Uncertainty relations and fluctuation theorems for Bayes nets

David H. Wolpert

Santa Fe Institute, Santa Fe, New Mexico
Complexity Science Hub, Vienna
Arizona State University, Tempe, Arizona
http://davidwolpert.weebly.com

Many physical scenarios are naturally modeled as a set of multiple co-evolving systems. Recently, research in stochastic thermodynamics has considered such scenarios, e.g., by modeling the co-evolution of the systems as a Bayes net. In particular, we now have a fluctuation theorem relating the entropy production of one of the systems in a Bayes net to the overall structure of the Bayes net. Here I extend this recent research in four ways. First, I derive fluctuation theorems concerning arbitrary subsets of the systems in the Bayes net. Second, I derive "conditional" fluctuation theorems, governing the probability distribution of entropy production in an arbitrary subset of the systems, conditioned on the entropy production in a different subset of the systems. I then derive thermodynamic uncertainty relations relating the total entropy production of all the systems in the Bayes net to the set of all the precisions of probability currents within the individual systems. I end with an example.

I. INTRODUCTION

Much of the work in stochastic thermodynamics considers a single non-equilibrium system executing a specified continuous-time or discrete-time evolution. Examples include analyses of a system undergoing bit erasure [52, 56], or more generally a system undergoing an arbitrary discrete-time dynamics [31, 51, 66], as well as a system maintaining a non-equilibrium steady state (NESS [62]). In particular, there has been groundbreaking work on fluctuation theorems (FTs [8, 10, 25, 54, 62]) and thermodynamic uncertainty relations (TURs [5, 13, 16, 17, 19, 29]) for single systems.

Other research has considered the thermodynamics of a pair of interacting systems [18], in some cases where the first system measures the second one [20, 57], or performs a sequence of measurements and manipulations of the second one [1, 22, 30, 64]. In particular, there has been research on FTs for a system under the feedback control of another system [22, 60].

However, many physical scenarios are most naturally viewed as sets of more than two interacting systems, with their joint discrete-time dynamics described by a probabilistic graphical model [26]. For example, such scenarios are very common in biology [3, 6, 14, 15, 27, 28, 32]. The extension of stochastic thermodynamics to study such scenarios was pioneered in [23, 24, 50], who modeled the joint discrete-time dynamics of the systems as Bayesian networks (BNs [26, 33]) [34]. The major result of [23] was an FT governing the relationship between the total entropy production (EP) of an arbitrary single one of the systems in the BN, v, the initial and final mutual informations between v and the rest of the BN, and total transfer entropy from v to the rest of the BN. They then illustrated this FT for time-delayed adaptation of a biological system to its environment.

In this paper I also consider the thermodynamics of BNs. In contrast to [23] I derive FTs governing the amount of EP generated by an arbitrary subset of the systems in a BN. I also derive "conditional FTs". These govern the amount of EP generated by any subset of the systems in a BN, conditioned on how much EP is generated by a separate subset of those systems [42]. In addition I derive TURs that relate the total EP generated by running a BN to the precisions of currents defined separately for each of the systems in that BN [2, 13, 16, 17, 19]. I end with an example of these TURs.

II. STOCHASTIC THERMODYNAMICS OF BAYESIAN NETWORKS

Although many of the results below are more general, to fix thinking the reader can presume that the entire set of interacting systems evolves according to a continuous-time Markov chain (CTMC). I will sometimes use "(forward) protocol", or "process", to refer to a sequence of Hamiltonians and rate matrices in such a CTMC. The results below hold even if the system is coupled with multiple reservoirs [10, 13]. However, as usual, to assign thermodynamic meaning to those results we typically need to assume local detailed balance (LDB [12, 62]).

I write the Kronecker delta as $\delta(.,.)$. I assume the reader knows basic terminology concerning Bayesian networks, e.g., directed acyclic graphs (DAGs), their roots, leaves, children, parents, ancestors etc. [43]. I indicate the entropy of a distribution $p(X)$ as $S(p(X))$ or $S(p)$ [7], and sometimes indicate the entropy at time t as $S_t(X)$. I write the mutual information of a distribution $p(X, Y)$ as $I_p(X; Y)$, or just $I(X; Y)$ for short. In addition, I write $|A|$ for the cardinality of any set A.

A. Path-wise solitary processes

I will use the term **semi-fixed** process to refer to any process involving two distinct systems, the **evolving** and **fixed** systems, with states x_A and x_B, respectively, where x_B does not change during the interval $[t_0, t_1]$. (Such processes have been analyzed before in [21, 58, 60].)

As is conventional [62, 65], write the **(path-wise) global entropy production** incurred if the joint system in a semi-fixed process follows trajectory $\boldsymbol{x}$ as

$$\sigma(\boldsymbol{x}, p^{t_0}, p^{t_1}) := \left(\ln[p^{t_0}(\boldsymbol{x}^{t_0})] - \ln[p^{t_1}(\boldsymbol{x}^{t_1})] \right) - \mathcal{Q}(\boldsymbol{x}, p^{t_0}) \quad (1)$$

where p^{t_0} and p^{t_1} are the marginals of the joint system at t_0 and t_1, respectively, and $\mathcal{Q}(\cdot, \cdot)$ is the total entropy flow (EF) into the joint system during the process. Also define the **(path-wise) local EP** for the evolving system as

$$\overline{\sigma}_A(\boldsymbol{x}, p_A^{t_0}, p_A^{t_1}) := \left(\ln[p_A^{t_0}(\boldsymbol{x}_A^{t_0})] - \ln[p_A^{t_1}(\boldsymbol{x}_A^{t_1})] \right) - \mathcal{Q}(\boldsymbol{x}, p^{t_0})$$
$$(2)$$

where p_A refers to the marginal distribution over the states of the evolving system [44]. So in a semi-fixed process,

$$\sigma(\boldsymbol{x}, p^{t_0}, p^{t_1}) = \overline{\sigma}_A(\boldsymbol{x}_A, p_1^{t_0}, p_A^{t_1}) - \Delta I_{p^{t_0}, p^{t_1}}(\boldsymbol{x}_A; \boldsymbol{x}_B) \quad (3)$$

where $\Delta I_{p^{t_0}, p^{t_1}}(\boldsymbol{x}_A; \boldsymbol{x}_B)$ is the difference between the ending and starting (path-wise) mutual information between the two systems [59, 61].

An important special type of semi-fixed process is one in which we can write

$$\mathcal{Q}(\boldsymbol{x}, p^{t_0}) = \mathcal{Q}_1(\boldsymbol{x}_A, p_A^{t_0}) \quad (4)$$

for some appropriate function $\mathcal{Q}_1(\cdot, \cdot)$. I refer to this as a **(path-wise) solitary process**, and write the associated local EP as

$$\sigma_A(\boldsymbol{x}_A, p_A^{t_0}, p_A^{t_1}) := \left(\ln[p_A^{t_0}(\boldsymbol{x}_A^{t_0})] - \ln[p_A^{t_1}(\boldsymbol{x}_A^{t_1})] \right) - \mathcal{Q}_1(\boldsymbol{x}_A, p_1^{t_0})$$
$$(5)$$

I refer to the evolving system of a semi-fixed process as an **independent** system, and refer to it as the **solitary** system in the special case of a solitary process.

To be concrete, we can assume that the master equation of a solitary process is of the form

$$\frac{dp(x_A, x_B)}{dt} = \sum_{v, x_1'} W_{x_A, x_A'}^v(t) \delta(x_B', x_B) p(x_A', x_B') \quad (6)$$

where $W_{x_A', x_A}^m(t)$ is a properly normalized rate matrix, giving the influence on the evolving system of reservoir m at time t. Applying standard arguments to this rate matrix establishes that the expected value of σ_A,

$\int d\boldsymbol{x}_A \, \mathbf{P}(\boldsymbol{x}_A) \sigma_A(\boldsymbol{x}_A, p_A)$, is non-negative in a solitary process [4, 68, 69]. Therefore by Eq. (3), the expected global EP generated by running the joint system is lower-bounded by the expected drop in mutual information, $-\Delta I(X_1; X_2) = -\int d\boldsymbol{x} \mathbf{P}(\boldsymbol{x}) \Delta I(\boldsymbol{x}_A; \boldsymbol{x}_B)$. The data-processing inequality [7] confirms that this lower bound on global EP is non-negative. In fact, typically $-\Delta I(X_1; X_2)$ is strictly positive, which translates to a strictly positive lower bound on expected global EP [45].

B. Entropy production in Bayesian networks

Suppose we have a physical system comprising a finite set $\mathcal{S}$ of $|\mathcal{S}|$ subsystems, and write the (finite) space of that joint system as X, with elements $x = (x_1, x_2, \ldots, x_{|\mathcal{S}|})$. We also have a BN with a set of nodes $\mathcal{V}$, with non-root nodes $V \subset \mathcal{V}$, and root nodes $R := \mathcal{V} \setminus V$. In addition, we have a function $g : \mathcal{V} \to \mathcal{S}$ that maps each $v \in \mathcal{V}$ to one of the subsystems. (See Appendix A in the Supplemental Material at [URL will be inserted by publisher] for some technical details on the required relation between $g(\cdot)$ and the BN's DAG.)

Write the initial distribution of the joint state of the subsystems in terms of the distribution over the root nodes of the BN, as $p^0(x_{g(R)})$ [46]. The set of conditional distributions is $\pi = \{\pi_v(x_{g(v)}|x_{g(\mathrm{pa}[v])}) : v \in V\}$. So the distribution over $X_{g(v)}$ after any non-root node v runs is

$$\sum_{x_{g(R)}} p^0(x_{g(R)}) \left[\left(\sum_{x_{g(\mathrm{Anc}[v])}} \prod_{v' \in \mathrm{Anc}[v]} \pi_{v'}(x_{g(v')}|x_{g(\mathrm{pa}(v'))}) \right) \right.$$
$$\left. \times \pi_v(x_{g(v)}|x_{g(\mathrm{pa}[v])}) \right] \quad (7)$$

where $Anc[v]$ is the set of ancestor nodes of v that are not root nodes.

