Entropy production in communication channels

Farita Tasnim[*]

Massachusetts Institute of Technology, Cambridge, 02139 Massachusetts, USA

Nahuel Freitas [†]

Departamento de Fisica, FCEyN, UBA, Pabellon 1, Ciudad Universitaria, 1428 Buenos Aires, Argentina

David H. Wolpert [‡]

Santa Fe Institute, Santa Fe, New Mexico, USA;
Complexity Science Hub, Vienna, Austria;
Arizona State University, Tempe, Arizona, USA;
International Center for Theoretical Physics, Trieste 34151, Italy;
and Albert Einstein Institute for Advanced Study, New York, New York, USA

In many complex systems, whether biological or artificial, the thermodynamic costs of communication among their components are large. These systems also tend to split information transmitted between any two components across multiple channels. A common hypothesis is that such inverse multiplexing strategies reduce total thermodynamic costs. So far, however, there have been no physics-based results supporting this hypothesis. This gap existed partially because we have lacked a theoretical framework that addresses the interplay of thermodynamics and information in off-equilibrium systems. Here we present the first study that rigorously combines such a framework, stochastic thermodynamics, with Shannon information theory. We develop a minimal model that captures the fundamental features common to a wide variety of communication systems, and study the relationship between the entropy production of the communication process and the channel capacity, the canonical measure of the communication capability of a channel. In contrast to what is assumed in previous works not based on first principles, we show that the entropy production is not always a convex and monotonically increasing function of the channel capacity. However, those two properties are recovered for sufficiently high channel capacity. These results clarify when and how to split a single communication stream across multiple channels.

I. INTRODUCTION

One of the major thermodynamic costs of many complex systems arises from communication among their separate subsystems. Examples include cellular sensing systems [1,2], the human brain [3,4], ecosystems [5,6], wireless sensor networks [7–9], hardware implementations of machine learning algorithms [10], and digital computers [11–13]. Similarly, reducing the thermodynamic costs of communication between processing units and memory units of conventional von Neumann computational architectures is one of the primary motivations of the field of neuromorphic computing [14–16]. As emphasized in the latest "Physics of Life" report from the National Academy of Sciences, it is crucial to understand the *common* physical principles underlying communication in biological systems, whether that be a set of interacting

bacteria, a colony of insects, a flock of birds, or a human social group [17]. It is important to appreciate that the information transmission in different communication systems are subject to different physical constraints associated to the transmission process and media [e.g., diffusive, electronic, acoustic, etc.; see Figs. 1(a)–1(c)]. In turn, each of these different constraints impose their own, system-specific thermodynamic costs, since they each limit what theoretical efficiencies the system can exploit [18]. However, in order to uncover and investigate the principles common to all types of communication systems, we need a minimal model grounded in the features shared by all of those systems.

One feature common to all of these systems is that they involve an "output" component that is biased to change its state to reflect the state of a separate "input" component, which is set exogenously [Fig. 1(d)]. A minimal model that can apply to diverse types of communication systems should include this shared aspect of copying between two separate components. Another feature common to many biological and artificial communication systems is that they often split a stream of information over multiple channels, i.e., they inverse multiplex. Sometimes this occurs even when a single channel could

---

[*]Contact author: farita@mit.edu, web: farita.me
[†]Contact author: nahuel.freitas@uni.lu
[‡]Contact author: dhw@santafe.edu, web: davidwolpert.weebly.com
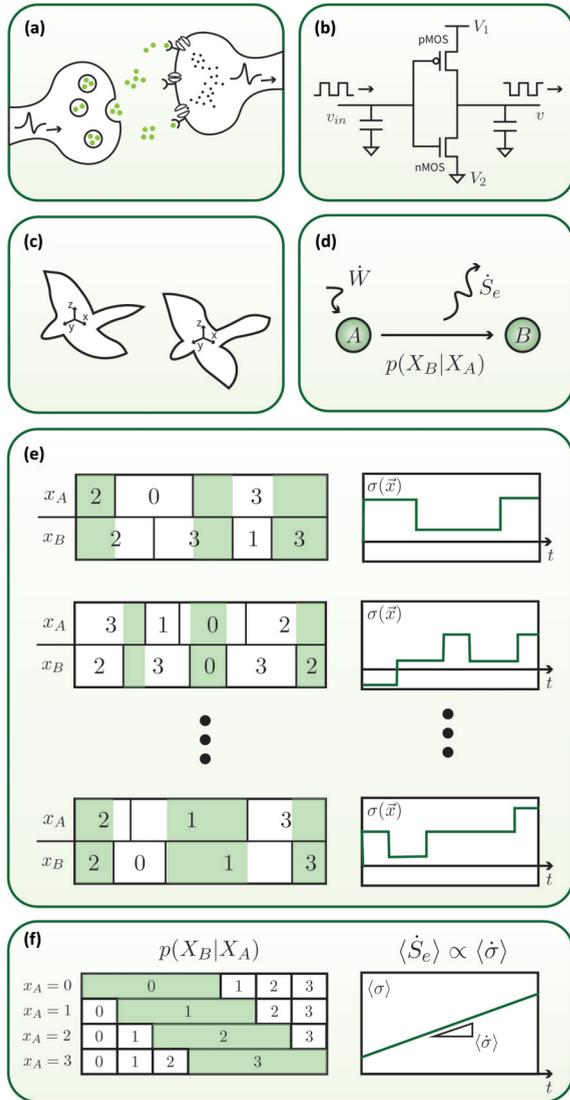
FIG. 1. Examples of communication channels include (a) neuronal synapses, (b) CMOS inverters, or (c) adjacent animals in a flock. These communication channels effectively d) consist of two nodes: an input A and an output B, which attempts to copy the state of A. The conditional probability distribution of the output state given the input state, $p(X_B|X_A)$, reflects the noisiness of the copying process. Since this process must occur out of equilibrium, it requires external work at a rate $\dot{W}$ and dissipates some amount of that work in terms of entropy flow to the surrounding environment at a rate $\dot{S}_e$. (e) Such a communication channel can be modeled as a stochastic process. For a fixed time period of length t, the system can have many different realizations, or trajectories $\vec{x}$. These trajectories capture how the state x_A of the input and the state x_B of the output coevolve. In each trajectory, the output's attempts to copy the state of the input result in an entropy production (EP) $\sigma(\vec{x})$. Boxes highlighted in green represent the time periods for which the state of the output matches the state of the input. Note that there is stochasticity both in the values of the states as well as in the timing of the state transitions. (f) Averaged over all possible trajectories, the state occupancies of the input and output define the conditional distribution $p(X_B|X_A)$, and an average EP $\langle\sigma\rangle$ of the process. Typically the average EP rate, $\langle\dot{\sigma}\rangle$, accounts for most of the rate of entropy flow to the environment, $\langle\dot{S}_e\rangle$.

handle the entire communication load. For example, multiple synapses tend to connect adjacent neurons [19–22] and multiple neuronal pathways tend to connect brain regions [23–25]. In engineering, spatial multiplexing techniques have made multiple-input-multiple-output technology the gold standard for modern wireless communication systems [26,27]. Therefore, analyzing a minimal model of communication should also result in greater insights into the thermodynamics inverse multiplexing in many scenarios. It may even provide guidance for the design of artificial inverse multiplexing systems.

Until now, it has been difficult to construct such a minimal model using a physics-based formalism. The difficulty is that communication systems operate very far from thermodynamic equilibrium, which rules out conventional arguments based on the quasistatic limit [28] or the linear-response regime [29]. However, the development of stochastic thermodynamics [30] since the early 2000s has supplied a theoretical framework that relates the dynamics of information in a system to the free energy it dissipates in a process arbitrarily far from equilibrium. Stochastic thermodynamics can allow us, therefore, to investigate how this thermodynamic cost varies with communication rate in a two-component copying system. That is the goal of this article.

We begin by providing a background on communication theory and stochastic thermodynamics. We then build on this background to motivate a minimal model of communication with a well-defined thermodynamics. In this simple model, an exogenous agent modulates the energy levels of an output system in order to transmit information. While the output system is in contact with a single heat bath, the switching of its energy levels acts as an external work source that takes it out of equilibrium. Next, we present a preliminary analysis of the thermodynamics of two types of communication encompassed by more general models. The first one is also a energy-switching scenario, but relaxing some of the assumptions made in the simple model mentioned above (namely, we consider several thermal baths and arbitrary relaxation timescales). In the second one, called reservoir switching, the exogenous agent does not change the energy levels of the output system but modulates their coupling to different thermal baths. These two communication types correspond roughly to (i) wireless communication or intercellular communication and (ii) wired electronic communication. We show that in many examples of both types of communication, the thermodynamic cost is a convex function of the communication rate above a critical value of that rate. Regardless of the convexity of this function, one might expect that increasing the speed of information transmission through a channel requires a monotonic increase in the thermodynamic costs [3,31]. However, we find that in many cases this function is *not* monotonic.

We then investigate the consequences of this result for when and how one should split a single information stream across multiple physical channels. In particular, we derive a Pareto front representing the trade-off between minimizing total thermodynamic cost and maximizing total information transmission rate shared across a fixed number of communication channels. We end by discussing our model in the broader context of the thermodynamics of computation and by suggesting future work.

II. BACKGROUND ON SHANNON INFORMATION THEORY AND STOCHASTIC THERMODYNAMICS

Claude Shannon's channel coding theorem states that one can pass messages through a noisy communication channel with vanishingly low error [32]. A channel approaches this error-free limit if the messages sent through it are encoded with an appropriate codebook into strings of letters from a finite, discrete alphabet. These encoded strings can be decoded to recover the original message at the output of the channel. Importantly, one can implement error-free communication in noisy channels in this way only up to a finite maximum baud rate, called the channel capacity. Exceeding that rate imposes a nonzero probability of error. A channel's capacity therefore serves as the primary measure of its communication capabilities. This capacity equals the maximal mutual information between input and output attainable by varying the input alphabet distribution [32,33].

