I. INTRODUCTION

Perhaps the most famous result in non-equilibrium statistical physics is the classical physics formulation of Landauer’s bound [2, 14]. In its fully modern, generalized form [6, 9, 20, 23, 29, 31, 36], it says that for any thermodynamically closed system which evolves from an initial distribution \( p_0(x) \) to an ending distribution \( p_f(x) \), the amount of heat dissipated to the heat bath is at least as large as the associated drop in entropy of the system, \( S(p_0) - S(p_f) \) (in units where \( k_B T = 1 \)). Stated differently, it says that the minimal integrated entropy production (EP) in the system as it evolves in any thermodynamically closed process is non-negative.

Some strengthened versions of Landauer’s bound have been derived by substantially restricting the process under consideration, e.g., by simultaneously imposing limits on how long the process takes and on (integrals of) certain time-varying properties of the process dynamics [10, 18, 19, 24, 27]. In addition, a version of Landauer’s bound has been derived that applies quite broadly, to any composite system that comprises two interacting subsystems [21]. However, this particular version of Landauer’s bound was derived by redefining terms, so that Landauer’s bound applied to the composite system results in an apparent violation of the bound for a subsystem. It does not continuously produce any EP generated by evolving the full, composite system.

More recently, some strengthened versions of Landauer’s bound for the EP generated by full, composite systems have been derived [4, 34, 36]. However, these results only concern the special case where the subsystems in the composite system evolve independently of one another. This is a major restriction on the results. In a more general setting, while the dynamics of each subsystem only depends on the states of some of the other subsystems, the subsystems are not independent. As an illustration of such a dependency structure, consider a composite system with three subsystems \( A, B \) and \( C \). \( B \) evolves independently of \( A \) and \( C \). However, \( B \) is continually observed by \( C \) as well as \( A \). Moreover, suppose that \( A \) is really two subsystems, 1 and 2. Only subsystem 2 directly observes \( B \), whereas subsystem 1 observes subsystem 2, e.g., to record a running average of the values of subsystem 2 (see Fig. 1). These new, strengthened Landauer bounds do not apply to such scenarios.

Physically, such a scenario where the subsystems interact with one another arises whenever any of the many stochastic thermodynamics models of one classical system observing another classical system without any back-action [13, 20, 22, 26, 32, 33] are “chained together”. As an example, [3, 8] considers a tripartite system where receptors in the wall of a cell observe the concentration level of a ligand in a surrounding medium, with no back-action on that concentration level, while a memory observes the state of those receptors, again with no back-action. This is exactly the scenario considered in Fig. 1 just without subsystem 4; subsystem 3 is the concentration level in the medium, subsystem 2 is the set of receptors in a cell observing that concentration level, and subsystem 1 is the memory within the cell observing the state of the ligand receptors.

I use the term “dependency constraint” to mean the specification for a subsystem \( i \) in the composite system of which other subsystems can affect its dynamics. So it is a constraint on the rate matrix of the composite system. In this paper I use stochastic thermodynamics to analyze the minimal heat that a composite system must dissipate if it obeys a given set of dependency constraints (each constraint concerning the dynamics of a different subsystem). The analysis holds for arbitrary sets of dependency constraints, allowing the subsystems to interact with one another as they all evolve, e.g., as in Fig. 1.

To begin, I translate a given set of dependency constraints into a convex polytope, \( P \), in a unit simplex. I also translate the combination of the dependency constraints, an initial distribution \( p_0(x) \), and a final distribution \( p_f(x) \), into \( F \), a linear function over \( P \). My main
result is to show that for any point \( \pi \in \mathcal{P} \), \( F(\pi) \) is a lower bound on the total heat that must be dissipated into the baths by running any process that is consistent with the constraints and maps \( p_0 \) into \( p_{t_f} \). Finding the strongest such bound on the dissipated heat is just a linear programming problem. The results in \([41, 34, 36]\) are recovered as the special case where the constraints are maximal, in that no interactions among the subsystems are allowed whatsoever.

The new lower bound given by my main result is not always stronger than the conventional Landauer bound. However, for any set of constraints on the rate matrix, there is an initial distribution \( p(x(0)) \) and conditional distribution \( p(x(t)|x(0)) \) such that any rate matrix that obeys those constraints and implements that conditional distribution will result in a non-negative lower bound on the EP, i.e., a lower bound at least as strong as the conventional Landauer bound. Indeed, for some sets of rate matrix constraints, the new EP bound is stronger than the conventional Landauer bound no matter what \( p(x(0)) \) and \( p(x(t)|x(0)) \) are (so long as that conditional distribution is consistent with the constraints on the rate matrix).

As an example, suppose we modify the scenario considered in \([3, 22]\) by introducing a second cell, which is observing the same external medium as the first cell. Assume that the cells are far enough apart physically so that their dynamics are independent of one another. This gives us the precise scenario in Fig. 1 where subsystem 4 is the state of the receptors of that second cell.

Ex. 4 below analyzes an example of this scenario. In this example \( |X_2| = |X_4| = 2 \) (i.e., the state of each cell’s receptors is coarse-grained into two bins, which I assume are always internally thermalized). It is also assumed that \( x_2 = x_4 \) with probability 1 at the start of the process, e.g., because some molecules were flushed through the medium at the start of an experiment in order to reset those receptors, or because the cells were both just born. Finally, it is assumed that the composite system evolves long enough so that both receptors lose all information about their initial states.