Below I will often write x_v as shorthand for $x_{g(v)}$. I will write a trajectory through X as $\boldsymbol{x}$, and write the associated trajectory of states of subsystem i as $\boldsymbol{x}_i$, with its value at time t written as $\boldsymbol{x}_i^t$. (So for example, given any node v in the BN, the trajectory of states of subsystem $g(v)$ is written as $\boldsymbol{x}_v$.) I will write p_i^t for the marginal distribution for the subsystem variable i at time t.

Following the analysis in [23], I assume that the continuous-time dynamics of $\mathcal{S}$ implements the conditional distributions of the BN's nodes one after another, in a sequence specified by a topological order of the BN [47]. So whenever some node v in the BN runs, none of the subsystems other than $g(v)$ are allowed to change their state. This allows [23] to assume that for any node v, the associated physical process updating the state of the subsystem $g(v)$ can be treated as a solitary process, where the evolving system is $A = g(v) \cup g(pa[v])$ [48]. I adopt the same convention here [49]. (See Fig. 1 for an illustration of a BN and the sequence of solitary processes it represents.)

In light of this convention, for any $V' \subset V$ I define $[V']$ to mean either the set of nodes $V' \cup \mathrm{pa}[V'] :=$

$\cup_{v \in V'} v \cup \mathrm{pa}[v]$, or the set of associated subsystems, with the context making the meaning clear. In addition, for any $V' \subset V$, I define $-V' := V \setminus V'$. I index the nodes by their (integer-valued) position in the topological order, with an index $v \in \{1 - |R|, \ldots, |V|\}$, so that the non-root nodes start with $v > 0$. I write the distribution over all subsystems after the process implementing node v has run as p^v [35], and unless indicated otherwise, assume that it runs in the time interval $[v - 1, v]$. $\boldsymbol{x}$ is the trajectory of joint states of the subsystems starting at time 0, after the root nodes have been jointly sampled, and for any $v \in V$, I write that segment of $\boldsymbol{x}$ corresponding to the time interval when node v runs as $\boldsymbol{x}^v$. As an example of this notation, $\boldsymbol{x}_{[V']}$ is the full trajectory of the components of $\boldsymbol{x}$ specified by $g([V'])$, i.e., $\boldsymbol{x}_{[V']} = \boldsymbol{x}_{g([V'])} = \boldsymbol{x}_{\{g(v): v \in V' \cup \mathrm{pa}(V')\}}$.

When v runs, the set of subsystems $[v] = g(v) \cup g(\mathrm{pa}(v))$ form a solitary system, while the subsystem $g(v)$ is only an independent system. Accordingly, I write the local EP generated by $[v]$ when node v runs as $\sigma_v(\boldsymbol{x})$, and write the local EP generated by just $g(v)$ when node v runs as

$$\overline{\sigma}_v(\boldsymbol{x}) = \sigma_v(\boldsymbol{x}) + \Delta^v \ln \left[p(\boldsymbol{x}_{\mathrm{pa}(v)} | \boldsymbol{x}_v) \right] \tag{8}$$

where $\Delta^v f[p(\boldsymbol{x})]$ is shorthand for $f[p^v(\boldsymbol{x})] - f[p^{v-1}(\boldsymbol{x})]$. (See Eqs. (2), (4) and (5).)

The expected value of $\overline{\sigma}_v$ can be negative, in contrast to the expected value of σ_v. In addition, while the usual FTs and TURs apply to the EP σ_v, they do not apply to $\overline{\sigma}_v$ in general. For these reasons, I formulate results in terms of σ_v. However, if desired they can be recast in terms of $\overline{\sigma}_v$, by using Eq. (8). (This allows the results of this paper to be connected with those in [23], which involve $\overline{\sigma}_v$ rather than σ_v — see Appendix C in the Supplemental Material at [URL will be inserted by publisher].)

Since EP is cumulative over time, by repeated application of Eq. (3), 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

$$\sigma(\boldsymbol{x}, \pi, p^0) = \sum_{v=1}^{|V|} \left[\sigma_v(\boldsymbol{x}_{[v]}^v, \pi_v, p^{v-1}) \right.$$
$$\left. + \left(I_{p^{v-1}}(\boldsymbol{x}_{[v]}^v; \boldsymbol{x}_{-[v]}^v) - I_{p^v}(\boldsymbol{x}_{[v]}^v; \boldsymbol{x}_{-[v]}^v) \right) \right]$$
$$:= \sum_{v \in V} \sigma_v(\boldsymbol{x}^v) - \Delta I^v(\boldsymbol{x}) \tag{9}$$

This expression for global EP is independent of the topological order, so long as the physical process that implements each conditional distribution π_v is fixed. (See Appendix D in the Supplemental Material at [URL will be inserted by publisher].) So $\langle \sum_{v \in V} -\Delta I^v(\boldsymbol{x}) \rangle$, the sum of expected values of the mutual information drops, provides a lower bound on the expected global EP which is independent of the topological order. This

lower bound is independent of the precise physical system that implements the BN, and typically is strictly positive.

III. FT'S AND TUR'S FOR BAYESIAN NETWORKS

Appendices E and F in the Supplemental Material at [URL will be inserted by publisher] contain proofs of all results in this section, along with some intermediate theorems.

The first result is an FT involving the EPs of the evolving system when each node runs, and the associated changes in mutual information between the states of the evolving and fixed systems:

$$\left\langle \exp \sum_v (\Delta I^v - \sigma_v) \right\rangle = 1 \tag{10}$$

where the expectation is over all trajectories. Eq. (10) means that the larger the sum of the drops in mutual information when the nodes run, the larger must be the EPs generated by running each of those nodes, considered as isolated systems.

In addition we have a set of conditional FTs, which hold for all nodes $v \in V$, for all associated joint values $(\sigma_v, \Delta I^v)$ that occur with nonzero probability under that CTMC:

$$\left\langle \exp \sum_{v' \neq v} (\Delta I_{v'} - \sigma_{v'}) \right\rangle_{\mathbf{P}(.|\sigma_v, \Delta I^v)} = 1 \tag{11}$$

where the subscript indicates that we evaluate the expectation by averaging conditioned on the given pair of values $(\sigma_v, \Delta I^v)$.

Concretely, Eq. (11) concerns the case where we are able to measure the EP generated by the evolving system $[v]$ when node v runs, together with the associated change in mutual information between the state of the evolving and fixed systems. It says that if we average over all instances where that EP and mutual information drop have some specific value, then the associated (exponential of the sum of the) EPs and drops of mutual information when all the *other* nodes besides v are run must average to 1.

As mentioned above, σ_v obeys the usual FT when considered by itself, in isolation from the rest of the subsystems. This gives us a second set of conditional FTs, appropriate for the case where the experimentalist can observe the local EP σ_v but cannot observe the value of ΔI_v and so treat it as a random variable:

$$\left\langle \exp \left(\Delta I_v + \sum_{v' \neq v} (\Delta I_{v'} - \sigma_{v'}) \right) \right\rangle_{\mathbf{P}(.|\sigma_v)} = 1 \tag{12}$$

Eqs. (10) to (12) all generalize to FTs that concern arbitrary subsets of the nodes in the BN and / or arbitrary subsets of all the subsystems. Eqs. (11) and (12)

both generalize to probability distributions conditioned on simultaneous properties of multiple nodes v, not just one.

Also as mentioned above, each σ_v bounds currents generated in the associated evolving system $[v]$, in accordance with the usual TURs. We can exploit this to derive TURs which relate the global EP generated by running an entire BN to the currents in each evolving system when it runs.

As an illustration, slightly generalize our scenario so that the time that each node v takes time to run is τ_v, and therefore each node v starts to run at time $t_{v-1} = \sum_{v' \leq v-1} \tau_{v'}$. Let $j_v^t(\boldsymbol{x})$ be any instantaneous current over X_v at time t [36]. Write the time-integrated current over the associated time interval as $J_v(\boldsymbol{x}) = \int_{t_{v-1}}^{t_v} dt\, j_v^t(\boldsymbol{x})$. Assume that the CTMC does not change during each of the intervals when a node runs (although it will change from one such interval to another, in general). Then we can exploit a recently derived TUR [29] to establish that the expected global EP bounds the precisions of the currents in the subsystems along with the associated drops in mutual information:

$$\langle \sigma \rangle \geq \sum_v \left(\frac{2\tau_v^2 \langle j_v^{t_v} \rangle^2}{\mathrm{Var}(J_v)} - \langle \Delta I^v \rangle \right) \tag{13}$$

This TUR holds without any restrictions on the beginning and ending distributions when each node v runs. However, suppose that for each node v, $p^v = p^{v-1}$, i.e., the beginning and ending distributions of the evolving system are the same. Also allow the rate matrix when node v runs to vary rather than stay constant, so long as it is symmetric about the time $[t^{v-1}, t^v]/2$. In this situation we can exploit the generalized TUR [13, 16] to establish a different TUR for the full BN:

$$\langle \sigma \rangle \geq \sum_v \left(\ln\left[\frac{2\langle J_v \rangle^2}{\mathrm{Var}(J_v)} + 1 \right] - \langle \Delta I^v \rangle \right) \tag{14}$$

If in point of fact every evolving system $[v]$ is in an NESS when node v runs, then we get

$$\langle \sigma \rangle \geq \sum_v \left(\frac{2\langle J_v \rangle^2}{\mathrm{Var}(J_v)} - \langle \Delta I^v \rangle \right) \tag{15}$$

This formula holds even if the distribution of the global system is continually changing, so long as during each solitary process the marginal distribution of the associated solitary system does not change.