The channel coding theorem further states that one can achieve the channel capacity with negligible error using any one of *many* possible optimal encodings of information at the source [34]. These encodings produce sequences that look as if they were identically and independently drawn from the input distribution that maximizes the mutual information between input and output [35]. This theorem led to massive engineering efforts toward developing error-correcting codes that achieve this bound [36,37]. Such coding strategies are now widespread in digital communication systems [38–42]. Moreoever, the study of codes that approach the Shannon bound has been one of the major areas of research in modern information theory for several decades.

The duality between information and Shannon entropy creates a tight link between information theory and thermodynamics [43–45]. Stochastic thermodynamics has revealed that the entropy produced over any finite time interval in a single realization of a nonequilibrium process [46] *can* be negative, in defiance of the second law. This entropy production (EP) recovers non-negativity only when averaged over all possible realizations (trajectories) [47,48]. Throughout the paper, we refer to the EP as the thermodynamic cost. This measure is crucial in many powerful results, including bounds on the precision of generalized currents [49,50] and the speed of changes in a system's probability distribution over states [51]. The EP has also been applied to study the dynamics of information [52].

Unfortunately, most of the prior work in stochastic thermodynamics involving mutual information and "information processing" fails to relate the EP to the channel capacity. In fact, most of these results do not consider a system dedicated to communication. Some studies analyze how components influence one another in multipartite systems of multiple feedback controllers by quantifying the "information flow" between subsystems [53–56]. Other studies, e.g., on Bayes' nets [57], use the "transfer entropy" rather than the channel capacity to characterize information transferred from inputs to outputs. Studies of specific models for biomolecular copying processes [58] and cellular sensing [18,59,60] have made analogies with information theory and the channel coding theorem, but are limited in scope to only biological systems. Similarly, the analysis of channel capacity in Ref. [61] applies only to the very restricted domain of electronic circuits comprising a set of transistors running subthreshold.

An early connection between channel capacity and thermodynamic cost appeared naturally in the study of continuous Gaussian channels. These channels are an idealization of the information transmission through a telegraphic or telephonic line or radio link. Without additional constraints such continuous channels have infinite capacity, which is not physical. A natural and realistic constraint is to impose a maximum power on the physical signal used to transmit information. Since the medium through which the signal propagates is linear, the power is just a quadratic function of the signal intensity. Under the previous assumptions, it was understood early on that the capacity C of a Gaussian channel is related to its bandwidth W, its noise spectral density N_0 and the power P of the input signal as [31]

$$C = W \ln \left(1 + \frac{P}{N_0 W} \right). \tag{1}$$

In the infinite bandwidth limit we have $C \simeq P/N_0$. Also, for a medium in thermal equilibrium at temperature T the power spectral density of the noise is $N_0 = k_b T$, which implies that one must transmit a signal with an energy of at least $E = n k_b T$ in order to communicate n bits. However, one must note that the input power P is not necessarily dissipated in the transmission medium, and cannot be directly related with the entropy produced in the transmission process. Therefore, until now the literature has lacked an investigation of the quantitative relationship between the channel capacity and the EP that could apply to all types of communication, even those with a finite set of discrete symbols. We conduct exactly this investigation, justifying the features of our minimal model with the rigor of Shannon information theory. As a final comment, note that according to Eq. (1) the transmitted power P is an increasing convex function of the channel capacity C. Those two properties, the monoticity and the convexity of the thermodynamic cost for any C, are sometimes assumed in some works that are not based on first principles [62]. Interestingly, as we will see, they are not preserved in general models of communication channels with discrete set of symbols. They are, however, recovered for sufficiently large channel capacity.

III. STOCHASTIC THERMODYNAMICS OF COMMUNICATION CHANNELS

We analyze the thermodynamics of communication channels operating at their channel capacity. In Shannon's model of communication, a process external to the system sets the input state so that its dynamics reflect samplings from the input distribution that achieves the channel capacity. This external system effectively acts as a work reservoir, and could follow arbitrary (potentially non-Markovian) dynamics. The input could be set deterministically or stochastically, periodically according to a clock, according to a continuous-time Markov chain (CTMC), etc. This feature is physically motivated by the fact that in many real communication channels the input is set by an exogenous process. Therefore, we ignore the thermodynamics of the external system that sets the input. Finally, the coupling between input and output is nonreciprocal, in that the output dynamics depends on the state of the input but not

vice versa. This common assumption is sometimes called "no back-action" [18,55,60,63–66].

We consider a communication channel in which the input A sends letters x_A from the alphabet $X_A = \{1, 2, \ldots, L\}$ in any way that reflects the distribution π_{X_A}. The communication channel is defined by how the output B changes its state, $x_B \in X_B = \{1, 2, \ldots, L\}$, in response to the input state. At any given time, the rates of state transitions in the output depend on the current state of the input. The energies of different output states may depend on the state of the input according to a Hamiltonian function $H(x_B|x_A)$. In addition, the coupling of the output system to different thermodynamic reservoirs may also depend on the state of the input A.

A. A minimax bound on the channel capacity

To begin, we derive a general bound on the capacity of a channel that will be useful for the subsequent thermodynamic analysis. The joint distribution for input-output pairs can be written as $p_{x_B,x_A} = p_{x_B|x_A}\pi_{x_A}$ in terms of the conditional distribution $p_{x_B|x_A}$. The mutual information between input and output can be written as:

$$I(A : B) = \sum_{x_A} \left(p_{x_B|x_A} \big\| p_{x_B} \right) \pi_{x_A}, \tag{2}$$

where $p_{x_B} = \sum_{x_A} p_{x_B,x_A}$ is the unconditional distribution over outputs and $(p\|q)$ is the Kullback-Leibler divergence or relative entropy between distributions p and q. The channel capacity is defined as the maximum of $I(B : A)$ over all possible input distributions π_{x_A}. Thus, we do a unconstrained optimization of the function $G \equiv I(B : A) - \alpha(\sum_{x_A} \pi_{x_A} - 1)$ over $\{\pi_{x_A} \geqslant 0\}$ and the Lagrange multiplier α. After some algebra, we find:

$$\frac{\partial G}{\partial \pi_{x_A}} = -1 - \alpha + D\left(p_{x_B|x_A} \big\| p_{x_B} \right). \tag{3}$$

Then, for the optimal solution satisfying $\partial G/\partial \pi_{x_A} = 0$, the relative entropy $D \equiv D(p_{x_B|x_A}||p_{x_B})$ must be independent of x_A. Also, in that case we have $I(A : B) = D$. Therefore, the channel capacity is $C = D$.

This analysis does not provide an explicit expression for the channel capacity, since that would require finding the distribution π_{x_A} solving $\partial G/\partial \pi_{x_A} = 0$. However, it allows to obtain a useful upper bound. Indeed, for any x_A, we have:

$$C = D\left(p_{x_B|x_A} \big\| p_{x_B} \right) = \sum_{x_B} p_{x_B|x_A} \log\left(\frac{p_{x_B|x_A}}{p_{x_B}} \right)$$

$$\leqslant \sum_{x_A', x_B} \pi_{x_A'} \, p_{x_B|x_A} \log\left(\frac{p_{x_B|x_A}}{p_{x_B|x_A'}} \right)$$

$$= \sum_{x_A'} \pi_{x_A'} \left(p_{x_B|x_A} \big\| p_{x_B|x_A'} \right), \tag{4}$$

where in the second line we used Jensen's inequality. It follows that

$$C \leqslant \max_{x_A'} \left(p_{x_B|x_A} \big\| p_{x_B|x_A'} \right) \tag{5}$$

for any x_A. Then

$$C \leqslant \min_{x_A} \max_{x_A'} \left(p_{x_B|x_A} \big\| p_{x_B|x_A'} \right). \tag{6}$$

The meaning of this inequality is clear if we interpret the relative entropy $\left(p_{x_B|x_A} \big\| p_{x_B|x_A'} \right)$ as a distance between those conditional output distributions. Indeed, if that distance is small, then it means that the output distribution does not change too much for different inputs, and therefore the mutual information between input and output must also be small.

B. Energy-switching communication: The case of a single reservoir and timescale separation

Before studying the problem in full generality, we examine what probably is the simplest model of communication channel with a well-defined thermodynamics. As we will see, this model is already nontrivial, and displays many features shared by more general settings examined later.

In this model, the output system B is in contact with a single thermal reservoir at temperature T, and the only effect of the input A is to control the energies of the output states according to the Hamiltonian:

$$H(x_B|x_A) = \begin{cases} -\epsilon & x_B = x_A \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

Thus, the output state matching the input state is energetically favored. Eq. (7) can also be understood as the time-independent Hamiltonian of the global system. For a constant input x_A, the steady state of the output is given by the equilibrium distribution:

$$p_{x_B|x_A} = \begin{cases} \dfrac{e^{\beta\epsilon}}{e^{\beta\epsilon}+L-1} & x_B = x_A \\ \dfrac{1}{e^{\beta\epsilon}+L-1} & \text{otherwise} \end{cases}, \tag{8}$$

where $\beta = 1/T$ (we take $k_b = 1$). Recall that L is the number of distinct states in both the input and output systems. The average energy and entropy of the output are therefore independent of the input and read, respectively

$$U = -\epsilon \frac{e^{\beta\epsilon}}{e^{\beta\epsilon}+L-1} \quad \text{and} \quad S = \ln(e^{\beta\epsilon}+L-1) + \beta U. \tag{9}$$

Also, we suppose that the waiting times between successive changes of the input are much longer than the relaxation time of the output. Under this timescale separation assumption, the entropy produced during the relaxation of the output following a change $x_A \to x_A'$ in the input is given by the relative entropy between the initial and the final distributions [45]. Thus, we have:

$$\sigma_{x_A \to x_A'} \equiv \left(p_{x_B|x_A} \big\| p_{x_B|x_A'} \right) = -\beta \frac{\epsilon + LU}{L - 1}, \tag{10}$$

where the notation indicates that the average defining the KL divergence is only over values of x_B, conditioned on $X_A = x_A$, not over values of x_A.

