Santa Fe Institute Working Paper 2017-01-01
arxiv.org:1612.08616 [cond-mat.stat-mech]

Transient Dissipation and Structural Costs of Physical Information Transduction

Alexander B. Boyd,[1,*] Dibyendu Mandal,[2,†] Paul M. Riechers,[1,‡] and James P. Crutchfield[1,§]

[1] *Complexity Sciences Center and Physics Department, University of California at Davis, One Shields Avenue, Davis, CA 95616*

[2] *Department of Physics, University of California, Berkeley, CA 94720, U.S.A.*

(Dated: April 15, 2017)

A central result that arose in applying information theory to the stochastic thermodynamics of nonlinear dynamical systems is the Information-Processing Second Law (IPSL): the physical entropy of the universe can decrease if compensated by the Shannon-Kolmogorov-Sinai entropy change of appropriate information-carrying degrees of freedom. In particular, the asymptotic-rate IPSL precisely delineates the thermodynamic functioning of autonomous Maxwellian demons and information engines. How do these systems begin to function as engines, Landauer erasers, and error correctors? Here, we identify a minimal, inescapable transient dissipation engendered by physical information processing not captured by asymptotic rates, but critical to adaptive thermodynamic processes such as found in biological systems. A component of transient dissipation, we also identify an implementation-dependent cost that varies from one physical substrate to another for the same information processing task. Applying these results to producing structured patterns from a structureless information reservoir, we show that "retrodictive" generators achieve the minimal costs. The results establish the thermodynamic toll imposed by a physical system's structure as it comes to optimally transduce information.

PACS numbers: 05.70.Ln 89.70.-a 05.20.-y 05.45.-a
Keywords: transducer, intrinsic computation, Information-Processing Second Law of Thermodynamics

Introduction. Classical thermodynamics and statistical mechanics appeal to various reservoirs—reservoirs of heat, work, particles, and chemical species—each characterized by unique, idealized thermodynamic properties. A heat reservoir, for example, corresponds to a physical system with a large specific heat and short equilibration time. A work reservoir accepts or gives up energy without a change in entropy. Arising naturally in recent analyses of Maxwellian demons and information engines [1–17], *information reservoirs* have come to play a central role as idealized physical systems that exchange information but not energy [18–20]. Their inclusion led rather directly to an extended Second Law of Thermodynamics for complex systems: The total physical (Clausius) entropy of the universe and the Shannon entropy of its information reservoirs cannot decrease in time [4, 18, 21–23]. We refer to this generalization as the Information Processing Second Law (IPSL) [24].

A specific realization of an information reservoir is a tape of symbols where information is encoded in the symbols' values [25]. To understand the role that information processing plays in the efficiencies and bounds on thermodynamic transformations the following device has been explored in detail: a "ratchet" slides along a tape and interacts with one symbol at a time in presence of heat and work reservoirs [26]. By increasing the tape's Shannon entropy, the ratchet can steadily transfer energy from the heat to the work reservoirs [4]. This violates the conventional formulation of the Second Law of Thermodynamics but is permitted by the IPSL.

An information ratchet can leverage the structure of the information reservoir by traversing the tape in many different ways [7]. However, we consider ratchets that run along the tape unidirectionally, where there's a clear transformation from input to output domain. This is effectively a transformation of the information encoded in the tape from input to output, so the ratchet is an *information transducer*. Recent models of autonomous Maxwellian demons and information engines are specific examples of information transducers. From an information-theoretic viewpoint, these transducers are memoryful communication channels from input to output symbol sequences [27]. Information transducers are also similar to Turing machines in design [28], except that a Turing machine need not move unidirectionally. More importantly, an information transducer is a physical thermodynamic system and so is typically stochastic [29]. Despite this difference, like a Turing machine a transducer can perform any computation, if allowed any number of internal states.

Previous analyses on the thermodynamics of information processing largely focused on the minimal asymptotic entropy production *rate* for a given information transduction; see Eq. (2) below. The minimal rate is completely specified by the information transduction; there is no mention of any cost due to the transducer itself. In contrast, this Letter first derives an exact expression for the minimal *transient* entropy production required for information transduction; see Eq. (3). This transient dissipation is the cost incurred by a system as it adapts to its environment. It is related to the excess heat in transitions between nonequilibrium steady states [30–32]. Moreover, hidden in this minimal transient dissipation, we identify the minimal cost associated with the transducer's construction; Eq. (4) below. Among all possible constructions that support a given computational task, there is a minimal, finite cost due to the physical *implementation*.

The Letter goes on to consider the specific case of structured pattern generation from a structureless information reservoir—a tape of independent and identically distributed (IID) symbols. While the transducer formalism for information ratchets naturally includes inputs with temporal structure, most theory so far has considered structureless inputs [4, 5, 7, 26, 33, 34]. This task requires designing a transducer that reads a tape of IID symbols as its input and outputs a target pattern. Employing the algebra of Shannon measures [35] and the structure-analysis tools of computational mechanics [36, 37], we show that the minimum implementation-dependent cost is determined by the mutual information between the transducer and the output's "past"—that portion of the output tape already generated. The result is that a maximally efficient implementation is achieved with a "retrodictive" model of the structured pattern transducer. Since the retrodictor's states depend only on the output future, it only contains as much information about the output's past as is required to generate the future. As a result it has a minimal cost proportional to the tape's *excess entropy* [37]. Such thermodynamic costs affect information processing in physical and biological systems that undergo finite-time transient processes when adapting to a complex environment.

Information Processing Second Law. Consider a discrete-time Markov process involving the transducer's current state X_N and the current state of the information reservoir $\mathbf{Y}_N$ it processes. The latter is a semi-infinite chain of variables over the set $\mathcal{Y}$ that the transducer processes sequentially. Y_N is the Nth tape element, if the transducer has not yet processed that symbol; it is denoted Y'_N, if the transducer has. We call Y_N an input and Y'_N an output. The current tape $\mathbf{Y}_N = Y'_{0:N} Y_{N:\infty}$ concatenates the input tape $Y_{N:\infty} = Y_N Y_{N+1} Y_{N+2} \ldots$ and output tape $Y'_{0:N} = Y'_0 Y'_1 \ldots Y'_{N-2} Y'_{N-1}$. The information ratchet performs a computation by steadily transducing the input tape process $\Pr(Y_{0:\infty})$ into the output tape process $\Pr(Y'_{0:\infty})$.