IV.   EXAMPLE OF TUR'S FOR A BN

Suppose we have three subsystems, A, B and C, and that the full system evolves according to the BN in Fig. 1, which is physically implemented by the sequence of solitary processes indicated in that figure. So in the first
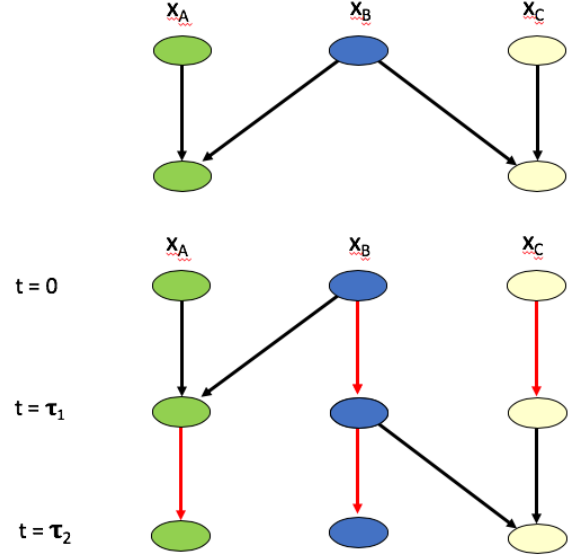

FIG. 1. The upper figure illustrates the BN discussed in Section IV. The lower figure illustrates that same BN, expanded into a sequence of two solitary processes. Red arrows indicate the identity map. In the first solitary process the evolving system is subsystems A and B, while in the second one it is subsystems C and B.

solitary process the evolving system is the composite system AB, where C is the fixed system while in the second solitary process the evolving system is the composite system BC, where A is the fixed system. Let $J_A(.)$ be any net current among subsystem A's states during the first solitary process, and similarly for $J_C(.)$, with $j_A^t(.)$ and $j_C^t(.)$ being the associated instantaneous currents at time t. So using obvious notation, subsystem A gets updated in the Bayes net by the conditional distribution

$$P(x_A^{\tau_1} | x_A^0, x_B^0) = \exp\left(\tau_1 W_A^{x_B^0} \right)\Big|_{x_A^{\tau_1}, x_A^0} \tag{16}$$

and similarly for subsystem C [37].

Plugging into Eq. (13) and then using the fact that x_B never changes,

$$\langle \sigma \rangle \geq \frac{2\tau_1 \langle j_A^{\tau_1} \rangle^2}{\mathrm{Var}(J_A)} + \frac{2\tau_2 \langle j_C^{\tau_1 + \tau_2} \rangle^2}{\mathrm{Var}(J_C)} - \langle \Delta I_A \rangle - \langle \Delta I_C \rangle \tag{17}$$

$$= \frac{2\tau_1 \langle j_A^{\tau_1} \rangle^2}{\mathrm{Var}(J_A)} + \frac{2\tau_2 \langle j_C^{\tau_1 + \tau_2} \rangle^2}{\mathrm{Var}(J_C)}$$
$$+ \Delta S(X_A, X_B, X_C)) - \Delta S(X_A, X_B) - \Delta S(X_C, X_B) \tag{18}$$

where Δ indicates the value of a quantity at $t = \tau_1 + \tau_2$ minus its value at $t = 0$. Eq. (18) provides a trade-off among global EP, three entropies (concerning the distributions at the beginning and the end of the full BN), and the instantaneous currents at the precise moments when each of the two subsystems finishes its update. Note that the RHS of Eq. (18) would not change if the two solitary processes had been run in the reverse order.

Now consider the case where X_B is binary, both matrices W^{x_B} have a (unique) NESS over X_A, but that the NESS differs for the two x_B values. Suppose further that the initial distribution is

$$p^0(x_A, x_B, x_C) = p^0(x_B)p^0(x_A|x_B)\delta(x_C, x_A) \qquad (19)$$

Next, assume that for both values of x_B, $p^0(x_A|x_B)$ is the NESS of the associated rate matrix W^{x_B}. This guarantees that $p^0(x_A, x_B)$ is a NESS during the first solitary process, regardless of $p^0(x_B)$. For completeness, assume that the second solitary process proceeds the same way, just with subsystem C substituted for subsystem A. Therefore we can apply Eq. (15).

Since there are nonzero probability currents in an NESS, there is nonzero probability that the *ending* state of x_A after the first solitary process runs differs from the state of x_C then. However, with probability 1 their initial states were identical. Accordingly, there must be a drop in (expected) mutual information between the evolving and fixed systems of the first solitary process. The same is true for the second solitary process. These two drops in mutual information mean that the global system is not in an NESS throughout the full process. In addition, they increase the RHS of Eq. (15). That increase can either be paid for by an increase in the global EP or by a reduction in the precisions of the two solitary processes.

As a final, technical note, in general it is not possible to implement an arbitrary conditional distribution π_v with a CTMC, without the dynamics being expanded to involve some "hidden" states whose evolution is not specified by π_v [51, 67]. However, as discussed in Appendix G in the Supplemental Material at [URL will be inserted by publisher], this turns out not to affect the applicability of the results above.

V. DISCUSSION

In this paper I derive new FTs and TURs for systems that comprise multiple interacting subsystems. Following [23], I formulate the interactions of those subsystems as a Bayesian network. However, in contrast to [23], I choose the subsystems to ensure that their EP has the conventional thermodynamic properties of EP (e.g., that its expectation is non-negative). This is crucial to the derivation of the TURs. It also allows me to derive conditional FTs, involving probabilities of global EP conditioned on a given EP value of one of the subsystems.

[1] Andre C Barato and Udo Seifert, *Stochastic thermodynamics with information reservoirs*, Physical Review E **90** (2014), no. 4, 042150.

[2] Andre C. Barato and Udo Seifert, *Thermodynamic uncertainty relation for biomolecular processes*, Phys. Rev. Lett. **114** (2015), 158101.

[3] Concha Bielza and Pedro Larrañaga, *Bayesian networks in neuroscience: a survey*, Frontiers in computational neuroscience **8** (2014), 131.

[4] Alexander B Boyd, Dibyendu Mandal, and James P Crutchfield, *Thermodynamics of modularity: Structural costs beyond the landauer bound*, Physical Review X **8** (2018), no. 3, 031036.

[5] Davide Chiuchiù and Simone Pigolotti, *Mapping of uncertainty relations between continuous and discrete time*, Phys. Rev. E **97** (2018), 032109.

[6] Dong-Yeon Cho, Yoo-Ah Kim, and Teresa M Przytycka, *Network biology approach to complex diseases*, PLoS computational biology **8** (2012), no. 12.

[7] Thomas M. Cover and Joy A. Thomas, *Elements of information theory*, John Wiley & Sons, 2012.

[8] Gavin E Crooks, *Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems*, Journal of Statistical Physics **90** (1998), no. 5-6, 1481–1487.

[9] Frank Dondelinger, Sophie Lèbre, and Dirk Husmeier, *Non-homogeneous dynamic bayesian networks with bayesian regularization for inferring gene regulatory networks with gradually time-varying structure*, Machine Learning **90** (2013), no. 2, 191–230.

[10] Massimiliano Esposito, Upendra Harbola, and Shaul Mukamel, *Entropy fluctuation theorems in driven open systems: Application to electron counting statistics*, Phys. Rev. E **76** (2007), 031132.

[11] Massimiliano Esposito and Christian Van den Broeck, *Three detailed fluctuation theorems*, Physical review letters **104** (2010), no. 9, 090601.

[12] _____, *Three faces of the second law. i. master equation formulation*, Physical Review E **82** (2010), no. 1, 011143.

[13] Gianmaria Falasco, Massimiliano Esposito, and Jean-Charles Delvenne, *Unifying thermodynamic uncertainty relations*, arXiv preprint arXiv:1906.11360 (2019).

[14] Nir Friedman, *Inferring cellular networks using probabilistic graphical models*, Science **303** (2004), no. 5659, 799–805.

[15] Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er, *Using bayesian networks to analyze expression data*, Journal of computational biology **7** (2000), no. 3-4, 601–

620.

[16] Yoshihiko Hasegawa and Tan Van Vu, *Generalized thermodynamic uncertainty relation via fluctuation theorem*, arXiv preprint arXiv:1902.06376 (2019).

[17] J.M. Horowitz and T.R. Gingrich, *Thermodynamic uncertainty relations constrain non-equilibrium fluctuations*, Nature Physics (2019).

[18] Jordan M Horowitz and Massimiliano Esposito, *Thermodynamics with continuous information flow*, Physical Review X **4** (2014), no. 3, 031015.

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

[20] Jordan M Horowitz and Juan MR Parrondo, *Designing optimal discrete-feedback thermodynamic engines*, New Journal of Physics **13** (2011), no. 12, 123019.

[21] _____, *Thermodynamic reversibility in feedback processes*, EPL (Europhysics Letters) **95** (2011), no. 1, 10005.

[22] Jordan M. Horowitz and Suriyanarayanan Vaikuntanathan, *Nonequilibrium detailed fluctuation theorem for repeated discrete feedback*, Phys. Rev. E **82** (2010), 061120.

[23] Sosuke Ito and Takahiro Sagawa, *Information thermodynamics on causal networks*, Physical review letters **111** (2013), no. 18, 180603.

[24] _____, *Information flow and entropy production on Bayesian networks*, arXiv:1506.08519 [cond-mat] (2015), arXiv: 1506.08519.

[25] Christopher Jarzynski, *Nonequilibrium equality for free energy differences*, Physical Review Letters **78** (1997), no. 14, 2690.

[26] Daphne Koller and Nir Friedman, *Probabilistic graphical models: principles and techniques*, MIT press, 2009.

[27] Harri Lähdesmäki, Sampsa Hautaniemi, Ilya Shmulevich, and Olli Yli-Harja, *Relationships between probabilistic boolean networks and dynamic bayesian networks as models of gene regulatory networks*, Signal processing **86** (2006), no. 4, 814–834.

[28] Antti Larjo and Harri Lähdesmäki, *Active learning for bayesian network models of biological networks using structure priors*, 2013 IEEE International Workshop on Genomic Signal Processing and Statistics, IEEE, 2013, pp. 78–81.

[29] Kangqiao Liu, Zongping Gong, and Masahito Ueda, *Thermodynamic uncertainty relation for arbitrary initial states*, 2019.

[30] Dibyendu Mandal and Christopher Jarzynski, *Work and information processing in a solvable model of maxwell's demon*, Proceedings of the National Academy of Sciences **109** (2012), no. 29, 11641–11645.