Applying the main result of this paper to this scenario establishes that the minimal total EP of the composite system is at least \( \ln 2 \). In contrast, the conventional Landauer bound on EP is 0. This new, strengthened bound holds no matter how the state of the medium evolves as it is being observed, and no matter how the two sets of cell receptors observe the medium.

In the next section I formalize dependency constraints for composite systems as restrictions on the rate matrix of a continuous-time Markov chain (CTMC). I then use this formalization to derive an expression for the EP of a composite system that involves the dependency constraints of its rate matrix, but also involves other factors. In the following section I derive a lower bound on that expression for EP purely in terms of those dependency constraints, the initial distribution \( p(0) \), and the ending distribution \( p(t_f) \). This lower bound is my main result. In the ending section I present examples.

II. RATE MATRIX DEPENDENCY STRUCTURES

I begin by defining notation. First, I write \( \mathcal{N} \) for a particular 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, \( x_A \) is the vector of all components of \( x \) other than those in \( A \). For any set \( L, \Delta_L \) is the associated unit-simplex. In addition, for any function \( f(p) \), I write \( \Delta f := f(p_{t_f}) - f(p_0) \).

The set of bits is \( \mathbb{B} = \{0, 1\} \). I write the Kronecker delta as \( \delta(a, b) \). For any family of sets, \( A = \{a_1, a_2, \ldots\} \), I define \( \cup A = a_1 \cup a_2 \cup \ldots \).

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 \( p_x(t) \). \( p(x(t)) \) or \( p(x(t)) \) is more convenient. Similarly, I write \( p((x(t)|x(t'))) \) for the conditional distribution of the state at time \( t \) given the state at time \( t' \), etc. I write Shannon entropy as \( S(p_x(t)), S_I(X), \) or \( S^X(t) \), as convenient.

A (dependency) unit \( \omega \) at time \( t \) is a set of subsystems such that

\[
\frac{dp_{\omega}(t)}{dt} = \sum_{x_{\omega}} K_{\omega}(\omega; t)p_{\omega}(t)
\]

for some associated rate matrix \( K(\omega; t) \). Intuitively, a unit is any set of subsystems whose evolution is inde-
dependent of the states of the subsystems outside the unit. Note though that in general, the evolution of the subsystems outside of a unit may depend on the states of subsystems inside the unit. Note as well that any nonempty intersection of units is a unit, as is any union of units.

As an example, [1, 2, 3, 5] investigate a special type of bipartite system, where the "internal" subsystem $B$ observes the "external" subsystem $A$, but cannot affect the dynamics of that external subsystem. So $A$ is its own unit, evolving independently of $B$, while $B$ is not its own unit; its dynamics depends on the state of $A$ as well as its own state. Another example of these definitions is illustrated in Fig. 1.

I will call a set of units defined over a set of subsystems $\mathcal{N}$ a dependency (unit) structure, and write it as $\mathcal{N}^*$, if has the following properties:

1. The union of the units in $\mathcal{N}^*$ equals all of $\mathcal{N}$.
2. $\mathcal{N}^*$ is closed under intersections of its units;
3. For simplicity I don’t allow $\mathcal{N}^*$ to contain any “orphan” unit $\omega$ where every subsystem $i \in \omega$ is already contained in some unit $\omega' \subset \omega$.

I will sometimes say that $\mathcal{N}^*$ represents the set of subsystems $\mathcal{N}$. In general, a given process can be represented with more than one dependency structure, since for example any process over $\mathcal{N}$ can be represented as a dependency structure with a single unit, $\mathcal{N}$ itself. Also, in general there are sets of subsystems $A \subset \mathcal{N}$ which are not units of $\mathcal{N}$, and so cannot be represented by any dependency structure. For simplicity, from now on I assume that the set of units doesn’t change with $t$.

The dynamics of any two units $\omega, \alpha \subset \omega$ must be consistent with one another, i.e., for all $p_{x_{\omega}, t} = p_{x_{\alpha}, x_{\omega \setminus \alpha}, t}$, it must be that

$$\sum_{x_{\omega \setminus \alpha}} \frac{dp_{x_{\alpha}, x_{\omega \setminus \alpha}, t}}{dt} = \frac{dp_{x_{\alpha}}}{dt} \quad (2)$$

Using Eq. (1) to evaluate the two derivatives and then setting $p_{x_{\alpha}}(t)$ to a delta function, we see that for all $x_{\alpha}, x_{\alpha}', x_{\omega \setminus \alpha}, x_{\omega \setminus \alpha}'$, it must be that

$$k_{x_{\alpha}}^{x_{\alpha}'}(\alpha; t) = \sum_{x_{\omega \setminus \alpha}} k_{x_{\alpha}, x_{\omega \setminus \alpha}, x_{\omega \setminus \alpha}'}(\omega; t) \quad (3)$$

Note that this must hold whether or not the units $\omega, \alpha$ are elements of some particular dependency structure. As an example, this self-consistency condition always holds (and therefore so does Eq. (5)) in a multipartite process, in which the rate matrix of the overall system can be represented as a sum of rate matrices of the individual subsystems [11, 32].

