Many systems are naturally modeled as multiple co-evolving subsystems. In this paper I analyze the stochastic thermodynamics of such systems, formalizing them as multipartite processes. In many multipartite processes the rate matrix of each subsystem $i$ only depends on a proper subset of the remaining subsystems, known as the neighborhood of $i$. The intersections among the neighborhoods of the different subsystems forms a network. Here I derive fluctuation theorems (FTs) for multipartite processes, formulated in terms of this neighborhood network. I first derive several vector FTs, governing the joint probability of the entropy production (EP) of all of the neighborhoods. These vector FTs in turn imply a set of conditional FTs, for the probability of the overall system’s EP conditioned on the EP of any single neighborhood. Some of these FTs involve extensions of mutual information, to characterize the statistical coupling among the neighborhoods induced by the neighborhood network.

I. INTRODUCTION

Some of the most important results in stochastic thermodynamics are the so-called fluctuation theorems (FTs) \[4, 18, 23, 24\]. These govern the probability distribution of the total entropy production (EP) in a system over some fixed time interval if it evolves according to a continuous-time Markov chain (CTMC).

The early work on FTs did not take into account how the evolving system might decompose into a set of co-evolving subsystems. An important recent set of papers has started to go beyond this early work, deriving FTs specifically for bipartite processes, i.e., for systems that are composed of two co-evolving subsystems, which have zero probability of making a state transition simultaneously \[1, 2, 5, 7, 8, 15–17, 19, 21\].

However, many systems are naturally modeled as having more than two interacting subsystems. For example, a biological cell is naturally modeled as composed of many separate organelles and biomolecule species. Accordingly, recent research has started to extend stochastic thermodynamics from bipartite processes to fully multipartite processes \[6, 8\]. No research has yet explicitly considered FTs of multipartite processes, however.

In general, the dynamics of each subsystem $i$ in a multipartite process will only directly depend on a proper subset of the other subsystems, which is called the “neighborhood” of $i$ \[6\]. For example, each organelle in a biological cell will only directly interact with a subset of the biomolecules in the cell and / or the other organelles in the cell. These neighborhoods will collectively form a network, based on their intersections with one another, as illustrated in Fig. 1. As noted in \[6\], such neighborhood networks have important thermodynamic consequences.

Here I extend this previous work to derive FTs of multipartite processes, expressed in terms of the process’ neighborhood network. Formulating the FTs this way has several advantages. First, the joint dynamics of the subsystems in any single neighborhood obey all the usual thermodynamic properties of autonomous systems. In particular, the EP of neighborhoods cannot decrease with time. (This is not the case with the EPs of the individual subsystems, which are the focus of the analyses in \[2, 19, 21\].) This allows me to derive a FT of the full vector of EPs of all the neighborhoods in the multipartite system, rather than just for the EP of the overall system, or of a single one of the subsystems. In turn, this vector FT can be used to derive conditional FTs, governing the probability distribution of the EP of the full system conditioned on the EP of one of the neighborhoods. It also means allows me to extend earlier, information-theoretic formulations of FTs of bipartite systems, involving the change in mutual information between the two systems during the process, to information-theoretic formulations of FTs of multipartite systems, involving the change in the “in-ex” information among the neighborhoods during the process.

In the next section I introduce terminology. In the section after that I introduce the trajectory-level stochastic thermodynamics of multipartite processes. I then derive the vector-valued FT for multipartite processes, and present several examples of the associated conditional FTs. I illustrate these FTs with the system in Fig. 1.

II. LOCALIZED NEIGHBORHOOD SETS

$\mathcal{N}$ is a set of $N$ subsystems, with finite state spaces $\{X_i : i = 1, \ldots, N\}$. $x$ indicates a vector in $X$, the joint space of $\mathcal{N}$. For any $A \subset \mathcal{N}$, I write $-A := \mathcal{N} \setminus A$. So for example $x\cdot A$ is the vector of all components of $x$ other than those in $A$.

$X$ is the set of all possible trajectories of the system, i.e., all possible combinations of a finite partition of the time interval $[0, t_f]$, along with states of the joint system in every element of that partition.

A distribution over a set of values $x$ at time $t$ is written as $p^x(t)$, with its value for $x \in X$ written as $p^x_x(t)$, or just $p_x(t)$ for short. The simplex of all such distributions is
ΔX. Similarly, \( x_{yi}^X(t) \) is the conditional distribution of \( X \) given \( Y \) at time \( t \), evaluated for the event \( X = x, Y = y \) (which I sometimes shorten to \( p_{xy}(t) \)). A distribution over trajectories \( x \) is written as \( P(x) \). I write Shannon entropy as \( S(p_X(t)), S_t(X) \), or \( S^X(t) \), as convenient. I also write the conditional entropy of \( X \) given \( Y \) at time \( t \) as \( S^XY(t) \), etc., and write relative entropy between \( p(x) \) and \( q(x) \) as \( D(p||q) \).

