Esposito. Thermodynamics of a physical model implementing a Maxwell demon. *Phys. Rev. Lett.*, 110:040601, 2013.
- [20] A. C. Barato and U. Seifert. An autonomous and reversible Maxwell’s demon. *Europhysics Lett.*, 101(6):60001, 2013.
- [21] J. M. Horowitz, T. Sagawa, and J. M. R. Parrondo. Imitating chemical motors with optimal information motors. *Phys. Rev. Lett.*, 111:010602, 2013.
- [22] A. C. Barato and U. Seifert. Stochastic thermodynamics with information reservoirs. *Phys. Rev. E*, 90(4):042150, 2014.
- [23] J. Hoppenau and A. Engel. On the energetics of information exchange. *Europhysics Lett.*, 105:50002, 2014.
- [24] Z. Lu, D. Mandal, and C. Jarzynski. Engineering Maxwell’s demon. *Physics Today*, 67(8):60–61, January 2014.
- [25] J. Um, H. Hinrichsen, C. Kwon, and H. Park. Total cost of operating an information engine. *New J. Phys.*, 17(8):085001, 2015.
- [26] N. Shiraishi, T. Matsumoto, and T. Sagawa. Measurement-feedback formalism meets information reservoirs. *New J. Phys.*, 18(1):013044, 2016.
- [27] A. J. P. Garner, J. Thompson, V. Vedral, and M. Gu. Thermodynamics of complexity and pattern manipulation. *Physical Review E*, 95(4):042140, 2017.
- [28] A. B. Boyd, D. Mandal, P. M. Riechers, and J. P. Crutchfield. Transient dissipation and structural costs of physical information transduction. *Phys. Rev. Lett.*, 118(22):220602, 2017.
- [29] 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.
- [30] A. C. Barato and U. Seifert. An autonomous and reversible Maxwell’s demon. *Europhysics Lett.*, 101:60001, 2013.
- [31] S. Kullback. *Information Theory and Statistics*. Dover, New York, 1968.
- [32] R. W. Yeung. *Information Theory and Network Coding*. Springer, New York, 2008.
- [33] B. McMillan. The basic theorems of information theory. *Ann. Math. Stat.*, 24:196–219, 1953.
- [34] L. Breiman. The individual ergodic theorem of information theory. *Ann. Math. Stat.*, 28(3):809–811
- [35] R. L. Adler, A. G. Konheim, and M. H. McAndrew. Topological entropy. *Trans. Amer. Math. Soc.*, 114:309–319, 1965.
- [36] Except for the special class of maximum-entropy processes, which we do not consider directly.

- [37] C. Aghamohammadi and J. P. Crutchfield. Minimum memory for generating rare events. *Phys. Rev. E*, 95(3):032101, 2017.
- [38] H. Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478:1–69, 2009.
- [39] K. Young and J. P. Crutchfield. Fluctuation spectroscopy. *Chaos, Solitons, and Fractals*, 4:5 – 39, 1994.
- [40] R. L. Jack and P. Sollich. Large deviations and ensembles of trajectories in stochastic models. *Prog. Theo. Physics Suppl.*, 184:304–317, 2010.
- [41] J. P. Garrahan and I. Lesanovsky. Thermodynamics of quantum jump trajectories. *Phys. Rev. Lett.*, 104(16):160601, 2010.
- [42] R. Chetrite and H. Touchette. Nonequilibrium Markov processes conditioned on large deviations. *Annales Henri Poincaré*, 16(9):2005–2057, 2015.
- [43] G. Verley, M. Esposito, T. Willaert, and C. Van den Broeck. The unlikely Carnot efficiency. *Nature communications*, 5:4721, 2014.
- [44] T. R. Gingrich, G. M. Rotskoff, S. Vaikuntanathan, and P. L. Geissler. Efficiency and large deviations in time-asymmetric stochastic heat engines. *New J. Physics*, 16(10):102003, 2014.
- [45] H. Vroylandt, A. Bonfils, and G. Verley. Efficiency fluctuations of small machines with unknown losses. *Physical Review E*, 93(5):052123, 2016.
- [46] D. J. Evans, E. G. D. Cohen, and G. P. Morriss. Probability of second law violations in shearing steady flows. *Phys. Rev. Lett.*, 71:2401–2404, 1993.
- [47] D. J. Evans and D. J. Searles. Equilibrium microstates which generate second law violating steady states. *Phys. Rev. E*, 50:1645, 1994.
- [48] G. Gallavotti and E. G. D. Cohen. Dynamical ensembles in nonequilibrium statistical mechanics. *Phys. Rev. Lett.*, 74:2694–2697, 1995.
- [49] J. Kurchan. Fluctuation theorem for stochastic dynamics. *J. Phys. A: Math. Gen.*, 31:3719, 1998.
- [50] G. E. Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.*, 90(5/6):1481–1487, 1998.
- [51] J. L. Lebowitz and H. Spohn. A Gallavotti-Cohen-type symmetry in the large deviation functional for stochastic dynamics. *J. Stat. Phys.*, 95:333, 1999.
- [52] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr., and C. Bustamante. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, 437:231, 2005.
- [53] R. Bowen. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, volume 470 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1975.
- [54] J. A. Bucklew. *Large Deviation Techniques in Decision, Simulation, and Estimation*. Wiley-Interscience, New York, 1990.
- [55] G. Verley, M. Esposito, T. Willaert, and C. Van den Broeck. The unlikely carnot efficiency. *Nature Comm.*, 5, 2014.
- [56] G. Verley, T. Willaert, C. Van den Broeck, and M. Esposito. Universal theory of efficiency fluctuations. *Phys. Rev. E*, 90(5):052145, 2014.
- [57] C. Aghamohammadi and J. P. Crutchfield. Minimum memory for generating rare events. *Phys. Rev. E*, 95(3):032101, 2017.
- [58] P. M. Riechers and J. P. Crutchfield. Fluctuations when driving between nonequilibrium steady states. *J. Stat. Phys.*, 168(4):873–918, 2017.