The IPSL sets a bound on the average heat dissipation $Q_{0\to N}$ into the thermal reservoir over the time interval $t \in [0, N]$ in terms of the change in state uncertainty of the information ratchet and information reservoir [26, App. A]:

$$\frac{\langle Q_{0\to N}\rangle}{k_{\mathrm{B}} T \ln 2} \geq \mathrm{H}[X_0, \mathbf{Y}_0] - \mathrm{H}[X_N, \mathbf{Y}_N], \tag{1}$$

where k_{B} is Boltzmann's constant, T the absolute temperature of the reservoirs, $\mathrm{H}[Z]$ the Shannon (information) entropy of the random variable Z.

To date, studies of such information engines developed the IPSL's asymptotic-rate form:

$$\lim_{N\to\infty} \frac{1}{N} \frac{\langle Q_{0\to N}\rangle}{k_{\mathrm{B}} T \ln 2} \geq -(h'_\mu - h_\mu) \ , \tag{2}$$

where h'_μ (h_μ) is the Shannon entropy rate of the output (input) tape [38] and, in addition, we assume the transducer has a finite number of states [26].

The asymptotic IPSL in Eq. (2) says that thermal fluctuations from the environment can be rectified to either perform work or refrigerate (on average) at the cost of randomizing the information reservoir ($\langle Q\rangle < 0$ when $h'_\mu > h_\mu$). Conversely, an information reservoir can be refueled or 'charged' back to a clean slate by erasing its Shannon-entropic information content at the cost of emitting heat.

A crucial lesson in the physics of information is that Eq. (2) takes into account *all orders* of temporal correlations present in the input tape as well as all orders of correlation that the transducer develops in the output tape. An approximation of Eq. (2), based on the inclusion of only lowest-order (individual symbol) statistics, had been used to interpret the thermodynamic functioning of the original models of autonomous Maxwellian Demons [4, 5]. Later, Eq. (2) itself was used to identify a region in an engine's phase diagram that is wrongly characterized as functionally useless by the approximation, but actually is a fully functional eraser. In turn, this motivated the construction of an explicit mechanism by which temporal correlations in the input sequence can be exploited as a thermodynamic resource [26]. Equation (2) also led to (i) a general thermodynamic framework for memory in sequences and in transducers and (ii) a thermodynamic instantiation of Asbhy's law of requisite variety—a cybernetic principle of adaptation [24].

Equation (2), however, does not account for correlations between input and output tapes nor those that arise between the transducer and the input and output. As we now show, doing so leads directly to predictions about the relative effectiveness of transducers that perform the same information processing on a given input, but employ different physical implementations. Specifically, the retrodictive generator is the thermodynamically simplest implementation, *not* computational mechanics' optimal predictor—the ϵ-machine [36] [39].

Subtracting the IPSL's asymptotic-rate version (Eq. (2)) from the IPSL's original (Eq. (1)) leads to a lower bound on the *transient* thermodynamic cost $\langle Q^{\mathrm{tran}} \rangle$ of information transduction, the Letter's central result:

$$\begin{aligned} \frac{\langle Q^{\mathrm{tran}} \rangle_{\min}}{k_{\mathrm{B}} T \ln 2} &\equiv \lim_{N \to \infty} \left[\frac{\langle Q_{0 \to N} \rangle_{\min}}{k_{\mathrm{B}} T \ln 2} + N(h'_\mu - h_\mu) \right] \\ &= -\mathbf{E}' + \mathrm{I}[\overleftarrow{Y}'; \overrightarrow{Y}] + \mathrm{I}[X_0; \overleftarrow{Y}', \overrightarrow{Y}] \; . \end{aligned} \tag{3}$$

Here, $\mathbf{E}'$ is the output sequence's *excess entropy* [40], which is the mutual information $I[\overleftarrow{Y}'; \overrightarrow{Y}']$ between the output past and output future. In these expressions we use arrows to describe past and future random variables. $\overleftarrow{Y}'$ is the random variable for output past—the sequence of output symbols that have been produced by the transducer—and $\overrightarrow{Y}'$ is the output future—the sequence of output symbols that have not yet been produced. Similarly, the input past random variable $\overleftarrow{Y}$ is the sequence of input variables that have already interacted with the transducer and have been transformed into an output. While the input future random variable $\overrightarrow{Y}$ is the sequence of input variables that have yet to interact. Finally, X_0 is the random variable for the transducer's state after sufficiently long time, such that $\overleftarrow{Y}'$ and $\overleftarrow{Y}$ are both effectively semi-infinite. The expression itself comes from shifting to the ratchet's reference frame, so that at time N state X_N becomes X_0 and the currently interacting tape symbol is relabeled Y_0, rather than Y_N. (Equation (3) is proved in the Supplementary Materials.)

From it we conclude that the minimum transient cost has three components. However, they are subtly interdependent and so we cannot minimize them piece-by-piece to maximize thermodynamic efficiency. For instance, the first term in the transient cost is a benefit of having correlation between the output past and output future, qualified by $\mathbf{E}'$. Without further thought, one infers that outputs that are more predictable from their past, given a fixed entropy production rate, are easier to produce thermodynamically. However, as we see below when analyzing process generation, the other terms cancel this benefit, regardless of the output process. Perhaps counterintuitively, the most important factor is the output's intrinsic structure.

The remaining two terms in the transient cost are the cost due to correlations between the input and the output, quantified by $\mathrm{I}[\overleftarrow{Y}'; \overrightarrow{Y}]$, and the cost due to correlations between the transducer and the entire input-output sequence, quantified by $\mathrm{I}[X_0; \overleftarrow{Y}', \overrightarrow{Y}]$. The last term, which through X_0 depends explicitly on the transducer's structure, shows how different implementations of the same computation change energetic requirements.