Since the joint system evolves as a multi-partite process, there is a set of time-varying stochastic rate matrices, \( \{K^X_{yi}(i;t) : i = 1, \ldots, N\} \), where for all \( i \), \( K^X_{yi}(i;t) = 0 \) if \( x_i' \neq x_i \), and where the joint dynamics over \( X \) is governed by the master equation

\[
\frac{dp_x(t)}{dt} = \sum_{x'} K^X_x(t)p_{x'}(t)
\]

(1)

\[
= \sum_{x'} \sum_{i \in N} K^X_x(i;t)p_{x'}(t)
\]

(2)

The rate matrix given by \textbf{windowing} \( A \) onto \( K(t) \) for some \( A \subseteq N \) is

\[
K^X_A(i;t) := \sum_{i \in A} K^X_x(i;t)
\]

(3)

For each subsystem \( i \), I write \( r(i;t) \) for any set of subsystems at time \( t \) that includes \( i \) where we can write

\[
K^X_r(i;t) = K^X_{x(i);r(i)}(i;t)\delta(x_i',x_{r(i)};t)
\]

(4)

for an appropriate set of functions \( K^X_{x(i);r(i)}(i;t) \). The elements of \( r(i;t) \) are the \textbf{neighbors} of \( i \) at time \( t \), and the full set \( r(i;t) \) is a \textbf{neighborhood} of \( i \) at \( t \). For any neighborhood \( r \) of some subsystem \( i \) at time \( t \), and any \( x' \), \( \sum_{x_i} K^X_x(i;t) = 0 \) (see [26] for proof). Note that the neighbor relation is not symmetric in general. There is no implicit restriction to only considering minimal neighborhoods. (So for example, it may be that \( j \in r(i;t) \) even if \( K^x_{x_j}(i;t) \) is independent of \( x_j \).) As an example of these definitions, [1] considers a special type of bipartite system with an “internal” subsystem \( A \) and an “external” subsystem \( B \). \( B \) is its own neighborhood, i.e., it is autonomous. However, \( A \) is not its own neighborhood; its dynamics depends on the state of \( B \) as well as its own state.

I say that a set of neighborhoods is \textbf{localized} if for all subsystems \( i, j, k \) where subsystem \( k \) is a neighbor of subsystem \( j \) which in turn is a neighbor of \( i \), \( k \) is also a neighbor of \( i \). Equivalently, we have a localized set of neighborhoods if for all subsystems \( i \), the neighborhood of \( i \) contains the neighborhoods of all of \( i \)'s neighbors. Note that if we have a localized set of neighborhoods, then the intersection of any two neighborhoods is a neighborhood. See Fig. 1 for an example of these definitions.

We can always choose the neighborhoods of any CTMC so that they form a localized set [26], and so from now on I restrict attention to localized neighborhood sets. Also, for simplicity I assume that the neighborhoods of the rate matrix \( K(t) \) don’t depend on \( t \). Accordingly from now on I shorten \( r(i;t) \) to \( r(i) \).

I use the term \textbf{neighborhood structure} to refer to any family of sets, \( N_1 = \{\omega_1, \omega_2, \ldots, \omega|N_1|\} \), where each \( \omega_i \) is a union of neighborhoods, and where \( \cup_{i=1}^{|N_1|} \omega_i = N \). As shorthand, for any neighborhood union \( \omega \) I write

\[
K^X_{x(i);\omega}(\omega;t) := \sum_{i \in \omega} K^X_{x(i);r(i)}(i;t)
\]

(5)

So \( K^X_{x(i);\omega}(\omega;t) = K^X_{x(i);\omega}(\omega;t)\delta(x_{x\omega},x_{\omega}) \), by Eqs. (3) and (4).

At any time \( t \), for any neighborhood union \( \omega \), \( p_{x(i)}(t) \) evolves as a CTMC with rate matrix \( K^X_{x(i);\omega}(\omega;t) \):

\[
\frac{dp_{x(i)}(t)}{dt} = \sum_{x'} K^X_{x(i);\omega}(\omega;t)p_{x'}(t)
\]

(6)

(See App. A in [26] for proof.) So a neighborhood union evolves according to a self-contained CTMC, which is not the case for a single subsystem in general [26].

III. TRAJECTORY-LEVEL THERMODYNAMICS OF MULTIPARTITE PROCESSES

I assume that the Hamiltonian of the full system respects the neighborhood structure:

\[
H_x(t) = \sum_i H_{x(i)}(i;t)
\]

(7)
for some functions $H_{x_{t_0}}(i;t)$. In addition, as in [6], I assume that each subsystem is in contact with its own heat bath, and that none of the heat baths are directly coupled with one another [11]. I also assume that all heat baths have the same temperature. This scenario can also be viewed as a model of the case where the subsys-

tems are all coupled to different, statistically independent parts of a single macroscopic heat bath [12].

For convenience, choose units so that $k_B$ is the reciprocal of the (shared) temperature of all the baths.

Since each subsystem is in contact with its own heat bath, I also assume that the evolution of the system obeys local detailed balance (LDB). So for all $i$, $t$, $x$, $x'$, with $r$ a neighborhood of $i$ [13],

$$K^{x}(i;t)p_{x_i}(i;t) = K^{x}(i;t)p_{x_i}(i;t)$$

where

$$p_{x_i}(i;t) := \frac{e^{-H_{x_i}(i;t)}}{Z_{x_i}(i;t)}$$

with $Z_{x_i}(i;t) := \sum_{x'} e^{-H_{x'}(i;t)}$. Note that $p_{x_i}(i;t)$ is not the Boltzmann distribution over all of $x_r$, but just over $x$, with $x_r$ held fixed, so that it acts as a parameter. Although I require LDB, detailed balance may not hold for a multipartite process, as shown in Appendix A.