However, since the right-hand side of Eq. (10) is independent of the input transition, the average rate of entropy production is just $\dot{\sigma} = f_s \sigma$, where f_s is the average frequency of input transitions, and we define $\sigma := \sigma_{x_A \to x_A'}$. Note that Eq. (10) and the minimax bound of Eq. (6) allow us to relate
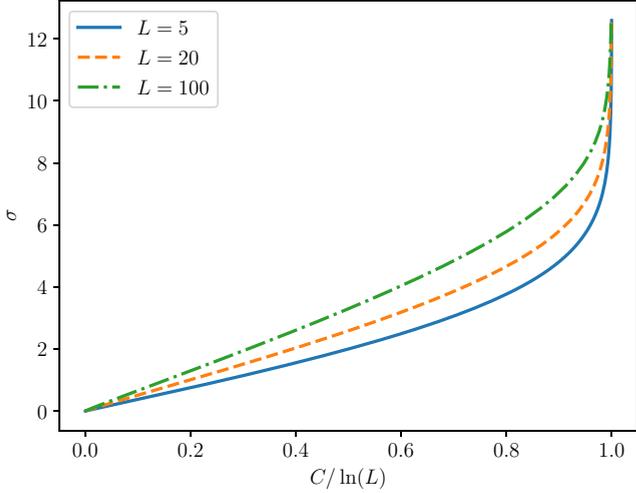
FIG. 2. Entropy production versus channel capacity for the channel given in Eq. (8).

FIG. 3. Efficiency versus channel capacity for the channel given in Eq. (8).

the entropy production with the channel capacity. This will be used later to define the thermodynamic efficiency of the channel.

In this symmetric case the channel capacity is achieved for a uniform input distribution and simply reads:

$$C = \ln(L) - S, \tag{11}$$

where S is the entropy of the output system given in Eq. (9).

We want to study the relationship between the channel capacity and the entropy production. From the previous equations it follows that:

$$\frac{d\sigma}{dC} = -\frac{d\sigma}{dS} = \frac{1}{L-1}\left[L + \frac{\epsilon + LU}{\beta U(\epsilon + U)}\right]. \tag{12}$$

Note that $d\sigma/dC \geqslant 0$, since $U \leqslant -\epsilon/L$. Then, the entropy production increases with the channel capacity.

After some algebra, the second derivative of the entropy production with respect to the channel capacity can be found to be

$$\frac{d^2\sigma}{d^2C} = \frac{1}{L-1}\frac{1}{\beta^2 U^2(\epsilon + U)}\left[\epsilon + (\beta^{-1} + U)\frac{\epsilon + LU}{\epsilon + U}\right]. \tag{13}$$

It can be seen that $d^2\sigma/d^2C \geqslant 0$ if and only if

$$e^{2\beta\epsilon}(\beta\epsilon - 1) - e^{\beta\epsilon}(L-2) + (\beta\epsilon + 1)(L-1) \geqslant 0, \tag{14}$$

which is always true for $L = 2$. For a given $L > 2$, we have $d^2\sigma/d^2C < 0$ for small $\beta\epsilon$ and $d^2\sigma/d^2C > 0$ for large $\beta\epsilon$. As a consequence, as a function of the channel capacity C, the entropy production σ is concave for small C and convex above a critical value. Finally, we note that in the limit $\beta\epsilon \gg 1$ in which C approaches $\ln(L)$, the entropy production diverges as

$$\sigma \simeq \ln\left(\frac{L-1}{\ln(L)-C}\right) + \ln\left(1 + \ln\left(\frac{L-1}{\ln(L)-C}\right)\right). \tag{15}$$

Figure 2 shows the entropy production as a function of the channel capacity. The concave region at low capacity is too flat and too small to notice in the plot. Another interesting

quantity to study is the ratio $\eta \equiv C/\sigma$ between channel capacity and entropy production, which can be considered the thermodynamic efficiency of the channel. It is shown in Fig. 3, where we see that we always have $\eta \leqslant 1$. That this must be the case is a direct consequence of Eqs. (6) and (10).

C. General case

We now study more general models where we relax the main assumptions made above, namely (i) that the output system is in contact with a single thermal bath, (ii) that its relaxation time is shorter than the typical time between input transitions, and (iii) that the only effect of the input is to modulate the energies of the output states. The dynamics of both input and output states is modeled as a CTMC [Figs. 1(e) and 1(f)]. Then, the joint distribution $p_{x_A,x_B}(t)$ evolves according to the master equation:

$$\dot{p}_{x_A,x_B}(t) = \sum_{x'_A,x'_B} R^{x'_A,x'_B}_{x_A,x_B} \, p_{x'_A,x'_B}(t), \tag{16}$$

where $R^{x'_A,x'_B}_{x_A,x_B}$ are the elements of a $L^2 \times L^2$ rate matrix. Each nondiagonal element of that matrix indicates the probability of observing a state transition $(x'_A, x'_B) \rightarrow (x_A, x_B)$ at any given time. The diagonal elements are such that the normalization is preserved. The input and output together evolve as a bipartite process [67], so only one subsystem changes state at any given time. Thus, each nondiagonal element of the rate matrix obeys

$$R^{x'_A,x'_B}_{x_A,x_B} = \delta^{x'_B}_{x_B} J^{x'_A}_{x_A}(x_B) + \delta^{x'_A}_{x_A} K^{x'_B}_{x_B}(x_A), \tag{17}$$

where $\delta^{x'}_x$ is the Kronecker delta function that equals 1 when $x' = x$, and equals 0 otherwise. Additionally, the output B changes its state due to interactions with a set of N equilibrium reservoirs $\mathcal{V} := \{v_1, v_2, \ldots, v_N\}$, at temperatures $\{T_1, T_2, \ldots, T_N\}$. We denote the temperature of reservoir v as T_v. This means that the overall rate matrix $K^{x'_B}_{x_B}(x_A)$ can be written as a sum over the rate matrices representing each

reservoir's effect on the output's state transitions:

$$K_{x_B}^{x_B'}(x_A) = \sum_{v=1}^{N} K_{x_B}^{x_B'}(x_A; v). \tag{18}$$

Each rate matrix specific to reservoir v follows local-detailed balance:

$$\ln \frac{K_{x_B}^{x_B'}(x_A; v)}{K_{x_B'}^{x_B}(x_A; v)} = -\frac{1}{T_v}(H(x_B|x_A) - H(x_B'|x_A)). \tag{19}$$

The previous condition allows to decompose each matrix $K_{x_B}^{x_B'}(x_A; v)$ as the product of a diagonal matrix encoding the equilibrium distribution corresponding to temperature T_v, and a symmetric matrix that contains the information about how the output system is coupled to that environment (and thus determines how the relaxation to the previous equilibrium takes place). Indeed one can write the nondiagonal elements $K_{x_B}^{x_B'}(x_A; v)$ as

$$K_{x_B}^{x_B'}(x_A; v) = \sum_{x_B''} \pi_{x_B}(x_A; v)\delta_{x_B}^{x_B''} S_{x_B''}^{x_B'}(x_A; v), \tag{20}$$

where $S_{x_B}^{x_B'}(x_A; v)$ are the elements of a symmetric matrix and $\pi_{x_B}(x_A; v) = e^{-H(x_B|x_A)/T_v}/Z$, with $Z = \sum_{x_B} e^{-H(x_B|x_A)/T_v}$, is the equilibrium output distribution corresponding to reservoir v. Note that although every reservoir-specific rate matrix is detailed balanced, the overall rate matrix is not detailed balanced in general unless there is only a single reservoir or all of them are at the same temperature. The bipartite nature of the overall system's evolution implies that the rate of entropy production $\dot{\sigma}$ can be split as [53,66]:

$$\dot{\sigma} = \dot{\sigma}_A + \dot{\sigma}_B, \tag{21}$$

whereas $\dot{\sigma}_A$ represents the cost of switching the input and $\dot{\sigma}_B$ represents the thermodynamic cost of the copying process. At any given time, the later can be computed as:

$$\dot{\sigma}_B = \sum_{\substack{v, x_A, \\ x_B', x_B}} K_{x_B}^{x_B'}(x_A; v)p_{x_A, x_B'}(t) \ln \frac{K_{x_B}^{x_B'}(x_A; v)p_{x_A, x_B'}(t)}{K_{x_B'}^{x_B}(x_A; v)p_{x_A, x_B}(t)} \tag{22}$$

A similar expression holds for $\dot{\sigma}_A$. Within this minimal model, we analyze two general methods that can modulate the output dynamics in real communication systems. In the "energy-switching" case, different values of the input modulate the energies of different output values, as in the simple case above. Intercellular chemical communication in biological organisms can be modeled in this way. For example in cellular sensing, the concentration of a ligand (input) modulates the free energy of receptor binding (output) [60,68]. Additionally, we can model wireless communication in terms of energy switching, where electromagnetic waves sent by the transmitter modify the potential energy function of electrons at the receiver. In the "reservoir-switching" case, different values of the input modulate the coupling of the output to different reservoirs. We can model wired electronic communication in circuits in this way, where the input voltage modulates the output wire's couplings with different chemical reservoirs of electrons [69].

