Thermodynamic speed limits for multiple, co-evolving systems

Farita Tasnim[*]
Massachusetts Institute of Technology, Media Lab, Cambridge, MA, USA

David H. Wolpert[†]
Santa Fe Institute, Santa Fe, NM, USA
Complexity Science Hub, Vienna, Austria and
Arizona State University, Tempe, AZ, USA
(Dated: April 6, 2024)

The thermodynamic speed limit theorems (SLTs) provide lower bounds on the time required for a system to evolve between any two given distributions, in terms of the system's total entropy production along the path with which its distribution evolves. Previously derived versions of the SLTs apply to a single physical system without regard to its internal structure. However, many systems of interest are multipartite, comprising a set of co-evolving subsystems. Here we derive three strengthened versions of the SLT that reflect the multipartite nature of such systems. The first is guaranteed to be at least as strong as the conventional "global" SLT. While the other two do not always have this guarantee, in many cases they are stronger than the first of our new SLTs. We demonstrate our results with a numerical example involving a cell sensing its environment.

Introduction— The classical speed limit theorems (SLTs) provide the minimum time required for a system evolving according to a continuous-time Markov chain (CTMC) to evolve between two probability distributions over its states. These bounds are growing functions of the total variation distance between the two distributions, and shrinking functions of the total entropy production (EP) along the path of system evolution [1–5]. Intuitively, the SLTs tell us that if we modify a CTMC to have it change a distribution by a greater amount and / or in less time, then we must "pay for" that increased speed by increasing (a lower bound on) the total EP.

Like most of the other theorems in (classical) stochastic thermodynamics, the previously derived versions of the SLTs hold regardless of the size of a system, so long as its dynamics can be modeled as a CTMC. Indeed, the previously derived SLTs ignore all aspects of how the system might decompose into a set of co-evolving subsystems. Such "global SLTs" are most appropriate for nanoscale systems.

However, in information-processing systems above the nanoscale, inevitably there exist many constraints on the possible CTMC and associated reservoirs, often due to their modular nature. As a practical matter, the minimal EP of any system that obeys those constraints while also implementing a given conditional distribution $P(x_f|x_i)$ to an initial distribution $P(x_i)$ is often the dominant thermodynamic cost. That lower bound on the EP often far exceeds the associated "generalized Landauer bound" on the thermodynamic cost of applying $P(x_f|x_i)$ to $P(x_i)$, which is given by the change in Shannon entropy between the initial and final distributions [6]. Since the SLTs can be formulated as lower bounds on EP, the fact that constraints on the allowed

CTMC contribute to the minimal EP suggests that accounting for such constraints could lead to tighter SLTs. In order to investigate this possibility, we must analyze how the constraints on the allowed dynamics in a system affect its stochastic thermodynamics.

One major class of dynamical constraints arises because many systems are most naturally modeled as a set of multiple co-evolving subsystems [7–14], where the interactions between subsystems impose restrictions on how the system can evolve. Here, we focus on systems for which this broadly-applicable type of constraint applies. Such a system evolves as a multipartite process (MPP) [13, 14]. An MPP can have any arbitrary collection of dependency constraints, i.e., restrictions on how the dynamics of each subsystem depends on the states of the other subsystems [15].

Formally, an MPP is any system with the property that only one subsystem can change state at a time. MPPs are extremely common [8, 9], including biological sensing, information engines, and Ising spin models that can be modeled with Glauber dynamics. Other examples include eukaryotic cells, which are naturally modelled as a set of multiple interacting organelles and biomolecular species [16]. The subcellular process of protein synthesis could also be modelled as an MPP, with interacting subsystems as ribosomal subunits, mRNA, and sets of tRNA-activated amino acids [17].

Roadmap— In this paper, we show that accounting for the multipartite nature of a system's dynamics results in strengthened forms of the SLTs. We begin by reviewing the relevant stochastic thermodynamics of MPPs, with emphasis on subsystem-indexed contributions to quantities such as the entropy production. In the next section, we present our main results, a set of three SLTs that explicitly account for the MPP nature of composite systems. The first of these new SLTs is guaranteed to be at least as strong as the global SLT, regardless of the details of the rate-matrix dependencies of the

subsystems in the MPP. The other two SLTs do not always have this guarantee; however, in many cases, they are stronger than the first of our new SLTs. In order to illustrate our results, we conduct numerical calculations for the example of a cell sensing its environment. We discuss this example and present the resultant speed limit bounds in the following section. We end the main text with a discussion of future work.

We derive our new SLTs in full detail in Appendices A and B. In addition, we discuss the thermodynamic extensibility of our results to systems whose dynamics are defined by local rather than global Hamiltonians, along with definitions of the unique types of local detailed balance possible in MPPs, in Appendix C.

We then present some auxiliary results in the remaining appendices. In Appendix D, we note that the global SLT also applies to subsets of subsystems, or units, that evolve according to their own CTMC; this leads to what we call "unit-local SLTs". In Appendix E, we derive that the strongest of a generalized form of the SLT that involves a linear combination of unit-local thermodynamic properties is given by the unit-local SLT for the system's slowest -evolving unit. In Appendix F, we consider the SLTs as bounds on the space of distributions that a system can access within a given time. In particular, we show that the unit-local SLTs together more tightly restrict (than the global SLT) this space of final joint distributions for a pair of independently evolving spins. In Appendix G, we derive maximum and minimum speeds of evolution for Bayes' Nets, which are a type of MPP for which the state transitions occur according to a global clock. Finally, in Appendix H, we derive lower bounds on the difference between the EP rate and the rate at which one unit "learns" about another using counterfactual thermodynamic quantities.

Stochastic Thermodynamics of Multipartite Processes—
We consider a composite system $\mathcal{N}$ as a particular set of N subsystems with finite state spaces $\{X_i : i = 1, \ldots, N\}$. x indicates a state vector in X, which is the joint state space represented by the Cartesian product $\bigtimes_{i \in \mathcal{N}} X_i$. For any set of subsystems $A \subset \mathcal{N}$, we write $-A := \mathcal{N} \backslash A$. x_A indicates a state vector in $X_A = \bigtimes_{i \in A} X_i$. There is a set of time-varying stochastic rate matrices governing the evolution of the joint system: $\{K_x^{x'}(i;t) : i = 1, \ldots, N\}$, each a $|X| \times |X|$ matrix. The main condition for a multipartite process is that $K_x^{x'}(i;t) = 0 \; \forall i, t, \{x', x \,|\, x'_{-i} \neq x_{-i}\}$, meaning that at any given time, only one subsystem can undergo a state transition. For any $B \subseteq \mathcal{N}$, we define $K_x^{x'}(B;t) = \sum_{i \in B} K_x^{x'}(i;t)$. The joint dynamics over X is governed by the master equation:

$$\frac{dp_x(t)}{dt} = \sum_{x'} K_x^{x'}(t) p_{x'}(t) = \sum_{x'} \sum_{i \in \mathcal{N}} K_x^{x'}(i;t) p_{x'}(t). \quad (1)$$

where the rate matrix governing state transitions in subsystem i is

$$K_x^{x'}(i;t) := K_{x_i, x'_{-i}}^{x'_i, x'_{-i}}(i;t) = K_x^{x'}(t)\delta(x'_{-i}, x_{-i}) \quad (2)$$
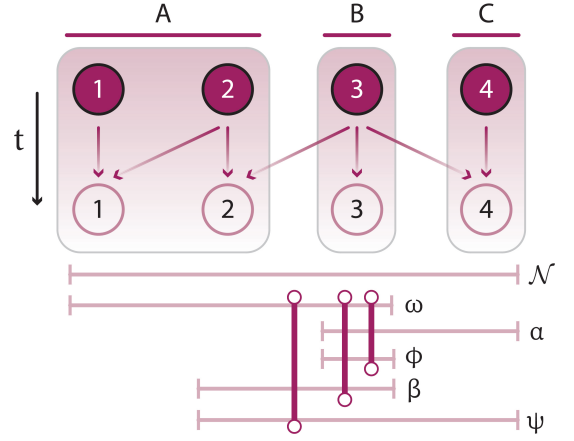

FIG. 1. An example of a multipartite process. The system consists of four subsystems, indicated by outlined circles. The arrows represent dependency constraints. Here, the state of subsystem 1 at time $t + \delta t$ depends on the state of subsystems 1 and 2 at time t. Subsystem 2 depends on 2 and 3. Subsystem 3 depends only on itself, and subsystem 4 depends on 3 and 4. These dependency constraints can be represented as the **dependency graph** ($1 \leftarrow 2 \leftarrow 3 \rightarrow 4$). The entire system, which is itself a unit, is denoted $\mathcal{N}$. This dependency graph defines the units $\omega, \alpha, \phi, \beta, \psi$ of the system. Intuitively, each unit is a set of subsystems that together evolve independently from other subsystems. Each unit structure, represented by circle-capped vertical lines, is a set of units that is closed under intersections and that encompasses all the subsystems (covers $\mathcal{N}$). There are four possible unit structures, $\mathcal{N}_1^* = \{\omega, \alpha, \phi\}$, $\mathcal{N}_2^* = \{\omega, \alpha, \phi, \beta\}$, $\mathcal{N}_3^* = \{\omega, \psi, \beta\}$, or $\mathcal{N}_4^* = \{\omega, \psi, \alpha, \beta, \phi\}$, that emerge from these units. The labels, A, B, and C, represent a grouping (of subsystems) that corresponds to particular physical systems, and has no effect on the units or unit structure.

We define a **dependency constraint** for any subsystem i as the specification of which other subsystems can affect i's dynamics. This constrains the system's rate matrix. For each subsystem i, we write $r(i;t)$ for any set of subsystems at time t that include i and for which

$$K_x^{x'}(i;t) = K_{x_{r(i;t)}}^{x'_{r(i;t)}}(i;t)\,\delta(x'_{-r(i;t)}, x_{-r(i;t)}) \quad (3)$$

holds for an appropriate set of $\left|X_{r(i;t)}\right| \times \left|X_{r(i;t)}\right|$ rate matrices $K_{x_{r(i;t)}}^{x'_{r(i;t)}}(i;t)$. In general, for any given i there could be multiple such sets $r(i;t)$.

We define a **unit** ω (at an implicit time t) as a set of subsystems i such that $i \in \omega$ implies that $r(i;t) \subseteq \omega$. Any intersection of two units is a unit, as is any union of two units. We write $\mathcal{N}^{**}$ for the set of all units in $\mathcal{N}$, other than the global system itself. For the example system shown in Fig. 1, $\mathcal{N}^{**} = \{\omega, \alpha, \phi, \beta, \psi\}$.

For any unit ω, we write the $|X_\omega| \times |X_\omega|$ rate matrix $K_{x_\omega}^{x'_\omega}(\omega;t) = \sum_{i \in \omega} K_{x_\omega}^{x'_\omega}(i;t)$. So, by Eq. (3), $K_x^{x'}(\omega;t) = \sum_{i \in \omega} K_x^{x'}(i;t) = K_{x_\omega}^{x'_\omega}(\omega;t)\delta(x'_{-\omega}, x_{-\omega})$. At any time t and for any unit ω, the local marginal distribution $p_{x_\omega}(t)$ evolves as a self-contained CTMC with the local rate ma-

trix $K_{x_\omega}^{x'_\omega}(\omega;t)$ (proved in Appendix A of [14]):

$$\frac{dp_{x_\omega}(t)}{dt} = \sum_{x'_\omega} K_{x_\omega}^{x'_\omega}(\omega;t)p_{x'_\omega}(t) = \sum_{x'_\omega}\sum_{i\in\omega} K_{x_\omega}^{x'_\omega}(i;t)p_{x'_\omega}(t) \tag{4}$$

Any unit, therefore, obeys all the usual stochastic thermodynamic theorems for Markov processes, e.g. the second law, FTs, TURs, and SLTs. In general, this is not true for a single subsystem in an MPP, due to that subsystem's dependency constraints [14]. Instead, the marginal distribution of subsystem i changes as

$$\frac{d}{dt}p_{x_i}(t) = \sum_{x'} K_x^{x'}(i;t)p_{x'}(t) \tag{5}$$

$$= \sum_{x'\neq x} K_x^{x'}(i;t)p_{x'}(t) - K_{x'}^{x}(i;t)p_x(t) \tag{6}$$

$$= \sum_{x'_{-i}}\sum_{x'_i\neq x_i} K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) - K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t) \tag{7}$$

Throughout the text, we refer to the bound derived in [1] as the global SLT. It states that, for any time τ,

$$\tau \geq \frac{\left(\mathcal{L}(p_x(0),p_x(\tau))\right)^2}{2\langle\sigma^{\mathcal{N}}(\tau)\rangle\langle\mathcal{A}^{\mathcal{N}}\rangle_\tau} \tag{8}$$

where $\mathcal{L}(p_x(0),p_x(\tau)) = \sum_x|p_x(0)-p_x(\tau)|$ is the total variation (L1) distance between the initial and final distributions, $\langle\mathcal{A}^{\mathcal{N}}\rangle_\tau = \frac{1}{\tau}\int_0^\tau\sum_{x'\neq x}dt K_x^{x'}(t)p_{x'}(t)$ is the global time-averaged dynamical activity (often interpreted as the frequency of state transitions), and $\langle\sigma^{\mathcal{N}}(\tau)\rangle = \int_0^\tau dt\sum_{x',x}K_x^{x'}(t)p_{x'}(t)\ln\left[\frac{K_x^{x'}(t)p_{x'}(t)}{K_{x'}^{x}(t)p_x(t)}\right]$ is the global EP accrued during the evolution.