We can now define trajectory-level thermodynamic quantities. First, define the local stochastic entropy as

$$s^\omega(x_w(t)) := -\ln p_{x_w(i)}(t)$$

In general I will use the prefix $\Delta$ to indicate the change of a variable's value between $t_0$ and $t_f$, e.g.,

$$\Delta s^\omega(x_w) := \ln p_{x_w(i)}(0) - \ln p_{x_w(i)}(t_f)$$

Let $M(x)$ be the total number of state transitions during the time interval $[0,t_f]$ by all subsystems (which might equal 0). If $M(x) \geq 1$, define $\eta_x : \{1,\ldots,M(x)\} \to N$ as the function that maps any integer $j \in \{1,\ldots,M(x)\}$ to the subsystem that changes its state in the $j$th transition. Similarly, let $\tau_x : \{0,\ldots,M(x)\} \to \eta$ be the function that maps any integer $j \in \{0,\ldots,M(x)\}$ to the time of the $j$th transition, and maps 0 to the time 0.

From now on, I leave the subscript $x$ on the maps $\eta_x$ and $\tau_x$ implicit. So for example, $\eta^{-1}(i)$ is the set (of indices specifying) all state transitions at which subsystem $i$ changes state in the trajectory $x$. More generally, for any set of subsystems $A$, $\eta^{-1}(A) := \cup_{i \in A} \eta^{-1}(i)$ is the set of all state transitions at which a subsystem $i \in A$ changes state in the trajectory $x$.

Given these definitions, the total heat flow into subsystem $i$ from its heat bath during $[0,t_f]$ is

$$Q^\omega(x) := \sum_{j \in \eta^{-1}(i)} \left[ H_{x_r}(i;t) - H_{x_r}(i;t-1) \right]$$

where I interpret the sum on the RHS to be zero if subsystem $i$ never undergoes a state transition in trajectory $x$. The heat flow into the subsystems in any neighborhood union $\omega$ from the associated heat baths during $[0,t_f]$ is

$$Q^\omega(x) := \sum_{i \in \omega} Q^\omega(i;\tau)$$

I refer to $Q^\omega(x)$ as the local entropy flow (EF) into neighborhood union $\omega$ for trajectory $x$. In Appendix B it is shown that

$$Q^\omega(x) = \sum_{i \in \omega} \sum_{j \in \eta^{-1}(i)} \ln \frac{K_{x_i}(\tau)}{K_{x_i}(\tau-1)}$$

The local EP of any neighborhood union $\omega$ is

$$\sigma^\omega(x) := \Delta s^\omega(x) - Q^\omega(x)$$

which can be evaluated by combining Eqs. [13] and [14]. The expected value of $\sigma^\omega$ is non-negative (in contrast to the analogous expression “$\alpha^\omega$” defined just before Eq. (21) in [20]). In addition, due to Eqs. [7] and [13] and the fact that the neighborhood set is localized, the heat flow into (the subsystems in) $\omega$ along trajectory $x$ is only a function of $x_w$. So we can write $Q^\omega(x) = Q^\omega(x_w)$. Since by definition $\Delta s^\omega(x)$ also only depends on $x_w$, we can also write $\sigma^\omega(x)$ as a function of just $x_w$.

Setting $\omega = N$ in Eqs. (11), (14) and (15) allows us to define global versions of those trajectory-level thermodynamic quantities. In particular,

$$\sigma(x) := \Delta s(x) - Q(x)$$

Note that $x_{\tau}(\tau(j-1)) = x_{\tau}(\tau(j))$ for all subsystems $i$, for all $j \in \eta^{-1}(i)$, since the process is multipartite. Therefore

$$Q(x) = \sum_i Q^i(x) = \sum_i \sum_{j=1}^{M(x)} \ln \frac{K_{x_r}(\tau(j))}{K_{x_r}(\tau(j-1))}$$

There are two decompositions of the global EP that will be used below. First, for any neighborhood union $\omega$, define $\chi^\omega(x) := \sigma(x) - \sigma^\omega(x)$. It is shown in Appendix C that

$$\chi^\omega(x) = \Delta s^X(x) - Q^\omega(x)$$

Note that in general $\chi^\omega$ will not be a neighborhood union. So the heat flow into the associated baths, $Q^\omega(x)$, may depend on the trajectory of subsystems outside of $\omega$, i.e., it may depend on $x_{\tau}(\tau(j))$, for all $j \in \eta^{-1}(i)$.

It is shown in Appendix C that $d(\chi^\omega)/dt$ is the sum of two terms. The first term is the expected global EP rate under a counterfactual rate matrix. The second term is (negative of) the derivative of the mutual information between $x_{\tau}(\tau(j))$ and $x_{\tau}(\tau(j-1))$.
between $X_{\omega}$ and $X_{-\omega}$, under a counterfactual rate matrix in which $X_{-\omega}$ never changes its state. (This second term is a generalization of what is called the “learning rate" in \cite{1}.) Both of these terms are non-negative.