D. Energy-switching scenario

In the energy-switching method of communication, the energy levels of different output states depend on the input state. We analyze the class of channels for which the output achieves its lowest energy state when it matches the current state of the input, and all other "mismatched states" have the same higher value of energy. Thus the Hamiltonian is also given by Eq. (7). As before, since the Hamiltonian is symmetric with respect to every letter in the input alphabet, the channel capacity-achieving input distribution π_{X_A} is the uniform distribution ($\pi_{x_A} = 1/L$). Additionally, we set all nonzero off-diagonal entries of the matrix $S(x_A; v)$ to a constant r_v, so that the couplings of the output to its different reservoirs cannot be modulated by the state of the input. Therefore, every nonzero off-diagonal rate matrix element equals

$$K_{x_B}^{x_B'}(x_A; v) = \begin{cases} r_v \frac{e^{\epsilon/T_v}}{L(e^{\epsilon/T_v}+(L-1))} =: \alpha_v & x_B' \neq x_A; x_B = x_A \\ r_v \frac{1}{L(e^{\epsilon/T_v}+(L-1))} =: \zeta_v & x_B \neq x_A \end{cases}. \tag{23}$$

In our simulations for the energy-switching case, the input dynamics follow a telegraph process [70], which is consistent with a CTMC. A telegraph process serves as an accurate model for any real communication channel (e.g., synaptic release [15]) for which the waiting times for the input in any given state follows an exponential distribution. The input switches its state with average rate f_s, which is a positive real number. This means that each state transition in the input is equally likely, so that $R_{x_A, x_B}^{x_A', x_B} = J_{x_A}^{x_A'}(x_B) = f_s$, for $x_A' \neq x_A$. Therefore, the nonzero off-diagonal elements of the overall channel's rate matrix are

$$R_{x_A, x_B}^{x_A', x_B'} = \begin{cases} f_s & x_A' \neq x_A; x_B' = x_B \\ \alpha := \sum_{v=1}^{N} \alpha_v & x_A' = x_A; x_B' \neq x_A; x_B = x_A \\ \zeta := \sum_{v=1}^{N} \zeta_v & x_A' = x_A; x_B \neq x_A \end{cases}. \tag{24}$$

We first analyze the situation where the channel is in a nonequilibrium steady-state (NESS), which means

$$\sum_{x_A', x_B'} R_{x_A, x_B}^{x_A', x_B'} \pi_{x_A', x_B'} = 0. \tag{25}$$

Solving for the joint steady-state distribution π_{X_A, X_B}, we obtain

$$\pi_{x_A, x_B} = \begin{cases} \frac{\alpha + f_s}{L[\alpha + (L-1)\zeta + Lf_s]} := \frac{p_m}{L} & x_A = x_B \\ \frac{\zeta + f_s}{L[\alpha + (L-1)\zeta + Lf_s]} := \frac{p_e}{L} & \text{otherwise} \end{cases}. \tag{26}$$

This means that the conditional steady-state distribution of the output state given the input state reads

$$\pi_{x_B|x_A} = \begin{cases} p_m & x_B = x_A \\ p_e & \text{otherwise} \end{cases} \tag{27}$$

and the marginal steady-state distribution over just the output states is uniform. Note that the previous conditional distribution differs from an equilibrium distribution even in the case there is a single thermal reservoir, except in the limit $f_s \to 0$. The reason is that the changing input plays the role of a work reservoir that parametrically changes the energy landscape of

(a) (b)
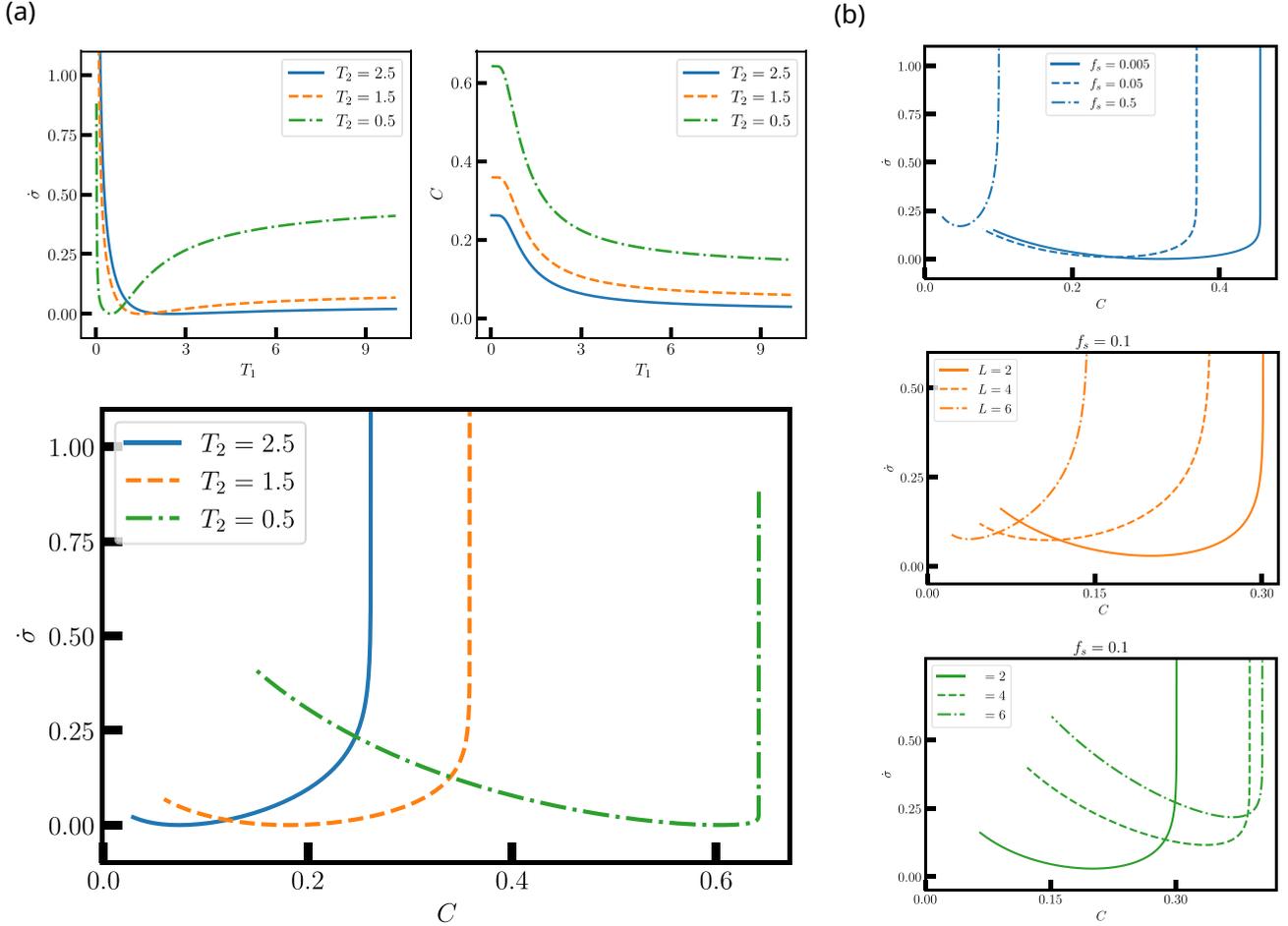
FIG. 4. Results for the case when the state energies of the communication channel's output vary with the input state. Here the output couples to two reservoirs with temperatures T_1 and T_2. We vary the noise in the channel by holding the temperature of one reservoir fixed and varying the temperature of the other reservoir. Black dots in each graph indicate $T_1 = T_2$. (a) These plots reflect the case of a binary alphabet ($L = 2$) and for energy bias $\epsilon = 2$. (a, top left) We first analyze the case of zero input signal rate: $f_s = 0$. We find that the EP rate $\langle \dot{\sigma} \rangle = \langle \dot{\sigma}_B \rangle$ as a function of T_1 has a single global root when $T_1 = T_2$. The derivative of the EP rate is negative (non-negative) to the left (right) of this root. (a, top right) The channel capacity C is a positive function of T_1 and its derivative is nonpositive. (a, bottom) Combining these two relationships, we find that the EP rate has at most one minimum with respect to the channel capacity. As a result, $\langle \dot{\sigma} \rangle(C)$ is a convex and nonmonotonic function (see proof in the Appendix). (b) Plots of the function $\langle \dot{\sigma}_B \rangle(C)$ for (top) different signaling rates $f_s > 0$, (middle) different alphabet lengths L, and (bottom) different energy biases ϵ. We observe in all plots that the EP rate retains a single global minimum (marked by black dots) with respect to the channel capacity. We considered $r_1 = r_2 = 1$ in all cases.

the output. In the opposite limit $f_s \to \infty$ the conditional distribution is just uniform, since the output cannot adapt to the rapidly changing input. Since the input distribution is chosen so that the system runs at its channel capacity, the value of the channel capacity in the NESS is simply given by the mutual information between the input and the output:

$$C = I(A : B) = S(B) - S(B|A), \quad (28)$$

$$= \ln L + p_m \ln p_m + (L - 1)p_e \ln p_e. \quad (29)$$

Figure 4 shows results of analyzing how the thermodynamic cost varies with the channel capacity when we adjust the amount of noise in the channel. We restrict attention to the case when the system is out of equilibrium due to the presence of two thermal reservoirs at different temperatures ($N = 2$). The coupling constants are taken to be $r_1 = r_2 = 1$. To adjust the noise level in the channel, we sweep the temperature of

one reservoir T_1 while holding fixed the temperature of the other at T_2 [71]. We find that $\langle \dot{\sigma}_B \rangle$ as a function of C is convex with a single global minimum that occurs at a positive value of the channel capacity. In particular, this means that the thermodynamic cost is not a monotonic function of the channel capacity. We prove this convex, nonmonotonic relationship in the limit of low signaling rate (see the Appendix).