Said differently, we can alter transducer states as well as their interactions with tape symbols, all the while preserving the computation—the joint-input output distribution—and this *only* affects the last term in Eq. (3). For this reason, we call it the *minimal implementation energy cost* Q^{impl} given a transducer:

$$(k_{\mathrm{B}} T \ln 2)^{-1} \langle Q^{\mathrm{impl}} \rangle_{\min} = \mathrm{I}[X_0; \overleftarrow{Y}', \overrightarrow{Y}] \; . \tag{4}$$

This bound on the cost applies beyond predictively generating an output process [33]—it applies to any type of input-output transformation. The minimum implementation energy cost is not guaranteed to be achievable, but, like Landauer's bound on the cost of erasure [22, 23], it provides a guidepost for an essential physical cost of information processing. By choosing an implementation with the least information shared between the transducer's state and the joint state of the output past and input future, we minimize the unavoidable cost of computation. Moreover, this allows us to show how to achieve the minimum dissipation by employing physical implementations of pattern generators [33].

Generating Structured Patterns. Paralleling Ref. [26], we now consider the thermodynamic cost of generating a sequential pattern of output symbols from a sequence of IID input symbols. Since the latter are uncorrelated and we restrict ourselves to nonanticipatory transducers (i.e., transducers with no direct access to future input [27]), the input future is statistically independent of both the current transducer state and the output past: $\mathrm{I}[X_0; \overleftarrow{Y}', \overrightarrow{Y}] = \mathrm{I}[X_0; \overleftarrow{Y}']$ and $\mathrm{I}[\overleftarrow{Y}'; \overrightarrow{Y}] = 0$. As a result, we have the following simplifications for the minimal transient dissipation and implementation costs (from Eq. (3)):

$$(k_{\mathrm{B}} T \ln 2)^{-1} \langle Q^{\mathrm{tran}} \rangle_{\min} = \mathrm{I}[X_0; \overleftarrow{Y}'] - \mathbf{E}' \tag{5}$$

$$(k_{\mathrm{B}} T \ln 2)^{-1} \langle Q^{\mathrm{impl}} \rangle_{\min} = \mathrm{I}[X_0; \overleftarrow{Y}'] \; . \tag{6}$$

The fact that the input is IID tells us that the transducer's states are also the internal states of the hidden Markov model (HMM) generator of the output process [26, 27]. This means that the transducer variable X_0 must contain all information shared between the output's past $\overleftarrow{Y}'$ and future $\overrightarrow{Y}'$ [40, 41], as shown in the *informa-*

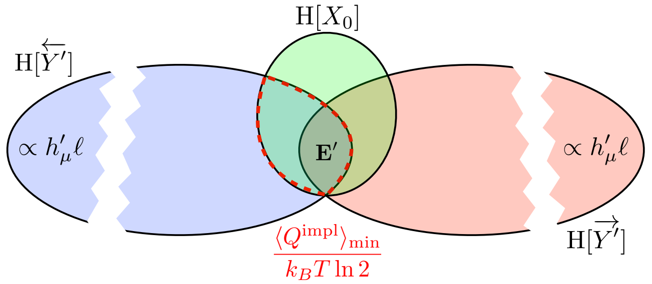

FIG. 1. Shannon measures for physical information transduction—general case of nonunifilar transducers: Transducer output past $\overleftarrow{Y}'$ and output future $\overrightarrow{Y}'$ left (blue) and right (red) ellipses, respectively; shown broken since the future and past entropies $\mathrm{H}[\overleftarrow{Y}']$ and $\mathrm{H}[\overrightarrow{Y}']$ diverge as $h_\mu \ell$, with ℓ being the length of past or future, respectively. $\mathrm{H}[X_0]$ illustrates the most general relationship the generating transducer state X_0 must have with the process future and past. Implementation cost $\mathrm{I}[X_0; \overleftarrow{Y}'] = \langle Q^{\mathrm{impl}}\rangle_{\min}/k_B T \ln 2$ is highlighted by a dashed (red) outline.

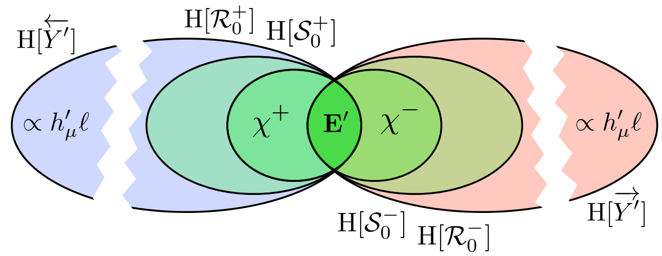

FIG. 2. Optimal physical information transducers—predictive and retrodictive process generators: Process generator variables, predictive states $\mathcal{R}$ and causal states $\mathcal{S}$, denoted with green ellipses. Being the minimal set of predictive states, causal states $\mathcal{S}$ are contained within the set of general predictive states $\mathcal{R}$. A given process has alternative unifilar ($\mathcal{R}_0^+$ or $\mathcal{S}_0^+$) and counifilar generators ($\mathcal{R}_0^-$ or $\mathcal{S}_0^-$). Component areas are the sigma-algebra *atoms*: conditional entropies—entropy rate h_μ and crypticities χ^+ and χ^-—and a mutual information—the excess entropy $\mathbf{E}'$. Since the state random variables $\mathcal{R}_0^+$ and $\mathcal{S}_0^+$ are functions of the output past $\overleftarrow{Y}'$, their entropies are wholly contained within the past entropy $\mathrm{H}[\overleftarrow{Y}']$. Similarly, counifilar generators, denoted by the random variables $\mathcal{R}_0^-$ and $\mathcal{S}_0^-$, are functions of output future $\overrightarrow{Y}'$. Thus, their entropies are contained within the output future entropy $\mathrm{H}[\overrightarrow{Y}']$. The ϵ-machine generator with causal states $\mathcal{S}_0^+$ is the unifilar generator with minimal Shannon entropy (area). The random variable $\mathcal{R}_0^-$ realizes the current state of the minimal counifilar generator, which is the time reversal of the ϵ-machine for the time-reversed process [42]. Transducers taking the form of any of these generators produce the same process, but structurally distinct generators exhibit different dissipations and thermodynamic implementation costs.