To present the second decomposition of global EP, we need to introduce more notation. Let $A$ be any cover of $\mathcal{N}$ (i.e., a multiset $(a : a \subseteq \mathcal{N}, \cup a = \mathcal{N})$) and suppose we have a set of $a$-indexed functions, $f^a : \mathcal{X} \rightarrow \mathbb{R}$. The associated inclusion-exclusion sum (or just “in-ex sum") is defined as

$$\sum_{a \in A} f^a(x) := \sum_{a \in A} f^a(x) - \sum_{a \in A \cap A'} f^a(x) + \sum_{a \in A \cap A \cap A} f^a(x) - \ldots$$

where $A \cap A := \{a \cap a' : a, a' \in A\}$, etc. It will also be useful to define the time-$t$ in-ex information:

$$I^A(x(t)) := \sum_{a \in A} s^a(x(t)) - s(x(t)) = -s(x) + \sum_{a \in A} s^a(x) - \sum_{a \in A \cap A} s^a(x) + \ldots$$

As an example, if $A = \mathcal{N}_1$, it consists of two neighborhood unions, $\omega_1, \omega_2$, with no intersection, then the expected in-ex information at time $t$ is just the mutual information between those neighborhood unions at that time. More generally, if there is another arbitrary number of neighborhood unions in $\mathcal{N}_1$ but none of them overlap, then the expected in-ex information is the “mutual information" (sometimes instead called “total correlation") among those neighborhood unions $\mathcal{N}_1[22, 23].$

Combining Eqs. \cite{14} and \cite{17} with the inclusion-exclusion principle gives

$$Q(x) = \sum_{\omega \in \mathcal{N}_1} Q^\omega(x)$$

Using Eqs. \cite{11}, \cite{15} and \cite{16}, this gives the second decomposition of the global EP, into the sum of an information-theoretic term and a thermodynamic term:

$$\sigma(x) = \sum_{\omega \in \mathcal{N}_1} \sigma^\omega(x) - \Delta I^{\mathcal{N}_1}(x)$$

As an example, if there are no overlaps between any two neighborhood unions (i.e., if those unions form a partition of $\mathcal{N}$), then Eq. \cite{23} reduces to

$$\sigma(x) = \sum_{\omega} \sigma^\omega(x_\omega) - \Delta I^{\mathcal{N}_1}(x)$$

(See \cite{23} for analysis of a special case of Eq. \cite{24}, involving the thermodynamics of Bayes nets.)

IV. MULTIPARTITE FLUCTUATION THEOREMS

Write $\tilde{x}$ for $x$ reversed in time, i.e., $\tilde{x}(t) = x(t_f - t)$. Also write $P$ for the probability density function generated by starting from the ending distribution $p_x(t_f)$, and evolving from there according to the time-reversed sequences of rate matrices, i.e., according to $\tilde{K}(t) = K(t_f - t)$. As conventional, I will refer to the process that results in $P$ as the “forward protocol" and the process that results in $\tilde{P}$ as the “reverse protocol".

For any $A$ that is a neighborhood structure over some subset of $\mathcal{N}$, define $\tilde{\sigma}^A$ as the vector whose components are the local EP values $\sigma^\omega$ for $\omega \in A$. It is shown in Appendix E that

$$\ln \left[ \frac{P(\tilde{\sigma}^A)}{\tilde{P}(\sigma^A)} \right] = \sigma^A$$

where $\sigma^A$ is the global EP generated by the subsystems in $A$ under the forward protocol, and the probability of EP values under the reverse protocol is

$$\tilde{P}(-\sigma^A) := \int \mathcal{D}x \tilde{P}(\tilde{x}) \prod_{\omega \in A} \delta \left( -\sigma^\omega - \ln \left[ \frac{\tilde{P}(\tilde{x}_\omega)}{\tilde{P}(x_\omega)} \right] \right)$$

(See \cite{23} for some comments about how to interpret $\tilde{P}(\sigma^A)$.) As an example, choosing $A = \mathcal{N}_1$, Eq. \cite{25} shows that $\sigma$ is a single-valued function of $\tilde{\sigma}$, and recovers the standard, unconditioned integral fluctuation theorem (IFT), $\langle e^{-\sigma} \rangle = 1$.

We can add and subtract instances of Eq. \cite{25} evaluated for different choices of $A$, to derive conditional DFTs. These in turn give conditional IFTs (see \cite{25}). Such IFTs are appropriate when we know the total EP generated by (the subsystems in) one or more neighborhood structures, and so can condition on those values. In the rest of this paper I illustrate such conditional IFTs.

As a first example, consider two neighborhood structures, $A = \mathcal{N}_1$ and $A' = \{\omega\}$ for any $\omega \in \mathcal{N}_1$. If we subtract Eq. \cite{25} evaluated for $A'$ from Eq. \cite{25} evaluated for $A$ and plug in Eq. \cite{24}, we see that for any value $\sigma^\omega$ with non-zero probability,

$$\mathbb{E}\left( \exp\left( \Delta I^{\mathcal{N}_1} - \sum_{\omega' \in \mathcal{N}_1} \sigma^{\omega'} + \sigma^\omega \right) \right) = 1$$

If we plug in Eq. \cite{18} instead of Eq. \cite{24}, we get

$$\mathbb{E}\left( e^{\lambda^\omega} | \sigma^\omega \right) = \mathbb{E}\left( \exp\left( Q^{-\omega} - \Delta X^\omega \right) \right) = 0$$

Eq. \cite{28} can often be refined by mixing and matching among all decompositions of $X^{\omega}$ given in Appendix C.