In order to derive SLTs that account for the multipartite nature of the system dynamics, we now define subsystem-indexed contributions to the EP and dynamical activity. We start by decomposing the global EP rate $\langle\dot\sigma^{\mathcal{N}}(t)\rangle$ for an MPP at time t as follows:

$$\langle\dot\sigma^{\mathcal{N}}(t)\rangle := \sum_{x',x}K_x^{x'}(t)p_{x'}(t)\ln\left[\frac{K_x^{x'}(t)p_{x'}(t)}{K_{x'}^{x}(t)p_x(t)}\right] \tag{9}$$

$$= \sum_{i\in\mathcal{N}}\sum_{x',x}K_x^{x'}(i;t)p_{x'}(t)\ln\left[\frac{K_x^{x'}(i;t)p_{x'}(t)}{K_{x'}^{x}(i;t)p_x(t)}\right] \tag{10}$$

$$= \sum_{i\in\mathcal{N}}\sum_{x'_{-i},x'_i,x_i}K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t)\ln\left[\frac{K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t)}{K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)}\right] \tag{11}$$

$$:= \sum_{i\in\mathcal{N}}\langle\dot\zeta^i(t)\rangle \tag{12}$$

where we used Eq. (B2) of [14] in the step from Eq. (9) to Eq. (10). Here we have defined $\langle\dot\zeta^i(t)\rangle$ as a subsystem-indexed contribution [18] to the global EP rate. For any

time τ, we write the time-integrated quantity $\langle\zeta^i(\tau)\rangle = \int_0^\tau dt\langle\dot\zeta^i(t)\rangle$. We derive detailed relations regarding the thermodynamic extensibility of $\langle\dot\zeta^i(t)\rangle$ in Appendix C.

Similarly, we define the dynamical activity due only to (state transitions in) subsystem i as:

$$\mathcal{A}^i(t) := \sum_{x'_{-i}}\sum_{x_i}\sum_{x'_i\neq x_i}K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) \tag{13}$$

where $\mathcal{A}^{\mathcal{N}}(t) := \sum_{x'\neq x}K_x^{x'}(t)p_{x'}(t) = \sum_{i\in\mathcal{N}}\mathcal{A}^i(t)$. The corresponding time-averaged dynamical activity (averaged over an interval $[0,\tau]$) is $\langle\mathcal{A}^i\rangle_\tau = \frac{1}{\tau}\int_0^\tau\mathcal{A}^i(t)$.

The subsystem-local marginal total variation distance is defined in the usual way, as $\mathcal{L}(p_{x_i}(0),p_{x_i}(\tau)) = \sum_{x_i}|p_{x_i}(0)-p_{x_i}(\tau)|$.

We indicate the Shannon entropy of a distribution over states X_A as S^{X_A} or S^A. We set $k_B = 1$. Finally, we emphasize that all thermodynamic speed limits hold for any chosen time τ. Indeed, since all EPs and activities depend on τ, the speed limit bounds are functions of τ. Throughout the paper, we formulate our equations as if each subsystem i is connected to a single heat bath (with inverse temperature β_i); however, all of our results extend naturally to the case of multiple reservoirs per subsystem, using the formalisms detailed in [13].

Strengthened Speed Limits for Multipartite Processes—

As our main result, we find that accounting for the multipartite nature of a system's dynamics allows us to strengthen the thermodynamic speed limit (see derivation in Appendix A):

$$\tau \geq \frac{\left(\mathcal{L}(p_x(0),p_x(\tau))\right)^2}{2\left(\sum_{i\in\mathcal{N}}\sqrt{\langle\zeta^i(\tau)\rangle\langle\mathcal{A}^i\rangle_\tau}\right)^2} \tag{14}$$

The bound in this speed limit is always at least as tight as the bound in the global SLT (Eq. (8)) derived in [1] because the Cauchy-Schwartz inequality ensures that $\left(\sum_{i\in\mathcal{N}}\sqrt{\langle\zeta^i(\tau)\rangle\langle\mathcal{A}^i\rangle_\tau}\right)^2 \leq \langle\sigma^{\mathcal{N}}(\tau)\rangle\langle\mathcal{A}^{\mathcal{N}}\rangle_\tau$.

Our second result is a speed limit using only the thermodynamic quantities for subsystems within a single unit $\omega\subset\mathcal{N}$ (see derivation in Appendix A):

$$\forall\,\omega\in\mathcal{N}^{**} : \tau \geq \frac{\left(\mathcal{L}(p_{x_\omega}(0),p_{x_\omega}(\tau))\right)^2}{2\left(\sum_{i\in\omega}\sqrt{\langle\zeta^i(\tau)\rangle\langle\mathcal{A}^i\rangle_\tau}\right)^2} \tag{15}$$

One such speed limit holds for every unit $\omega\in\mathcal{N}^{**}$. The bound in this SLT is always at least as tight as the bound in the unit-local SLT Eq. (D1) for the unit ω (also referred to as the ω-local SLT) because of the Cauchy-Schwartz inequality, which gives $\left(\sum_{i\in\omega}\sqrt{\langle\zeta^i(\tau)\rangle\langle\mathcal{A}^i\rangle_\tau}\right)^2 \leq \langle\sigma^\omega(\tau)\rangle\langle\mathcal{A}^\omega\rangle_\tau$.

Our third result bounds an MPP's speed of evolution using only the thermodynamic properties of any single

subsystem (see derivation in Appendix B):

$$\forall\, i \in \mathcal{N} \;\; : \tau \geq \frac{\left(\mathcal{L}(p_{x_i}(0), p_{x_i}(\tau))\right)^2}{2\langle \zeta^i(\tau)\rangle \langle \mathcal{A}^i \rangle_\tau} \tag{16}$$

One such speed limit holds for every subsystem $i \in \mathcal{N}$. Together, they form the set of subsystem-local SLTs.

Although these last two results are not guaranteed to be tighter than the global SLT, in many cases, at least one of each kind (i.e., Eq. (15) for for some unit ω and Eq. (16) some subsystem i) is stronger than even our first result, Eq. (14). *A priori*, one cannot rank the strength of our results (Eqs. (14) to (16)) among themselves. However, it is straightforward to write necessary and sufficient conditions for any one bound to dominate over any other. For example, a necessary and sufficient condition for Eq. (16) (for subsystem i) to outperform Eq. (14) is if $\frac{\left(\sum_{j\in\mathcal{N}} \sqrt{\langle\zeta^j(\tau)\rangle\langle\mathcal{A}^j\rangle_\tau}\right)^2}{\langle\zeta^i(\tau)\rangle\langle\mathcal{A}^i\rangle_\tau}(\mathcal{L}(p_{x_i}(0),p_{x_i}(\tau)))^2 \geq (\mathcal{L}(p_x(0),p_x(\tau)))^2$. The fulfillment of such conditions will vary with the system's dependency graph and control protocol.

Example of Cellular Sensing— We quantitatively illustrate our results for an example MPP in which a cell is sensing and storing information about its environment. Using these numerical calculations, we evaluate the strength of our results (Eqs. (14) to (16)) among each other and in comparison to the global SLT.

The scenario, which has been previously studied in the literature [14, 16], is captured abstractly in Fig. 1 and illustrated in Fig. 2. The system consists of two sets of cellular receptors, each of which (belong to a unique cell and) independently observe the concentration of a ligand in the medium. One set of receptors additionally affects a set of proteins that act as a subcellular memory of the cell's fraction of bound receptors. Subsystem 3
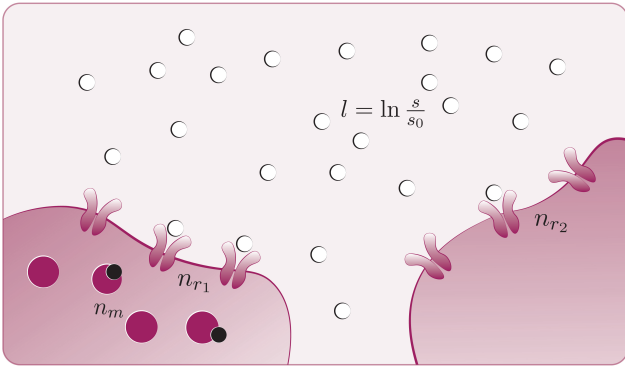

FIG. 2. The example scenario. Subsystem 3 is l, the concentration of the ligand in a medium. There are two "nearby" cells that sense this ligand concentration. The number of bound receptors, n_{r_1}, in Cell 1 comprise subsystem 2, and the number of bound receptors, n_{r_2}, in Cell 2 comprise subsystem 4. Subsystem 1 represents a memory of (phosphorylated) proteins, n_m, in Cell 1 that observes the state of its receptors.

represents the ligand concentration, which is indicated by the signal $l = \ln\frac{s}{s_0}$, where s_0 is some reference concentration value. Subsystem 2 represents the number of receptors bound by the ligand in the cell membrane of Cell 1. It is given by $n_{r_1} = 0, 1, 2, \ldots, N_{r_1}$, where N_{r_1} is the total number of receptors. Subsystem 4 represents the number bound receptors in Cell 2. It is given by $n_{r_2} = 0, 1, 2, \ldots, N_{r_2}$, where N_{r_2} is the total number of receptors. Subsystem 1 represents the number of phosphorylated proteins, serving as a memory of the bound receptors, in Cell 1. It is given by $n_m = 0, 1, 2, \ldots, N_m$, where N_m is the total number of proteins. In our calculations, we use $N_{r_1} = 3, N_{r_2} = 3$, and $N_m = 4$.

We construct the time-homogeneous rate matrices (THRMs) according to Section III of [16]. We set the initial joint distribution as $p_x(0) = p_{x_3}(0)p_{x_2|x_3}(0)p_{x_4|x_3}(0)p_{x_1|x_2}(0)$, where each of the conditional distributions are Gaussians with mean set by the value of the state on which it is dependent, and where p_{x_3} is $\sim \mathcal{N}(0, 0.01)$. For more details, see the code available at https://github.com/FaritaTasnim/MPP_SLTs_cellular_sensing.

We evolve the joint distribution over time according to the joint THRM $K_x^{x'}$ by solving the master equation to obtain $p_x(t) = p_x(0)e^{tK_x^{x'}}$. We calculate the distribution every 50 μs in the interval $[0, 55]$ ms. From the rate matrices and time-t distributions, we calculate all relevant thermodynamic quantities at each timestep.

Our numerical calculations (Fig. 3) confirm that our main result, the MPP SLT in Eq. (14) (pink), provides a tighter bound on the speed of system evolution than does the global SLT (Eq. (8)) from [1] (dark green). Additionally, we find that in this particular example, the MPP SLTs in Eq. (15) for every unit (lavender, sea green, violet, dark blue, olive green) and Eq. (16) for every subsystem (yellow, light blue, orange, lime green) are stronger than not only the global SLT, but also the MPP SLT in Eq. (14). Interestingly, even though the rate matrix for system evolution is time-homogeneous, we see that the strength of the subsystem-local SLT for subsystem 3 (orange) surpasses that for subsystem 2 (light blue) after 35 ms of system evolution.

Discussion— In this paper we have extended conventional classical speed limits to derive strengthened speed limits for the case of multiple, co-evolving subsystems — multipartite processes. These results can be useful for obtaining limits on the speed of state transformation for any composite system, e.g., chemical reaction networks [19, 20] or opinion dynamics [21, 22].