tion diagram in Fig. 1. (Graphically, the $\mathbf{E}'$ atom is entirely contained within $\mathrm{H}[X_0]$.) There, an ellipse depicts a variable's Shannon entropy, an intersection of two ellipses denotes the mutual information between variables, and the exclusive portion of an ellipse denotes a variable's conditional entropy. For example, $\mathbf{E}' = \mathrm{I}[\overleftarrow{Y}'; \overrightarrow{Y}']$ is the intersection of $\mathrm{H}[\overleftarrow{Y}']$ and $\mathrm{H}[\overrightarrow{Y}']$. And, the leftmost crescent in Fig. 1 is the conditional Shannon entropy $\mathrm{H}[\overleftarrow{Y}'|X_0]$ of the output past $\overleftarrow{Y}'$ conditioned on transducer state X_0. The diagram also notes that this information atom, which is in principle infinite, scales as $h_\mu \ell$, where ℓ is the sequence length.

As stated above, Fig. 1 also shows that the ratchet state statistically shields past from future, since the ratchet-state entropy $\mathrm{H}[X_0]$ (green ellipse) contains the information $\mathbf{E}'$ shared between the output past and future (overlap between (left) blue and right (red) ellipses). Thus, the implementation cost $\mathrm{I}[X_0; \overleftarrow{Y}']$, highlighted by dashed (red) outline, necessarily contains the mutual information between the past and future. The Supplementary Materials show that both the transient and asymptotic bounds—Eqs. (2) and (5), respectively—are achievable through an alternating adiabatic and quasistatic protocol. We are now ready to find the most efficient thermodynamic implementations for a given computation.

Consider first the class of predictive, *unifilar* information transducers; denote their states $\mathcal{R}_0^+$. Unifilarity here says that the current state $\mathcal{R}_0^+$ is restricted to be a function of the semi-infinite output past: the ratchet's next state $\mathcal{R}_0^+$ is unambiguously determined by $\overleftarrow{Y}'$.

A unifilar information transducer corresponds to the case where the transducer state entropy $\mathrm{H}[X_0 = \mathcal{R}_0^+]$ has no area outside that of the output past's entropy $\mathrm{H}[\overleftarrow{Y}']$. (See Fig. 2.) As evident there, the implementation cost $\mathrm{I}[X_0; \overleftarrow{Y}']$ is the same as the transducer's *state uncertainty*—the Shannon entropy $\mathrm{H}[X_0 = \mathcal{R}_0^+]$. Thus, according to Eq. (6) the thermodynamically most efficient unifilar transducer is that with minimal state-uncertainty $\mathrm{H}[X_0 = \mathcal{S}_0^+]$—the entropy of the ϵ-*machine causal states* $\mathcal{S}_0^+$ of computational mechanics [37], which comprise the minimal set of predictive states [43]. This confirms the result that, if one is restricted to *predictive* generators, simpler is better [33].

Critically, there are further connections with computational mechanics that, by removing the restriction, lead to substantial generalizations. For ϵ-machine information transducers with causal states $\mathcal{S}_0^+$, the mutual information between the transducer and the output past is the output process' statistical complexity: $\mathrm{I}[\mathcal{S}_0^+; \overleftarrow{Y}'] = C'_\mu$. In other words, the minimal implementation cost of a pattern generated by a unifilar information transducer is the pattern's statistical complexity. The transient dissipation that occurs when generating a structured pattern, given in Eq. (5), is then the output's crypticity $\chi^+ = C'_\mu - \mathbf{E}'$ [41], as Ref. [42] concluded previously and Ref. [33] more recently.

Now, consider the more general case in which we allow the transducer implementation to be nonunifilar; see Fig. 1 again. From the Data Processing Inequality [44], it follows that the mutual information between X_0 and $\overleftarrow{Y}'$ cannot be less than the output's excess entropy:

$$\mathrm{I}[X_0; \overleftarrow{Y}'] \geq \mathbf{E}' \ . \tag{7}$$

Thus, the minimum structural cost over alternate pattern-generator implementations is the output pattern's excess entropy.

Figure 1 suggests how to find this minimum. The implementation cost highlighted by the dashed (red) line can be minimized by choosing a transducer whose states are strictly functions of the future. In this case, the transducer's mutual information with the output past is simply $\mathbf{E}'$, achieving the bound on implementation cost given by Eq. (7). (Refer now to Fig. 2.) Constructed using states that are functions of the future, such a ratchet is a generator with retrodictive (as opposed to predictive) states, denoted $\mathcal{R}_0^-$ or $\mathcal{S}_0^-$ [45]. This means that the generator is *counifilar*, as opposed to unifilar [46, 47]. These generators have the same states as the unifilar generators of the time-reversed process, but generally are nonunifilar. These *retrodictive generators* produce the same output process by running along the information reservoir in the same way as the predictive generators, but rather than store all of the information in the past outputs required to *predict* the future, they only store just enough to *generate* it. This affords them a fundamental energetic advantage.

Critically, too, any such retrodictive implementation is maximally efficient, dissipating zero transient heat $\langle Q^{\mathrm{tran}} \rangle_{\min} = 0$, even though the state uncertainty varies across implementations: $\mathrm{H}[\mathcal{R}_0^-] > \mathrm{H}[\mathcal{S}_0^-]$. Unlike unifilar transducers, for a given output process there are infinitely many counifilar information transducers of varying state-complexity that are all maximally thermodynamically efficient. In other words, simpler is not necessarily thermodynamically better for optimized transducers, as we demonstrate using the example of the $(3, 2)$ Golden Mean Process in the Supplementary Materials. Figure S1 there demonstrates the distinct thermodynamic advantages of retrodictive representations. This shows, as a practical matter, that both the design and evolution of efficient biological computations have a wide latitude when it comes to physical instantiations.