E. Reservoir-switching scenario

The channel can alternatively modulate the coupling of the output system with its different reservoirs. For this reservoir-switching case, the energies $H(x_B|x_A) = H(x_B)$ of the output states do not depend on the input. So the equilibrium distributions $\pi_{x_B}(x_A; v)$ corresponding to each reservoir are independent of the input state. Then, for communication to be possible, the components of the symmetric matrix $S^{x_B'}_{x_B''}(x_A; v)$

in Eq. (20) must be input dependent. Importantly, this kind of communication protocol is used to transfer information between different components of modern electronic circuits. For example, consider a CMOS inverter [Fig. 1(b)], which is the electronic implementation of a NOT gate. The output voltage v of the inverter connects to two chemical reservoirs of electrons at fixed voltages V_1 and V_2 (with $V_1 > V_2$), through two complementary MOS transistors (a nMOS and a pMOS). Increasing v_{in} increases the conductivity of the nMOS transistor and decreases the conductivity of the pMOS transistor. So for high input voltage, the output's interaction with the reservoir at voltage V_2 dominates, making $v \simeq V_2$ at steady state. The situation reverses for low input voltages, resulting in $v \simeq V_1$. However, the energy associated to an output voltage v is always $E(v) = C_o v^2/2$ (C_o is the output capacitance), independently of the input. We consider the following minimal model of the reservoir-switching case. The input can fluctuate between $L = 2$ states. Transition $x_A : 0 \to 1$ has rate f_+ and the reverse transition has rate f_-. The output has a single ground state with zero energy, and M degenerate excited states with energy ϵ. Over that set of microscopic states, we define two macrostates $x_B = 0$ or $x_B = 1$ corresponding respectively to the ground state being occupied, or any of the excited states being occupied. The output is in contact with two thermal reservoirs at temperatures T_v ($v = 1, 2$) that can induce transitions between the ground and excited microstates but not between the excited microstates themselves. In this case, the coarse grained rates between macrostates satisfy the local detailed balance conditions [72]:

$$\ln \frac{\lambda_+(x_A; v)}{\lambda_-(x_A; v)} = -\frac{1}{T_v}(\epsilon - T_v \ln(M)), \qquad (30)$$

where $\lambda_+(x_A; v)$ is the rate of the transition $x_B : 0 \to 1$ induced by the reservoir v for input x_A, and $\lambda_-(x_A; v)$ is the rate of the reverse transition. The term $\ln(M)$ takes into account the internal entropy of the macrostate $x_B = 1$. The equilibrium distributions $\pi_{x_B}(x_A; v)$ are then:

$$\pi_{x_B}(x_A; v) = \begin{cases} \frac{M e^{-\epsilon/T_v}}{1 + M e^{-\epsilon/T_v}} & x_B = 1 \\ \frac{1}{1 + M e^{-\epsilon/T_v}} & x_B = 0 \end{cases}, \qquad (31)$$

that as noted before and in contrast to the energy-switching case, are actually independent of x_A. Let us consider now that $T_1 \ll \epsilon \ll T_2$ and that $\lambda_\pm(0; 1) \gg \lambda_\pm(0; 2)$ and $\lambda_\pm(1; 1) \ll \lambda_\pm(1; 2)$. That means that when the input is $x_A = 0$, the output system is more strongly coupled to the cold reservoir at temperature T_1, while for input $X_A = 1$ the output is more strongly coupled to the hot reservoir at T_2. As a consequence, if the frequency of input changes is sufficiently low, then for $x_A = 0$ the ground state $x_B = 0$ will be occupied with higher probability, and for $x_A = 1$ the system will be most probably excited. In this way, communication is possible without requiring energy modulation. The stationary input distribution always satisfies $\pi_{x_A=0}/\pi_{x_A=1} = f_-/f_+$. Note that in this case the channel capacity is not necessarily maximized by a uniform input distribution. Thus, for a fixed average input transition rate $\langle f \rangle \equiv f_+ f_-/(f_- + f_+)$, one needs to find the ratio f_-/f_+ that maximizes the channel capacity given all the other parameters. This is done numerically in what follows. Also, we parametrize the transition rates with the constants
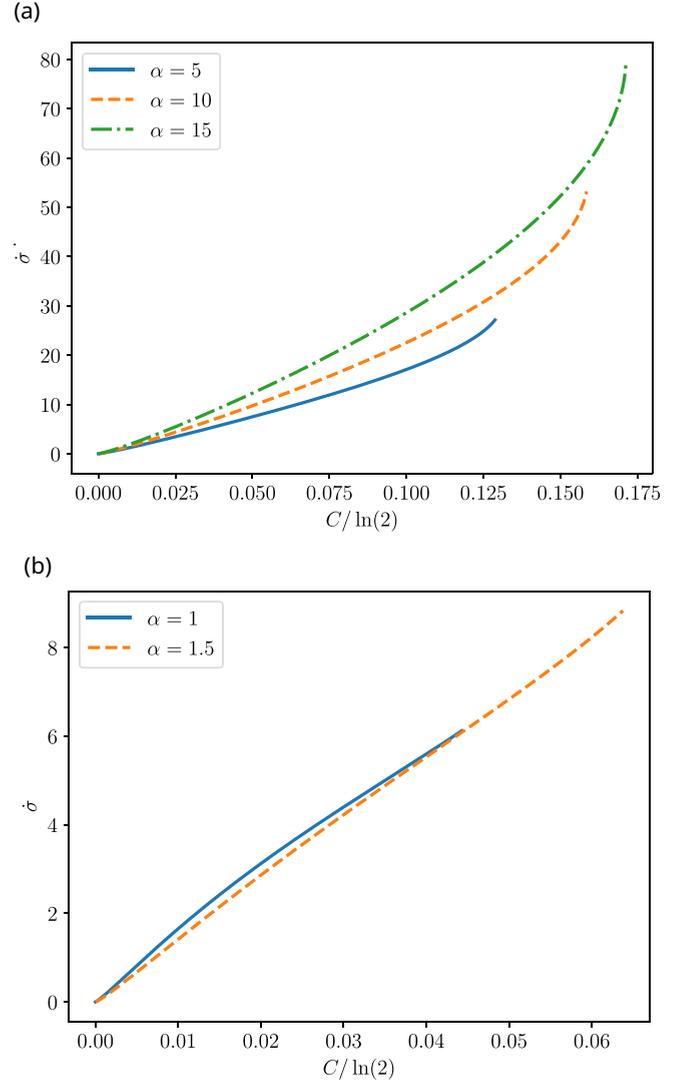
FIG. 5. Entropy production rate versus channel capacity for the minimal model of reservoir switching. The parameters are $T_1/\epsilon = 0.1$, $\beta(1) = \beta(0)^{-1} = 0.1$, and $M = 4$. The temperature T_2 was varied in the range $(10^{-1}, 10^2)$ to obtain each of the curves.

α, $\beta(0)$ and $\beta(1)$ by considering $\lambda_-(0; 1) = \lambda_-(1; 1) = \alpha \langle f \rangle$ and $\lambda_-(x_A; 2) = \beta(x_A)\lambda_-(x_A; 1)$. The rest of the rates are determined by Eq. (30). In Fig. 5 we show the entropy production rate versus the channel capacity for different values of *alpha*. All the other parameters are fixed as indicated in the caption with the exception of T_2, that was swept to obtain each curve. We see that as we increase α, which corresponds to making the output system faster compared to the changing input, it is possible to achieve higher channel capacities at the expense of a higher entropy production rate. For the parameters in Fig. 5(a) the entropy production rate is a convex and monotonic function of the channel capacity. However, the convexity is lost for lower values of α, as shown in Fig. 5(b).

IV. THERMODYNAMICS INFORMS WHEN AND HOW TO INVERSE MULTIPLEX

We have found that in most cases, for both types of communication analyzed above, the EP is a convex function of

the channel capacity. So by the theory of convex optimization, a thermodynamic benefit to inverse multiplexing must arise. In what follows, we analyze when one can reduce thermodynamic cost by splitting information streams across multiple channels. We additionally analyze how to distribute information transmission rates across multiple channels in order to reach the minimum achievable thermodynamic cost. In doing so, we derive the Pareto front that represents the set of "optimal" tuples $(C, \langle \dot{\sigma} \rangle)$ one can achieve for any given fixed number of channels.

A. M-channel capacity

Consider concurrently operating M channels with capacities $C_1, C_2, \ldots, C_M$; inputs $A_1, \ldots, A_M$; and outputs $B_1, \ldots, B_M$. We set the inputs simultaneously at the beginning of each channel use. We know that the mutual information between the inputs and outputs of M channels is upper bounded by a value that can only be achieved if the inputs are independent [31] (see the Appendix). So the capacity of this multi-channel setup equals the sum of the capacities of each channel.

$$C = \max_{p_{X_{A_1}, \ldots, X_{A_M}}} I(X_{A_1}, \ldots, X_{A_M}; X_{B_1}, \ldots, X_{B_M}), \quad (32)$$

$$= \sum_{i=1}^{M} \max_{p_{X_{A_1}, \ldots, X_{A_M}}} I(X_{A_i}; X_{B_i}), \quad (33)$$

$$= \sum_{i=1}^{M} \max_{p_{X_{A_i}}} I(X_{A_i}; X_{B_i}) = \sum_{i=1}^{M} C_i. \quad (34)$$

The only joint input distribution that achieves this maximum channel capacity is $\pi_{X_{A_1}, \ldots, X_{A_M}} = \prod_{i=1}^{M} \pi_{X_{A_i}}$. So by the channel-coding theorem, optimal codebooks make it appear as if the inputs were generated in each channel from its capacity-maximizing input distribution, independently from one another. Additionally, since the inputs to all of the channels are statistically independent, there is no unavoidable mismatch cost, as arises for example in the parallel bit erasure of statistically coupled bits [73].