We suggest several avenues for future work. The analyses in this paper as well as those in [11, 13, 14, 23] consider time-homogeneous dependency constraints. In many real-world scenarios, however, the dependency constraints may change with time. Integrating this time-dependence into our framework may further strengthen the speed limits for MPPs. Additionally, we expect that incorporating finer-grained information
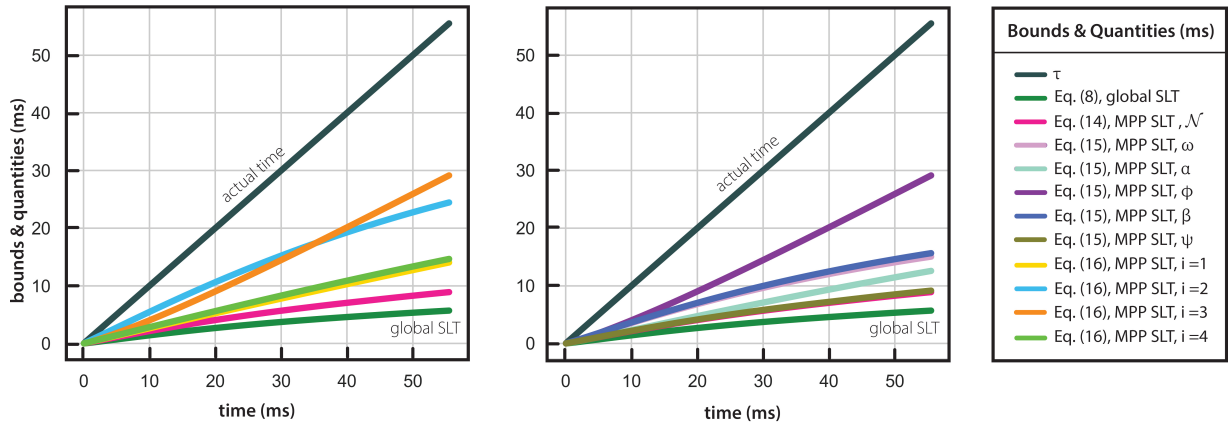
FIG. 3. Comparison of speed limit bounds for the example of cellular sensing shown in Fig. 2. The dark grey line represents the actual time of system evolution. All other lines represent the lower bounds on time provided by different SLTs. Note that bounds represented by the orange and violet lines are equivalent because the unit ϕ is composed of only subsystem 3.

about the topology of the dependency graph would lead to stronger SLTs. However, in many cases, one does not know the full details of the dependency graph, but instead might know certain of its properties, e.g., degree distribution or community structure. It would therefore be valuable to extend the stochastic thermodynamics of MPPs, including the SLTs, to cases where one has such summary statistics of the dependency graph.

[1] N. Shiraishi, K. Funo, and K. Saito, Speed limit for classical stochastic processes, Physical Review Letters **121**, 10.1103/physrevlett.121.070601 (2018).

[2] V. T. Vo, T. V. Vu, and Y. Hasegawa, Unified approach to classical speed limit and thermodynamic uncertainty relation, Physical Review E **102**, 10.1103/physreve.102.062132 (2020).

[3] Y. Zhang, Comment on "speed limit for classical stochastic processes", arXiV (2018).

[4] T. Van Vu and Y. Hasegawa, Geometrical bounds of the irreversibility in markovian systems, Physical Review Letters **126**, 010601 (2021).

[5] M. Okuyama and M. Ohzeki, Quantum speed limit is not quantum, Physical review letters **120**, 070402 (2018).

[6] The analysis of the thermodynamic implications of constraints on a system's allowed dynamics is in its infancy. One example of relevant research derives the stochastic thermodynamics under protocol constraints [24] There has also been some important work where the "constraint" on such a many-degree-of-freedom classical system is simply that it be some very narrowly defined type of system, whose dynamics is specified by many different kinds of parameters. For example, there has been analysis of the stochastic thermodynamics of chemical reaction networks [20, 25], of electronic circuits [26–28], of spin glasses where all spins are coupled to one another [29], of biological copying mechanisms [30], and of systems in which the state transitions of subsystems occur according to a synchronous global clock [10, 11]. The reader should note that resource theory and related research [31] also considers issues concerning broadly applicable constraints and thermodynamics, in the context of quantum mechanics.

[7] R. M. D'Souza, Structure comes to random graphs, Nature Physics **5**, 627 (2009).

[8] J. M. Horowitz and M. Esposito, Thermodynamics with continuous information flow, Physical Review X **4**, 031015 (2014).

[9] J. M. Horowitz, Multipartite information flow for multiple maxwell demons, Journal of Statistical Mechanics: Theory and Experiment **2015**, P03006 (2015).

[10] S. Ito and T. Sagawa, Information thermodynamics on causal networks, Physical review letters **111**, 180603 (2013).

[11] D. H. Wolpert, Uncertainty relations and fluctuation theorems for bayes nets, Physical Review Letters **125**, 10.1103/physrevlett.125.200602 (2020).

[12] D. H. Wolpert, Strengthened landauer bound for composite systems, arXiV (2020).

[13] D. H. Wolpert, Fluctuation theorems for multiple systems with interdependent dynamics, New Journal of Physics (2021).

[14] D. H. Wolpert, Minimal entropy production rate of interacting systems, New Journal of Physics **22**, 113013 (2020).

[15] This includes non-reciprocal couplings between subsystems [32], such as those that can be found in Ising networks represented by directed acyclic graphs (DAGs) [33], opinion networks [21, 22], and so on. This assumption of no back-action is used throughout literature [16, 34].

[16] D. Hartich, A. C. Barato, and U. Seifert, Sensory capacity: An information theoretical measure of the performance of a sensor, Physical Review E **93**, 10.1103/phys-

reve.93.022116 (2016).

[17] A. Dutta, G. M. Schütz, and D. Chowdhury, Stochastic thermodynamics and modes of operation of a ribosome: A network theoretic perspective, Physical Review E **101**, 032402 (2020).

[18] Note that, due to Eqs. (2) and (3), these subsystem-indexed quantities are mathematically equivalent whether they are formulated as contributions to a global quantity or a unit-local quantity. That is, for example, as long as $i \in \omega$, $\langle \dot{\zeta}^i(t) \rangle = \sum_{x',x} K_x^{x'}(i;t) p_{x'}(t) \ln \left[\frac{K_x^{x'}(t) p_{x'}(t)}{K_{x'}^{x}(t) p_x(t)} \right] = \sum_{x'_\omega, x_\omega} K_{x_\omega}^{x'_\omega}(i;t) p_{x'_\omega}(t) \ln \left[\frac{K_{x_\omega}^{x'_\omega}(t) p_{x'_\omega}(t)}{K_{x'_\omega}^{x_\omega}(t) p_{x_\omega}(t)} \right]$.

[19] R. Rao and M. Esposito, Conservation laws and work fluctuation relations in chemical reaction networks, The Journal of chemical physics **149**, 245101 (2018).

[20] R. Rao and M. Esposito, Nonequilibrium thermodynamics of chemical reaction networks: wisdom from stochastic thermodynamics, Physical Review X **6**, 041064 (2016).

[21] D. Acemoglu and A. Ozdaglar, Opinion dynamics and learning in social networks, Dynamic Games and Applications **1**, 3 (2011).

[22] P. Jia, A. MirTabatabaei, N. E. Friedkin, and F. Bullo, Opinion dynamics and the evolution of social power in influence networks, SIAM review **57**, 367 (2015).

[23] G. Kardeş and D. H. Wolpert, Thermodynamic uncertainty relations for multipartite processes, arXiv preprint arXiv:2101.01610 (2021).

[24] A. Kolchinsky and D. H. Wolpert, Entropy production and thermodynamics of information under protocol constraints, arXiv preprint arXiv:2008.10764 (2020).

[25] A. Wachtel, R. Rao, and M. Esposito, Thermodynamically consistent coarse graining of biocatalysts beyond michaelis–menten, New Journal of Physics **20**, 042002 (2018).

[26] C. Y. Gao and D. T. Limmer, Principles of low dissipation computing from a stochastic circuit model, arXiv preprint arXiv:2102.13067 (2021).

[27] D. H. Wolpert and A. Kolchinsky, Thermodynamics of computing with circuits, New Journal of Physics **22**, 063047 (2020).

[28] N. Freitas, J.-C. Delvenne, and M. Esposito, Stochastic thermodynamics of non-linear electronic circuits: A realistic framework for thermodynamics of computation, arXiv preprint arXiv:2008.10578 (2020).

[29] T. Herpich, T. Cossetto, G. Falasco, and M. Esposito, Stochastic thermodynamics of all-to-all interacting many-body systems, New Journal of Physics **22**, 063005 (2020).

[30] J. M. Poulton, P. R. Ten Wolde, and T. E. Ouldridge, Nonequilibrium correlations in minimal dynamical models of polymer copying, Proceedings of the National Academy of Sciences **116**, 1946 (2019).

[31] H. Wilming, R. Gallego, and J. Eisert, Second law of thermodynamics under control restrictions, Physical Review E **93**, 042126 (2016).

[32] S. A. Loos and S. H. Klapp, Irreversibility, heat and information flows induced by non-reciprocal interactions, New Journal of Physics **22**, 123051 (2020).

[33] H. Eissfeller and M. Opper, Mean-field monte carlo approach to the sherrington-kirkpatrick model with asymmetric couplings, Physical Review E **50**, 709 (1994).

[34] T. Sagawa and M. Ueda, Minimal energy cost for thermodynamic information processing: measurement and information erasure, Physical review letters **102**, 250602 (2009).

[35] K. Liu, Z. Gong, and M. Ueda, Thermodynamic uncertainty relation for arbitrary initial states, Physical Review Letters **125**, 10.1103/physrevlett.125.140602 (2020).

[36] R. A. Brittain, N. S. Jones, and T. E. Ouldridge, What we learn from the learning rate, Journal of Statistical Mechanics: Theory and Experiment **2017**, 063502 (2017).

Appendix A: Derivation of Eq. (14) and Eq. (15)

To begin, we note that subsystem i's contribution to the EP rate obeys the following inequality:

$$\langle \dot{\zeta}^i(t) \rangle = \sum_{x',x} K_x^{x'}(i;t) p_{x'}(t) \ln \left[\frac{K_x^{x'}(i;t) p_{x'}(t)}{K_{x'}^{x}(i;t) p_x(t)} \right] \tag{A1}$$

$$= \frac{1}{2} \sum_x \sum_{x' \neq x} \left(K_x^{x'}(i;t) p_{x'}(t) - K_{x'}^{x}(i;t) p_x(t) \right) \ln \left[\frac{K_x^{x'}(i;t) p_{x'}(t)}{K_{x'}^{x}(i;t) p_x(t)} \right] \tag{A2}$$

$$= \frac{1}{2} \sum_{x'_{-i}} \sum_{x_i} \sum_{x'_i \neq x_i} \left(K_{x_i, x_{-i}}^{x'_i, x'_{-i}}(i;t) p_{x'_i, x'_{-i}}(t) - K_{x'_i, x'_{-i}}^{x_i, x'_{-i}}(i;t) p_{x_i, x'_{-i}}(t) \right) \ln \left[\frac{K_{x_i, x'_{-i}}^{x'_i, x'_{-i}}(i;t) p_{x'_i, x'_{-i}}(t)}{K_{x'_i, x'_{-i}}^{x_i, x'_{-i}}(i;t) p_{x_i, x'_{-i}}(t)} \right] \tag{A3}$$

$$\geq \sum_{x'_{-i}} \sum_{x_i} \sum_{x'_i \neq x_i} \frac{\left(K_{x_i, x'_{-i}}^{x'_i, x'_{-i}}(i;t) p_{x'_i, x'_{-i}}(t) - K_{x'_i, x'_{-i}}^{x_i, x'_{-i}}(i;t) p_{x_i, x'_{-i}}(t) \right)^2}{K_{x_i, x'_{-i}}^{x'_i, x'_{-i}}(i;t) p_{x'_i, x'_{-i}}(t) + K_{x'_i, x'_{-i}}^{x_i, x'_{-i}}(i;t) p_{x_i, x'_{-i}}(t)} \tag{A4}$$