To summarize, we identified the transient and structural thermodynamic costs of physical information transduction, generalizing the recent Information Process Second Law. These bound the energetic costs incurred by any physically embedded adaptive system as it comes to synchronize with the states of a structured environment. Physical organization in the environment is a thermal resource for adaptive biological agents. To take advantage of that resource, however, the agent's internal state must reflect the hidden structure of its input [17, 24, 33]. Similarly, when producing an organized output, the agent must transition among the recurrent hidden states that are capable of generating the desired structure. Information transducing agents such as these involve a transient phase during which additional costs are incurred due to the agent adapting to its asymptotic behavior. When asking about which physical implementations are the most thermodynamically efficient we showed that they can be bounded by the information shared by the agent, output past, and input future. This led us to see that the most efficient generators of organization are retrodictive, not necessarily ϵ-machines which are generative but predictive.

Supplementary Materials: Derivations, further discussion and interpretation, and an explicit comparison of the thermodynamics of generators and predictors for an example system.

We thank Ryan James and John Mahoney for helpful discussions on prediction versus generation. As External Faculty, JPC thanks the Santa Fe Institute for its hospitality during visits. This work was supported in part by FQXi Grant number FQXi-RFP-1609 and the U.S. Army Research Laboratory and the U.S. Army Research Office under contract W911NF-13-1-0390.

* abboyd@ucdavis.edu
† dibyendu.mandal@berkeley.edu
‡ pmriechers@ucdavis.edu
§ chaos@ucdavis.edu

[1] K. Maruyama, F. Nori, and V. Vedral. *Rev. Mod. Phys.*, 81:1, 2009.

[2] T. Sagawa and M. Ueda. *Phys. Rev. Lett.*, 104:090602, 2010.

[3] J. M. Horowitz and S. Vaikuntanathan. *Phys. Rev. E*, 82:061120, 2010.

[4] D. Mandal and C. Jarzynski. *Proc. Natl. Acad. Sci. USA*, 109(29):11641–11645, 2012.

[5] D. Mandal, H. T. Quan, and C. Jarzynski. *Phys. Rev. Lett.*, 111:030602, 2013.

[6] P. Strasberg, G. Schaller, T. Brandes, and M. Esposito. *Phys. Rev. Lett.*, 110:040601, 2013.
[7] A. C. Barato and U. Seifert. *Europhys. Lett.*, 101:60001, 2013.
[8] J. Hoppenau and A. Engel. *Europhys. Lett.*, 105:50002, 2014.
[9] Z. Lu, D. Mandal, and C. Jarzynski. *Physics Today*, 67(8):60–61, 2014.
[10] A. B. Boyd and J. P. Crutchfield. *Phys. Rev. Lett.*, 116:190601, 2016.
[11] A. Chapman and A. Miyake. *Phys. Rev. E*, 92:062125, 2015.
[12] N. Merhav. *J. Stat. Mech.*, P06037, 2015.
[13] J. Um, H. Hinrichsen, C. Kwon, and H. Park. *arXiv:1501.03733*.
[14] J. V. Koski, A. Kutvonen, I. M. Khaymovich, T. Ala-Nissila, and J. P. Pekola. *Phys. Rev. Lett.*, 115(26):260602, 2015.
[15] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa. *Nat. Phys.*, 11:131, 2015.
[16] S. Rana and A. M. Jayannavar. *J. Stat. Mech.*, 2016:103207.
[17] A. B. Boyd, D. Mandal, and J. P. Crutchfield. *Phys. Rev. E*, 95(1):012152, 2017.
[18] S. Deffner and C. Jarzynski. *Phys. Rev. X*, 3:041003, 2013.
[19] T. Sagawa and M. Ueda. In R. Klages, W. Just, and C. Jarzynski, editors, *Noneq. Stat. Phys. Small Syst.*, pages 181–211. Wiley-VCH, Weinheim, 2013.
[20] A. C. Barato and U. Seifert. *Phys. Rev. E*, 90:042150, 2014.
[21] U. Seifert. *Phys. Rev. Lett.*, 95:040602, 2005.
[22] C. H. Bennett. *Intl. J. Theo. Phys.*, 21:905, 1982.
[23] R. Landauer. *IBM J. Res. Develop.*, 5(3):183–191, 1961.
[24] A. B. Boyd, D. Mandal, and J. P. Crutchfield. *J. Stat. Phys.*, in press, arxiv.org:1609.05353, 2016.
[25] See Supplementary Materials: Information reservoir implementations.
[26] A. B. Boyd, D. Mandal, and J. P. Crutchfield. *New J. Physics*, 18:023049, 2016.
[27] N. Barnett and J. P. Crutchfield. *J. Stat. Phys.*, 161(2):404–451, 2015.
[28] P. Strasberg, J. Cerrillo, G. Schaller, and T. Brandes. arXiv:1506.00894.
[29] See Supplementary Materials: Turing machines versus transducers.
[30] T. Hatano and S. Sasa. *Phys. Rev. Lett.*, 86:3463–3466, 2001.
[31] D. Mandal and C. Jarzynski. *J. Stat. Mech.*, 2016(6):063204.
[32] P. M. Riechers and J. P. Crutchfield. *arXiv:1610.09444*.
[33] A. J. P. Garner, J. Thompson, V. Vedral, and M. Gu. arXiv:1510.00010.
[34] A. C. Barato and U. Seifert. *Phys. Rev. Lett.*, 112:090601, 2014.
[35] R. W. Yeung. *Information Theory and Network Coding.* Springer, New York, 2008.
[36] J. P. Crutchfield and K. Young. *Phys. Rev. Let.*, 63:105–108, 1989.
[37] J. P. Crutchfield. *Nature Physics*, 8:17–24, 2012.
[38] See Supplementary Materials: Information generation in physical systems.
[39] The following explains this contrast with Ref. [33]'s optimality claim for the ϵ-machine.
[40] J. P. Crutchfield and D. P. Feldman. *CHAOS*, 13(1):25–54, 2003.
[41] J. P. Crutchfield, C. J. Ellison, and J. R. Mahoney. *Phys. Rev. Lett.*, 103(9):094101, 2009.
[42] C. J. Ellison, J. R. Mahoney, R. G. James, J. P. Crutchfield, and J. Reichardt. *CHAOS*, 21(3):037107, 2011.
[43] One might conclude that simpler (smaller forward causal-state entropy) is thermodynamically better (more efficient) [33]. Our development shows when this holds and when not, leading to a broader and more incisive view of optimal adaptive systems.
[44] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley-Interscience, New York, second edition, 2006.
[45] C. J. Ellison, J. R. Mahoney, and J. P. Crutchfield. *J. Stat. Phys.*, 136(6):1005–1034, 2009.
[46] J. R. Mahoney, C. J. Ellison, R. G. James, and J. P. Crutchfield. *CHAOS*, 21(3):037112, 2011.
[47] J. P. Crutchfield, C. J. Ellison, J. R. Mahoney, and R. G. James. *CHAOS*, 20(3):037105, 2010.