It will often be convenient to re-express a dependency structure as a directed graph. Define the dependency graph $\Gamma_{\mathcal{N}^*} = (\mathcal{N}^*, E)$ by the rule that there is an edge $e \in E$ from node $\omega \in \mathcal{N}^*$ to node $\omega' \in \mathcal{N}^*$ iff both: $\omega' \subseteq \omega$, and there is no intervening unit $\omega''$ such that $\omega' \subseteq \omega'' \subseteq \omega$. (Note that $\Gamma_{\mathcal{N}^*}$ is a directed acyclic graph (DAG).) For a dependency structure $\mathcal{N}^*$ where $\mathcal{N} \in \mathcal{N}^*$, there would be a single root of the dependency graph, but if $\mathcal{N} \notin \mathcal{N}^*$, the dependency graph has multiple roots.

I will abuse notation and sometimes treat a unit $\omega$ as a set of subsystems while at other times I treat it as a single node in $\Gamma_{\mathcal{N}^*}$. I write the set of parents of any node $\omega \in \Gamma_{\mathcal{N}^*}$ as $pa(\omega)$, and the set of its descendants as $desc(\omega)$, with $fa(\omega) := \omega \cup desc(\omega)$, the “family” of node $\omega$. The maximal number of nodes in any directed path that starts at $\omega$ is the height of $\omega$. So any unit $\omega$ which has no sub-units contained in it is a leaf node of $\Gamma_{\mathcal{N}^*}$, with height 1. (The maximal height of all nodes in $\Gamma_{\mathcal{N}^*}$ is simply called “the height of $\mathcal{N}^*$.”) I write $R_{\mathcal{N}^*}$ for the set of root nodes in $\Gamma_{\mathcal{N}^*}$.

As an example, the dependency graph of Fig. 1 has two root nodes, $\omega$ and $\alpha$, and one leaf node, $\omega'$, which is their common child. The height of the graph is 2.

Another, more complicated example is shown in Fig. 2. This example could be used to model a simple extension of the triparitite system considered in [3, 8]. To see this, label the joint concentration levels in the medium surrounding a cell of three separate ligands as subsystem 4. Label the ligand receptors in the wall of a cell that each observe a separate one of the three ligand concentrations, without any back-action, as subsystems 1, 3, 5. Finally, let subsystems 2 and 6 be two internal memories in the cell, each recording its own summary statistic concerning the states of the three sets of ligand receptors, again with no back-action. The associated dependency constraints are of the form given in Fig. 2.

Finally, define a conditional distribution for the ending joint state given an initial joint state, $p(x(t_f) | x(0))$. 

![Fig. 2](image-url)
to be consistent with a specified dependency structure if there is some rate matrix that obeys that dependency structure and that implements \( p(\{x(t_f)|x(0)\}) \).

III. THERMODYNAMICS OF DEPENDENCY STRUCTURES

As in the conventional Landauer bound, for simplicity I assume that the composite system is connected to a work reservoir and one heat bath, with units chosen so that \( k_B T = 1 \). (The results below also hold if different subsystems are connected to different heat baths all at the same temperature, as in analyses of multipartite systems [11].) So in the standard way, the expected entropy flow (EF) rate of any unit \( \omega \subseteq N \) at time \( t \) is [6,31]:

\[
\langle \dot{Q}^{\omega}(t) \rangle = \sum_{x_0,x_\omega} K^x_{x_\omega}(\omega; t)p_{x_\omega}(t) \ln \left[ \frac{K^x_{x_\omega}(\omega; t)}{K^x_{x_\omega}(\omega; t)} \right]
\]

(4)

(Note that this is entropy flow from \( \omega \) into the heat bath.) I refer to \( \langle \dot{Q}^{\omega}(t) \rangle \) as a local EF rate, and define the global EF rate as \( \langle \dot{Q}^N(t) \rangle \).

Make the associated definition that the expected EP rate of \( \omega \) at time \( t \) is

\[
\langle \dot{\sigma}^{\omega}(t) \rangle = \frac{dS^{\omega}(t)}{dt} + \langle \dot{Q}^{\omega}(t) \rangle
\]

(5)

\[
= \sum_{x_0,x_\omega} K^x_{x_\omega}(\omega; t)p_{x_\omega}(t) \ln \left[ \frac{K^x_{x_\omega}(\omega; t)}{K^x_{x_\omega}(\omega; t)} \right]
\]

(6)

I refer to \( \langle \dot{\sigma}^{\omega}(t) \rangle \) as a local EP rate, and define the global EP rate as \( \langle \dot{\sigma}(t) \rangle := \langle \dot{\sigma}^N(t) \rangle \). For any unit \( \omega \), \( \langle \dot{\sigma}^{\omega}(t) \rangle \geq 0 \), since \( \langle \dot{\sigma}^{\omega}(t) \rangle \) has the usual form of an EP rate of a single system. (See [32] for a discussion of the relation between local EP rates and similar quantities discussed in [11,12,28].)

Write the local EP generated by a unit \( \omega \) during the process as

\[
\sigma^{\omega} := \int_0^{T_f} \langle \dot{\sigma}^{\omega}(t) \rangle dt
\]

(7)

and similarly write \( \sigma = \sigma^N \) for the global EP. In Appendix A, Eq. (5) and the log sum inequality [5] are used to prove that for any two units \( \omega, \alpha \subset \omega \), not necessarily part of a dependency structure, \( \langle \dot{\sigma}^{\omega}(t) \rangle \geq \langle \dot{\sigma}^{\alpha}(t) \rangle \) at all times \( t \). Therefore

\[
\sigma^{\omega} \geq \sigma^{\alpha}
\]

(8)

In particular, this holds if \( \alpha \) is a union of units (since a union of units is itself a unit).