Applying Jensen’s inequality to these IFTs shows that for all values $\sigma^\omega$ with nonzero probability,

$$\mathbb{E}\left( \sum_{\omega' \in \mathcal{N}_1} \sigma^{\omega'} - \Delta X^{\mathcal{N}_1} \right) \geq \sigma^\omega$$

and

$$\mathbb{E}\left( \Delta X^{\mathcal{N}_1} - Q^{-\omega} \right) \geq 0$$
We can illustrate these results with the scenario depicted in Fig. 1. By Eq. (30), the expected global EP conditioned on the joint EP of just A and B is never smaller than that joint EP of A and B. Similarly, the expected joint EP of A and B conditioned on the EP of B is never smaller than that EP of B. As another example, note that \( N_1 \) contains \( \omega, \omega' \) and \( \alpha \), \( N_1 \cap N_1 \) contains three instances of \( \omega' \) (formed from \( \omega \cap \alpha, \omega \cap \omega', \) and \( \omega' \cap \alpha \), respectively) and \( N_1 \cap N_1 \cap N_1 \) contains another instance of \( \omega' \). Therefore by Eq. (29), using letter superscripts to indicate the systems in Fig. 1, we see that

\[
\mathbb{E}\left( \sigma^{AB} - \sigma^B + \sigma^C + \Delta s^{ABC} - \Delta s^{AB} + \Delta s^B - \Delta s^C \big| \sigma^C \right) \geq \sigma^C
\]

(32)

for any value of \( \sigma^C \) with nonzero probability. Similarly, Eq. (32) holds if we condition and lower-bound with a value of \( \sigma^B \) rather than a value of \( \sigma^C \), or if we condition and lower bound with a value of \( \sigma^{AB} \). As a final example, we can use Eq. (31) to upper-bound the total amount of heat flow out of the heat bath of system C:

\[
\mathbb{E}\left( \Delta s^{ABC} - \Delta s^{AB} \big| \sigma^{AB} \right) \geq \mathbb{E}(Q^C \big| \sigma^{AB})
\]

(33)

for all values of \( \sigma^{AB} \) with nonzero probability.

All of the results above hold for arbitrary \( P_\omega(0) \) and arbitrary \( t_f > 0 \). Therefore they translate into results concerning time derivatives. For example, in [26], it is shown that \( d\mathbb{E}(\chi^{\omega})/dt \geq 0 \). However, Eq. (30) tells us that in fact for any value of \( \sigma^{\omega} \) that has nonzero probability throughout an interval \( t \in [0, t_f > 0] \),

\[
d\mathbb{E}(\chi^{\omega} \big| \sigma^{\omega})/dt \geq 0 \text{ at } t = 0.
\]

Eq. (25) has other implications in addition to FTs and associated bounds. For example, since \( \sigma \) is a single-valued function of \( \tilde{\sigma} \), taking the average of both sides of Eq. (25) over all \( \tilde{\sigma} \) establishes that

\[
\langle \sigma \rangle = D\left( P(\tilde{\sigma}) \| P(-\tilde{\sigma}) \right)
\]

(34)

(This is in addition to the fact that expected global EP is the relative entropy between forward and backward trajectories.) Similarly, the conditional DFT associated with Eq. (28) means that for any \( \omega \) with nonzero probability,

\[
\mathbb{E}(\chi^{\omega} \big| \omega) = D\left( P(\omega \big| \sigma^{\omega}) \| \tilde{P}(\omega \big| -\sigma^{\omega}) \right)
\]

(35)

V. DISCUSSION

There are many directions for future work. For example, it may be possible to use the framework introduced in this paper to derive multipartite process refinements of Crooks’ theorem (and therefore also of the Jarzynski equality), e.g., for the case where the neighborhood structure forms a chordal graph, so that the marginalizations of the joint Boltzmann distribution onto the neighborhood unions are themselves Boltzmann distributions [9]. Another possibility is to integrate the FTs derived in this paper, which operate at the level of multiple neighborhoods, with the FTs derived in [20], which operate within individual neighborhoods. It may also be possible to adapt the approach used in [25] to derive novel thermodynamic uncertainty relations that apply to multipartite processes.

ACKNOWLEDGMENTS

This work was supported by the Santa Fe Institute, Grant No. CHE-1648973 from the US National Science Foundation and Grant No. FQXi-RFP-IPW-1912 from the FQXi foundation. The opinions expressed in this paper are those of the author and do not necessarily reflect the view of the National Science Foundation.
that the state changes, \( t \) is the vector of those transition times, and \( \mathbf{x} \) is the vector of states \( x \in X \) for each interval between two successive transition times. \( P(M) \) is a conventional probability distribution over a countably infinite set (the positive integers), \( P(\mathbb{I}M) \) is a density function over \( \mathbb{R}^M \), and \( P(3\mathbb{I}^T M) \) is again a conventional probability distribution (this time over a finite space, \( \chi^M \)). The \( \delta \) function over trajectories in the equations below is shorthand for a function that equals zero everywhere that its argument is nonzero, and such that its integral over \( \{M, \mathbb{R}^M, \chi^M; M \in \mathbb{Z}^+\} \) equals 1.