[31] O.J.E. Maroney, *Generalizing landauer's principle*, Physical Review E **79** (2009), no. 3, 031105.

[32] Stephen H Muggleton, *Machine learning for systems biology*, International Conference on Inductive Logic Programming, Springer, 2005, pp. 416–423.

[33] Richard E Neapolitan et al., *Learning bayesian networks*, vol. 38, Pearson Prentice Hall Upper Saddle River, NJ, 2004.

[34] Note in particular that we can take the BN to be a "dynamic BN", i.e., the same DAG and associated conditional distribution, infinitely iterated, with the leaves of each iteration being the roots of the next one, so that the dynamics is open-ended. See the analysis of biochemical adaption with time-delay in [23]) for an example.

[35] Note that $p^0(x_{g(R)})$ must have been sampled before such a node v runs, and that by definition of BN, this means that every subsystem has a well-defined state by the time node v has run.

[36] Here I allow this current to only depend on state transitions that arise due to energy transfers with some subset of the reservoirs that are coupled to the system.

[37] Note that in general there are conditional distributions $P(x_A^{\tau_1}|x_A^0, x_B^0)$ which do not obey Eq. (16) for *any* choice rate matrices $\{W_A^{x_B} : x_B = 0, 1\}$ [51, 67]. So we cannot implement an arbitrary BN using rate matrices that are time-homogeneous during each node's update.

[38] The reason for the name is that the "Landauer cost" of implementing $\pi(\boldsymbol{x}^{t_1}|\boldsymbol{x}^{t_0})$ — the minimal EF needed by any physical process that implements that conditional distribution — is increased by $\mathcal{L}(\boldsymbol{x}, p)$ if we add the requirement that the process obey condition (2) of path-wise subsystem processes.

[39] The special case of Landauer loss for solitary processes when there is a single bath and LDB holds was called "modularity dissipation" in [4]. However, condition (2) was never explicitly stated in that work. See discussion in [69].

[40] To see this, first expand $\mathbf{P}(\boldsymbol{x}) = \mathbf{P}(\boldsymbol{x}(0))P(\boldsymbol{x}^1|\boldsymbol{x}(0))P(\boldsymbol{x}^2|\boldsymbol{x}(1))\ldots$ where $\boldsymbol{x}^k$ means the time interval $[k-1, k]$ and and $\boldsymbol{x}^t$ means the slice of $\boldsymbol{x}$ at time t. Expand $\tilde{\mathbf{P}}(\tilde{\boldsymbol{x}})$ similarly. Then note that by construction, $\mathbf{P}(\boldsymbol{x}(t)) = \tilde{\mathbf{P}}(\tilde{\boldsymbol{x}}(|V|-t))$ for all t. Combine these facts to expand the ratio $\mathbf{P}(\boldsymbol{x})/\tilde{\mathbf{P}}(\tilde{\boldsymbol{x}})$, and then take take its logarithm.

[41] Note that that each solitary process transpires in continuous time, with the BN specifying the associated marginal state distributions at a sequence of discrete times. So the relevant TURs are the ones for continuous time dynamics, *not* the TURs for discrete time dynamics [5, 53].

[42] In the main text I only present integral FTs; in the usual way these integral FTs imply an associated set of detailed FTs [11].

[43] See the appendix in [23] for a succinct summary for physicists.

[44] Note that since $\boldsymbol{x}_B$ does not change, the terms in the rate matrix of the joint system that govern the dynamics of $\boldsymbol{x}_B$ have to be delta functions. In that sense, the usual stochastic thermodynamics formulas for EP rate and EF rate tell us that system B does not contribute to either.

[45] It is straight-forward to extend this analysis of solitary processes to processes where both systems evolve, and even to involve more than two systems co-evolving. In this case the lower bound on expected global EP becomes a drop in multi-information rather than a drop in mutual information. See Appendix B.

[46] For convenience, I put no *a priori* restrictions on this joint distribution. In contrast, conventional BNs require that the distribution over the root nodes be a product of a set of "prior" distributions, one for each of those nodes.

[47] However, for simplicity I modify the formulation in [23] by allowing the subsystems to have their initial states set in parallel, by sampling the joint distribution p^0 over the root nodes at once, rather than require that those nodes be sampled independently, one after the other. Ultimately, this flexibility of allowing the root nodes to be jointly sampled is just a modeling choice. An alternative would be to model the same physical system with an BN

with a different DAG, which included a single extra node that is a shared parent of what in my model is the set R of root nodes.

[48] To see that this assumption is made in [23], note that in Eq. 4 of [23], the two conditional probabilities on the RHS are conditioned only on $\mathcal{B}^{k+1}$. So it is being assumed that the entropy flow into the baths can be written as a function of the state of the physical variable x at the times k and $k + 1$, as well as the (identical) states of the variables corresponding to its parents at those two times, i.e., that condition 2 of a solitary process is being obeyed.

[49] Note that one can*not* when node v runs the subsystem v considered by itself is an independent system, but not a solitary system. The reason is that since the dynamics of $x_{g(v)}$ depends on the value of $x_{g(\text{pa}[v])}$ when v runs, Eq. (4) does not hold for any function $\mathcal{Q}_1$ if $\boldsymbol{x}_A$ is set to $x_{g(v)}$.

[50] Shun Otsubo and Takahiro Sagawa, *Information-thermodynamic characterization of stochastic boolean networks*, arXiv preprint arXiv:1803.04217 (2018).

[51] Jeremy A. Owen, Artemy Kolchinsky, and David H. Wolpert, *Number of hidden states needed to physically implement a given conditional distribution*, New Journal of Modern Physics (2018).

[52] Juan MR Parrondo, Jordan M Horowitz, and Takahiro Sagawa, *Thermodynamics of information*, Nature Physics **11** (2015), no. 2, 131–139.

[53] Karel Proesmans and Christian Van den Broeck, *Discrete-time thermodynamic uncertainty relation*, EPL (Europhysics Letters) **119** (2017), no. 2, 20001.

[54] R. Rao and M. Esposito, *Detailed fluctuation theorems: A unifying perspective*, Santa Fe Institute Press, 2019.

[55] Joshua W Robinson and Alexander J Hartemink, *Non-stationary dynamic bayesian networks*, Advances in neural information processing systems, 2009, pp. 1369–1376.

[56] Takahiro Sagawa, *Thermodynamic and logical reversibilities revisited*, Journal of Statistical Mechanics: Theory and Experiment **2014** (2014), no. 3, P03025.

[57] Takahiro Sagawa and Masahito Ueda, *Second law of thermodynamics with discrete quantum feedback control*, Physical review letters **100** (2008), no. 8, 080403.

[58] _____, *Generalized jarzynski equality under nonequilibrium feedback control*, Physical review letters **104** (2010), no. 9, 090602.

[59] _____, *Generalized Jarzynski equality under nonequilibrium feedback control*, Physical review letters **104** (2010), no. 9, 090602.

[60] _____, *Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics*, Phys. Rev. Lett. **109** (2012), 180602.

[61] _____, *Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics*, Physical review letters **109** (2012), no. 18, 180602.

[62] Udo Seifert, *Stochastic thermodynamics, fluctuation theorems and molecular machines*, Reports on Progress in Physics **75** (2012), no. 12, 126001.

[63] Le Song, Mladen Kolar, and Eric P Xing, *Time-varying dynamic bayesian networks*, Advances in neural information processing systems, 2009, pp. 1732–1740.

[64] Philipp Strasberg, Gernot Schaller, Tobias Brandes, and Massimiliano Esposito, *Quantum and information thermodynamics: A unifying framework based on repeated interactions*, Physical Review X **7** (2017), no. 2, 021003.

[65] Christian Van den Broeck and Massimiliano Esposito, *Ensemble and trajectory thermodynamics: A brief introduction*, Physica A: Statistical Mechanics and its Applications **418** (2015), 6–16.

[66] David H. Wolpert, *Extending Landauer's bound from bit erasure to arbitrary computation*, arXiv:1508.05319 [cond-mat.stat-mech], 2015.

[67] David H. Wolpert, Artemy Kolchinsky, and Jeremy A. Owen, *A space/time tradeoff for implementing a function with master equation dynamics*, Nature Communications (2019).

[68] David Hilton Wolpert and Artemy Kolchinsky, *The thermodynamic costs of circuits*, arXiv preprint arXiv:1806.04103v3 (2018).

[69] D.H. Wolpert, *The stochastic thermodynamics of computation*, Journal of Physics A: Mathematical and Theoretical (2019).

Appendix A: Technical details of the extension of Bayes nets considered in this paper.

Strictly speaking, the models considered here and in [23] are variants of conventional of BNs, which involve a function g that maps nodes to physical subsystems. For this model to give both a complete and consistent representation of the thermodynamics of a set of evolving subsystems, there must be exactly one root node r corresponds to each subsystem, i.e., for all subsystems i, $g(r) = i$ for exactly one $r \in R$. In addition, every non-root node in the DAG must represent an update of the state of one of the physical systems. This means that for each root node r, corresponding to subsystem $i = g(r)$, all other nodes $v' \in g^{-1}(i)$ representing states of that subsystem lie on a single path of connected edges leading out of r. This then implies that there is one (and only one) parent of v in the DAG, v', such that $g(v') = g(v)$; we interpret the value of $x_{g(v')}$ when node v starts to run as the initial state of $X_{g(v)}$ at the beginning of a process updating it, while the value of $x_{g(v)}$ when node v has finished running is the state of $X_{g(v)}$ when that update has completed.

As mentioned in the text, any BN can be implemented by executing the conditional distributions at the nodes of the associated DAG, in a sequence specified by a topological order of the underlying DAG. Here I add the requirement that the BN can in fact be implemented in a topological order where no subsystem's update at a node v depends on an old state of another subsystem which has been overwritten by the time that v is executed. Formally, this requirement means that for all nodes v, there is no $v' \in \text{pa}[v]$ and other node v'' such that: $g(v') = g(v'')$, v' is an ancestor of v'', and v'' occurs before v in the topological order. This ensures that the physical process implementing the BN is Markovian.