In addition, it is shown in [34,36] that in the special case where there is a set of units \( \{\alpha_j\} \) who have no overlap with another, for any unit \( \omega \supseteq \bigcup_j \alpha_j \),

\[
\sigma^{\omega} \geq \sum_j \sigma^{\alpha_j}
\]

(9)

(See also Eq. (26) below.)

Let \( N^* = \{\omega : j = 1,2,\ldots,n\} \) be a dependency structure. Suppose we have a real-valued vector \( f \) indexed by the sets of \( N^* \). The associated inclusion-exclusion sum (or just “in-ex sum”) is defined as

\[
\sum_{\omega \in N^*} f^{\omega} := \sum_{j=1}^n f^{\omega_j} - \sum_{1 \leq j < j' \leq n} f^{\omega_j \cap \omega_{j'}} + \ldots
\]

(10)

(Note that the precise assignment of integer indices to the units in \( N^* \) is irrelevant.)

The time-\( t \) in-ex information is defined in terms of this notation, as

\[
I^{N^*} := \sum_{\omega \in N^*} S^{\omega} - S^N
\]

(11)

where all the terms in the sums on the RHS are marginal entropies over the (distributions over the subsystems in) the indicated units. As an example, if \( N^* \) consists of two units, \( \omega_1, \omega_2 \), with no intersection, then the expected in-ex information at time \( t \) is just the mutual information between those units at that time. More generally, if there an arbitrary number of units in \( N^* \) but none of them overlap, then the expected in-ex information is what is called the “multi-information”, or “total correlation”, among those units [15,30,37].

When local detailed balance holds, we want to be able to interpret \( \langle \dot{Q}^{\omega}(t) \rangle \) in the usual way as the expected rate of heat flow due to joint transitions of the states of one or more of the subsystems in unit \( \omega \), with the state of all subsystems outside of \( \omega \) fixed and arbitrary. It will also be convenient to suppose that the dependency structure is rich enough so that no state transition can occur that simultaneously changes the state of all subsystems in a set \( a \) unless \( a \in N^* \). (So in particular, if \( N^* \notin N^* \), then no state transition can occur in which every subsystem simultaneously changes its state.) As an example, this supposition would hold if each unit were attached to its own heat bath, or equivalently, to statistically decoupled portions of the same heat bath [11]. (See [32] for a detailed example involving multipartite systems.)

In Appendix B, Rota’s extension of the inclusion-exclusion principle is combined with this supposition to show that

\[
\langle \dot{Q}^N(t) \rangle = \sum_{\omega \in N^*} \langle \dot{Q}^{\omega}(t) \rangle
\]

(12)

This implies that the global EP rate is

\[
\langle \dot{\sigma}(t) \rangle = \frac{dS^N(t)}{dt} + \langle \dot{Q}^N(t) \rangle
\]

(13)
a result first derived in [32] for the special case of multipartite processes. Integrating Eq. \( (14) \) from the beginning to the end of a process gives

\[
\sigma^N = \sum_{\omega \in \mathcal{M}^*} \sigma^\omega - \Delta I^{\mathcal{M}^*} \tag{15}
\]

Eq. \( (15) \) applies to any dependency structure, including dependency structures that are subsets of other dependency structures. In addition, for any dependency structure \( \mathcal{M}^* \) over a set of subsystems \( \mathcal{M} \subset \omega \), Eq. \( (8) \) and the fact that the union of a set of units is itself a unit means that \( \sigma^\omega - \sigma^{\mathcal{M}} \geq 0 \). Therefore using Eq. \( (15) \) to expand \( \sigma^\mathcal{M} \) gives

\[
\sigma^\omega - \sum_{\omega' \in \mathcal{M}^*} \sigma^{\omega'} \geq -\Delta I^{\mathcal{M}^*} \tag{16}
\]

(Note that Eq. \( (16) \) holds even if the units in \( \mathcal{M}^* \) are not in \( \mathcal{N}^* \), which can happen because in general any process obeying a given set of dependency constraints can be represented by more than one dependency structure.)

IV. A STRENGTHENED LANDAUER BOUND

In this paper I use Eqs. \( (5), (9), (15) \) and \( (16) \) and the inclusion-exclusion principle to derive lower bounds on EP. To do this I first define \( V(\mathcal{N}^*) \), a set of distributions over the Boolean hypercube, \( \mathbb{B}^{\mathcal{N}^*} \). Next, I define a “centering distribution” to be any convex combination of the elements of \( V(\mathcal{N}^*) \) which equals \( \{1/\mathcal{N}^*, 1/\mathcal{N}^*, \ldots \} \), the uniform distribution over the Boolean hypercube. My first main result, presented in Proposition 1, is a function taking each such centering distribution to a different lower bound on global EP.