Supplementary Materials
for
Transient Dissipation and Structural Costs of Physical Information Transduction

Alexander B. Boyd, Dibyendu Mandal, Paul M. Riechers, and James P. Crutchfield

SHANNON VERSUS KOLMOGOROV-SINAI ENTROPY RATES

On the one hand, it is now commonplace shorthand in physics to describe a symbol-based process in terms of Shannon's information theory and so measure its intrinsic randomness via the Shannon entropy rate. On the other, properly capturing the information processing behavior of physically embedded systems is more subtle. There are key conceptual problems. For one, the symbol-based Shannon entropy rate may not be well defined for a given continuous physical system. In the present setting we consider the entire engine as a physical system. Then, h_μ and h'_μ are the Kolmogorov-Sinai entropies of the associated reservoir dynamical systems [S1,S2]. They are well defined suprema over all coarse-grainings of the system's state space.

INFORMATION RESERVOIRS BEYOND TAPES

Rather than implement the information reservoir as a tape of symbols, one can simply employ a one-dimensional lattice of Ising spins. Moreover, the reservoir need *not* be 1D, but this is easiest to analyze since the total entropy of a 1D sequence is related to (but not equal to) the Shannon entropy rate and an engine simply accesses information by moving sequentially along the tape. Higher-dimension reservoirs, even with nonregular topologies connecting the information-bearing degrees-of-freedom, can be a thermodynamic resource when there is large total correlation among its degrees of freedom, at the cost of decorrelating the information-bearing degrees of freedom [S3, Ch. 9].

ACHIEVABILITY AND IMPLEMENTABILITY

A protocol that changes the joint Hamiltonian of the ratchet and information reservoir quasistatically in time is thermodynamically reversible. The joint distribution evolves according to the time-dependent Boltzmann distribution [S4]. It is possible to achieve the bounds of Eqs. (3) and (4) with the help of such protocols. While optimal protocols for general information transduction have not been formulated yet, a recent work describes a protocol for process generation that achieves the information-theoretic bounds, both the asymptotic and the transient [33]. It follows an alternating sequence of adiabatic (fast) and then quasistatic control of energy levels. It is noteworthy that, to take advantage of all correlations between the bits and ratchet, the efficient quasistatic protocol described there requires the ratchet to interact the entire sequence of bits simultaneously. If we limit ourselves to ratchets that interact with one bit at a time, as is the standard implementation, the bounds may not be achievable, even for process generation. We will explore the resulting difference in efficiency in greater detail in a later work.

THERMODYNAMICS OF GENERAL COMPUTATION

We focused on spatially-unidirectional information transduction due to the stronger thermodynamic results. However, the thermodynamic results are valid much more broadly, applicable to Turing-equivalent machines as well as non-1D information transduction, as just noted.

First, Turing machines that move unidirectionally, reading input tape cells once and writing results only once to an output tape, are equivalent to the transducers used here. However, unidirectional Turing machines employ internal tapes as scratch storage [S5] and this now-internal memory must be taken into account when assessing thermodynamic resources.

Second, the choice of *implementation* of a particular computation implies a transient thermodynamic cost above the asymptotic implementation-independent work rate. The general result is:

$$\langle Q^{\text{tran}}_{0\to N}\rangle / k_{\text{B}} T \ln 2 \geq \text{H}[X_0, \mathbf{Y}_0] - \text{H}[X_N, \mathbf{Y}_N] + N \Delta h \ ,$$

where $\mathbf{Y}_N$ is the random variable for the information-bearing degrees of freedom at time N and Δh is the difference in the extensive component of the entropy density of the output and input tape processes. In short, *the transient cost due to an implementation stems from the correlation built up between the device's state and the pattern on which it acts, discounted by the intensive part of the output pattern's entropy.*

ORIGIN OF TRANSIENT INFORMATION PROCESSING COSTS

We demonstrate how the transient IPSL of Eq. (3) arises. The steps give additional insight.

Assuming that we are able to achieve asymptotic IPSL bounds—say, using an alternating quasistatic-adiabatic protocol [33]—the cumulative transient cost of information processing over the interval $t \in [0, N]$ is given by:

$$\langle Q^{\text{tran}}_{0\to N}\rangle \equiv \langle Q_{0\to N}\rangle - N(h_\mu - h'_\mu) k_{\text{B}} T \ln 2 \ . \tag{S1}$$

Combining with Eq. (1), yields:

$$\begin{aligned}
\frac{\langle Q^{\text{tran}}_{0\to N}\rangle}{k_{\text{B}} T \ln 2} &\geq \text{H}[X_0, \mathbf{Y}_0] - \text{H}[X_N, \mathbf{Y}_N] + N(h'_\mu - h_\mu) \\
&= \text{H}[X_0, Y_{0:\infty}] - \text{H}[X_N, Y'_{0:N}, Y_{N,\infty}] + N(h'_\mu - h_\mu) \\
&= (\text{H}[X_0] + \text{H}[Y_{0:\infty}] - \text{I}[X_0; Y_{0:\infty}]) - (\text{H}[X_N] + \text{H}[Y'_{0:N}, Y_{N:\infty}] - \text{I}[X_N; Y'_{0:N}, Y_{N:\infty}]) + N(h'_\mu - h_\mu) \ .
\end{aligned}$$