This is because $(a-b)\ln\frac{a}{b} \geq \frac{(a-b)^2}{(a+b)}$ for all positive a and b, as pointed out in [1]. We can then use the fact that $\frac{d}{dt}p_x(t) = \sum_{i\in\mathcal{N}}\frac{d}{dt}p_{x_i}(t)$ (which follows from Eq. (5) and Eq. (2)) to bound the global total variation distance as follows:

$$\sum_x \left|\frac{d}{dt}p_x(t)\right| = \sum_x \left|\sum_{x'\neq x}\sum_{i\in\mathcal{N}}\left(K_x^{x'}(i;t)p_{x'}(t) - K_{x'}^x(i;t)p_x(t)\right)\right| \tag{A5}$$

$$\leq \sum_{i\in\mathcal{N}}\sum_{x_i,x_{-i}}\left|\sum_{x'_{-i}}\sum_{x'_i\neq x_i}\left(K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) - K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)\right)\right| \tag{A6}$$

$$= \sum_{i\in\mathcal{N}}\sum_{x_i}\left|\sum_{x'_{-i}}\sum_{x'_i\neq x_i}\left(K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) - K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)\right)\right|$$

$$\leq \sum_{i\in\mathcal{N}}\sum_{x_i}\sqrt{\left(\sum_{x'_{-i}}\sum_{x'_i\neq x_i}\frac{\left(K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) - K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)\right)^2}{K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) + K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)}\right)\left(\sum_{x'_{-i}}\sum_{x'_i\neq x_i}K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) + K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)\right)} \tag{A7}$$

$$\leq \sum_{i\in\mathcal{N}}\sqrt{\left(\sum_{x'_{-i}}\sum_{x_i}\sum_{x'_i\neq x_i}\frac{\left(K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) - K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)\right)^2}{K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) + K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)}\right)\left(\sum_{x'_{-i}}\sum_{x_i}\sum_{x'_i\neq x_i}K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) + K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)\right)} \tag{A8}$$

$$\leq \sum_{i\in\mathcal{N}}\sqrt{2\langle\dot{\zeta}^i(t)\rangle\mathcal{A}^i(t)} \tag{A9}$$

Then, integrating over time,

$$\mathcal{L}(p_x(0), p_x(\tau)) \leq \sum_x \int_0^\tau dt\left|\frac{d}{dt}p_x(t)\right| \tag{A10}$$

$$\leq \sum_{i\in\mathcal{N}}\int_0^\tau dt\sqrt{2\langle\dot{\zeta}^i(t)\rangle\mathcal{A}^i(t)} \tag{A11}$$

$$\leq \sum_{i\in\mathcal{N}}\sqrt{2\tau\langle\zeta^i(\tau)\rangle\langle\mathcal{A}^i\rangle_\tau} \tag{A12}$$

where we have used the Cauchy-Schwartz inequality in the step from Eq. (A7) to Eq. (A8) and in the step from Eq. (A11) to Eq. (A12). This leads to the speed limit in Eq. (14). *Mutatis mutandis*, we can follow a similar exercise for the subsystems in a unit (same steps as above derivation but with all $\mathcal{N}$ replaced with ω, and with the state space limited to x_ω instead of x) to obtain the speed limit in Eq. (15). One such speed limit holds for each unit $\omega \in \mathcal{N}^{**}$.

Appendix B: Derivation of Eq. (16)

We can bound the marginal total variation distance of subsystem i as follows:

$$\sum_{x_i}\left|\frac{d}{dt}p_{x_i}(t)\right| \leq \sum_{x_i}\left|\sum_{x'_{-i}}\sum_{x'_i\neq x_i}\left(K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) - K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)\right)\right| \tag{B1}$$

$$\leq \sum_{x_i}\sqrt{\left(\sum_{x'_{-i}}\sum_{x'_i\neq x_i}\frac{\left(K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) - K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)\right)^2}{K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) + K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)}\right)\left(\sum_{x'_{-i}}\sum_{x'_i\neq x_i}K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) + K_{x'_i,x'_{-i}}^{x_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)\right)} \tag{B2}$$

$$\leq \sqrt{\left(\sum_{x'_{-i}}\sum_{x_i}\sum_{x'_i \neq x_i} \frac{\left(K^{x'_i,x'_{-i}}_{x_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) - K^{x_i,x'_{-i}}_{x'_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)\right)^2}{K^{x'_i,x'_{-i}}_{x_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) + K^{x_i,x'_{-i}}_{x'_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)}\right)\left(\sum_{x'_{-i}}\sum_{x_i}\sum_{x'_i \neq x_i} K^{x'_i,x'_{-i}}_{x_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t) + K^{x_i,x'_{-i}}_{x'_i,x'_{-i}}(i;t)p_{x_i,x'_{-i}}(t)\right)}$$

<div align="right">(B3)</div>

$$\leq \sqrt{2\langle\dot{\zeta}^i(t)\rangle\mathcal{A}^i(t)} \tag{B4}$$

Then, integrating over time,

$$\mathcal{L}(p_{x_i}(0),p_{x_i}(\tau)) \leq \sum_{x_i}\int_0^\tau dt\left|\frac{d}{dt}p_{x_i}(t)\right| \tag{B5}$$

$$\leq \int_0^\tau dt\sqrt{2\langle\dot{\zeta}^i(t)\rangle\mathcal{A}^i(t)} \tag{B6}$$

$$\leq \sqrt{2\tau\langle\zeta^i(\tau)\rangle\langle\mathcal{A}^i\rangle_\tau} \tag{B7}$$

where we have used the Cauchy-Schwartz inequality in the step from Eq. (B2) to Eq. (B3) and in the step from Eq. (B6) to Eq. (B7). This leads to the speed limit in Eq. (16). One such speed limit like this holds for each subsystem $i \in \mathcal{N}$.

Appendix C: Thermodynamic Extensibility and the Different Types of Local Detailed Balance in MPPs

Local detailed balance (LDB), as is standard in literature, means that the entire system obeys the following relation for arbitrary state transitions (in any subsystem i) using the same global Hamiltonian:

$$\ln\left[\frac{K^{x'}_x(i;t)}{K^x_{x'}(i;t)}\right] = \beta_i\left(H_{x'}(t) - H_x(t)\right) \tag{C1}$$

We call this **global LDB (GLDB)**.

In a system consisting of multiple, co-evolving subsystems, however, it is often impossible to write down a global Hamiltonian in a thermodynamically consistent way, e.g., as could be caused by non-reciprocal dependency constraints. Therefore, we also define the following types of LDB that can establish thermodynamic consistency for MPPs whose rate matrices are constrained by non-global Hamiltonians.

Unit LDB (ULDB) means that each unit ω obeys LDB for state transitions (of any subsystem i in that unit) using a unit-specific local Hamiltonian, which can only depend on the states of subsystems within that unit:

$$\ln\left[\frac{K^{x'}_x(i;t)}{K^x_{x'}(i;t)}\right] = \beta_i\left(H_{x'_\omega}(\omega;t) - H_{x_\omega}(\omega;t)\right) \tag{C2}$$

ULDB does not follow from GLDB.

Subsystem LDB (SLDB) means that each subsystem i obeys LDB for its own state transitions using its own local Hamiltonian, which can only depend on the states of the subsystems ($j \in r(i)$) that affect i:

$$\ln\left[\frac{K^{x'}_x(i;t)}{K^x_{x'}(i;t)}\right] = \beta_i\left(H_{x'_{r(i)}}(i;t) - H_{x_{r(i)}}(i;t)\right) \tag{C3}$$

for some local Hamiltonians $H_{x_{r(i)}}(i;t)$.

Finally, **multi-scale LDB (MSLDB)** means that both ULDB and GLDB hold. It is shown in [13] how to use "Hamiltonian stubs" to build a global Hamiltonian that ensures MSLDB. The same paper also works through some conditions involving "Hamiltonian scaling" in order for MSLDB to approximately hold.

Mathematically, the global SLT from [1] is certainly valid for all classical stochastic processes. However, when a system obeys only SLDB, e.g., if it is defined using local Hamiltonians, there does not necessarily exist a global Hamiltonian $H_x(t)$ that is consistent with all of the local Hamiltonians. Therefore, the quantity $\langle\sigma^{\mathcal{N}}(t)\rangle$, and thus the global SLT is not *thermodynamically extensible* (Eq. (5) from [1] becomes nonphysical).

Here we show that the thermodynamic extensibility of our MPP SLTs does not require GLDB. Instead, SLDB is sufficient. Therefore, our results vastly open up the experimental applicability of the thermodynamic speed limits to scenarios in which systems are defined using local, rather than global, Hamiltonians. This includes the growing number of network-based models for nonequilibrium systems in which a node's dynamics are defined by its local interactions with neighboring nodes.

Let's say that each subsystem i obeys SLDB. We start by noting that $\langle\dot{\zeta}^i(t)\rangle$, the contribution to the system EP rate due to state transitions in subsystem i, can be split into two contributions as follows,

$$\langle\dot{\zeta}^i(t)\rangle = \sum_{x',x} K^{x'}_x(i;t)p_{x'}(t)\ln\left[\frac{p_{x'}(t)}{p_x(t)}\right]$$
$$+ \sum_{x',x} K^{x'}_x(i;t)p_{x'}(t)\ln\left[\frac{K^{x'}_x(i;t)}{K^x_{x'}(i;t)}\right] \tag{C4}$$

To simplify the first sum on the RHS of Eq. (C4), using Eq. (5), we expand

$$\sum_{x',x} K_x^{x'}(i;t)p_{x'}(t)\ln\left[\frac{p_{x'}(t)}{p_x(t)}\right] \tag{C5}$$

$$= \sum_{x'_{-i}}\sum_{x_i}\sum_{x'_i\neq x_i} K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t)\ln\left[\frac{p_{x'_i,x'_{-i}}(t)}{p_{x_i,x'_{-i}}(t)}\right] \tag{C6}$$

$$= -\sum_{x'_{-i}}\sum_{x_i}\sum_{x'_i\neq x_i} K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t)\ln p_{x_i,x'_{-i}}(t) \tag{C7}$$

$$= -\sum_{x'_{-i}}\sum_{x_i}\sum_{x'_i\neq x_i} K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t)\ln p_{x'_{-i}|x_i}(t)$$
$$\quad -\sum_{x'_{-i}}\sum_{x_i}\sum_{x'_i\neq x_i} K_{x_i,x'_{-i}}^{x'_i,x'_{-i}}(i;t)p_{x'_i,x'_{-i}}(t)\ln p_{x_i}(t) \tag{C8}$$

$$= \frac{d^i}{dt}S^{X|X_i}(t) + \frac{d}{dt}S^{X_i}(t) \tag{C9}$$

The first term on the RHS, $\frac{d^i}{dt}S^{X|X_i}(t)$, is called the the "$i$-windowed derivative of the conditional entropy". It was first defined in [14], and equals the negative of the information flow $\langle \dot{I}^{-i\to i}\rangle$. This is the information flow from subsystem i to the rest of the subsystems in the system. In the special case that the system is in a NESS, this term is referred to in the literature as the "learning rate" [9], i.e., the rate at which subsystem i "learns" about the other subsystems $-i$. The second term on the RHS, $\frac{d}{dt}S^{X_i}(t)$, is the rate of change of the Shannon entropy of subsystem i, defined by the marginal distribution over only subsystem i's states.

To simplify the second sum on the RHS of Eq. (C4), we note that the state transitions that contribute to $\langle \check{\zeta}^i(t)\rangle$ involve only a change in x_i, with the state x'_{-i} held fixed. This is because $K_x^{x'}(i;t) = 0$ if $x'_{-i}\neq x_{-i}$. Invoking SLDB, the entropy flow (EF) associated with such a transition $x'(= x'_i,x'_{-i})\to x(= x_i,x'_{-i})$ is $Q_{x'\to x}^i = H_{x_{r(i)}}(i;t)-H_{x_{r(i)}}(i;t)$, for some local Hamiltonians $H_{x_{r(i)}}(i;t)$. The EF rate into the subsystem i is $\dot{Q}^i(t) = \sum_{x'\neq x} K_x^{x'}(i;t)p_{x'}(t)Q_{x'\to x}^i$.

Therefore, SLDB allows us to write, in a similar fashion to Eq. (5) of [1],

$$\langle \check{\zeta}^i(t)\rangle = \frac{d^i}{dt}S^{X|X_i}(t) + \frac{d}{dt}S^{X_i}(t) + \beta_i\dot{Q}^i(t) \tag{C10}$$

$$= -\langle \dot{I}^{-i\to i}\rangle + \frac{d}{dt}S^{X_i}(t) + \beta_i\dot{Q}^i(t) \tag{C11}$$

We see, therefore, that subsystem i's contribution to the system's EP rate consists of the information flow between i and the rest of the system, the rate of change of Shannon entropy of the marginal distribution of i, and the EF rate into subsystem i from its reservoirs. This establishes thermodynamic extensibility of the quantity $\langle \check{\zeta}^i(t)\rangle$. Thus our MPP SLTs extend in a physically meaningful way to experimental scenarios in which the rate matrices are defined with local rather than global Hamiltonians.