Note that not all BNs can be represented in a topological order that respects this last requirement. Most simply, suppose we have two subsystems, X and Y, with states x and y. Suppose as well that under the BN, x^1 de-

pends on both x^0 and y^0, as does y^1. So if we update X first under the topological order, then we cannot properly update Y after that, since x^0 no longer describes the state of X. Similarly, if we update Y first, then we cannot properly update X after that, since y^0 no longer describes the state of Y. So this requirement is actually a restriction on the BNs being considered in this paper.

These requirements are all assumed in [23], implicitly or otherwise. On the other hand, as mentioned in the text, for convenience I relax the standard BN requirement that the distribution over the root nodes be a product distribution, whereas [23] does not.

The model defined above is similar to several graphical models in the literature, including non-stationary dynamic Bayesian networks [55], time-varying dynamic Bayesian networks [63], and non-homogeneous dynamic Bayesian networks [9], among others. Nonetheless, for simplicity, I will simply refer to this structure as a Bayes net, even though that is not exactly accurate.

Appendix B: Path-wise subsystem processes

Write the multi-information of a joint distribution over a set of random variables, $p(X_1, X_2, ...)$, as

$$\mathcal{I}(p) := \left[\sum_i S(p(X_i)) \right] - S(p(X)) \qquad \text{(B1)}$$

Multi-information is also sometimes called "total correlation". Mutual information is the special case of multi-information where there are exactly two random variables. I will use "path" and "trajectory" interchangeably, to mean a function from time into a state space. In the usual way, I use the argument list of a probability distribution to indicate what random variables have been marginalized, e.g., $P(x) = \sum_y p(x, y)$.

Consider a CTMC governing evolution during time interval $[t_0, t_1]$. That CTMC is a **(path-wise) subsystem process** if

1. The subsystems evolve independently of one another, i.e., the discrete-time conditional distribution over the joint state space is

$$\pi(\boldsymbol{x}^{t_1} | \boldsymbol{x}^{t_0}) = \prod_{i \in \mathcal{S}} \pi_i(\boldsymbol{x}_i^{t_1} | \boldsymbol{x}_i^{t_0}) \qquad \text{(B2)}$$

2. There are $|\mathcal{S}|$ functions, $\mathcal{Q}_i(.,.)$, such that the entropy flow (EF) into the joint system during the process if the full system follows trajectory $\boldsymbol{x}$ and the initial joint distribution is p^{t_0} can be written as

$$\mathcal{Q}(\boldsymbol{x}, p^{t_0}) = \sum_{i \in \mathcal{S}} \mathcal{Q}_i(\boldsymbol{x}_i, p_i^{t_0}) \qquad \text{(B3)}$$

for all trajectories $\boldsymbol{x}$ that have nonzero probability under the protocol for initial distribution p^{t_0}.

(Recall that the entire sequence of Hamiltonians and rate matrices is referred to as a "protocol" in stochastic thermodynamics.)

Intuitively, in a subsystem process the separate subsystems evolve in complete isolation from one another, with decoupled Hamiltonians and rate matrices. (See [68, 69] for explicit examples of CTMCs that implement subsystem processes, for the special case where there are two subsystems.)

I use the term **(path-wise, subsystem) Landauer loss** to refer to the extra, unavoidable EP generated by implementing the protocol due to the fact that we do so with a subsystem process:

$$\mathcal{L}(\boldsymbol{x}, p) := \sigma(\boldsymbol{x}, p) - \sum_{i=1}^{N} \sigma_i(\boldsymbol{x}_i, p_i)$$
$$= \mathcal{I}_{p^{t_0}}(\boldsymbol{x}^{t_0}) - \mathcal{I}_{p^{t_1}}(\boldsymbol{x}^{t_1})$$
$$:= -\Delta \mathcal{I}_p(\boldsymbol{x}) \qquad \text{(B4)}$$

where the second line uses condition (2) of path-wise subsystem processes to cancel the EFs [38]. Note that even though $\Delta \mathcal{I}_p(.)$ is written as a function of an entire trajectory, its value only depends and the initial and final states of the trajectory.

Both (expected) subsystem EP and global EP are always non-negative. Moreover, by Eq. (B4), if the expected multi-information among the subsystems decreases in a subsystem process, the Landauer loss must be strictly positive — and so the global EP has a strictly positive lower bound. This is true no matter how thermodynamically efficiently the individual subsystems evolve.

One way to understand this phenomenon is to note that in general the Shannon information stored in the initial statistical coupling among the subsystems will diminish (and maybe disappear entirely) as the process runs. However, for each subsystem i the rate matrix governing how x_i evolves cannot depend on the states of the rest of the subsystems, x_{-i}, due to condition (2) of subsystem processes. So that rate matrix cannot exploit the information in the statistical coupling between the initial states of the subsystems to reduce the total amount of entropy that is produced as the information about the initial coupling of the subsystem states disappears. In contrast, if it were not for condition (2), then the rate matrix governing the dynamics of x_i *could* depend on the value of x_{-i}, and therefore could exploit that value to reduce the amount of entropy that is produced as the information about the initial coupling of the subsystem states disappears. (See [68, 69].)

These results do not require that the underlying process generating trajectories is a CTMC. However, from now on I assume that in fact the dynamics is generated by a CTMC, so that the conventional fluctuation theorems and uncertainty relations all hold.

Appendix C: The differences between the thermodynamic properties of σ_A and $\overline{\sigma}_A$.

There are several specific thermodynamic properties of the EP of the evolving system σ_A in a solitary process that need not hold for the EP of the evolving system $\overline{\sigma}_A$ in a general semi-fixed process. Perhaps the most important is that in keeping with the conventional second law, in a solitary process the expected subsystem EP of the evolving system is non-negative. i.e., $\langle \sigma_A \rangle \geq 0$, whereas $\langle \overline{\sigma}_A \rangle$ can be strictly negative. As an explicit demonstration of such a case where the expected EP can be negative, suppose that the entire joint system evolves in a thermodynamically reversible process, so that $\langle \sigma \rangle = 0$. Then $\mathcal{Q}(\boldsymbol{x}, p^{t_0}) = \ln[p^{t_0}(\boldsymbol{x})] - \ln[p^{t_1}(\boldsymbol{x})]$. Therefore

$$\langle \overline{\sigma}_A \rangle = S_{t_1}(X_A) - S_{t_0}(X_A) - \left[S_{t_1}(X_A, X_B) - S_{t_0}(X_A, X_B) \right]$$
$$= S_{t_0}(X_A | X_B) - S_{t_1}(X_A | X_B) \qquad (C1)$$

A priori, this drop in the conditional entropy of the evolving system's state given the fixed system's state can be positive or negative.

A second important difference arises if we consider the minimal amount of work required to send the ending joint distribution of a semi-fixed process, p_f, back to the initial one, p_i. The difference between the amount of work expended in getting from p_i to p_f in the first place and this minimal amount of work to go back is sometimes called the "dissipated" work in going from p_i to p_f, because it is the minimal amount of work lost to the heat bath if one were to run a full cycle $p_i \rightarrow p_f \rightarrow p_i$. Much of the stochastic thermodynamics literature presumes that dissipated work can be treated as a synonym for EP. In agreement with this, the dissipated work in a solitary process always equals the expected subsystem EP, $\langle \sigma_A \rangle$. In contrast, dissipated work does not equal $\langle \overline{\sigma}_i \rangle$ in semi-fixed processes, in general.

Another difference is that in a solitary process the conventional FT holds for the evolving system with state space X_A, considered in isolation from the fixed system, if the EP in that FT is identified as $\sigma_A(\boldsymbol{x}_1, p_1)$. However, in general the conventional FT will not hold in general for a semi-fixed process with state space X_A in an arbitrary semi-fixed process, if the EP in that FT is identified as $\overline{\sigma}_A(\boldsymbol{x}, p)$.

As a final example, the expected EP of the evolving system in a solitary process bounds the precision of any current defined over the state of the subsystem, in the usual way given by the thermodynamic uncertainty relations [13]. However, the expected EP of the evolving system in a semi-fixed process need not have so simple a relationship with the current in that subsystem.

Example 1. *Consider a process involving three subsystems, A, B and C. Only subsystem A changes its state in this process, and the dynamics of subsystem A depends on the state of subsystem B, but not on the state of subsystem C. We can formulate this process as a semi-fixed process where either subsystem A or the joint subsystem AB is the evolving system.*

*Note though that we can*not *identify subsystem A as the evolving system of a solitary process, with the joint subsystem BC being the fixed subsystem. (Since the evolution of subsystem A depends on the state of subsystem B, condition (1) would be violated.)*

On the other hand, the situation is not so clear-cut if we ask whether the process is a solitary process with AB the evolving system. If the joint subsystem AB is physically decoupled from subsystem C, with no interaction Hamiltonian coupling C to the other subsystems, and no coupling of C with AB in the rate matrix for the full system ABC, then we have a solitary process, with the evolving system identified as the joint subsystem AB. However, as an alternative, we could run the entire process in a way that is globally thermodynamically reversible, incurring zero global EP. (N.b., in general this would require an interaction Hamiltonian coupling C to the other subsystems, and require that the rate matrix for the full system ABC couples the dynamics of C to that of AB.) In this case, the bound in Eq. (3) would be violated in general if we identify AB as the evolving system (e.g., if the expected drop in mutual information is strictly nonzero). So the overall process would not a solitary process with AB the evolving system.

This demonstrates that in general, just specifying the joint dynamics of a co-evolving set of subsystems does not determine whether we can view a particular physical process that implements that dynamics as a solitary process, for some appropriately identified evolving system [39].

Appendix D: Independence of global EP from the topological order of the nodes in the BN

By hypothesis, for all $v \in V$, the physical process that implements the conditional distribution $\pi_v(x_{g(v)} | x_{g(\mathrm{pa}(v))})$ is the same in both topological orders. So changing the topological order doesn't change any of the values σ_v. Therefore to establish the claim, we need to establish that $-\sum_{v \in V} \Delta I^v(\boldsymbol{x})$ is independent of the topological order.