The last line used the standard identity $\text{H}[A, B] = \text{H}[A] + \text{H}[B] - \text{I}[A; B]$ for random variables A and B. Since we are interested in the purely transient cost and not spurious costs arising from arbitrary initial conditions, we start the engine in its stationary state, resulting in stationary behavior, so that $\text{H}[X_0]$ is the same as $\text{H}[X_N]$. Furthermore, we assume that the engine's initial state is uncorrelated with the incoming symbols and so disregard $\text{I}[X_0; Y_{0:\infty}]$. We then decompose the terms $\text{H}[Y'_{0:N}, Y_{N:\infty}]$ and $\text{H}[Y_{0:\infty}] \equiv \text{H}[Y_{0:N}, Y_{N:\infty}]$ according to the above. These assumptions and decompositions lead to:

$$\begin{aligned}
\frac{\langle Q^{\text{tran}}_{0\to N}\rangle}{k_{\text{B}} T \ln 2} \geq\ & \text{H}[Y_{0:N}] + \text{H}[Y_{N:\infty}] - \text{I}[Y_{0:N}; Y_{N:\infty}] - \text{H}[Y'_{0:N}] - \text{H}[Y_{N:\infty}] + \text{I}[Y'_{0:N}; Y_{N:\infty}] \\
& + \text{I}[X_N; Y'_{0:N}, Y_{N:\infty}] + N(h'_\mu - h_\mu) \ .
\end{aligned} \tag{S2}$$

In the limit of large N, in which the transducer has interacted with a sufficiently large number of input symbols, we can invoke the following definitions of excess entropy:

$$\begin{aligned}
\mathbf{E} &= \lim_{N\to\infty} (\text{H}[Y_{0:N}] - N h_\mu) \\
&= \lim_{N\to\infty} I[Y_{0:N}; Y_{N:\infty}] \\
\mathbf{E}' &= \lim_{N\to\infty} (H[Y'_{0:N}] - N h'_\mu) \ .
\end{aligned}$$

Upon shifting to the ratchet's reference frame and switching back to the more intuitive notation: $X_N \to X_0$, $Y_{0:N} \to \overleftarrow{Y}$, $Y'_{0:N} \to \overleftarrow{Y}'$, and $Y_{N:\infty} \to \overrightarrow{Y}$, in which a random variable with a left arrow means the past and a right arrow the future, and invoking the above definitions, new notation, and these definitions, we rewrite the inequality Eq. (S2), after some cancellation, as:

$$\frac{\langle Q^{\text{tran}}\rangle}{k_{\text{B}} T \ln 2} \geq -\mathbf{E}' + \text{I}[\overleftarrow{Y}'; \overrightarrow{Y}] + \text{I}[X_0; \overleftarrow{Y}', \overrightarrow{Y}] \ ,$$

where $\langle Q^{\text{tran}} \rangle$ is the total transient cost over infinite time: $\langle Q^{\text{tran}} \rangle = \lim_{N\to\infty} \langle Q^{\text{tran}}_{0\to N} \rangle$. This is our main result, Eq. (3), and the starting point for the other results.

THERMODYNAMIC EFFICIENCY OF ALTERNATIVE GENERATORS OF THE SAME PROCESS

We now explore the consequences of the bound on implementation energy cost of generating a process via an example. The $(3, 2)$ Golden Mean Process consists of strings where 0s only appear in sequences of three bookended by 1s and 1s only appear in sequences of two or more bookended by 0s. There are infinitely many HMMs that generate this process, four of which are shown in the first row in Fig. S1. The circles there denote an HMM's hidden states and the arrows denote the transitions between states. Each transition is associated with an output. Using this Mealy HMM representation, computational mechanics describes a wide variety of process properties and gives ways to calculate HMM information-theoretic properties [36]. State transitions in Mealy HMMs are specified by symbol-labeled transition matrices, with elements $T^{(y')}_{x\to x'}$ being the probability of transitioning to hidden state x' and outputting y' given the current hidden state x:

$$T^{(y')}_{x\to x'} = \Pr(Y'_N = y', X_{N+1} = x' | X_N = x) \ .$$

This is depicted by labeling the arrow from x to x' with $T^{(y')}_{x\to x'}$. For instance, in Fig. S1 we see that the ϵ-machine generator (the HMM in the first row, second column) has a transition from hidden state A to hidden state B with probability $T^{(0)}_{A\to B} = 0.5$. This means, if the ϵ-machine is in hidden state A, it has the probability 0.5 to make a transition to hidden state B and produce an output 0.

It is easy to see how the HMMs in the first row lead to the $(3, 2)$ Golden Mean Process. Consider, again, the $\epsilon-$machine HMM. If we happen to be in state A, with probability 0.5 we make a transition to state B and output a 0. Afterwards, we are forced to output two more 0s as we make the only possible transitions $B \to C \to D$. Then, we produce 1s until we arrive at A again. Clearly, 0s are produced in threes and 1's in pairs or more.

Each distinct HMM corresponds to a different information ratchet operating on an IID input process. As described in Ref. [26], a ratchet is characterized by a Markov transition dynamic $M_{x\otimes y\to x'\otimes y'}$ over the joint state space of the ratchet and interaction bit:

$$M_{x\otimes y\to x'\otimes y'} = \Pr(Y'_N = y', X_{N+1} = x' | X_N = x, Y_N = y) \ .$$

To implement a particular HMM generator of a process from an IID input, we choose a ratchet which ignores the input y:

$$M_{x\otimes y\to x'\otimes y'} = T^{(y')}_{x\to x'} \ .$$

The result is an output process generated by $T^{(y')}_{x\to x'}$, where the ratchet states $\mathcal{X}$ take the place of hidden states. For a device that quasistatically implements [33] the HMM process generator and takes advantage of all correlations to achieve the information-theoretic bounds, $\langle Q^{\text{impl}} \rangle_{\min}$ is its implementation cost when generating the process. In the following, we calculate $\langle Q^{\text{impl}} \rangle_{\min} = \mathrm{I}[X_0; \overleftarrow{Y}']$ for each of the four HMMs.