Appendix D: Unit-Local Speed Limits for MPPs

In order for the global system to reach a certain distribution $p_x(\tau)$ by time τ, each unit ω must have also reached the corresponding marginal distribution $p_{x_\omega}(\tau) = \sum_{x_{-\omega}} p_x(\tau)$ by time τ. Therefore, by applying the SLT to each of the units in $\mathcal{N}^{**}$ we get a set of additional lower bounds on the speed of system evolution:

$$\forall \omega \in \mathcal{N}^{**} : \tau \geq \frac{\left(\mathcal{L}(p_{x_\omega}(0),p_{x_\omega}(\tau))\right)^2}{2\langle\sigma^\omega(\tau)\rangle\langle\mathcal{A}^\omega\rangle_\tau} \tag{D1}$$

These bounds form the set of **unit-local SLTs** for an MPP. Here, $\mathcal{L}(p_{x_\omega}(0),p_{x_\omega}(\tau)) = \sum_{x_\omega}|p_{x_\omega}(0) - p_{x_\omega}(\tau)|$ is the ω-local marginal total variation distance, $\langle\mathcal{A}^\omega\rangle_\tau = \frac{1}{\tau}\int_0^\tau \sum_{x'_\omega\neq x_\omega} dt K_{x_\omega}^{x'_\omega}(\omega;t)p_{x'_\omega}(t)$ is the ω-local time-averaged dynamical activity, and $\langle\sigma^\omega(\tau)\rangle = \int_0^\tau dt \sum_{x'_\omega,x_\omega} K_{x_\omega}^{x'_\omega}(\omega;t)p_{x'_\omega}(t)\ln\left[\frac{K_{x_\omega}^{x'_\omega}(\omega;t)p_{x'_\omega}(t)}{K_{x'_\omega}^{x_\omega}(\omega;t)p_{x_\omega}(t)}\right]$ is the ω-local EP. We note in Appendix E that the strongest of a generalized form of the SLT that involves a linear combination of unit-local thermodynamic properties is given by the unit-local SLT for the system's slowest-evolving unit.

Appendix E: Strongest Version of the SLT Involving Linear Combinations of Unit-Local Thermodynamic Properties is Given by a Unit-Local SLT

From the global and unit-local speed limits Eq. (D1), it follows immediately that we can bound the time of evolution of a system by applying the SLTs to any linear combination of units. Define $\mathcal{N}^\dagger = \mathcal{N}^{**}\cup\mathcal{N}$. Let $\vec{m}$ be an arbitrary, non-negative, unit-indexed vector $\vec{m}\in\{\mathbb{R}_0^+\}^{|\mathcal{N}^\dagger|}$. For any such $\vec{m}$, the following **generalized SLT** holds in an MPP:

$$\tau \geq \frac{\sum_{\omega\in\mathcal{N}^\dagger} m_\omega c^\omega(\tau)}{\sum_{\omega'\in\mathcal{N}^\dagger} m_{\omega'}\langle\sigma^{\omega'}\rangle} \tag{E1}$$

where, for each unit ω, we have defined the function $c^\omega(\tau) = \frac{\left(\mathcal{L}(p_{x_\omega}(0),p_{x_\omega}(\tau))\right)^2}{2\langle\mathcal{A}^\omega\rangle_\tau}$. To see this, weight the unit-local SLT Eq. (D1) for each $\omega\in\mathcal{N}^\dagger$ by m_ω, add these inequalities together, and rearrange into a lower bound on τ. Intuitively, a specific value of $\vec{m}$ corresponds to using a particular weighting of the local thermodynamic properties of different units in bounding the speed of system evolution.

We will now use the generalized mediant inequality to derive the tightest of these bounds. It states that the weighted median of n fractions, $\frac{a_1}{b_1},\frac{a_2}{b_2},\ldots,\frac{a_n}{b_n}$, with weights $w_i\geq 0$, lies between the largest and smallest values of the fractions $\frac{a_i}{b_i}$:

$$\min_i \frac{a_i}{b_i} \leq \frac{\sum_i w_i a_i}{\sum_j w_j b_j} \leq \max_i \frac{a_i}{b_i} \tag{E2}$$

Taking the i as units, w_i becomes m_ω, each a_i becomes $c^\omega(\tau)$, and each b_i becomes $\langle \sigma^\omega \rangle$. It immediately follows that

$$\min_{\omega \in \mathcal{N}^+} \frac{c^\omega(\tau)}{\langle \sigma^\omega \rangle} \leq \frac{\sum_{\omega \in \mathcal{N}^+} m_\omega c^\omega(\tau)}{\sum_{\omega' \in \mathcal{N}^+} m_{\omega'} \langle \sigma^{\omega'} \rangle} \leq \max_{\omega \in \mathcal{N}^+} \frac{c^\omega(\tau)}{\langle \sigma^\omega \rangle} \quad \text{(E3)}$$

The tightest (loosest) bound in the set of bounds defined by Eq. (E1) is achieved when $\vec{m}$ is a Dirac-delta function over the space of units ($\mathcal{N}^+$) such that the only non-zero value of m_ω is the one for the slowest-evolving (fastest-evolving) unit, i.e., the unit with the largest (smallest) value of the bound in its local SLT. Intuitively, the slowest-evolving (fastest-evolving) unit (SEU, FEU) corresponds to the unit for which the local thermodynamic and dynamical properties of its evolution are low (high) enough that in the best case scenario, out of all the units, this unit would take the most (least) time to reach its desired final distribution.

Appendix F: Unit-Local Speed Limits More Tightly Constrain Space of Accessible Distributions than Global Speed Limit (Shown for the Case of Two Independent Bits)

For simplicity of illustration, consider the general case of an MPP with n non-overlapping units ω, each with $|X_\omega| = k$ possible states. The total number of joint states is then $|X| = k^n$. All valid joint probability distributions comprise the $(k^n - 1)$-simplex, Δ^{k^n-1} (with side length $\sqrt{2}$). For our calculations we embed this simplex in k^n-space, where it is an $(k^n - 1)$-dimensional manifold. The global SLT, which states $\sum_x |p_x(\tau) - p_x(0)| \leq \sqrt{2\tau \langle \sigma \rangle \langle \mathcal{A} \rangle_\tau}$, restricts the space of accessible distributions to the intersection between Δ^{k^n-1} and $B_1^{k^n}(p_x(0))$. $B_1^{k^n}(p_x(0))$ is the k^n-cross polytope ($l_1^{k^n}$ ball) with edge length $\sqrt{2} \times \sqrt{2\tau \langle \sigma \rangle \langle \mathcal{A} \rangle_\tau}$ centered at the point on Δ^{k^n-1} that represents the initial joint distribution, $p_x(0)$.

Each unit-local SLT Eq. (D1), which states $\sum_{x_\omega} |p_{x_\omega}(\tau) - p_{x_\omega}(0)| \leq \sqrt{2\tau \langle \sigma^\omega \rangle \langle \mathcal{A}^\omega \rangle_\tau}$, further restricts this space of accessible distributions to the collection of points in Δ^{k^n-1} that are consistent with the intersection of the marginal simplex Δ^{k-1} with $B_1^k(p_{x_\omega}(0))$. $B_1^k(p_{x_\omega}(0))$ is the k-cross polytope with edge length $\sqrt{2} \times \sqrt{2\tau \langle \sigma^\omega \rangle \langle \mathcal{A}^\omega \rangle_\tau}$) centered at the point on the Δ^{k-1} that represents its initial marginal distribution, $p_{x_\omega}(0) = \sum_{x_{-\omega}} p_x(0)$. Similar geometric relations hold for overlapping units.

To demonstrate a calculation, we proceed by considering an MPP consisting of n independent bits that don't necessarily evolve according to the same rate matrix. Thus, any bit i generates an EP of $\langle \sigma^i \rangle \leq \langle \sigma \rangle$ in the interval $[0, \tau]$. Additionally, the time-averaged dynamical activity of any bit i is $\langle \mathcal{A}^i \rangle_\tau \leq \langle \mathcal{A} \rangle_\tau$.

Let us consider the case of $n = 2$ independent bits, i and j. This means that each bit comprises its own unit, which doesn't overlap with the other. For accompany-

ing illustrations, see Fig. 4. Let's also say that the initial joint distribution is a Kronecker-delta function centered at p_{11}, i.e. $p_{ij}(0) : p_{00}(0) = 0, p_{01}(0) = 0, p_{10}(0) = 0, p_{11}(0) = 1$. Due to the symmetry of the problem, the following calculations hold for a delta function initial distribution centered at any one of the four joint states. All final joint distributions $p_{ij}(\tau)$ must lie on the simplex Δ^3, which can be visualized as a tetrahedron. The total volume V_t of this space of distributions is $1/3$ (remember that a probability simplex has side length $\sqrt{2}$).

We can write out the global SLT as:

$$\sqrt{2\tau \langle \sigma \rangle \langle \mathcal{A} \rangle_\tau} = c \geq \\ |p_{00}(\tau)| + |p_{01}(\tau)| + |p_{10}(\tau)| + |p_{11}(\tau) - 1| \quad \text{(F1)}$$

The RHS is representable by the $B_1^4(p_{ij}(0))$, which is a 16-cell centered at one corner of Δ^3. The intersection of this off-centered l_1^4 ball with Δ^3 is Δ_g^3, a scaled-down version of the 3-simplex with side length $\frac{c}{\sqrt{2}}$, and this is the set of accessible distributions as restricted by the global SLT alone. The volume V_g of (the convex hull of) this set of points reachable in time τ by the global SLT is the volume of a tetrahedron of side length $\frac{c}{\sqrt{2}}$. Thus $V_g = c^3/24$. Therefore, the fraction of distributions reachable in time τ by the global SLT is $f_g = V_g/V_t = c^3/8$. Note, by extension, that for any system with a total of $|X|$ states and delta-function initial distribution, the fraction of distributions reachable in time τ as constrained by the global SLT is $(\frac{c}{2})^{|X|-1}$.

Now we consider the restrictions placed by the unit-local SLTs for each bit. Since the bits are independent, it must hold that $\langle \sigma \rangle = \langle \sigma^i \rangle + \langle \sigma^j \rangle - \langle \Delta \mathcal{I}^{\mathcal{N}^*} \rangle$ (where $\langle \Delta \mathcal{I}^{\mathcal{N}^*} \rangle$ is the change in inclusion-exclusion information defined in [14]) and $\langle \mathcal{A} \rangle_\tau = \langle \mathcal{A}^i \rangle_\tau + \langle \mathcal{A}^j \rangle_\tau$. We write out the unit-local SLTs as:

$$\sqrt{2\tau \langle \sigma^i \rangle \langle \mathcal{A}^i \rangle_\tau} = b \geq \\ |p_{00}(\tau) + p_{01}(\tau)| + |p_{10}(\tau) + p_{11}(\tau) - 1| \quad \text{(F2)}$$

$$\sqrt{2\tau \langle \sigma^j \rangle \langle \mathcal{A}^j \rangle_\tau} = d \geq \\ |p_{00}(\tau) + p_{10}(\tau)| + |p_{01}(\tau) + p_{11}(\tau) - 1| \quad \text{(F3)}$$

where, without loss of generality, we set $b \leq d$. Using these restrictions, we can draw, as shown in Fig. 4, the convex hull of the set of points reachable in time τ by the set of these two local SLTs. To obtain this convex hull, simply note that p_{01} and p_{10} can increase together by $\frac{b}{\sqrt{2}}$ and $\frac{d}{\sqrt{2}}$, respectively, but that for any other pair of joint probabilities, the *sum* of their probabilities (i.e. $p_{00} + p_{10}$ or $p_{00} + p_{01}$) can at maximum change by $\min[\frac{b}{\sqrt{2}}, \frac{d}{\sqrt{2}}] = \frac{b}{\sqrt{2}}$. The volume V_u of this convex hull, as shown in Fig. 4b, is $\frac{b^2(d-b)}{8} + \frac{b^3}{12}$. Therefore, the fraction of distributions reachable in time τ by the global SLT is $f_u = \frac{3b^2(d-b)}{8} + \frac{b^3}{4}$. Therefore, for two identical but
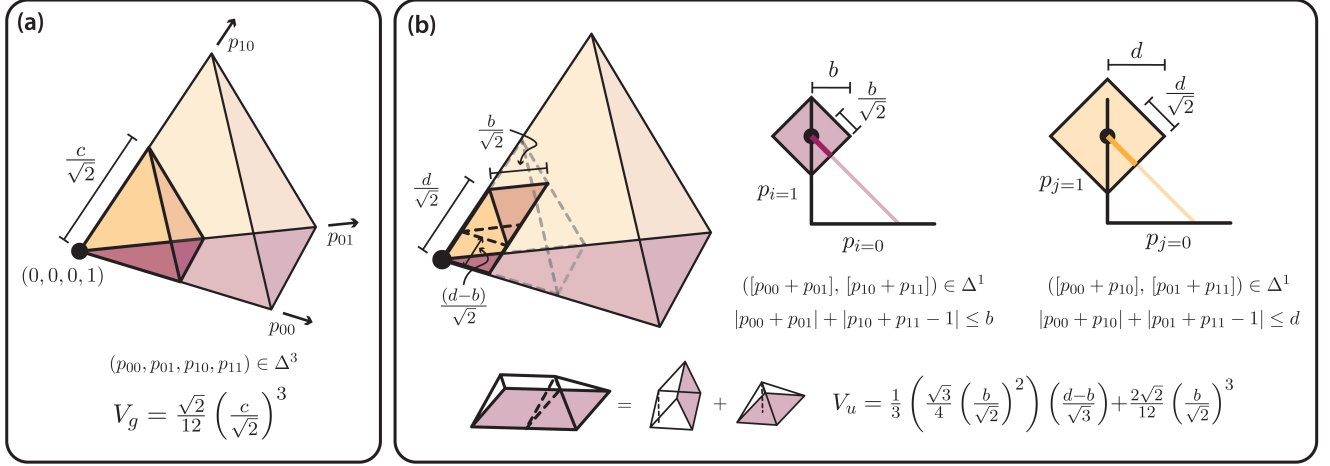
FIG. 4. For two independent bits, the set of unit-local SLTs more tightly restricts the space of accessible final joint distributions than does the global SLT. Geometrical representations and calculations of the space of accessible distributions as restricted by (a) the global SLT alone and by (b) the set of unit-local SLTs are shown for the case of two independently evolving bits. In both panels, the large tetrahedron represents the 3-simplex that constrains all probability distributions with 4 possible states, and the highlighted convex hull within its interior represents the space of joint distributions accessible by time τ. In (b) right, the 2D graphs show the space of accessible marginal distributions for each individual bit (shaded squares), as given by the unit-local SLTs. In (b) left, the joint distributions that are consistent with these marginal distributions form the highlighted convex hull. The dotted grey lines represent the convex hull of the space of accessible joint distributions as constricted by the global SLT, and are provided for visual comparison to the space consistent with the unit-local SLTs.