B. M-channel EP rate and thermodynamic benefits of inverse multiplexing

Since we model the multi-channel setup as if it is running at its capacity, each of the constituent channels is an independent CTMC, governed by its own master equation. Therefore, the total EP rate of the combined system is simply the sum of the EP rates of each channel [66]:

$$\dot{\sigma} = \sum_{i=1}^{M} \dot{\sigma}_i = \sum_{i=1}^{M} g_i(C_i), \quad (35)$$

where $g_i(C_i)$ expresses channel i's EP rate $\dot{\sigma}_i$ as a convex function g_i of its channel capacity, C_i. We can use the water-filling algorithm to minimize the total EP rate of a set of M channels subject to a desired total channel capacity $C_d = \sum_{i=1}^{M} C_i$, via

the Lagrangian [74]:

$$\mathcal{L} = \sum_{i=1}^{M} g(C_i) + \lambda \left(C_d - \sum_{i=1}^{M} C_i \right). \quad (36)$$

Differentiating with respect to C_j,

$$\frac{\partial \mathcal{L}}{\partial C_j} = g_i'(C_j) - \lambda = 0, \quad (37)$$

which means

$$C_j = (g_i')^{-1}(\lambda), \quad (38)$$

where λ is chosen to satisfy

$$\sum_{i=1}^{M} C_i = \sum_{i=1}^{M} (g_i')^{-1}(\lambda) = C_d. \quad (39)$$

In particular, this suggests that if $g_i = g \, \forall \, i$, then it is optimal to use M channels with identical capacity $\frac{C_d}{M}$. This finding leads to important insights regarding how to choose between different inverse multiplexing setups. For example, a setup of two channels with capacities C_A and C_B is identical in an information-theoretic sense to a setup with two identical channels each with capacity $\frac{1}{2}(C_A + C_B)$. However, from a thermodynamic viewpoint, it would be better to use the two identical channels because that would minimize EP rate due to the convexity of the EP rate with respect to the channel capacity. If we know the functional form of $g(C)$, then we can also use the method outlined in Ref. [3] to calculate the energetically optimal number of independent, identical channels to use in order to achieve a desired total information rate. More generally, for any set of functions $\{g_i\}$, we can find the λ that satisfies Eq. (39) (in addition to the positivity of the channel capacity), and plug it into the Lagrangian. In this way we can obtain the optimal distribution $\{C_i\}$ of information transmission rates across M channels that are not necessarily identical.

C. Pareto-optimal fronts for inverse multiplexing

These results suggest that for any number M of channels, there exists a Pareto front of points $(C, \dot{\sigma})$ that each minimizes $\dot{\sigma} = \sum_{i=1}^{M} \dot{\sigma}_i$ and simultaneously maximizes $C = \sum_{i=1}^{M} C_i$. We plot examples of these Pareto fronts in Fig. 6. The plots reveal that higher channel capacities benefit from splitting information rates across more channels. More specifically, there exist thresholds, $C^{(m)}$, above which it reduces the total EP to split information transmission across $m + 1$ channels instead of m channels.

V. DISCUSSION

A convex relationship between thermodynamic cost and channel capacity, as well as the energetic benefits of inverse multiplexing that follow, has previously been demonstrated in communication channel models with Gaussian noise [31]. However, for those models, the "power" does not correspond to actual dissipated energy per unit time. This relationship has also been justified semiformally for biological systems [3,4,62]. Due to their assumptions, both of these kinds of
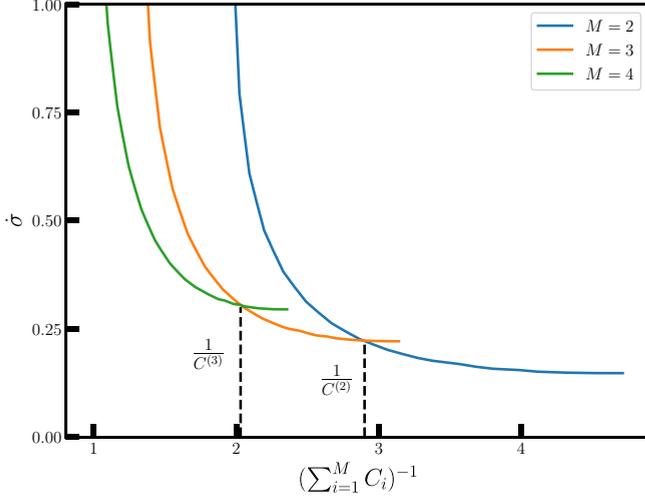
FIG. 6. For a collection of M channels implementing energy switching with capacities $\{C_1, \ldots, C_M\}$, these Pareto fronts are collections of points $(C, \langle \dot{\sigma} \rangle)$ that each minimizes $\langle \dot{\sigma} \rangle = \sum_{i=1}^{M} \langle \dot{\sigma}_i \rangle$ and simultaneously maximizes $C = \sum_{i=1}^{M} C_i$. All other possible (but nonoptimal) combinations $\{C_1, \ldots, C_M\}$ lie to the top right of each front. Above each threshold $C^{(m)}$ splitting information transmission across $m + 1$ channels instead of m channels reduces the total EP. In these plots, $f_s = 0.1$, $L = 4$, and $\epsilon = 2$.

studies derive that thermodynamic cost increases monotonically with channel capacity. To the best of our knowledge, ours is the first study to use physics to demonstrate that, in many cases, the thermodynamic cost is actually not necessarily a monotonic nor convex function of the channel capacity. However, we find that those two properties are recovered for sufficiently large channel capacity, so the thermodynamic benefits of splitting an information stream among multiple channels still arises. This paper also presents the first analysis of inverse multiplexing treated with the rigor provided by stochastic thermodynamics. Our results help illuminate observations in biological communication, and may provide heuristics for an engineer designing communication systems. There are several stochastic thermodynamics formulations one can use to continue to investigate the thermodynamic costs of communication. In the main text we use the one in which the dynamics of the system of interest (SOI) is Markovian, due to its coupling with infinite reservoirs that are at thermal equilibrium. Another formulation, called the "inclusive Hamiltonian," or "strong-coupling" model, treats the case where the reservoirs are finite. There, the joint system comprising the SOI and the reservoirs evolves with unitary Hamiltonian dynamics [75–78]. So the dynamics of the SOI is non-Markovian in general. There has been a very preliminary analysis of thermodynamically motivated rate distortion functions for communication channels using this formulation [79]. In the main text we have also assumed that the output has "no back-action" onto the input, so they each obey LDB separately. This nonreciprocity approximates most real-world observable dynamics well [64]. However, strictly speaking, many systems violate this approximation when modeled at finer scales due to microreversibility and the consequent property that the *entire system* obeys LDB [80]. For the purposes

of these investigations, we have not considered the thermodynamic cost of the input encoding function and the output decoding function. However since different coding strategies ("codebooks") would likely incur different thermodynamic costs, we suggest that future work should investigate their stochastic thermodynamics. Such analyses would naturally extend to investigations of the thermodynamic costs of different error-correcting codes.

APPENDIX: MATERIALS AND METHODS

1. Proof of the form of EP rate versus channel capacity in the energy-switching case in the nonequilibrium steady state

This proof is for the case when $N = 2$, $L = 2$, and $f_s = 0$. Zero signaling rate corresponds to communication protocols in which the input is set once, at time $t = 0$, by sampling from π_{X_A} remains fixed for the rest of the trajectory, during which only the output can change state. While holding T_2 fixed and varying T_1, we seek to derive the properties of

$$\frac{d\langle \dot{\sigma} \rangle}{dC} = \frac{d\langle \dot{\sigma} \rangle / dT_1|_{T_2}}{dC/dT_1|_{T_2}}. \tag{A1}$$

First, we note that C is always non-negative due to the non-negativity of the mutual information. Its derivative with respect to the temperature of one reservoir

$$\frac{\partial C}{\partial T_1} = \frac{\partial p_m}{\partial T_1}(1 + \ln p_m) + \frac{\partial p_e}{\partial T_1}(1 + \ln p_e), \tag{A2}$$

$$= \frac{\partial p_m}{\partial T_1} \ln \frac{p_m}{p_e}, \tag{A3}$$

$$= \frac{1}{r_1 + r_2} \frac{\partial \alpha_1}{\partial T_1} \ln \frac{\alpha}{\zeta}, \tag{A4}$$

$$= -\frac{\epsilon \alpha_1 \zeta_1}{T_1^2 r_1} \ln \frac{\alpha}{\zeta}, \tag{A5}$$

$$\leqslant 0 \,\forall\, T_1 \in (0, \infty), \tag{A6}$$

is always nonpositive. So the channel capacity is a nonincreasing function of the temperature of (any) reservoir. We then find that the EP rate can be factorized (assuming the r_v are fixed, positive reals):

$$\langle \dot{\sigma} \rangle = \left(\frac{1}{r_1} + \frac{1}{r_2} \right)^{-1} \left(e^{\frac{\epsilon}{T_1}} - e^{\frac{\epsilon}{T_2}} \right) \left(\frac{\epsilon}{T_1} - \frac{\epsilon}{T_2} \right)$$
$$\times \ln \left(\frac{r_1 e^{\frac{\epsilon}{T_1}} + r_2 e^{\frac{\epsilon}{T_2}}}{r_1 + r_2} \right). \tag{A7}$$

Since $(e^{\frac{\epsilon}{T_1}} - e^{\frac{\epsilon}{T_2}})(\frac{\epsilon}{T_1} - \frac{\epsilon}{T_2}) \geqslant 0$ and the argument of the logarithm is a function that is always $\geqslant 1$, the EP rate is always non-negative. (This is a fact of stochastic thermodynamics as well—that the ensemble-average EP rate is non-negative.) Furthermore, it is clear that the EP rate has a singular root that is achieved when $T_1 = T_2$, which gives an EP rate of zero.