An advantage of this assumption is that it ensure that with probability 1, at any time \( t \) at which some subsystem \( i \) changes its state, no other subsystem does so – which is exactly what’s required to have a multipartite process.

This scenario is similar to the scenario commonly considered in the literature where there are multiple “mechanisms” coupled to a single system. (The main difference is that here, in essence, each “mechanism” is coupled to only a subset of the degrees of freedom of that system.) In particular, \( x_{(i)}(\cdot) \) determines the energy levels of the possible states of \( i \), but the coupling of \( i \) with an external heat bath does not cause \( x_{(i)}(\cdot) \) to fluctuate. So the joint state of the subsystems in \( r(i) \) other than \( i \), \( x_{(i)}(\cdot) \), plays a role in the dynamics of \( x_i \) analogous to that played by the parameter vector of an external work reservoir acting on \( i \).

Note that since we have a multipartite process, both \( K_1^x(i; t) \) and \( K_2^x(i; t) \) equal zero if \( x_{(i)} \neq x_{(i)}^{eq} \) and so \( \Delta \) cannot be defined in terms of ratios.

Since each subsystem is coupled to its own bath, we can uniquely identify which bath was involved in each state transition in any given \( x \). Note that this is not the case for trajectory-level analyses of systems which are coupled to multiple mechanisms, e.g., [1]; to identify what bath is involved in each transition in that setting we need to know more than just \( x \).

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

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

Minimal energy cost for thermodynamic information processing: measurement and information erasure, Physical review letters 102 (2009), no. 25, 250602.

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

Ito S. Kawaguchi K. Shiraishi, N. and T. Sagawa, Role of measurement-feedback separation in autonomous maxwellEs demons, New Journal of Physics (2015).

Naoto Shiraishi and Takahiro Sagawa, Fluctuation theorem for partially masked nonequilibrium dynamics, Physical Review E 91 (2015), no. 1, 012130.

Sagawa T. Shiraishi, N., Fluctuation theorem for partially masked nonequilibrium dynamics, Physical Review E (2015).

Hu Kuo Ting, On the amount of information, Theory of Probability & Its Applications 7 (1962), no. 4, 439–447.

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.

David H. Wolpert, The stochastic thermodynamics of computation, Journal of Physics A: Mathematical and Theoretical (2019).

David H Wolpert, Uncertainty relations and fluctuation theorems for bayes nets, arXiv preprint arXiv:1911.02700v4 (2019).

Minimum entropy production in multipartite processes due to neighborhood constraints, arXiv preprint arXiv:2001.02205 (2020).

Appendix A: Demonstration that detailed balance can be violated

To see this, note that for almost all trajectories \( x \), and almost all times \( t \) at which the joint state of the system changes in that trajectory, there is some time \( t^- \) earlier than \( t \) and some subsystem \( i \) such that \( x_{(i)}(t^-) = x_{(i)}(t) \) but \( x_{(i)}(t^-) \neq x_{(i)}(t) \), and so

\[
\ln K_{x_{(i)}(t^-)}(t) - \ln K_{x_{(i)}(t)}(t) = \ln K_{x_{(i)}(t^-)}(i; t) - \ln K_{x_{(i)}(t)}(i; t) = H_{x_{(i)}(t^-)}(i; t^-) - H_{x_{(i)}(t)}(i; t) \tag{A1}
\]

However, in general

\[
H_{x_{(i)}(t^-)}(i; t^-) - H_{x_{(i)}(t)}(i; t) \neq H_{x_{(i)}(t^-)}(t) - H_{x_{(i)}(t)}(t) \tag{A2}
\]

since there may be Hamiltonians \( H_{j}(j; t) \) for \( j \neq i \) which depend on the value of \( x_{j} \).

Appendix B: Proof of Eq. (14)

Plug in Eq. (12) to get

\[
Q^{eq}(x) = \sum_{i \in \omega} \sum_{j \in \eta^{-1}(i)} \left[ \ln p^{eq}_{x_{(i-1)}(i; \tau(j))} \right] - \ln p^{eq}_{x_{(i)\tau(j))}}(i; \tau(j)) \right]
\]

\[
+ \sum_{i \in \omega} \sum_{j \in \eta^{-1}(i)} \left[ \ln Z_{x_{(i-1)}(i; \tau(j))} \right] - \ln Z_{x_{(i)\tau(j))}}(i; \tau(j)) \tag{B1}
\]

Since we have a multipartite process, we know that \( x_{(i)}(\tau(j)) = x_{(i-1)}(\tau(j)) \) for all terms in the sum in the second line in Eq. (B1). So that sum equals zero, and therefore total EF \( \phi \) along trajectory \( x \) is

\[
Q^{eq}(x) = \sum_{i \in \omega} \sum_{j \in \eta^{-1}(i)} \left[ \ln p^{eq}_{x_{(i-1)}(i; \tau(j))} \right] - \ln p^{eq}_{x_{(i)\tau(j))}}(i; \tau(j)) \right] \tag{B2}
\]

Plugging in Eq. (6) then establishes the claim.