The set \( V(\mathcal{N}^*) \) is the union of two sets of distributions over \( \mathbb{B}^{\mathcal{N}^*} \), defined as follows:

I) For any \( \omega \in \mathcal{N}^* \), write \( \delta^\omega \) for the distribution in \( \mathbb{B}^{\mathcal{N}^*} \) which is all 0’s except for a 1 in its \( \omega \) component. Using this notation, define \( V^1(\mathcal{N}^*) \) as the set of all distributions \( \alpha \) over \( \mathbb{B}^{\mathcal{N}^*} \) which obey at least one of the following three conditions:

1. \( \alpha = \delta^\omega \) for some \( \omega \in \mathcal{N}^* \) whose height \( \leq 2 \);  
2. \( \alpha = \frac{\sum_{\omega' \in \text{fa}(\omega)} \delta^{\omega'}}{\sum_{\omega' \in \text{fa}(\omega)}} \) for some \( \omega \in \mathcal{N}^* \);  
3. \( \alpha = \frac{\sum_{\omega' \in \text{fa}(\omega)} \delta^{\omega'}}{\sum_{\omega' \in \text{fa}(\omega)}} \) for a \( \omega \in \mathcal{N}^* \), \( v \in \text{desc}(\omega) \);

I will refer to any distributions that obey these conditions as type-1, type-2, and type-3 distributions, respectively.

To provide examples, consider the dependency structure illustrated in the left panel in Fig. 2. There are four associated type-1 distributions, \((0,0,1,0,0), (0,0,0,1,0,0), (0,0,0,0,1,0), (0,0,0,0,0,1)\), corresponding to units \( C,D,E,F \), respectively. The (unique) type-2 distribution for unit \( \omega = C \) is \( \alpha = (0,0,1/2,0,0,1/2) \), and for unit \( \omega = A \) it is \( \alpha = (1/5,0,1/5,1/5,1/5,1/5) \).

Next, note that units \( A,C,D,E,F \) are all units contained in unit \( A \) (i.e., those nodes are contained in family of node \( A \)). Similarly, units \( F,C \) are all units contained in \( C \). Therefore the units that are in \( A \) but not in \( C \) are \( A,D,E \). Accordingly, the (unique) type-3 distribution for the pair of units \( \omega = A,v = C \) is \( \alpha = (1/3,0,0,1/3,0) \).

II) For any unit \( \omega \), abuse notation and define \( \omega^* \) as the dependency structure \( \omega \cup \text{desc}(\omega) \). Using this shorthand, write \( \mathcal{M}(\omega) \subset \omega \) for any set of subsystems which can be represented by a dependency structure \( \mathcal{M}(\omega)^* \subset \omega^* \). Define \( \mathcal{V}^2(\mathcal{N}^*) \) as the set of all distributions over \( \mathbb{B}^{\mathcal{N}^*} \) of the form

\[
\alpha = \frac{\sum_{\omega' \in \omega^* \setminus \mathcal{M}(\omega)^*} \delta^{\omega'}}{\sum_{\omega' \in \omega^* \setminus \mathcal{M}(\omega)^*}} \tag{17}
\]

for some \( \omega \in \mathcal{N}^* \) and associated dependency structure \( \mathcal{M}(\omega)^* \subset \omega^* \). I will refer to any such distribution as a type-4 distribution. Note that any type-4 distribution \( \alpha \) uniquely specifies both \( \omega \) and \( \mathcal{M}(\omega)^* \). Given this, I will sometimes abuse notation and write \( \alpha^* \) for the dependency structure \( \mathcal{M}(\omega)^* \) specified by a type-4 distribution \( \alpha \).

To illustrate type-4 distributions, again consider the left panel in Fig. 2. Choose \( \omega = A \). So \( \omega^* = \{A,C,D,E,F\} \). Choose \( \mathcal{M}(\omega) \) to be the set of subsystems \( \{3,4,5\} \) with the associated dependency structure \( \mathcal{M}(\omega)^* = \{C,D,F\} \). So \( \omega^* \setminus \mathcal{M}(\omega)^* = \{A,E\} \). The associated type-4 distribution is \( \alpha = (1/2,0,0,1/2,0) \).

As shorthand, write \( V(\mathcal{N}^*) := V^1(\mathcal{N}^*) \cup \mathcal{V}^2(\mathcal{N}^*) \). As a final piece of terminology, let \( U = \{u\} \) be any set of distributions over some shared space. I will say that \( U \) is centered if there exists a centering distribution \( \pi \in \Delta_U \) such that \( \pi_u = \sum_{v \in U} \pi_{u,v} \) equals the uniform distribution. Note that the set of all centering distributions of any \( U \) is a convex polytope.

The following is proven in Appendix C.

**Proposition 1.** If \( \pi \) is a centering distribution for \( V(\mathcal{N}^*) \), then

\[
\sigma \geq -\Delta I^{\mathcal{N}^*} - \sum_{\alpha \in \mathcal{V}^2(\mathcal{N}^*)} \pi_{\alpha} \Delta I^{\alpha^*}
\]
Note that the values of the in-ex informations $I^N$ and $\Delta I^\alpha$ at the beginning and end of the process are fully specified by $p_0$ and $p_{t_f}$. So given a centering distribution, Proposition 1 provides a lower bound on global EP defined purely in terms of $p_0$ and $p_{t_f}$. The precise rate matrix is irrelevant, so long as it has $N^*$ as a dependency structure and maps $p_0$ to $p_{t_f}$.