To do that, given a topological order, label the nodes $v \in V$ in the sequence they occur in that topological order as $1, 2, \ldots |V|$, so that they occur in corresponding time intervals $[0, 1], [1, 2], \ldots, [|V|-1, |V|]$. (Note that in general it may be that more than one of those nodes change the state of the same subsystem.) Express $g(.)$ accordingly. Next, use the fact that while the marginal entropy of the evolving system changes during a solitary process, the marginal entropy of the semi-fixed system doesn't change, to expand

$$
\begin{aligned}
-\sum_{v\in V}\Delta I^v(\boldsymbol{x}) &= -\Big[\big(\ln(p^0_{g(1)}(\boldsymbol{x}))+\ln(p^0_{-g(1)}(\boldsymbol{x}))-\ln(p^0(\boldsymbol{x}))\big)-\big(\ln(p^1_{g(1)}(\boldsymbol{x}))+\ln(p^0_{-g(1)}(\boldsymbol{x}))-\ln(p^1(\boldsymbol{x}))\big)\Big]\\
&\quad -\Big[\big(\ln(p^1_{g(2)}(\boldsymbol{x}))+\ln(p^1_{-g(2)}(\boldsymbol{x}))-\ln(p^1(\boldsymbol{x}))\big)-\big(\ln(p^2_{g(2)}(\boldsymbol{x}))+\ln(p^1_{-g(2)}(\boldsymbol{x}))-\ln(p^2(\boldsymbol{x}))\big)\Big]\\
&\quad -\dots\\
&\quad -\Big[\big(\ln(p^{|V|-1}_{g(|V|)}(\boldsymbol{x}))+\ln(p^{|V|-1}_{-g(|V|)}(\boldsymbol{x}))-\ln(p^{|V|-1}(\boldsymbol{x}))\big)-\big(\ln(p^{|V|}_{g(|V|)}(\boldsymbol{x}))+\ln(p^{|V|-1}_{-g(|V|)}(\boldsymbol{x}))-\ln(p^{|V|}(\boldsymbol{x}))\big)\Big]\\
&= \ln(p^0(\boldsymbol{x}))-\ln(p^{|V|}(\boldsymbol{x}))-\sum_v \ln(p^{v-1}_{g(v)}(\boldsymbol{x}))-\ln(p^v_{g(v)}(\boldsymbol{x}))\\
&= \ln(p^0(\boldsymbol{x}))-\ln(p^{|V|}(\boldsymbol{x}))-\sum_i\sum_{v\in g^{-1}(i)} \ln(p^{v-1}_{g(v)}(\boldsymbol{x}))-\ln(p^v_{g(v)}(\boldsymbol{x}))\\
&= \ln(p^0(\boldsymbol{x}))-\ln(p^{|V|}(\boldsymbol{x}))-\sum_i \ln(p^0_i(\boldsymbol{x}))-\ln(p^{|V|}_i(\boldsymbol{x})) \qquad\text{(D1)}
\end{aligned}
$$

where the sums in last two equations are over the set of subsystems, and for simplicity I assume that each subsystem i occurs in at least one node, i.e., for all subsystems i, $g^{-1}(i)\neq\varnothing$.

The RHS of Eq. (D1) is fully specified by the combination of the BN and the initial distribution. So it cannot depend on the topological order. This establishes the claim.

As an aside, note that the expected value of the RHS of Eq. (D1) can be written as the drop from the beginning to the end of the BN in the total correlation (also called multi-information) among all the subsystems,

$$
\left[\sum_i S(p_i)\right]-S(p) \qquad\text{(D2)}
$$

Total correlation is an extension of mutual information to more than two systems. So loosely speaking, Eq. (D1) says that the minimal expected global EP is the drop in "how much information the subsystems jointly provide about one another" from the beginning to the end of the full BN.

Appendix E: Derivations of fluctuation theorems for Bayesian networks

Let $\tilde{\boldsymbol{x}}$ indicate the time-reversal of the trajectory $\boldsymbol{x}$. (For simplicity, I restrict attention to spaces X_i whose elements are invariant under time-reversal.) Let $\mathbf{P}(\boldsymbol{x})$ indicate the probability (density) of $\boldsymbol{x}$ under the forward protocol running the entire BN. Let $\tilde{\mathbf{P}}(\boldsymbol{x})$ indicate the probability of the same trajectory if we run the protocol in time-reversed order, where the ending distribution over X under $\mathbf{P}$ is the same as the starting distribution under $\tilde{\mathbf{P}}$. Also write $\tilde{\boldsymbol{x}}^v$ to indicate the time-reversal of the trajectory segment $\boldsymbol{x}^v$.

In the next subsection, I derive fluctuation theorems concerning probabilities of trajectories, and in the following subsection, I derive fluctuation theorems concerning the joint probability that each of the subsystem EPs has some associated specified value.

1. Fluctuation theorems for trajectories

Plugging Eq. (9) into the usual detailed fluctuation theorem (DFT) [65] gives the DFT,

$$
\ln\left[\frac{\mathbf{P}(\boldsymbol{x})}{\tilde{\mathbf{P}}(\tilde{\boldsymbol{x}})}\right]=\sum_{v=1}^{|V|}\sigma_v(\boldsymbol{x}^v)-\Delta I^v(\boldsymbol{x}) \qquad\text{(E1)}
$$

for all $\boldsymbol{x}$ with nonzero probability under $\mathbf{P}$. Exponentiating both sides of Eq. (E1) and then integrating results in the integrated fluctuation theorem (IFT),

$$
\begin{aligned}
\left\langle e^{-\sum_{v=1}^{|V|}\sigma_v-\Delta I^v}\right\rangle &:= \int d\boldsymbol{x}\,\mathbf{P}(\boldsymbol{x})e^{\left(-\sum_{v=1}^{|V|}\sigma_v(\boldsymbol{x}^v)-\Delta I^v(\boldsymbol{x})\right)}\\
&= 1 \qquad\text{(E2)}
\end{aligned}
$$

In addition to applying to runs of the entire BN, the usual DFT applies separately to successive time intervals, i.e., to each successive interval at which exactly one node and its parents co-evolve as that node's conditional distribution is executed. Therefore for all v,

$$
\ln\left[\frac{\mathbf{P}(\boldsymbol{x}^v)}{\tilde{\mathbf{P}}(\tilde{\boldsymbol{x}}^v)}\right]=\sigma_v(\boldsymbol{x}^v)-\Delta I^v(\boldsymbol{x}) \qquad\text{(E3)}
$$

which results in an IFT analogous to Eq. (E2).

Combining Eqs. (E1) and (E3) gives

$$
\mathcal{I}(\mathbf{P}(\boldsymbol{x}))=\mathcal{I}(\tilde{\mathbf{P}}(\tilde{\boldsymbol{x}})) \qquad\text{(E4)}
$$

where I define

$$
\mathcal{I}(\mathbf{P}(\boldsymbol{x})):=\ln\left[\frac{\mathbf{P}(\boldsymbol{x})}{\prod_{v=1}^{|V|}\mathbf{P}(\boldsymbol{x}^v)}\right] \qquad\text{(E5)}
$$

and similarly for $\mathcal{I}(\tilde{\mathbf{P}}(\tilde{\boldsymbol{x}}))$. Note that Eq. (E4) can be rewritten as

$$\frac{\mathbf{P}(\boldsymbol{x})}{\tilde{\mathbf{P}}(\tilde{\boldsymbol{x}})} = \prod_v \frac{\mathbf{P}(\boldsymbol{x}^v)}{\tilde{\mathbf{P}}(\tilde{\boldsymbol{x}}^v)} \tag{E6}$$

(This equality can also be derived directly, without invoking DFTs [40].)

The quantity $\mathcal{I}(\mathbf{P}(\boldsymbol{x}))$ defined in Eq. (E5) is an extension of multi-information to concern probabilities of entire trajectories of the joint system. So loosely speaking, Eq. (E4) means that the amount of information that the set of all the trajectory segments $\{\boldsymbol{x}^v\}$ provide about one another (under $\mathbf{P}$) equals the amount of information that the set of all the trajectory segments $\{\tilde{\boldsymbol{x}}^v\}$ provide about one another (under $\tilde{\mathbf{P}}$). (Note that the same subsystem may evolve in more than trajectory segment.) In this sense, there is no arrow of time, as far as probabilities of trajectory segments are concerned.

We can also combine Eqs. (E1) and (E3) to derive DFTs and IFTs involving conditional probabilities, in which the trajectories of one or more of the subsystems are given. To illustrate this, pick any $V' \subset V$, and use Eq. (E3) to evaluate the terms on the RHS of Eq. (E1) for all of the nodes $v \in V'$. Define $\boldsymbol{x}^{V'} := \{\boldsymbol{x}^{v'} : v' \in V'\}$, i.e., the "partial trajectory" given by all segments $v' \in V'$ of the trajectory $\boldsymbol{x}$. Then after clearing terms and using Eq. (E6), we get the following conditional DFT, which must hold for all partial trajectories $\boldsymbol{x}^{V'}$ with nonzero probability under $\mathbf{P}$:

$$\ln\left[\frac{\mathbf{P}\left(\boldsymbol{x}|\boldsymbol{x}^{V'}\right)}{\tilde{\mathbf{P}}\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{x}}^{V'}\right)}\right] = \mathcal{I}(\tilde{\mathbf{P}}(\tilde{\boldsymbol{x}}^{V'})) - \mathcal{I}(\mathbf{P}(\boldsymbol{x}^{V'})) + \sum_{v \in V \setminus V'} \sigma_v(\boldsymbol{x}^v) - \Delta I^v(\boldsymbol{x}) \tag{E7}$$

In turn, Eq. (E7) gives the following conditional IFT which must hold for each partial trajectory $\boldsymbol{x}^{V'}$ with nonzero probability under $\mathbf{P}$:

$$\left\langle e^{\left(\mathcal{I}^{V'} - \tilde{\mathcal{I}}^{V'} + \sum_{v \in V \setminus V'} (\Delta I_v - \sigma_v)\right)} \right\rangle_{\mathbf{P}(.|\boldsymbol{x}^{V'})} = 1 \tag{E8}$$

where $\mathcal{I}^{V'}$ is shorthand for the random variable $\mathcal{I}(\mathbf{P}(\boldsymbol{x}^{V'}))$ and similarly for $\tilde{\mathcal{I}}^{V'}$. (Note that all terms in the exponent in Eq. (E8) are defined in terms of the full joint distributions $\mathbf{P}(\boldsymbol{x})$), but that $\boldsymbol{x}$ is averaged according to $\mathbf{P}(\boldsymbol{x}|\boldsymbol{x}^{V'})$.)