Appendix C: Expansions of \( x^{eq} \)

As shorthand, I leave the function \( \tau \) implicit, so that for example, \( x(\tau(j)) \) gets shortened to \( x(j) \). (Note though
that with slight abuse of notation, I still take \( \mathbf{x}(t_f) \) to
mean the state of the system at \( t = t_f \) under trajectory

\[
\sigma(\mathbf{x}) = \ln p_{\mathbf{x}(0)} - \ln p_{\mathbf{x}(t_f)} + \sum_{j=1}^{M(\mathbf{x})} \ln \frac{K_{x(j)}^{x(j-1)}(\eta(j); j)}{K_{x(j)}^{x(j-1)}(\eta(j); j)}
\]

\[
= \ln p_{\mathbf{x}^\omega(0)} - \ln p_{\mathbf{x}^\omega(0)}(t_f) + \sum_{j\in\eta^{-1}(\omega)} \ln \frac{K_{x(j)}^{x(j-1)}(\eta(j); j)}{K_{x(j)}^{x(j-1)}(\eta(j); j)}
\]

\[
+ \ln p_{\mathbf{x}^\omega(0)}(0) - \ln p_{\mathbf{x}^\omega(0)}(t_f) + \sum_{j\in\eta^{-1}(\omega)} \ln \frac{K_{x(j)}^{x(j-1)}(\eta(j); j)}{K_{x(j)}^{x(j-1)}(\eta(j); j)}
\]

\[
= \sigma^\omega(\mathbf{x}) + \ln p_{\mathbf{x}^\omega(0)} - \ln p_{\mathbf{x}^\omega(0)}(t_f) + \sum_{j\in\eta^{-1}(\omega)} \ln \frac{K_{x(j)}^{x(j-1)}(\eta(j); j)}{K_{x(j)}^{x(j-1)}(\eta(j); j)}
\]

Now define

\[
\Delta_{\mathbf{x}^\omega}(\mathbf{x}) = \Delta_{\mathbf{x}^\omega}(\mathbf{x})
\]

\[
:= \ln p_{\mathbf{x}^\omega(0)} - \ln p_{\mathbf{x}^\omega(0)}(t_f)
\]

This allows us to rewrite Eq. (C1) more succinctly as

\[
\chi^\omega(\mathbf{x}) = \Delta_{\mathbf{x}^\omega}(\mathbf{x}) + \sum_{j\in\eta^{-1}(\omega)} \ln \frac{K_{x(j)}^{x(j-1)}(\eta(j); j)}{K_{x(j)}^{x(j-1)}(\eta(j); j)}
\]

\[
= \Delta_{\mathbf{x}^\omega}(\mathbf{x}) - Q^\omega(\mathbf{x})
\]

which establishes Eq. (18) in the main text.

To gain insight into Eq. (C4), define the counterfactual rate matrix \( K(t) := K(\omega, t) \), and let \( \mathbf{P} \) be what the density over trajectories \( \mathbf{x} \) would have been if the system had evolved from the initial distribution \( p_\omega(0) \) under \( K(t) \) rather than \( K(t) \). Define \( \Delta_{\mathbf{P}}(\mathbf{x}) \) and \( \sigma(\mathbf{x}, \mathbf{P}) \) accordingly. Then we can expand the second term on the RHS of Eq. (C4) as

\[
Q^\omega(\mathbf{x}) = \sigma(\mathbf{x}, \mathbf{P}) - \Delta_{\mathbf{P}}(\mathbf{x})
\]

So the heat flow from the baths connected to \(-\omega\) into the associated subsystems is the difference between a (counterfactual) global EP and a (counterfactual) change in the entropy of those subsystems.

We can iterate these results, to get more refined decompositions of global EP. For example, let \( \mathcal{N}^1 \) be a localized neighborhood structure of \( K \), the counterfactual rate matrix defined just before Eq. (C5). Let \( \omega \) be a neighborhood union in \( \mathcal{N}^1 \) while \( \alpha \) is a neighborhood union in \( \mathcal{N}^1 \). Then we can insert Eq. (C5) into Eq. (18), and then apply Eq. (18) to the resulting term \( \sigma(\mathbf{x}, \mathbf{P}) \), to get

\[
\sigma(\mathbf{x}) = \sigma^\omega(\mathbf{x}) + \left[ \sigma^\alpha(\mathbf{x}) + \chi^\omega(\mathbf{x}) + \Delta_{\mathbf{P}}(\mathbf{x}) - \Delta_{\mathbf{P}}(\mathbf{x}) \right]
\]

Note that in general, \( \alpha \) might contain subsystems outside of \( \mathcal{N} \setminus \omega \). As a result, it need not be a localized neighborhood of the full rate matrix \( K \). In addition, both (counterfactual) rates \( d(\langle \sigma^\omega \rangle / dt) \) and \( d(\langle \chi^\omega \rangle / dt) \) are non-negative. However, if we evaluate those two expectations under the actual density \( \mathbf{P} \) rather than the counterfactual \( \mathbf{P} \), this may not be the case. This is just like how the expected values of those “EP” terms in [8][20] which concern a single subsystem may decrease with time.