As an example of Proposition 1, suppose that $\Gamma_{N^*}$ has height 2. (So there are no units $\omega, \omega', \omega'' \in N^*$ such that $\omega'' \subset \omega' \subset \omega$.) Then all delta function distributions over $R[N^*]$ are type-1 distributions, and so contained in $V^1(N^*)$. Accordingly, the dependency structure is centered by a distribution $\pi$ that is uniform over all $\alpha \in V^1(N^*)$ and equals 0 for all $\alpha \in V^2(N^*)$. Plugging this into Proposition 1 establishes that the EP of any process with a height-2 dependency structure $N^*$ is lower-bounded by $-\Delta I^{N^*}$.

Note that in general, we can represent any process by using a dependency structure of height 2. (For example, we can do that by combining all subsystems that are members of some unit $\omega$ that is not a root node of $\Gamma_{N^*}$, into one, overarching unit.) Accordingly, for any process, we can always find an associated dependency structure $N^*$ for which $\sigma \geq -\Delta I^{N^*}$.

In addition, it is proven in Appendix 2 that for any dependency structure $N^*$, no matter what its height, $V(N^*)$ is centered. (The set of all associated centering distributions of $V(N^*)$ is the convex polytope discussed in the introduction.) In general, finding the optimal such centering distribution — the one that maximizes the bound in Proposition 1 and so provides the strongest lower bound on global EP — only requires solving a linear programming problem.

Unfortunately, as illustrated below, there are some dependency structures $N^*$ where the bound in Proposition 1 is negative for an appropriate initial distribution $p_0(x)$ and conditional distribution $p(x(t_f)|x(0))$ consistent with $N^*$, no matter what centering distribution we use. In such cases, Proposition 1 does not provide a stronger bound on EP than the conventional Landauer’s bound.

On the other hand, as also illustrated below, typically the bound in Proposition 1 will be stronger than the conventional Landauer’s bound. Indeed, for every dependency structure $N^*$, and every associated centering distribution, there are initial distributions $p_0(x)$ and conditional distributions $p(x(t_f)|x(0))$ that are consistent with $N^*$ where the EP bound in Proposition 1 at least as strong as Landauer’s bound:

**Proposition 2.** Let $N^*$ be any dependency structure that does not have $N$ itself as a member. Then there exists an initial joint distribution $p_0(x)$ and a conditional distribution $p(x(t_f)|x(0))$ consistent with $N^*$ such that for any associated centering distribution $\pi_{\alpha^*}$,

$$-\Delta I^{N^*} - \sum_{\alpha \in V^2(N^*)} \pi_{\alpha} \Delta I^{\alpha^*} \geq 0$$  

for every rate matrix that both implements that $p(x(t_f)|x(0))$ and obeys the dependency structure.

(See Appendix 3 for proof.)

Write the polytope of all centering distributions of the dependency structure $N^*$ as $P(N^*)$, and write the strongest lower bound on EP given by Proposition 1 as

$$\mathcal{F}(p_0, p_{t_f}, N^*) := -\min_{\pi \in P(N^*)} \left[ \Delta I^{N^*} + \sum_{\alpha \in V^2(N^*)} \pi_{\alpha} \Delta I^{\alpha^*} \right]$$  

Combining Propositions 1 and 2 establishes that for any dependency structure $N^*$ there are pairs of $p_0(x)$ and a conditional distribution $p(x(t_f)|x(0))$ consistent with $N^*$ such that

$$Q + \Delta S \geq \mathcal{F}(p_0, p_{t_f}, N^*) \geq 0$$

In contrast, the conventional Landauer’s bound says only that $Q + \Delta S \geq 0$, no matter what $p_0, p_{t_f}$, or $N^*$ are.

Summarizing, suppose we are given a dependency structure $N^*$, an initial distribution $p_0(x)$, and a conditional distribution $p(x(t_f)|x(0))$ consistent with $N^*$. Then we know that $-\Delta I^{N^*} - \sum_{\alpha \in V^2(N^*)} \pi_{\alpha} \Delta I^{\alpha^*}$ is a lower bound on the EP, for any $\pi$ that is a centering distribution for $N^*$. These lower bounds are simple to evaluate, and are often stronger than the Landauer bound. We can find the strongest such lower bound on EP due to the dependency structure $N^*$ by solving a linear programming problem.

Furthermore, in general, it is possible to represent any given set of constraints with more than one dependency structure. Each one of them results in its own strongest lower bound on EP, given by solving the associated linear programming problem. So to find the strongest Landauer bound for a given set of constraints, we should solve all the linear programming problems specified by the dependency structures that can represent those constraints.

This result extends the previous strengthenings of Landauer’s bound derived in [4, 34, 36], which all assume that the units have no overlap, to the case where the units may overlap with one another in arbitrary ways, even if none of the subsystems are fixed in the dynamics. In addition to this result, going from a set of constraints, $p_0$ and $p_{t_f}$ to a lower bound on EP, a secondary result is that for any set of constraints, there is a $p_0$ and $p_{t_f}$ that results in a strictly positive lower bound on EP, stronger than the conventional Landauer bound.