independently-evolving bits,

$$\frac{f_u}{f_g} = \frac{3b^2(d-b)+2b^3}{c^3} \tag{F4}$$

Note that in the case that $b = d = \frac{c}{2}$, this reduces to $\frac{1}{4}$. Also note that for the case of $n = 1$ bit, or for a set of fully connected bits (where each bit interacts with all other bits), the only unit is the global system itself, and thus there exist no local SLTs apart from the global SLT. Therefore, in this case, $\frac{f_u=f_g}{f_g} = 1$.

We suspect that it is true in general (for any type of unit structure and for any initial distribution) that the unit-local SLTs combined with the global SLT more tightly restrict the space of accessible final distributions than does the global SLT alone. We posit this is because the unit-local SLTs, since they involve marginal distributions, provide multiple constraints on the same joint probabilities. Note that if all of the transitions in a system occur in a single unit, the unit-local SLTs reduce to the global SLT and no additional constraints are possible. We also suspect that the MPP SLTs we show in the main text (and derived in Appendices A and B) taken together would more tightly constrain the space of accessible final distributions than the global SLTs, and potentially even the unit-local SLTs taken together. We leave these investigations for future work.

Appendix G: Maximum and Minimum Speeds for Bayes' Nets

We can run an MPP as a Bayes' Net (BN) [11]. This means that state transitions are no longer stochastically timed, but occur according to a global, synchronous clock. Therefore, for the results in this section, we are concerned with discrete-time thermodynamics instead of continuous-time dynamics as we were for the more general case of MPPs. We emphasize that although a BN is a special case of an MPP, the SLTs for MPPs do not reduce simply to the SLTs for BNs. We derive these results from relations specific to BNs (and which don't apply to MPPs). We therefore introduce these BN-specific relations here.

In a BN, For any times t, $t' > t$, the joint distribution at t' is

$$p(x_1^{t'}, x_2^{t'}, \ldots) = \sum_{x_1^t, x_2^t, \ldots} p(x_1^{t'}, x_2^{t'}, \ldots | x_1^t, x_2^t, \ldots) p(x_1^t, x_2^t, \ldots) \tag{G1}$$

Typically, there are conditional independencies in how each of the subsystems evolve. In general, this means that $p(x_1^{t'}, x_2^{t'}, \ldots | x_1^t, x_2^t, \ldots)$ can be decomposed into a product of conditional distributions, each of which captures one of those conditional independencies. If we consider the example in Fig. 1 and take instead as subsystems A, B, and C, retaining the same dependency constraints, and set $x_B^{t'} = x_B^t$ with probability
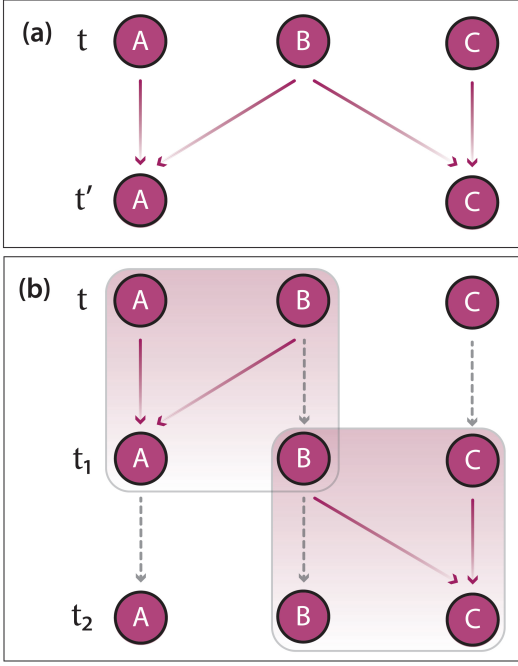
FIG. 5. (a) An example BN showing the conditional independencies between subsystems A, B, and C. The top three nodes are the root nodes of the DAG, representing the time-t states of the three subsystems. The bottom nodes are the leaf nodes, representing the time-t' states of two of the subsystems. (Subsystem B does not evolve, so we dispose of its leaf node.) The directed edges into the bottom-left leaf node indicate that $x_A^{t'}$ depends only on x_A^t and x_B^t, the two parents of that node. Similarly, the edges into the bottom-right leaf node indicate that $x_C^{t'}$ depends only on x_B^t and x_C^t, the two parents of that node. (b) An example of running the same BN as a sequence of solitary processes. Dashed gray arrows indicate the identity map. In the first solitary process, the evolving system consists of subsystems A and B, while in the second solitary process it consists of B and C.

1, then

$$
\begin{aligned}
p(x_A^{t'}, x_B^{t'}, x_C^{t'} | x_A^t, x_B^t, x_C^t) = \\
p(x_A^{t'} | x_A^t, x_B^t) p(x_C^{t'} | x_B^t, x_C^t) \delta(x_B^{t'}, x_B^t).
\end{aligned}
\tag{G2}
$$

These conditional (in)dependencies can be represented as a directed acyclic graph (DAG), as shown in Fig. 5. This representation of a distribution is an example of a Bayes' net. BNs can be generalized to represent the dynamics over an arbitrary number of subsystems. In addition, they can represent dynamics over an arbitrary number of times, not just the two times illustrated in Fig. 5(a), simply by adding more layers to the DAG.

Following the convention in [10, 11], we restrict attention to BNs where only one node's conditional distribution is implemented at a time. This means the evolution of any node v in the BN occurs in a solitary process, where the evolving system a is the union of the subsystem corresponding to v and the subsystems corresponding to the parents of v. For a solitary process occurring

over the time interval $[0, \tau]$, the local EP of the evolving system a along a trajectory $\boldsymbol{x}$ is

$$
\sigma_a(\boldsymbol{x}, p_a^0, p_a^\tau) := \left(\ln[p_a^0(\boldsymbol{x}_a^0)] - \ln[p_a^\tau(\boldsymbol{x}_a^\tau)] \right) - Q_1(\boldsymbol{x}_a) \tag{G3}
$$

$Q(\boldsymbol{x})$ is the entropy flow (EF) from the heat bath into the joint system during the process. For a solitary process, $Q(\boldsymbol{x}) = Q_1(\boldsymbol{x}_a)$ for some function Q_1. The evolving system of a solitary process is called a solitary system, and the local EP of a solitary process is only a function of $\boldsymbol{x}_a$. The global EP generated by the evolving system a and the fixed system b is

$$
\sigma(\boldsymbol{x}, p^0, p^\tau) = \left(\ln[p^0(\boldsymbol{x}^0)] - \ln[p^\tau(\boldsymbol{x}^\tau)] \right) - Q(\boldsymbol{x}) \tag{G4}
$$
$$
= \sigma_a(\boldsymbol{x}, p_a^0, p_a^\tau) - \Delta I_{p^0, p^\tau}(\boldsymbol{x}_a; \boldsymbol{x}_b) \tag{G5}
$$

where $\Delta I_{p^0, p^\tau}(\boldsymbol{x}_a; \boldsymbol{x}_b)$ is the difference between the ending and starting (stochastic) mutual information between the evolving and the fixed systems.

We write the set of nodes in a BN as V, and label them with successive integers. In general, more than one $v \in V$ might refer to evolution of the same subsystem, just at different times. In light of the fact that the evolving system in each solitary process will be the union of a node and its parents, for any $V' \subset V$ we define $[V'] := V' \cup pa(V') := \cup_{v \in V'} v \cup pa(v)$, where $pa(v)$ indicates the parents of node v. In addition, for any subset $V' \subset V$, we define $-V' = V \backslash V'$. We write the distribution over all subsystems after the process implementing node v has run as p^{t_v}, and unless indicated otherwise, assume that it runs in the time interval $[t_{v-1}, t_v]$. The root nodes are jointly sampled at $t = 0$ according to distribution $p^0(x)$, and the BN is allowed to run through all its nodes, resulting in full trajectory $\boldsymbol{x}$. For any $v \in V$, we write the segment of $\boldsymbol{x}$ corresponding to the time interval when node v runs as $\boldsymbol{x}^v$, reserving superscripts for time slices and subscripts for specification of particular subsystems. As an example of this notation, $\boldsymbol{x}_{[v]}$ is the full trajectory of the components of $\boldsymbol{x}$ specified by $[v]$, while $\boldsymbol{x}_{[v]}^v$ is the segment of that trajectory in the time interval $[t_{v-1}, t_v]$.

Let $\pi = \{\pi_v(x_v | x_{pa(v)})\}$ indicate the set of the conditional distributions at the nodes of the BN. We write the local EP generated in the solitary process associated with running node v as $\sigma_{[v]}$. Since EP is cumulative over time, by repeated application of Eq. (G5), once for each node in the BN, we see that the global EP incurred by running all nodes in the BN if the joint system follows trajectory $\boldsymbol{x}$ is

$$
\begin{aligned}
\sigma(\boldsymbol{x}, \pi, p^0) = \sum_{v=1}^{|V|} \Big[& \sigma_{[v]}(\boldsymbol{x}_{[v]}^v, \pi_v, p^{v-1}) \\
& + \Big(I_{p^{v-1}}(\boldsymbol{x}_{[v]}^{v-1}; \boldsymbol{x}_{-[v]}^{v-1}) - I_{p^v}(\boldsymbol{x}_{[v]}^v; \boldsymbol{x}_{-[v]}^v) \Big) \Big] \\
:= \sum_{v \in V} & \Big(\sigma_{[v]}(\boldsymbol{x}) - \Delta I^v(\boldsymbol{x}) \Big)
\end{aligned}
$$
$$
\tag{G6}
$$

where $\Delta I^v(\boldsymbol{x})$ is the change in (stochastic) multi-information between $[v]$ and $-[v]$. Taking the ensemble average over all trajectories $\boldsymbol{x}$, we arrive at the decomposition of the global EP in a BN,

$$\langle\sigma\rangle := \sum_{v\in V}\left(\langle\sigma_{[v]}\rangle - \langle\Delta I^v\rangle\right) \tag{G7}$$

We define the local dynamical activity of running the node v of a BN,

$$\mathcal{A}^v_{[v]} = \sum_{x'_{[v]}\neq x_{[v]}}\sum_{i\in[v]}K^{x'_{[v]}}_{x_{[v]}}(i)p_{x'_{[v]}}(t_{v-1}), \tag{G8}$$

where $t_{v-1} = \sum_{v'=1}^{v-1}\tau_{v'}$, $t_v = t_{v-1} + \tau_v$, $x'_{[v]} = x_{[v]}(t_{v-1})$, $x_{[v]} = x_{[v]}(t_v)$, and we drop the time dependence of the rate matrix elements $K^{x'_{[v]}}_{x_{[v]}}(i)$, because they don't change while node v runs. The global and local dynamical activities are related as

$$\langle\mathcal{A}\rangle = \sum_v\langle\mathcal{A}^v_{[v]}\rangle_{\tau_v}. \tag{G9}$$

1. Speed Limits for Bayes' Net Evolution

We can plug in lower bounds analogous to the unit-local SLT for each local EP $\langle\sigma_{[v]}\rangle$ into Eq. (G7) to obtain a new SLT applicable to a BN (this follows from Eq. (G6)):

$$\langle\sigma\rangle \geq \sum_{v\in V}\left(\frac{\left(\mathcal{L}_{[v]}(p^{t_{v-1}},p^{t_v})\right)^2}{2\tau_v\langle\mathcal{A}^v_{[v]}\rangle_{\tau_v}} - \langle\Delta I^v\rangle\right) \tag{G10}$$

where $\mathcal{L}_{[v]}(p^{t_{v-1}},p^{t_v})$ is the total variation distance in the joint distribution and $\langle\mathcal{A}^v_{[v]}\rangle_{\tau_v}$ is the time-averaged dynamical activity produced due to running node v, which occurs over a span of time $\tau_v = t_v - t_{v-1}$. Eq. (G10)'s lower bound on global EP is different from that obtained by treating the process as the evolution of a single physical system $\langle\sigma\rangle \geq \left(\mathcal{L}(p^0,p^\tau)\right)^2/(2\tau\langle\mathcal{A}\rangle_\tau)$, where $\mathcal{L}(p^0,p^\tau)$ is the total variation distance of the system distribution and $\langle\mathcal{A}\rangle_\tau$ is the time-averaged dynamical activity produced incurred by running the entire BN during the time interval $[0,\tau]$.