This singular root must be a minimum since the EP rate is everywhere non-negative. The EP rate has no other roots, so by Rolle's theorem,

$$\frac{d\langle\dot{\sigma}\rangle}{dT_1} \begin{cases} < 0 & T_1 < T_2 \\ > 0 & T_1 > T_2 \end{cases}. \quad (A8)$$

Therefore,

$$\frac{d\langle\dot{\sigma}\rangle}{dC} \begin{cases} > 0 & T_1 < T_2 \\ < 0 & T_1 > T_2 \end{cases}, \quad (A9)$$

which is exactly what we observe in Fig. 4. We also note the limits

$$\lim_{T_1\to 0} C = \ln 2 - S\left\{\eta_1 + \eta_2\left(\frac{e^{\frac{\epsilon}{T_2}}}{e^{\frac{\epsilon}{T_2}}+1}\right), \eta_2\left(\frac{1}{e^{\frac{\epsilon}{T_2}}+1}\right)\right\}, \quad (A10)$$

$$:= C_{\max}(T_2) \geqslant 0, \quad (A11)$$

$$\lim_{T_1\to\infty} C = \ln 2 - S\left\{\frac{\eta_1}{2} + \eta_2\left(\frac{e^{\frac{\epsilon}{T_2}}}{e^{\frac{\epsilon}{T_2}}+1}\right), \frac{\eta_1}{2} + \eta_2\left(\frac{1}{e^{\frac{\epsilon}{T_2}}+1}\right)\right\}, \quad (A12)$$

$$:= C_{\min}(T_2) \geqslant 0, \quad (A13)$$

where $\eta_1 = \frac{r_1}{r_1+r_2}$ and $\eta_2 = \frac{r_2}{r_1+r_2}$. So, the channel capacity has minimum and maximum values. We also find that, at the minimum, the channel capacity equals

$$C(T_1 = T_2 = T) = \ln 2 - S\left(\left\{\frac{e^{\frac{\epsilon}{T}}}{e^{\frac{\epsilon}{T}}+1}, \frac{1}{e^{\frac{\epsilon}{T}}+1}\right\}\right), \quad (A14)$$

which is strictly positive unless $T = \infty$. Furthermore, we write the limits on the entropy production rate

$$\lim_{T_1\to 0}\langle\dot{\sigma}\rangle = \infty, \quad (A15)$$

$$\lim_{T_1\to\infty}\langle\dot{\sigma}\rangle = \left(\frac{1}{r_1} + \frac{1}{r_2}\right)^{-1}\frac{\epsilon}{T_2}\left(e^{\frac{\epsilon}{T_2}} - 1\right)\ln\left(\frac{r_1 + r_2 e^{\frac{\epsilon}{T_2}}}{r_1 + r_2}\right), \quad (A16)$$

$$:= \langle\dot{\sigma}\rangle^{\dagger}(T_2) \geqslant 0. \quad (A17)$$

Putting Eq. (A9) together with these limit analyses results in the fact that the EP rate as a function of the channel capacity is convex with a single global minimum that occurs at a positive value of the channel capacity. (The channel capacity is zero when $T_1 = T_2 = \infty$.) Additionally, the EP rate diverges as the channel capacity approaches its maximum value $C_{\max}(T_2)$, as depicted in Fig. 4. So for this energy-switching case, the EP is a convex, nonmonotonic function of the channel capacity.

2. Mutual information is maximized when channels are independent

For any number $M \geqslant 2$ of channels [31],

$$I(X_{A_1}, \ldots, X_{A_M}; X_{B_1}, \ldots, X_{B_M}), \quad (A18)$$

$$= H(X_{B_1}, \ldots, X_{B_M}) - H(X_{B_1}, \ldots, X_{B_M}|X_{A_1}, \ldots, X_{A_M}), \quad (A19)$$

$$= H(X_{B_1}, \ldots, X_{B_M}) - \sum_{i=1}^{M} H(X_{B_i}|X_{A_1}, \ldots, X_{A_M}), \quad (A20)$$

$$= H(X_{B_1}, \ldots, X_{B_M}) - \sum_{i=1}^{M} H(X_{B_i}|X_{A_i}), \quad (A21)$$

$$\leqslant \sum_{i=1}^{M} H(X_{B_i}) - \sum_{i=1}^{M} H(X_{B_i}|X_{A_i}), \quad (A22)$$

$$= \sum_{i=1}^{M} I(X_{A_i}; X_{B_i}). \quad (A23)$$

Equation (A22) achieves equality when the M channels are all independent.

[1] C. C. Govern and P. R. Ten Wolde, Optimal resource allocation in cellular sensing systems, Proc. Natl. Acad. Sci. USA **111**, 17486 (2014).

[2] P. R. ten Wolde, N. B. Becker, T. E. Ouldridge, and A. Mugler, Fundamental limits to cellular sensing, J. Stat. Phys. **162**, 1395 (2016).

[3] V. Balasubramanian, Heterogeneity and efficiency in the brain, Proc. IEEE **103**, 1346 (2015).

[4] V. Balasubramanian, Brain power, Proc. Natl. Acad. Sci. USA **118**, e2107022118 (2021).

[5] T. Guilford and M. S. Dawkins, Receiver psychology and the evolution of animal signals, Anim. Behav. **42**, 1 (1991).

[6] J. M. Smith, D. Harper *et al.*, *Animal Signals* (Oxford University Press, Oxford, 2003).

[7] M. Broxton, J. Lifton, and J. Paradiso, Localizing a sensor network via collaborative processing of global stimuli, in *Proceeedings of the 2nd European Workshop on Wireless Sensor Networks, 2005* (IEEE, Los Alamitos, CA, 2005), pp. 321–332.

[8] J. A. Paradiso and T. Starner, Energy scavenging for mobile and wireless electronics, IEEE Perv. Comput. **4**, 18 (2005).

[9] Y. Wu, J. A. Stankovic, T. He, and S. Lin, Realistic and efficient multi-channel communications in wireless sensor networks, in *Proceedings of the 27th Conference on Computer Communications (INFOCOM'08)* (IEEE, Los Alamitos, CA, 2008), pp. 1193–1201.

[10] Y. Li, J. Park, M. Alian, Y. Yuan, Z. Qu, P. Pan, R. Wang, A. Schwing, H. Esmaeilzadeh, and N. S. Kim, A network-centric hardware/algorithm co-design to accelerate distributed training of deep neural networks, in *Proceedings of the 51st Annual IEEE/ACM International Symposium on Microarchitecture (MICRO'18)* (IEEE, Los Alamitos, CA, 2018), pp. 175–188.

[11] P. Christie and D. Stroobandt, The interpretation and application of rent's rule, IEEE Trans VLSI Syst. **8**, 639 (2000).

[12] M. Moses, G. Bezerra, B. Edwards, J. Brown, and S. Forrest, Energy and time determine scaling in biological and computer designs, Philos. Trans. R. Soc. B **371**, 20150446 (2016).

[13] Y. Demir, Y. Pan, S. Song, N. Hardavellas, J. Kim, and G. Memik, Galaxy: A high-performance energy-efficient multichip architecture using photonic interconnects, in *Proceedings of the 28th ACM International Conference on Supercomputing* (ACM, New York, 2014), pp. 303–312.

[14] M. Prezioso, F. Merrikh-Bayat, B. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov, Training and operation of an integrated neuromorphic network based on metal-oxide memristors, Nature (Lond.) **521**, 61 (2015).

[15] D. Marković, A. Mizrahi, D. Querlioz, and J. Grollier, Physics for neuromorphic computing, Nat. Rev. Phys. **2**, 499 (2020).

[16] A. E. Lombo, J. E. Lares, M. Castellani, C.-N. Chou, N. Lynch, and K. K. Berggren, A superconducting nanowire-based architecture forneuromorphic computing, Neuromorph. Comput. Eng. **2**, 034011 (2022).

[17] National Academies of Sciences, Engineering, *Physics of Life* (National Academies Press, 2022).

[18] S. J. Bryant and B. B. Machta, Physical constraints in intracellular signaling: The cost of sending a bit, Phys. Rev. Lett. **131**, 068401 (2023).

[19] T. A. Jones, A. Y. Klintsova, V. L. Kilman, A. M. Sirevaag, and W. T. Greenough, Induction of multiple synapses by experience in the visual cortex of adult rats, Neurobiol. Learn. Memory **68**, 13 (1997).

[20] I. Meinertzhagen and K. Sorra, Synaptic organization in the fly's optic lamina: Few cells, many synapses and divergent microcircuits, Prog. Brain Res. **131**, 53 (2001).

[21] K. Sorra and K. M. Harris, Occurrence and three-dimensional structure of multiple synapses between individual radiatum axons and their target pyramidal cells in hippocampal area Ca1, J. Neurosci. **13**, 3736 (1993).