Note that in addition to these results which hold when considering the entire system X, since each subsystem process is a solitary process, the usual DFT and IFT must hold for each subsystem v considered in isolation, in the interval during which it runs. So for example,

$$\ln\left[\frac{\mathbf{P}\left(\boldsymbol{x}^v_{[v]}\right)}{\tilde{\mathbf{P}}\left(\tilde{\boldsymbol{x}}^v_{[v]}\right)}\right] = \sigma_v(\boldsymbol{x}^v) \tag{E9}$$

(Compare to Eq. (E3).) Eq. (E9) gives us an additional set of conditional DFTs and IFTs. For example, it gives the following variant of Eq. (E7)

$$\ln\left[\frac{\mathbf{P}\left(\boldsymbol{x}|\boldsymbol{x}^v_{[v]}\right)}{\tilde{\mathbf{P}}\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{x}}^v_{[v]}\right)}\right] = -\Delta I_v(\boldsymbol{x}) + \sum_{v' \neq v} \left(\sigma_{v'}(\boldsymbol{x}^{v'}) - \Delta I_{v'}(\boldsymbol{x})\right) \tag{E10}$$

and the following variant of Eq. (E8)

$$\left\langle e^{-I_v + \sum_{v' \neq v}(\Delta I_{v'} - \sigma_{v'})} \right\rangle_{\mathbf{P}(.|\boldsymbol{x}^v_{[v]})} = 1 \tag{E11}$$

Note that the numerator of the expression inside the logarithm on the LHS of Eq. (E10) is a distribution conditioned on the joint trajectory of (the subsystem corresponding to) node v and its parents when node v runs. In contrast, the numerator inside the logarithm on the LHS of Eq. (E7) is a distribution conditioned on the joint trajectory of *all* of the subsystems when node v runs (not just the joint trajectory of v and its parents).

2. Fluctuation theorems for EP

We can use the DFTs of the previous subsection, which concern probabilities of trajectories, to construct "joint DFTs", which instead concern probabilities of vectors of the joint amounts of EP generated by all of the subsystems. (See Sec. 6 in [65]).

To begin, define $\tilde{\Delta}I_v(\tilde{\boldsymbol{x}}) := -\Delta I^v(\boldsymbol{x})$. Similarly define

$$\tilde{\sigma}_v(\tilde{\boldsymbol{x}}) := \ln\left[\frac{\tilde{\mathbf{P}}(\tilde{\boldsymbol{x}}^v_{[v]})}{\mathbf{P}(\boldsymbol{x}^v_{[v]})}\right] \tag{E12}$$

In the special case that $p^v = p^{v-1}$, we can rewrite this as $\sigma_v(\tilde{\boldsymbol{x}}_{[v]}, \tilde{\pi}_v, p^v)$, the EP generated by running (the part of the protocol that implements) the conditional distribution at node v backwards in time, starting from the distribution over $X_{[v]}$ that is the *ending* distribution when node v is implemented going forward in time. We cannot rewrite it that way in general though; see discussion of Eq. 85 in [65].

Using this notation, for any set of real numbers $\{\alpha_v, \gamma_v : v \in V\}$,

$$\mathbf{P}\left(\boldsymbol{x} : \left\{\sigma_v(\boldsymbol{x}_{[v]}) = \alpha_v, \Delta I^v(\boldsymbol{x}) = \gamma_v : v \in V\right\}\right) = \int d\boldsymbol{x} P(\boldsymbol{x}) \prod_v \delta\left(\sigma_v(\boldsymbol{x}^v_{[v]}) - \alpha_v\right) \delta\left(\Delta I^v(\boldsymbol{x}) - \gamma_v\right)$$

$$= e^{\sum_v \alpha_v - \gamma_v} \int d\boldsymbol{x} \tilde{\mathbf{P}}(\tilde{\boldsymbol{x}}) \prod_v \delta\left(\sigma_v(\boldsymbol{x}_{[v]}^v) - \alpha_v\right) \delta\left(\Delta I^v(\boldsymbol{x}) - \gamma_v\right)$$

$$= e^{\sum_v \alpha_v - \gamma_v} \int d\boldsymbol{x} \tilde{\mathbf{P}}(\tilde{\boldsymbol{x}}) \prod_v \delta\left(\ln\left[\frac{\mathbf{P}(\boldsymbol{x}_{[v]}^v)}{\tilde{\mathbf{P}}(\tilde{\boldsymbol{x}}_{[v]}^v)}\right] - \alpha_v\right) \delta\left(\Delta I^v(\boldsymbol{x}) - \gamma_v\right) \tag{E13}$$

$$= e^{\sum_v \alpha_v - \gamma_v} \int d\tilde{\boldsymbol{x}} \tilde{\mathbf{P}}(\tilde{\boldsymbol{x}}) \prod_v \delta\left(\ln\left[\frac{\tilde{\mathbf{P}}(\tilde{\boldsymbol{x}}_{[v]}^v)}{\mathbf{P}(\boldsymbol{x}_{[v]}^v)}\right] + \alpha_v\right) \delta\left(\tilde{\Delta} I_v(\tilde{\boldsymbol{x}}) + \gamma_v\right)$$

$$= e^{\sum_v \alpha_v - \gamma_v} \tilde{\mathbf{P}}\left(\left\{\tilde{\sigma}_v(\tilde{\boldsymbol{x}}_{[v]}) = -\alpha_v, \tilde{\Delta} I_v(\tilde{\boldsymbol{x}}) = -\gamma_v : v \in V\right\}\right) \tag{E14}$$

We can write Eq. (E14) more succinctly as

$$\ln\left[\frac{P(\{\sigma_v = \alpha_v, \Delta I^v = \gamma_v : v \in V\})}{\tilde{P}\left(\left\{\tilde{\sigma}_v = -\alpha_v, \tilde{\Delta} I_v = -\gamma_v : v \in V\right\}\right)}\right] = \sum_{v'} \alpha_{v'} - \gamma_{v'} \tag{E15}$$

or just

$$\ln\left[\frac{P(\{\sigma_v, \Delta I^v\})}{\tilde{P}\left(\left\{-\tilde{\sigma}_v, -\tilde{\Delta} I_v\right\}\right)}\right] = \sum_{v'} (\sigma_{v'} - \Delta I_{v'}) \tag{E16}$$

for short. (Since the arguments of the probabilities in these equations are not full trajectories, I am indicating those probabilities with P rather than the density function $\mathbf{P}$.) This establishes Eq. (10) of the main text:

$$\langle e^{\sum_v \Delta I^v - \sigma_v}\rangle := 1 \tag{E17}$$

In addition to Eq. (E16), which concerns the entire BN, the conventional extension of the DFT must hold separately for the time interval when each evolving system $[v] = \text{pa}[v] \cup \{v\}$ runs:

$$\ln\left[\frac{P(\sigma_v, \Delta I_v)}{\tilde{P}\left(-\tilde{\sigma}_v, -\tilde{\Delta} I_v\right)}\right] = \sigma_v - \Delta I^v \tag{E18}$$

Combining Eqs. (E16) and (E18) establishes that

$$\frac{P(\{\sigma_v, \Delta I^v\})}{\tilde{P}\left(\left\{-\tilde{\sigma}_v, -\tilde{\Delta} I_v\right\}\right)} = \prod_v \frac{P(\sigma_v, \Delta I^v)}{\tilde{P}\left(-\tilde{\sigma}_v, -\tilde{\Delta} I_v\right)} \tag{E19}$$

(Note that it is *not* true that $P(\{\sigma_v, \Delta I^v\}) = \prod_v P(\sigma_v, \Delta I^v)$ in general.) Eq. (E19) should be compared to Eq. (E4).

Taking logarithms of both sides of Eq. (E19) gives

$$\mathcal{I}(P(\{\sigma_v, \Delta I^v\})) = \mathcal{I}(\tilde{P}(\{-\tilde{\sigma}_v, -\tilde{\Delta} I_v\})) \tag{E20}$$

Loosely speaking, Eq. (E20) equates two "amounts of information". One is the amount of information that the set of all pairs, {EP generated by running a particular subsystem, associated drop in mutual information between that subsystem's state and all other variables in the full system} provide about one another. The other is the amount of information that the set of all

pairs, {EP generated by running a particular subsystem time-reversed, associated gain in mutual information between that subsystem's state and all other variables in the full system} provide about one another.

Combining Eqs. (E16) and (E18) also gives a set of conditional fluctuation theorems, analogous to Eqs. (E7) and (E8), only conditioning on values of EP and drops in mutual information rather than on components of a trajectory. For example, subtracting Eq. (E18) from Eq. (E16) gives the conditional DFT,

$$\ln\left[\frac{P(\{\sigma_{-v}, \Delta I_{-v}\} \mid \sigma_v, \Delta I_v)}{\tilde{P}\left(\left\{-\tilde{\sigma}_{-v}, -\tilde{\Delta} I_{-v}\right\} \mid -\tilde{\sigma}_v, -\tilde{\Delta} I_v\right)}\right] = \sum_{v' \neq v} (\sigma_{v'} - \Delta I_{v'}) \tag{E21}$$

which must hold for all pairs $(\sigma_v, \Delta I^v)$ that have nonzero probability under P. This in turn establishes Eq. (11) in the main text:

$$\left\langle e^{\sum_{v' \neq v}(\Delta I_{v'} - \sigma_{v'})}\right\rangle_{P(.|\sigma_v, \Delta I^v)} = 1 \tag{E22}$$

which must hold for all $(\sigma_v, \Delta I^v)$ with nonzero probability under P.