For a given global EP, Eq. (G10) serves as a trade-off between changing the speed, inverse activity, and total variation distance of one step in the process (e.g. that in which node v runs) and changing those of other steps (e.g. in which other nodes $v' \neq v$ run). We can see that slowing down any particular node in the BN could potentially give a smaller lower bound on EP. Therefore, one can exploit Eq. (G10) to identify the thermodynamically costlier nodes, i.e., those with the largest value of inverse activity-weighted total variation distance $\left(\mathcal{L}_{[v]}(p^{t_{v-1}},p^{t_v})\right)^2/\langle\mathcal{A}^v_{[v]}\rangle_{\tau_v}$. Slowing down these

particular nodes provides the most effective way to try to lower the dissipative cost of running a BN while making the fewest possible changes to the process. As a result, Eq. (G10) offers fine-grained control over a BN. In the example of a digital circuit, this translates to knowing which gates, if slowed down enough, could drastically reduce the circuit's dissipative cost. Finally, this result also shows the role of the multi-information in connecting the local run-time of a single node in the BN to the total EP of the entire BN.

We can also derive a bound for the time τ of running the entire BN. We first rearrange the speed limit for each local EP to get a local speed limit for each step in the process of running the BN:

$$\tau_v \geq \frac{\left(\mathcal{L}_{[v]}(p^{t_{v-1}},p^{t_v})\right)^2}{2\langle\sigma_{[v]}\rangle\langle\mathcal{A}^v_{[v]}\rangle_{\tau_v}} \tag{G11}$$

Upon summing this over all steps of the BN, we obtain a system-wide speed limit:

$$\tau = \sum_v\tau_v \geq \sum_v\frac{\left(\mathcal{L}_{[v]}(p^{t_{v-1}},p^{t_v})\right)^2}{2\langle\sigma_{[v]}\rangle\langle\mathcal{A}^v_{[v]}\rangle_{\tau_v}} \tag{G12}$$

which is different from the speed limit obtained by treating the process as the evolution of a single physical system: $\tau \geq \left(\mathcal{L}(p^0,p^\tau)\right)^2/(2\langle\sigma\rangle\langle\mathcal{A}\rangle_\tau)$. For a given desired speed of evolution of the entire system, Eq. (G12) serves as a trade-off between the changing the local thermodynamic quantities of one step in the process (e.g. that in which node v runs) versus changing those of other steps (e.g. in which other nodes $v' \neq v$ run).

Importantly, we remark that in the special case that the system begins and ends in the same non-equilibrium steady state (NESS) but can stray away from the NESS during the system evolution, the global SLTs applied to BNs provide lower bounds of zero. However, our results, Eqs. (G10) and (G12), provide strictly positive lower bounds.

2. Minimum Speeds of Bayes' Net Evolution

We now derive a minimum speed for the evolution of a BN. We start by noting that in each step of running a BN, the local EP $\langle\sigma_{[v]}\rangle$ obeys all the regular TURs. We can rearrange the thermodynamic uncertainty relation (TUR) for arbitrary initial states, derived in [35], to obtain a lower bound on the speed of evolution of for each step (i.e. running of a node v of the BN)

$$\tau_v \leq \sqrt{\frac{\langle\sigma_{[v]}\rangle\text{Var}(J_{[v]})}{2\langle j^{\tau_v}_{[v]}\rangle}} \tag{G13}$$

Upon summing this relation across all steps, we obtain a system-wide minimum speed (or maximum time) re-

quired to evolve the BN according to some set of instantaneous currents $\{\langle j^{\tau_v}_{[v]}\rangle\}$ and with a variance in the accumulated current $\text{Var}(J_{[v]})$:

$$\tau = \sum_v \tau_v \leq \sum_v \sqrt{\frac{\langle \sigma_{[v]}\rangle \text{Var}(J_{[v]})}{2\langle j^{\tau_v}_{[v]}\rangle}} \qquad \text{(G14)}$$

where for a local current $J_{[v]}$ produced by running a node v, $\text{Var}(J_{[v]})$ is a measure of the bulk local current fluctuations and $\langle j^{\tau_v}_{[v]}\rangle$ is the ensemble-average local instantaneous current observed right when node v has finished running. This minimum speed (Eq. (G14)) is different from that obtained by treating the BN as a single, undecomposed physical system: $\tau \leq \sqrt{(\langle\sigma\rangle\text{Var}(J))/(2\langle j^\tau\rangle)}$, where for a global current J produced by running the entire BN, $\text{Var}(J)$ is a measure of the bulk global current fluctuations and $\langle j^\tau\rangle$ is the instantaneous final current observed after running the entire BN.

For an engineer with a maximum allowable time of BN evolution, Eq. (G14) can be interpreted as a trade-off between changing the precisions of local bulk currents, the local EPs, and the local instantaneous currents of each step in the process (i.e. running node v) versus changing those quantities for another step (i.e. running node $v' \neq v$).

Appendix H: Lower Bounds on the Difference between the EP Rate and the Rate at which One Unit Learns about Another Based on Counterfactual Thermodynamic Quantities

Here we utilize the decomposition of the global expected EP rate as derived in [14]:

$$\langle\dot\sigma(t)\rangle = \langle\dot\sigma^\phi(t)\rangle + \langle\dot\sigma_{K(\omega\backslash\phi;t);\omega}(t)\rangle + \langle\dot\sigma_{K(\mathcal{N}\backslash\omega;t)}(t)\rangle$$
$$+ \frac{d^\phi}{dt}S^{X_\omega|X_\phi}(t) + \frac{d^\omega}{dt}S^{X|X_\omega}(t) \qquad \text{(H1)}$$

where ω and $\phi \subset \omega$ are both units under $K^{x'}_x(t)$. This equation applies to the example MPP depicted in Fig. 1.

The first term on the RHS, $\langle\dot\sigma^\phi(t)\rangle$, is the local expected EP rate of the unit ϕ at time t. The second term, $\langle\dot\sigma_{K(\omega\backslash\phi;t);\omega}(t)\rangle$ is the expected EP rate of the unit ω at time t if at that instant it were to evolve for an infinitesimal time δt under the counterfactual rate matrix $K(\omega\backslash\phi;t)$. Similarly, the third term, $\langle\dot\sigma_{K(\mathcal{N}\backslash\omega;t)}(t)\rangle$ is the expected EP rate of the entire system at time t if at that instant it were to evolve for an infinitesimal time δt under the counterfactual rate matrix $K(\mathcal{N}\backslash\omega;t)$. The counterfactual rate matrix $K(A;t)$ indicates the rate matrix if at time t only the subsystems in A are allowed to change state, and the states of the remaining subsystems in the system are held fixed. In a physical system, this corresponds to the counterfactual case of removing the reservoirs coupled to the subsystems in $\mathcal{N}\backslash A$ at time t.

The fourth term on the RHS is the windowed ϕ-derivative of the conditional entropy in X_ω given X_ϕ.

It is a measure of how quickly the statistical coupling between X_ω and $X_{\omega\backslash\phi}$ changes at time t, if rather than evolving under the actual rate matrix, the system were to evolve under a counterfactual rate matrix $K(\phi;t)$ in which $x_{\omega\backslash\phi}$ is not allowed to change. Similarly, the fifth term is the windowed ω-derivative of the conditional entropy in X given X_ω. We note that the general quantity $\frac{d^A}{dt}S^{X|X_A}(t)$ is the derivative of the negative mutual information between X_A and X_{-A} under the counterfactual rate matrix $K(A;t)$, because $S^{X_{-A}}(t)$ doesn't change under that counterfactual rate matrix. In the special case of two subsystems, both leading each other, with A as one of those subsystems, $\frac{d^A}{dt}S^{X|X_A}(t)$ is the same as the "information flow", termed the learning rate, analyzed in [8, 36]. See Eq. 4 in [9] for a generalization of information flows to multiple subsystems that is similar to the general expression $\frac{d^A}{dt}S^{X|X_A}(t)$. We can therefore identify $\frac{d^\phi}{dt}S^{X_\omega|X_\phi}(t)$ as the rate at which unit ϕ learns about the set of subsystems, $\omega\backslash\phi$, in unit ω but not in ϕ.

To proceed, we will now derive relations to instantaneous changes in distributions. We begin with the definition for a generalized counterfactual EP rate:

$$\langle\dot\sigma_{K(A;t);B}(t)\rangle =$$
$$\sum_{x_B,x'_B} K^{x'_B}_{x_B}(A;t)p_{x'_B}(t)\ln\frac{K^{x'_B}_{x_B}(A;t)p_{x'_B}(t)}{K^{x_B}_{x'_B}(A;t)p_{x_B}(t)}, \qquad \text{(H2)}$$

where $A \subseteq B \subseteq \mathcal{N}$, and B is a unit, but A is not necessarily a unit. $K^{x'_B}_{x_B}(A;t)p_{x'_B}(t)$ is the infinitesmal probability flow from $x'_B \to x_B$ in the distribution p_{X_B} if at that instant in time, p_{X_B} were to evolve under $K(A;t)$ instead of $K(B;t)$. $K^{x_B}_{x'_B}(A;t)p_{x_B}(t)$ is the infinitesmal probability flow from $x_B \to x'_B$ in that same situation. These quantities can also be viewed as the frequency of jumps between those states x'_B and x_B if the entire distribution over X_B were to evolve under the counterfactual rate matrix at that instant in time. This means that these quantities are related to the instantaneous counterfactual dynamical activity,

$$\mathcal{A}^{B;K(A;t)}(t) := \sum_{x'_B\neq x_B} K^{x'_B}_{x_B}(A;t)p_{x'_B}(t). \qquad \text{(H3)}$$

Following the analysis in [2], we define $\pi^{x'_B}_{x_B}$ as the forward transition from $x'_B \to x_B (\neq x'_B)$ and introduce a probability distribution for the forward transition under the counterfactual rate matrix:

$$\overrightarrow{P}_{K(A;t)}(\pi^{x'_B}_{x_B})(t) := (\mathcal{A}^{B;K(A;t)}(t))^{-1}(K^{x'_B}_{x_B}(A;t)p_{x'_B}(t)). \quad \text{(H4)}$$

Similarly, the probability distribution of the backward transition $\pi^{x_B}_{x'_B}$ from $x_B \to x'_B$ under the counterfactual rate matrix is

$$\overleftarrow{P}_{K(A;t)}(\pi^{x_B}_{x'_B})(t) := (\mathcal{A}^{B;K(A;t)}(t))^{-1}(K^{x_B}_{x'_B}(A;t)p_{x_B}(t)). \quad \text{(H5)}$$

The counterfactual EP rate can thus be rewritten as

$$\langle \dot{\sigma}_{K(A;t);B}(t) \rangle =$$

$$\mathcal{A}^{B;K(A;t)}(t) * D[\overrightarrow{P}_{K(A;t)}(\pi_{x_B}^{x'_B})(t) \| \overleftarrow{P}_{K(A;t)}(\pi_{x'_B}^{x_B})(t)]. \quad \text{(H6)}$$

We can then apply the same analysis as in Eqs. (16) - (19) of [2] to arrive at

$$\sqrt{2\mathcal{A}^{B;K(A;t)}(t) \langle \dot{\sigma}_{K(A;t);B}(t) \rangle} \geq \sum_{x_B} \left| \frac{d^{K(A;t)}}{dt} p_{x_B}(t) \right|, \quad \text{(H7)}$$

where $\frac{d^{K(A;t)}}{dt} p_{x_B}(t)$ is the instantaneous counterfactual change in the probability $p_{x_B}(t)$ if at that moment the distribution p^{X_B} evolved according to the counterfactual rate matrix $K(A;t)$ instead of the rate matrix $K(B;t)$ (under which it will actually evolve). In Fig. 6, this counterfactual infinitesemal change is represented by the dotted lines, whereas the actual evolution of the probability distributions $\dot{p}_{x_B}(t)$ and $\dot{p}_{x_B}^A(t)$ are represented by the solid lines.