[22] N. Toni, P.-A. Buchs, I. Nikonenko, C. Bron, and D. Muller, LTP promotes formation of multiple spine synapses between a single axon terminal and a dendrite, Nature (Lond.) **402**, 421 (1999).

[23] A. Fornito, A. Zalesky, and M. Breakspear, The connectomics of brain disorders, Nat. Rev. Neurosci. **16**, 159 (2015).

[24] D. S. Bassett and E. Bullmore, Small-world brain networks, The Neuroscientist **12**, 512 (2006).

[25] D. S. Bassett and E. T. Bullmore, Small-world brain networks revisited, The Neuroscientist **23**, 499 (2017).

[26] S. Gao, X. Cheng, and L. Yang, Spatial multiplexing with limited RF chains: Generalized beamspace modulation (GBM) for mmwave massive MIMO, IEEE J. Sel. Areas Commun. **37**, 2029 (2019).

[27] S. Chen, S. Sun, G. Xu, X. Su, and Y. Cai, Beam-space multiplexing: Practice, theory, and trends, from 4G TD-LTE, 5G, to 6G and beyond, IEEE Wireless Commun. **27**, 162 (2020).

[28] O. Shental and I. Kanter, The second law of thermodynamics for communication channels, in *Proceedings of the IEEE 25th Convention of Electrical and Electronics Engineers in Israel* (IEEE, Los Alamitos, CA, 2008), pp. 784–788.

[29] Y. Zhou, D. Segal *et al.*, Minimal model of a heat engine: Information theory approach, Phys. Rev. E **82**, 011120 (2010).

[30] C. P. Broedersz and P. Ronceray, Twenty-five years of nanoscale thermodynamics, Nature **604**, 46 (2022).

[31] T. M. Cover, *Elements of Information Theory* (John Wiley & Sons, New York, 1999).

[32] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[33] A. Feinstein, *A New Basic Theorem of Information Theory*, Ph.D. thesis, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, 1954.

[34] Although there are countably infinitely many such encodings, it is very difficult for a human to find one [81].

[35] L. Slot and S. Zur, Shannon's Noisy-Channel Coding Theorem, Tech. Rep., Universiteit van Amsterdam, 2015.

[36] C. Berrou and A. Glavieux, Near optimum error correcting coding and decoding: Turbo-codes, IEEE Trans. Commun. **44**, 1261 (1996).

[37] S.-Y. Chung, G. D. Forney, T. J. Richardson, and R. Urbanke, On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit, IEEE Commun. Lett. **5**, 58 (2001).

[38] T. M. Duman and A. Ghrayeb, *Coding for MIMO Communication Systems* (John Wiley & Sons, New York, 2008).

[39] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding* (Courier Corporation, North Chelmsford, MA, 2013).

[40] G. C. Clark, Jr. and J. B. Cain, *Error-correction Coding for Digital Communications* (Springer Science & Business Media, New York, 2013).

[41] K. D. Rao, *Channel Coding Techniques for Wireless Communications* (Springer, Berlin, 2015).

[42] M. V. Patil, S. Pawar, and Z. Saquib, Coding techniques for 5G networks: A review, in *Proceedings of the 3rd International Conference on Communication System, Computing and IT Applications (CSCITA'20)* (IEEE, Los Alamitos, CA, 2020), pp. 208–213.

[43] L. Brillouin, Maxwell's Demon cannot operate: Information and entropy. I, J. Appl. Phys. **22**, 334 (1951).

[44] R. Landauer, Irreversibility and heat generation in the computing process, IBM J. Res. Dev. **5**, 183 (1961).

[45] M. Esposito and C. Van den Broeck, Second law and Landauer principle far from equilibrium, Europhys. Lett. **95**, 40004 (2011).

[46] In particular, stochastic thermodynamics describes any nonequilibrium system whose state transitions are mediated by energy exchanges with infinite reservoirs.

[47] C. Van den Broeck and M. Esposito, Ensemble and trajectory thermodynamics: A brief introduction, Physica A **418**, 6 (2015).

[48] Moreoever, as the number of degrees of freedom in the model of the system increases, the probability that the entropy production is negative decreases [82]. This explains the observation that the inviolability of the Second law at the macroscopic scale is matched by a complete absence of the second law at the microscopic scale [83,84].

[49] A. C. Barato and U. Seifert, Thermodynamic uncertainty relation for biomolecular processes, Phys. Rev. Lett. **114**, 158101 (2015).

[50] J. M. Horowitz and T. R. Gingrich, Proof of the finite-time thermodynamic uncertainty relation for steady-state currents, Phys. Rev. E **96**, 020103(R) (2017).

[51] N. Shiraishi, K. Funo, and K. Saito, Speed limit for classical stochastic processes, Phys. Rev. Lett. **121**, 070601 (2018).

[52] J. M. Parrondo, J. M. Horowitz, and T. Sagawa, Thermodynamics of information, Nat. Phys. **11**, 131 (2015).

[53] J. M. Horowitz and M. Esposito, Thermodynamics with continuous information flow, Phys. Rev. X **4**, 031015 (2014).

[54] J. M. Horowitz, Multipartite information flow for multiple maxwell demons, J. Stat. Mech. (2015) P03006.

[55] T. Sagawa and M. Ueda, Nonequilibrium thermodynamics of feedback control, Phys. Rev. E **85**, 021104 (2012).

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

[57] S. Ito and T. Sagawa, Information thermodynamics on causal networks, Phys. Rev. Lett. **111**, 180603 (2013).

[58] J. M. Poulton, P. R. Ten Wolde, and T. E. Ouldridge, Nonequilibrium correlations in minimal dynamical models of polymer copying, Proc. Natl. Acad. Sci. USA **116**, 1946 (2019).

[59] S. Ito and T. Sagawa, Maxwell's demon in biochemical signal transduction with feedback loop, Nat. Commun. **6**, 7498 (2015).

[60] D. Hartich, A. C. Barato, and U. Seifert, Sensory capacity: An information theoretical measure of the performance of a sensor, Phys. Rev. E **93**, 022116 (2016).

[61] C. Y. Gao and D. T. Limmer, Principles of low dissipation computing from a stochastic circuit model, Phys. Rev. Res. **3**, 033169 (2021).

[62] V. Balasubramanian, D. Kimber, and M. J. Berry II, Metabolically efficient information processing, Neural Comput. **13**, 799 (2001).

[63] T. Sagawa and M. Ueda, Minimal energy cost for thermodynamic information processing: Measurement and information erasure, Phys. Rev. Lett. **102**, 250602 (2009).

[64] G. Verley, C. Van den Broeck, and M. Esposito, Work statistics in stochastically driven systems, New J. Phys. **16**, 095001 (2014).

[65] S. J. Bryant and B. B. Machta, Energy dissipation bounds for autonomous thermodynamic cycles, Proc. Natl. Acad. Sci. USA **117**, 3478 (2020).

[66] D. H. Wolpert, Minimal entropy production rate of interacting systems, New J. Phys. **22**, 113013 (2020).

[67] A bipartite process is a special case of a multipartite process (MPP) [53,66] involving only two subsystems.

[68] The fraction of bound receptors on the cellular membrane reflects the external ligand concentration.

[69] N. Freitas, J.-C. Delvenne, and M. Esposito, Stochastic thermodynamics of nonlinear electronic circuits: A realistic framework for computing around kT, Phys. Rev. X **11**, 031064 (2021).

[70] M. Meijers, S. Ito, and P. R. ten Wolde, Behavior of information flow near criticality, Phys. Rev. E **103**, L010102 (2021).

[71] The code used for the simulations can be found at https://github.com/FaritaTasnim/comm/blob/main/R_parametrization_sweeps.ipynb.

[72] M. Esposito, Stochastic thermodynamics under coarse graining, Phys. Rev. E **85**, 041125 (2012).

[73] T. E. Ouldridge and D. H. Wolpert, Thermodynamics of deterministic finite automata operating locally and periodically, New J. Phys. **25**, 123013 (2023).

[74] This is a convex optimization problem that loosely speaking is the inverse of the one discussed in Section 9.4 of Ref. [31], which maximizes total channel capacity subject to a "power constraint."

[75] M. Esposito, K. Lindenberg, and C. Van den Broeck, Entropy production as correlation between system and reservoir, New J. Phys. **12**, 013013 (2010).

[76] C. Jarzynski, Hamiltonian derivation of a detailed fluctuation theorem, J. Stat. Phys. **98**, 77 (2000).

[77] K. Ptaszyński and M. Esposito, Entropy production in open systems: The predominant role of intraenvironment correlations, Phys. Rev. Lett. **123**, 200603 (2019).

[78] A. M. Timpanaro, G. Guarnieri, J. Goold, and G. T. Landi, Thermodynamic uncertainty relations from exchange fluctuation theorems, Phys. Rev. Lett. **123**, 090604 (2019).

[79] G. Kardeş and D. Wolpert, Inclusive thermodynamics of computational machines, arXiv:2206.01165.

[80] D. H. Wolpert, Combining lower bounds on entropy production in complex systems with multiple interacting components, in *Frontiers in Entropy across the Disciplines: Panorama of Entropy: Theory, Computation, and Applications* (World Scientific, Singapore, 2023), pp. 405–453.

[81] J. T. Coffey and R. M. Goodman, Any code of which we cannot think is good, IEEE Trans. Inf. Theory **36**, 1453 (1990).

[82] R. Rao and M. Esposito, Detailed fluctuation theorems: A unifying perspective, Entropy **20**, 635 (2018).

[83] 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 (London) **437**, 231 (2005).

[84] S. Joubaud, N. Garnier, and S. Ciliberto, Fluctuations of the total entropy production in stochastic systems, Europhys. Lett. **82**, 30007 (2008).