Appendix D: Calculation of \( \frac{d(\langle \chi^\omega \rangle)}{dt} \)

As shorthand replace \( t_f \) with \( t \), and then expand

\[
\left\langle -\ln p_{\mathbf{x}(t)} \right\rangle = -\sum_{x} p_x(t) \left( \ln p_x(t) - \ln p_{\mathbf{x}(t)} \right)
\]

Therefore,

\[
\frac{d}{dt} \left( -\ln p_{\mathbf{x}(t)} \right) = -\sum_{x,x'} K_{x,x'}(t) p_x(t) \ln p_x(t)
\]

In addition, the sum in Eq. (C4) is just the total heat flow from the subsystems in \(-\omega\) into their respective heat
baths, during the interval \([0, t]\), if the system follows trajectory \(x\). Therefore the derivative with respect to \(t\) of the expectation of that sum is just the expected heat flow rate at \(t\) from those subsystems into their baths,

\[
- \sum_{x,x'} K^r_x (-\omega; t)p_{x'}(t) \ln \left[ \frac{K^r_{x'}(-\omega; t)}{K^r_x(-\omega; t)} \right]
\]  \hspace{1cm} \text{(D3)}

In addition, \(K(t) = K(\omega; t) + K(-\omega; t)\). So if we add Eq. (D3) to Eq. (D2), and use the fact that rate matrices are normalized, we get

\[
d \langle \chi^\omega \rangle = - \sum_{x,x'} K^r_x (\omega; t)p_{x'}(t) \ln p_{x|x_\omega}(t)
+ \sum_{x,x'} K^r_x (-\omega; t)p_{x'}(t) \ln \left[ \frac{K^r_{x'}(-\omega; t)}{K^r_x(-\omega; t)} \right]  \hspace{1cm} \text{(D4)}
\]

The first sum in Eq. (D4) is what was called the “windowed derivative” \(d\langle \chi^\omega \rangle / dt\) in [26]. Since \(\omega\) is a neighborhood union, it is the (negative) of the derivative of the mutual information between \(X_\omega\) and \(X_{-\omega}\), under a counterfactual rate matrix in which \(x_{-\omega}\) is held fixed. As discussed in [26], by the data-processing inequality, this term is non-negative.

The second sum in Eq. (D4) is what was called \(\langle \dot{\sigma}_{K(N(\omega,t))} \rangle\) in [26]. Since it is the expected rate of EP for a properly normalized, counterfactual rate matrix, it is non-negative.

This decomposition of \(d \langle \sigma - \sigma^\omega \rangle / dt\) was first derived in [26]. However, that derivation did not start from a trajectory-level definition of local and global EPs.

**Appendix E: Proof of Eq. (25)**

Proceeding in the usual way [3, 18, 23], we want to calculate

\[
r(x_A) := \ln \left[ \frac{P(x_A)}{P(x_A)} \right] \hspace{1cm} \text{(E1)}
\]

Paralleling the development in App. A of [3], we reduce \(r(x_A)\) to a sum of two terms. The first term is a sum, over all transitions in \(x\), of the log of the ratio of entries in the rate matrix of the subsystem that changes state in that transition [14]. Due to Eq. (14), that sum equals \(Q_A(x)\). The second term is just \(\Delta s^A(x)\). Therefore by Eq. (15), we have a global DFT over trajectories,

\[
\ln \left[ \frac{P(x_A)}{P(x_A)} \right] = \sigma(x_A) \hspace{1cm} \text{(E2)}
\]

Plugging in for the special case of \(A = \{\omega\}\) for any neighborhood union \(\omega \in N_1\) gives

\[
\ln \left[ \frac{P(x_\omega)}{P(x_\omega)} \right] = \sigma(x_\omega) \hspace{1cm} \text{(E3)}
\]

Therefore, paralleling [23], we can combine Eqs. (E2) and (E3) to get a DFT for the probability density function of values of \(\sigma(x)\):

\[
P(\tilde{\sigma}^A) = \int Dx P(x) \prod_{\omega \in A} \delta(\sigma^\omega - \ln \left[ \frac{P(x_\omega)}{P(x_\omega)} \right]) \hspace{1cm} \text{(E4)}
\]

\[
= e^{\sigma^A} \int D\tilde{x} \tilde{P}(\tilde{x}) \prod_{\omega \in A} \delta(\sigma^\omega - \ln \left[ \frac{P(x_\omega)}{P(x_\omega)} \right]) \hspace{1cm} \text{(E5)}
\]

\[
= e^{\sigma^A} \int D\tilde{x} \tilde{P}(\tilde{x}) \prod_{\omega \in A} \delta(-\sigma^\omega - \ln \left[ \frac{P(x_\omega)}{P(x_\omega)} \right]) \hspace{1cm} \text{(E6)}
\]

\[
is e^{\sigma^A} P(-\tilde{\sigma}) \hspace{1cm} \text{(E7)}
\]

i.e.,

\[
\ln \left[ \frac{P(\sigma^\omega)}{P(-\tilde{\sigma}^A)} \right] = \sigma^A \hspace{1cm} \text{(E8)}
\]

which establishes the claim.