As usual, since each subsystem process is a solitary process the usual DFTs and IFTs must hold for each subsystem v considered in isolation, in the interval during which it runs. So for example,

$$\ln\left[\frac{P(\sigma_v)}{\tilde{P}(-\tilde{\sigma}_v)}\right] = \sigma_v \tag{E23}$$

(Compare to Eq. (E18).) Combining Eqs. (E16) and (E23) gives us an additional set of DFTs and IFTs. For example, it gives the following variant of Eq. (E21):

$$\ln\left[\frac{P(\{\sigma_{-v}, \Delta I_{-v}\} \mid \sigma_v)}{\tilde{P}\left(\left\{-\tilde{\sigma}_{-v}, -\tilde{\Delta} I_{-v}\right\} \mid -\tilde{\sigma}_v\right)}\right] = -\Delta I_v + \sum_{v' \neq v} (\sigma_{v'} - \Delta I_{v'}) \tag{E24}$$

This immediately gives Eq. (12) in the main text.

Appendix F: TUR's for Bayes nets

In this subsection I show how to combine Eq. (9) with several thermodynamic uncertainty relations to derive

bounds on the precision of time-integrated currents in BNs [41].

A real-valued function $J(\boldsymbol{x})$ is called a time-integrated **current** if it is time-antisymmetric, i.e., if $J(\tilde{\boldsymbol{x}}) = -J(\boldsymbol{x})$. Since $\boldsymbol{x}$ is a random variable, so is $J(\boldsymbol{x})$. The **precision** of J is defined as $\langle J \rangle_{\mathbf{P}}^2 / \mathrm{Var}_{\mathbf{P}}(J)$, and measures the average size of the fluctuations in the value of $J(\boldsymbol{x})$. Note that there are an infinite number of current functions $J(.)$. Nonetheless, recently there has been a flurry of results in the literature that upper-bound the precision of every current $J(\boldsymbol{x})$ with an J-independent function of $\sigma[\mathbf{P}]$, the expected EP during the process that generates the trajectories $\boldsymbol{x}$ [13]. These results — called "thermodynamic uncertainty relations" (TURs) — mean that we cannot increase the precision of any current beyond a certain point without paying for it by increasing EP. Alternatively, they mean that if we can experimentally measure the precision of a current, then we can lower-bound the sum of all contributions to *EP* that are not directly experimentally measurable.

The TURs differ from one another in the restrictions they impose on $\mathbf{P}$. As an example, if the system is in a NESS when it runs, any associated current $J(\boldsymbol{x})$ obeys the bound [19],

$$\frac{2\langle J \rangle^2}{\mathrm{Var}(J)} \le \sigma \qquad (\mathrm{F1})$$

where from now on I will often leave the density function $\mathbf{P}$ implicit.

More generally, suppose only that the the starting and ending distributions over X are identical, and that the driving protocol is time-symmetric, i.e., both the trajectories of Hamiltonians and the trajectory of rate matrices are invariant if we replace all times t with $1 - t$. Under such circumstances,

$$\frac{2\langle J \rangle^2}{\mathrm{Var}(J)} \le e^{\sigma} - 1 \qquad (\mathrm{F2})$$

which is known as the "generalized thermodynamic uncertainty relation" (GTUR [16]), and being more general, is weaker than the finite time thermodynamic uncertainty relation.

More recently, a variant of the TURs was derived which upper-bounds the instantaneous current at the end of the process rather than the integrated current across the entire process [29]. Choose units of time so that the duration of the full process is τ, and write the ending instantaneous current as $j_\tau(\boldsymbol{x}) = \sum_{x \ne x'} W_x^{x'}(t)\mathbf{P}(\boldsymbol{x}_t')d_{x',x}$, where $d_{x,x'}$ is any antisymmetric matrix, $W(t)$ is the rate matrix of the underlying CTMC at time t, and $\mathbf{P}(\boldsymbol{x}_t')$ is the probability that the state at time t is $\boldsymbol{x}_t'$. Then this new TUR says that if in fact the rate matrix and Hamiltonian are both constant during the process, then

$$\frac{2\tau\langle j_\tau \rangle^2}{\mathrm{Var}(J)} \le \sigma \qquad (\mathrm{F3})$$

(Note that this holds regardless of the forms of the beginning and ending distributions.)

In light of these results, consider a BN formulated as a sequence of solitary processes, where the protocol of each separate solitary process is time-symmetric about the middle of the interval in which that solitary process takes place. Assume as well that the beginning marginal distribution of every joint subsystem $[v]$ when node v begins to run is the same as the ending marginal distribution of that joint subsystem after that node finishes running. (Formally, this means that for all $v \in V$, $P(x_{[v]}^{v-1}) = P(x_{[v]}^{v})$.)

Since we're implementing the BN as a sequence of solitary processes, this assumption of invariant joint marginals over set of nodes $[v]$ means that the marginal distribution over the state of any one subsystem i in the BN, $P(x^i)$, is the same at the beginning of the running of the entire BN as at the end of the running of the entire BN. This is true even if that variable corresponds to multiple nodes in the Bayes net, i.e., if $g^{-1}(i)$ contains more than one element of V. However, in general for any node v in the BN, the *joint* distribution over the states of $[v]$, $P(x_{g[[v]]})$, can differ arbitrarily between the beginning and the end of the running of the entire BN, since each of the subsystems in $[v]$ can also change state in the implementation of some other nodes of the Bayes net besides v.

Applying first Eq. (9) and then Eq. (F2) to this scenario, we get

$$\sigma[\boldsymbol{P}] = \sum_v \left(\sigma_v[\mathbf{P}_v] - \Delta I^v(\mathbf{P}) \right)$$

$$\ge \sum_v \left(\ln\left[\frac{2\langle J_v \rangle^2}{\mathrm{Var}(J_v)} + 1 \right] - \Delta I^v(\mathbf{P}) \right) \qquad (\mathrm{F4})$$

where the random variable J_v is any time-asymmetric function of the components of the trajectory segment $\boldsymbol{x}^v$ specified by $[v]$. In the special case that each subsystem v is actually in a NESS when it runs, we can apply Eq. (F1) to establish the stronger bound,

$$\sigma[\boldsymbol{P}] \ge \sum_v \left(\frac{2\langle J_v \rangle^2}{\mathrm{Var}(J_v)} - \Delta I^v(\mathbf{P}) \right) \qquad (\mathrm{F5})$$

Finally, if in fact the process is time-homogeneous, then no matter what the beginning and ending distributions are, we can apply Eq. (F4) to establish

$$\sigma[\boldsymbol{P}] \ge \sum_v \left(\frac{2\langle \tau_v j_v^{t_v} \rangle^2}{\mathrm{Var}(J_v)} - \Delta I^v(\mathbf{P}) \right) \qquad (\mathrm{F6})$$

where the duration of the process updating node v is τ_v, t_v is the time that that process ends, $j_v^{t_v}$ is the instantaneous current over $X_{[v]}$ evaluated at t_v, and $J_v = \int_{t_v - \tau_v}^{t_v} dt\, j_v^t$.

Eqs. (F4) to (F6) illustrate a trade-off among the precisions of (instantaneous) currents of the various subsystems, the sum of the drops in mutual information, and the total dissipated work of the joint system.

Example 2. *Return to the NESS scenario considered at the end of Section IV. Consider implementing that same BN — but with a different physical process. Just like the process described above, this alternative process would first update the state of X_A, and then when that was done it would update the state of X_C. However, those two updates would not be done with solitary processes. Instead, the state of X_A would be updated with a CTMC whose rate matrix evolves x_A based on the current state of all three variables, x_A, x_B, and x_C. (In contrast, as discussed in Appendix B in the Supplemental Material at [URL will be inserted by publisher], if x_{AB} runs with a solitary process, then the associated rate matrix can only involve x_A and x_B.) This allows the CTMC to exploit the initial coupling of x_A and x_C in order to reduce the total EP that is generated by updating x_A. Similarly, the state of X_C would be updated with a CTMC whose rate matrix evolves x_C based on the then-current state of all three variables, x_A, x_B, and x_C, and thereby reduce the total EP generated by updating x_C. The end result is that Eq. (15) would still hold, only with the ΔI^v terms removed. Since those terms are both negative, this would (in theory) allow the process to generate the same global EP as the original process but with greater precisions of both of the currents.*

Appendix G: The role of hidden states

In general it is not possible to implement an arbitrary discrete-time conditional distribution $\pi_v(x_v^{v+1}|x_v^v)$ using a CTMC over $X_{g([v])}$. Instead, in order to use a CTMC to model the dynamical process that results in that conditional distribution, there must be a set of extra "hidden" states of subsystem $g(v)$, in addition to the "visible" states in $X_{g(v)}$, and the CTMC must couple those two sets of states when it runs that conditional distribution to update that subsystem [51, 67].

However, at both $t = v - 1$ and $t = v$, the beginning and end of when node v runs, the state of subsystem $g(v)$ must be visible, i.e., it must lie in $X_{g(v)}$ at those two times. (If that weren't the case, then we could not be sure that the discrete time dynamics is actually given by $\pi_v(x_v^{v+1}|x_v^v)$ operating on $p^v(x_{\mathrm{pa}[v]}^v)$.) Accordingly, any hidden states can be ignored in calculating the drop in mutual information as node v runs. For the same reason, the change in $\ln[p^t(\boldsymbol{x})]$ from $t = v$ to $t = v + 1$ doesn't depend on whether hidden states exist. So the only effect of such states on Eq. (5) is to modify the EF function of the process updating node v, $\mathcal{Q}_v(.,.)$ (and similarly for Eq. (1)). Therefore their only effect on Eq. (9) — which is the starting point for the results in this paper concerning solitary process formulations of BNs — is to change the EF function implicitly defining $\sigma_A(.,.)$. Since the detailed form of the EF function is irrelevant for the results in this paper in the way that they are stated (the EF function for updating node v is subsumed in the function $\sigma_v(.,.)$), we can ignore hidden states for the purposes of this paper.