Figure S1 compares the implementation costs for both unifilar (predictive) and counifilar (retrodictive) generators, as discussed in the main text. The table shows that for the unifilar generators (shown in the first two columns), which have states that are a function of the output past, the implementation energy cost is proportional to the state uncertainty $\mathrm{H}[X_0]$ of the ratchet-generator states. That is, a simpler ratchet with fewer states is more thermodynamically efficient. Thus, the most efficient predictive generator is the ϵ-machine (second column), which dissipates $k_B T \ln 2$ less when synchronizing than the larger unifilar model presented in the first column. However, both the simple and more complex counifilar generators (shown in the last two columns) outperform the ϵ-machine in terms of efficiency. They achieve the theoretical bound on implementation cost, which is $k_B T \ln 2$ times the excess entropy of the output process $\mathbf{E}' = 1.5850$. Thus, even extremely complex generators can be maximally efficient as long as they are counifilar and their internal states are a function of future outputs.

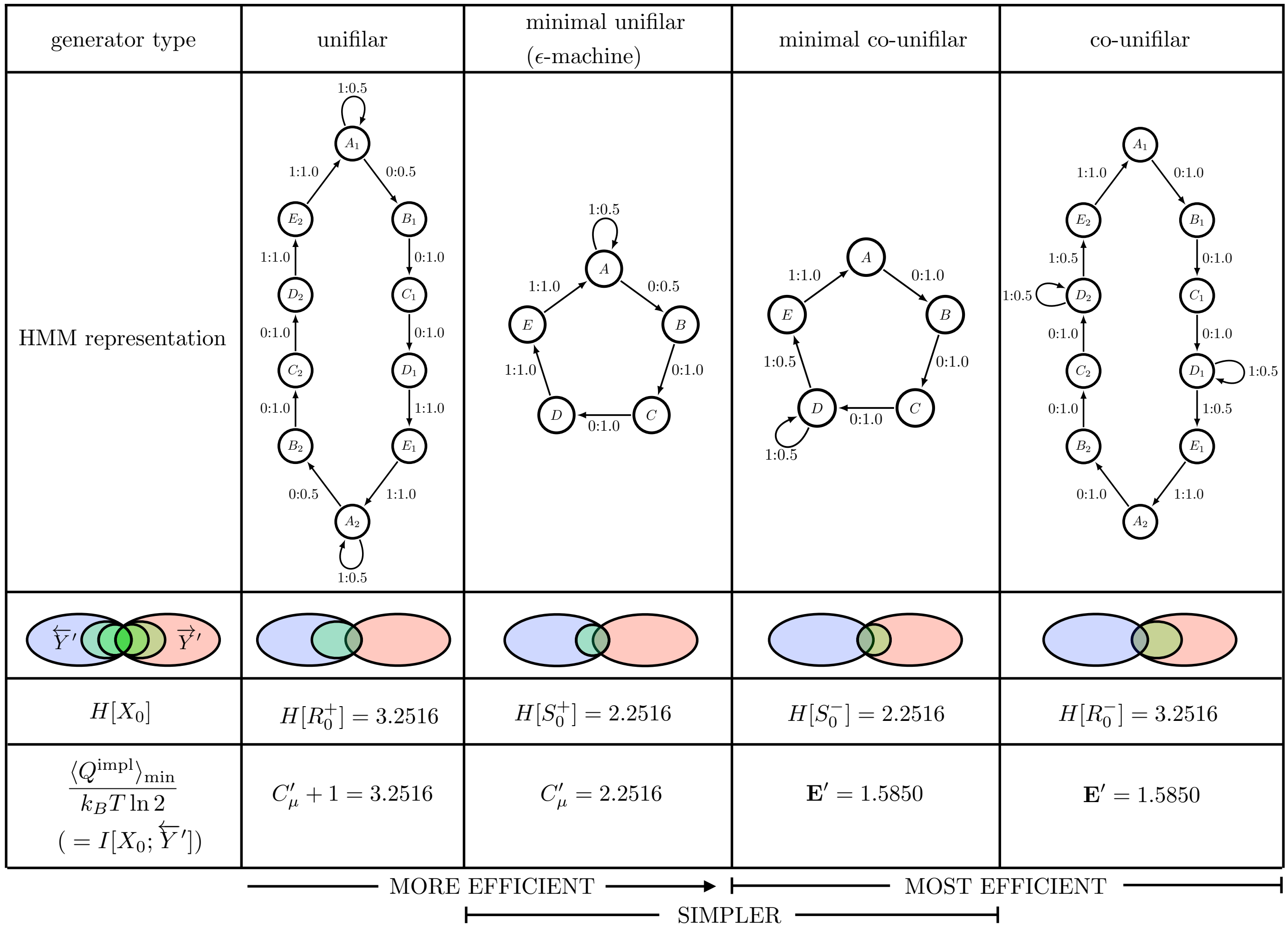

FIG. S1. Four hidden Markov models that generate the same $(3, 2)$ Golden Mean Process. On the one hand, in the first two columns the representations are unifilar and, thus, have states that are functions of the past. For these two HMMs, the ratchet state uncertainty is proportional to the representation dissipation. Thus, the HMM in the first column, with an extra bit of state uncertainty above the ϵ-machine in the second column, also has additional representation dissipation. On the other hand, the two counifilar models in the last two columns minimize dissipation, but have different state uncertainties. They are more thermodynamically efficient than the first two HMMs.

REFERENCES

[S1] A. N. Kolmogorov. Entropy per unit time as a metric invariant of automorphisms. *Dokl. Akad. Nauk. SSSR*, 124:754, 1959. (Russian) Math. Rev. vol. 21, no. 2035b.

[S2] Ja. G. Sinai. On the notion of entropy of a dynamical system. *Dokl. Akad. Nauk. SSSR*, 124:768, 1959.

[S3] P. M. Riechers. *Exact Results Regarding the Physics of Complex Systems via Linear Algebra, Hidden Markov Models, and Information Theory.* PhD thesis, University of California, Davis, 2016.

[S4] D. Mandal and C. Jarzynski, Analysis of slow transitions between nonequilibrium steady states, J. Stat. Mech.: Theor. Exp. 2016:063204.

[S5] J. E. Hopcroft, R. Motwani, and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation.* Prentice-Hall, New York, third edition, 2006.