**V. EXAMPLES**

I now present several examples of the main result. In the first example and some of the others, there are
three subsystems, each with two possible states, 0 and 1. I will consider two dependency structures: \( A^* = \{1, 2\}, \{2, 3\} \) and \( B^* = \{\{1\}, \{2, 3\}\} \). So under dependency structure \( A^* \), subsystem 2 evolves independently, while the dynamics of both subsystems 1 and 3 depend on the state of subsystem 2. In contrast, under \( B^* \), all three subsystems evolve independently. The height of both of the associated dependency graphs is \( \leq 2 \). Accordingly, we can choose a centering distribution that puts zero weight on all elements in \( V^2(N^*) \), which will be assumed for these examples.

**Example 1:** Suppose that the conditional distribution implemented by the process erases both subsystems 1 and 3 and doesn’t change the state of subsystem 2. So

\[
p\left(x(t_f) | x(t_0)\right) = \delta(x_1(t_f), 0) \delta(x_3(t_f), 0) \delta(x_2(t_f), x_2(0))
\]

Suppose as well that \( p_0(x) = (1/2) \sum_{i=0}^1 \prod_j \delta(x_i(0), j) \). So the three bits must initially have the same state — but that shared state is uniformly randomly set.

Evaluating, since the \( t = 0 \) joint entropy of the full system is \( \ln 2 \), \( \mathcal{I}(0) = - \ln 2 + 3 \ln 2 = 2 \ln 2 \). Similarly, \( \mathcal{I}(t_f) = - \ln 2 + \ln 2 = 0 \). So plugging into Proposition \[1\] shows that the minimal EP is at least \( 2 \ln 2 \) for dependency structure \( B^* \), in which all subsystems evolve independently while the discrete-time conditional distribution \( p\left(x(t_f) | x(0)\right) \) is implemented. In contrast, \( \mathcal{I}(0) = - \ln 2 + 3 \ln 2 - 3 \ln 2 + \ln 2 = 0 \), and similarly \( \mathcal{I}(t_f) = 0 \). So the minimal EP is only lower-bounded by \( 0 \) for dependency structure \( A^* \), in which the same conditional distribution is implemented, but both subsystems that are being erased are able to observe the (unchanging) state of subsystem 2.

In this particular scenario, we can evaluate the minimal EP for both dependency structures exactly \[17\]. Doing so confirms that \( \Delta \mathcal{I} \) in fact equals the minimal EP exactly, for both dependency structures. The difference between those two minimal EP values illustrates that implementing the exact same conditional distribution starting from the exact same initial distribution can result in different minimal heat dissipations, depending on the dependency structure.

We can generalize this scenario, to consider any process where no subsystem in the intersection of two or more distinct units ever changes its state. Suppose we represent this process with a dependency graph of height 2, e.g., like the one illustrated in Fig. 3. Since the leaf nodes are intersections of units, by our hypothesis none of the subsystems in those leaf nodes ever changes its state. Accordingly,

\[
-\Delta \mathcal{I}^{N^*} = \left[ \sum_{\omega \in \mathcal{N}^{N^*}} S(p_\omega(0)) - S(p_\omega(t_f)) \right] - \left[ S(p(0)) - S(p(t_f)) \right]
\]

where the sum runs only over the root nodes.

Moreover, since any unit that never changes its state generates no EP, if no subsystem in the intersection of two units ever changes its state, then

\[
\sum_{\omega} \sigma^{\omega} = \sum_{\omega} \sigma^{\omega}
\]

Since the lowest each \( \sigma^{\omega} \) can be is zero (when each unit \( \omega \) evolves semi-statically), we can use this with Eqs. \[15\] and \[23\] to establish that minimal EP equals

\[
\left[ \sum_{\omega} S(p_\omega(0)) - S(p_\omega(t_f)) \right] - \left[ S(p(0)) - S(p(t_f)) \right]
\]

exactly, i.e., Eq. \[23\] is a strict lower bound on the EP. This lower bound holds no matter what \( p_0(x) \) and \( p(x(t_f)|x(0)) \) are (so long as \( p(x(t_f)|x(0)) \) is consistent with the dependency structure).

As an illustration of this result, suppose that in fact no two units have nonempty intersection. Then the minimal EP is

\[
\left[ \sum_{i} S(p_i(0)) - S(p_i(t_f)) \right] - \left[ S(p(0)) - S(p(t_f)) \right]
\]

This quantity is the drop among the subsystems in their multi-information, a formula for the minimal EP previously derived in \[34\] \[36\]. By repeated application of the data-processing inequality, it is easy to confirm that this lower bound on the EP is non-negative.

Note though that Eq. \[23\] holds for any process where the dependency structure is of the type shown in Fig. 3 so long as the ending entropies of (the joint subsystems in the units corresponding to) the leaf nodes equals the starting entropies. This is true even if the subsystems in the leaf nodes do change state during the process. Since that dependency graph has height 2, Proposition \[1\] tells us that the expression in Eq. \[23\] is a lower bound on the EP of such a process, and is non-negative. However, in general, if the subsystems in the leaf nodes change their states during the process, that lower bound may not be tight.

**Example 2:** In light of the fact that subsystem 2 never changes its state in Ex. 1, suppose that the conditional

![FIG. 3. An arbitrary example of a dependency graph of height 2. The first process considered in Ex. 1 can be represented with this kind of dependency structure, where the subsystems in the leaf nodes never change their states. The last process considered in that example relaxes this requirement.](image)
distribution of that example is implemented using only subsystems 1 and 3. In other words, suppose we have a modified process where both 1 and 3 are erased, and are perfectly correlated under the initial joint distribution, but subsystem 2 doesn’t exist, so that subsystems 1 and 3 evolve independently of each other.