We also define the following relation for the derivative of the total variation distance of a distribution p_{x_B} under a counterfactual rate matrix $K(A;t)$

$$\lim_{\delta t \to 0} \frac{\delta \mathcal{L}^{B;K(A;t)}(p_{x_B}(t))}{\delta t} =$$

$$\lim_{\delta t \to 0} \frac{\sum_{x_B} \left| p_{x_B}(t) - p_{x_B}^{K(A;t)}(t + \delta t) \right|}{\delta t} \quad \text{(H8)}$$

$$\dot{\mathcal{L}}^{B;K(A;t)}(p_{x_B}(t)) = \sum_{x_B} \left| \frac{d^{K(A;t)}}{dt} p_{x_B}(t) \right| \quad \text{(H9)}$$

where $\frac{d^{K(A;t)}}{dt} p_{x_B}(t)$ is the distribution at time $t + \delta t$ if $p_{x_B}(t)$ were evolved using the rate matrix $K^B(A;t)$ for a time δt.

We can then arrive at the desired lower bound on a generalized counterfactual EP rate:

$$\langle \dot{\sigma}_{K(A;t);B}(t) \rangle \geq$$

$$\frac{\left(\sum_{x_B} \left| \frac{d^{K(A;t)}}{dt} p_{x_B}(t) \right| \right)^2}{2\mathcal{A}^{B;K(A;t)}(t)} = \frac{\left(\dot{\mathcal{L}}^{K(A;t)}(p_{x_B}(t)) \right)^2}{2\mathcal{A}^{B;K(A;t)}(t)}, \quad \text{(H10)}$$

where $\dot{\mathcal{L}}^{K(A;t)}(p_{x_B}(t))$ is the instantaneous rate of change of the distribution p_{X_B} and $\mathcal{A}^{B;K(A;t)}(t)$ is the instantaneous dynamical activity in unit B if, at that time t, p_{X_B} were to evolve under the counterfactual rate matrix $K(A;t)$. Eq. (H10) provides an upper bound on the (inverse activity-weighted) instantaneous change in the probability distribution over X_B if the subsystems in $B \backslash A$ were not allowed to change state. Note that it would be mathematically incorrect to integrate these results over time to obtain speed limits for system evolution. The trouble arises from trying to integrate the
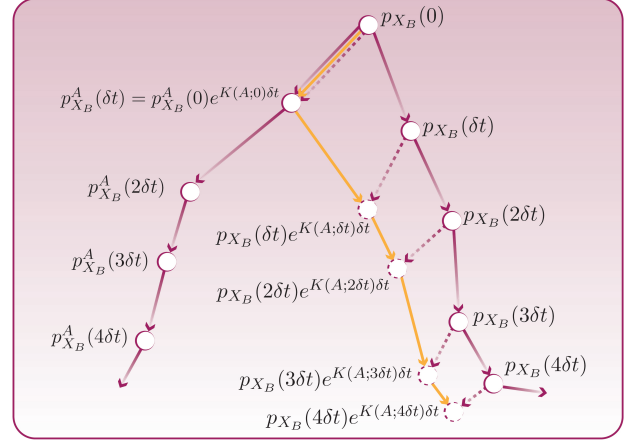

FIG. 6. An illustration of the mathematical complication involved in calculating the time-integral of the instantaneous counterfactual EP rate. The white circular dots represent instantaneous probability distributions. The arrows represent the evolution of probability distributions for infinitesimal times. The leftward path of solid magenta arrows represents the evolution of the probability distribution over unit B under a rate matrix $K(A;t)$ which at every time t only allows the subsystems in $A \subset B$ to change state. The rightward path of solid magenta arrows represents the evolution of the probability distribution over unit B under the actual rate matrix $K(B;t)$. Each dotted magenta arrow represents the evolution of the actual probability distribution $p_{x_B}(t)$ if at time t it were to evolve for an infinitesimal time period δt under the counterfactual rate matrix $K(A;t)$. This means that the states of subsystems in $B \backslash A$ would be held fixed for that infinitesimal time period δt. Since the set of dotted arrows do not provide a continuous path integral over which to integrate thermodynamic quantities, such time integrals cannot be consistently evaluated *unless* it would be mathematically equivalent to integrate instead over the yellow path. The yellow path represents an experimentally inaccessible evolution between disjoint probability distributions.

quantities in Eq. (H7) over time, since such a time integral would equate to path-integrating over the disjointed dotted paths in Fig. 6. For example, integrating expected counterfactual EP rate $\langle \dot{\sigma}_{K(A;t);B}(t) \rangle$ over time to get $\langle \sigma_{K(A;t);B}(t) \rangle$, the "total EP if at each instantaneous point in time the probability distribution *were to* evolve under a counterfactual rate matrix" wouldn't make sense *unless* it is acceptable to equate such a time integral to a path integral over the yellow path in Fig. 6.

We also note that in Appendix C of [14], it is shown that the fourth and fifth terms on the RHS of Eq. (H1) are non-negative. Using the non-negativity of the fourth term and plugging in relevant versions of Eq. (H10), we

obtain the following result. For any time t,

$$\langle \dot{\sigma}(t) \rangle - \frac{d^{\omega}}{dt} S^{X|X_{\omega}}(t) \geq$$

$$\frac{\left(\dot{\mathcal{L}}(p_{x_{\phi}}(t))\right)^2}{2\mathcal{A}^{\phi}(t)} + \frac{\left(\dot{\mathcal{L}}^{K(\omega\backslash\phi;t)}(p_{x_{\omega}}(t))\right)^2}{2\mathcal{A}^{\omega;K(\omega\backslash\phi;t)}(t)} + \frac{\left(\dot{\mathcal{L}}^{K(\mathcal{N}\backslash\omega;t)}(p_x(t))\right)^2}{2\mathcal{A}^{\mathcal{N};K(\mathcal{N}\backslash\omega;t)}(t)}$$

$$\text{(H11)}$$

This inequality provides, for any time t, a lower bound on how much greater the global EP rate must be than the rate at which the subsystems in ω learns about the subsystems in $-\omega$ (the $\omega - \mathcal{N}$ learning rate). This bound involves counterfactual instantaneous dynamical activities and counterfactual instantaneous total variation distances of the probability distributions over units $\phi \subset \omega$, ω, and $\mathcal{N}$. These quantities are either evaluated under the actual rate matrix or under an instantaneous (i.e. evaluated at time t) counterfactual rate matrix. This inequality can also be interpreted as an upper bound on the thermodynamically allowable instantaneous (inverse activity-weighted) change in the distributions of these units under actual or counterfactual rate matrices.

If we also incorporate the non-negativity of the fifth term on the RHS of Eq. (H1), we obtain

$$\langle \dot{\sigma}(t) \rangle \geq \frac{\left(\dot{\mathcal{L}}(p_{x_{\phi}}(t))\right)^2}{2\mathcal{A}^{\phi}(t)} + \frac{\left(\dot{\mathcal{L}}^{K(\omega\backslash\phi;t)}(p_{x_{\omega}}(t))\right)^2}{2\mathcal{A}^{\omega;K(\omega\backslash\phi;t)}(t)}$$

$$+ \frac{\left(\dot{\mathcal{L}}^{K(\mathcal{N}\backslash\omega;t)}(p_x(t))\right)^2}{2\mathcal{A}^{\mathcal{N};K(\mathcal{N}\backslash\omega;t)}(t)}$$

$$\text{(H12)}$$

Eq. (H12) states that, for the evolution of ϕ during $[0, \tau]$, there is a greater entropic cost than would be expected by the classical speed limit derived from properties only related to ϕ. This is due to the way in which the subsystems in ϕ interact with other subsystems within a larger dependency structure. The classical speed limit indirectly (i.e., if the last application of the Cauchy-Schwartz inequality to the time integrals in its derivation is omitted) gives a lower bound on the EP rate for unit ϕ that equals the first term on the RHS of Eq. (H12). When considering the composite system within which ϕ resides, however, we realize that there is an increased instantaneous minimal thermodynamic cost (in terms of EP rate) that must be paid. This increased cost is equal to the second and third terms on the RHS. The second term equals the minimal cost in EP rate of evolving unit ω, of which ϕ is a subset, during $[0, \tau]$ if the subsystems in ϕ remain fixed. The third term equals the minimal cost in EP rate of evolving the global system $\mathcal{N}$, of which ω is a subset, during $[0, \tau]$ if the subsystems in ω remain in a fixed state. Since these two terms are positive, Eq. (H12) thus provides a lower bound on the instantaneous global EP rate that is tighter than the lower bound obtained from the classical speed limit derived from the dynamics of one unit. Thus, we show that knowing the details of the larger dependency structure provides a stronger lower bound on the global EP rate.

Instead of replacing the local EP rate of ϕ with its lower bound, we could also keep the ϕ-local EP rate in Eq. (H12) and rearrange to obtain

$$\langle \dot{\sigma}(t) \rangle - \langle \dot{\sigma}^{\phi}(t) \rangle \geq$$

$$\frac{\left(\dot{\mathcal{L}}^{K(\omega\backslash\phi;t)}(p_{x_{\omega}}(t))\right)^2}{2\mathcal{A}^{\omega;K(\omega\backslash\phi;t)}(t)} + \frac{\left(\dot{\mathcal{L}}^{K(\mathcal{N}\backslash\omega;t)}(p_x(t))\right)^2}{2\mathcal{A}^{\mathcal{N};K(\mathcal{N}\backslash\omega;t)}(t)} \quad \text{(H13)}$$

Eq. (H13) explicitly provides the minimum additional EP rate that must be paid due to the interactions ϕ has within a larger dependency structure. It is a lower bound on the minimum difference in the EP rate for the entire system and the EP rate for a unit $\phi \subset \omega$ in that system using thermodynamic quantities under counterfactual rate matrices. This can also be interpreted as an upper bound on the thermodynamically allowable instantaneous (inverse activity-weighted) changes in the distributions of unit ω and the entire system $\mathcal{N}$ under counterfactual rate matrices.

These results relate *actual* EP rates and learning rates to entirely *counterfactual* thermodynamic quantities, which implies that in multipartite systems, what could be the case *but is not actually happening* has an effect on what *is actually happening* in the system. We note that, while the LHS of Eqs. (H12) and (H13) can be integrated over time, the RHS cannot, due to the reasoning discussed in Fig. 6.

Finally, we note that because the decomposition considered here is iterative, many valid equations such as Eq. (H1) exist in an MPP [14]. Therefore, in general, for any given MPP, there exist very many results analogous to those shown in Eqs. (H11), (H12), and (H13). As one example, the analogous result to Eq. (H11) as applied to the decomposition in Eq. (15) from [14] is

$$\langle \dot{\sigma}^{\omega}(t) \rangle - \frac{d^{\phi}}{dt} S^{X_{\omega}|X_{\phi}}(t) \geq$$

$$+ \frac{\left(\dot{\mathcal{L}}(p_{x_{\phi}}(t))\right)^2}{2\mathcal{A}^{\phi}(t)} + \frac{\left(\dot{\mathcal{L}}^{K(\omega\backslash\phi;t)}(p_{x_{\omega}}(t))\right)^2}{2\mathcal{A}^{\omega;K(\omega\backslash\phi;t)}(t)}$$

$$\text{(H14)}$$

which is a lower bound on how much greater the local EP rate of ω must be than the rate at which the subsystems in ϕ learn about the subsystems in $\omega\backslash\phi$ (the $\phi - \omega$ learning rate) based on counterfactual instantaneous thermodynamic quantities.

A wealth of other lower bounds can also be derived. As a final example, instead of replacing the ϕ-local EP rate with its lower bound, we can move it to the LHS to obtain a lower bound on how much greater the *difference* between the ω-local and ϕ-local EP rates must be than the $\phi - \omega$ learning rate:

$$\left(\langle \dot{\sigma}^{\omega}(t) \rangle - \langle \dot{\sigma}^{\phi}(t) \rangle\right) - \frac{d^{\phi}}{dt} S^{X_{\omega}|X_{\phi}}(t) \geq \frac{\left(\dot{\mathcal{L}}^{K(\omega\backslash\phi;t)}(p_{x_{\omega}}(t))\right)^2}{2\mathcal{A}^{\omega;K(\omega\backslash\phi;t)}(t)}$$

$$\text{(H15)}$$