The drop in in-ex information in this case is ln2. Again, one can confirm that this lower bound on the minimal EP is actually an equality. (This parallel bit erasure scenario was previously investigated in [39].)

Note that this minimal EP of parallel bit erasure using only two subsystems is less than the minimal EP for parallel bit erasure using three subsystems, investigated in Ex. 1 for the analogous dependency structure, \(B^*\). So even though neither subsystem 1 nor subsystem 3 observe subsystem 2 in the process analyzed in Ex. 1 for dependency structure \(B^*\), and even though subsystem 2 never changes its state in that process, simply including it in the model increases the minimal total entropy production.

\[
p(x(t_f) | x(0)) = p(x_1(t_f), x_3(t_f) | x_2(t_f), x(0)) p(x_2(t_f) | x(0)) = p(x_1(t_f) | x_2(t_f), x(0)) p(x_3(t_f) | x_2(t_f), x(0)) p(x_2(t_f))
\]

Therefore \(S(X(t_f)) = S(X_1(t_f)|X_2(t_f)) + S(X_3(t_f)|X_2(t_f)) + S(X_2(t_f))\), and so

\[
\mathcal{I}^{A^*}(p(t_f)) = \left[ (S(X_1(t_f)|X_2(t_f)) + S(X_2(t_f))) + (S(X_3(t_f)|X_2(t_f)) + S(X_2(t_f))) - S(X_1(t_f)) \right] - S(X(t_f)) = 0
\]

Combining Eqs. [29] and [34] establishes that the minimal EP is lower-bounded by ln2. Note that we can derive this lower bound on the EP even though both subsystems 1 and 3 are continually observing subsystem 2 during the process, and even if subsystem 2’s state is changing throughout the process. In fact, this lower bound holds no matter what the ending distribution of \(p_{f_i}(x)\) is, so long it can be written as in Eq. [32]. (So in particular, as discussed in the introduction, it applies to a simple extension of the cell-sensing scenario analyzed in [31].)

**Example 4:** This example presents a case where the new bound is not stronger than the Landauer bound. Again consider the three bits of Ex. 1, with dependency structure \(A^*\), with the same initial distribution \(p(x_0)\). However, now suppose that during the dynamics, neither \(x_1\) nor \(x_3\) change their state, while \(x_2\) completely uniformly randomizes.

As in Ex. 1, \(\mathcal{I}^{A^*}(p(0)) = 0\). However, with this different dynamics \(S(p(t_f)) = 2\ln 2\), and so \(\mathcal{I}^{A^*}(t_f) = 3\ln 2 - 2\ln 2 = \ln 2\). So in this scenario, \(\Delta \mathcal{I}^{A^*} > 0\). (Since the actual minimal possible EP of the full system cannot be less than zero, this means that \(\sum_{x} \sigma^\omega > 0\) in this scenario.) On the other hand, by Proposition [2] we know that for some other initial distribution \(p(x(0))\) and / or some other conditional distribution \(p(x(t_f)|x(0))\) that is consistent with the dependency structure, \(-\Delta \mathcal{I}^{A^*}\) is a strictly positive lower bound on EP.

As an aside, note that in this example, the composite system’s initial distribution is a Markov random field, with cliques equal to the units in \(A^*\):

\[
p_0(x_1, x_2, x_3) = p_0(x_1, x_2)p_0(x_2, x_3)/p_0(x_2)
\]

This is why the entropy of the initial distribution equals the in-ex sum of the entropies of the units, i.e., why the initial in-ex information equals 0. In addition, of course, the conditional dependencies of the dynamics is given by the units in \(A^*\). Nonetheless, the ending distribution is not given by a Markov random field with cliques equal to the units in \(A^*\). This is why the entropy of the final distribution can differ from the in-ex sum of the entropies of the units, which in turn is what allows the
The green arrows in the right panel illustrate the interaction among six subsystems, \{1,2,3,4,5,6\}. These interactions are the same as in Fig. 2 except that subsystem 2 no longer observes subsystem 1, and subsystem 6 no longer observes subsystem 3. A choice for the associated dependency structure is the set of six units, \{(4),(1,4),(3,4),(4,5),(2,3,4,5),(1,4,5,6)\}. The left panel shows the associated dependency graph, identifying those six units as F, C, E, D, A, B, respectively.

endin in-ex information to differ from the initial in-ex information.

\[
-\Delta \mathcal{I}^N = \Delta S(123456) - \sum_{\omega \in N^*} \Delta S(\omega)
\]

where \(\Delta S(ab...)\) is shorthand for the change in the joint entropy of the subsystems \(ab...\) between \(t=0\) and \(t=t_f\). By Proposition 2, we are guaranteed that for some initial distribution \(p(0)\), and some final distribution \(p(t_f)\) that is generated from \(p(0)\) by a process consistent with the dependency structure, the lower bound on \(\mathcal{I}^p\) given by Eq. (40) is strictly positive.

**Example 6:** Finally, consider again the set of six subsystems, interacting with the dependency structure shown in Fig. 2. As in Ex. 5 this dependency structure has height 3, and so we cannot construct a centering distribution only using type-1 distributions over \(\mathbb{B}^5\). However, consider the following four distributions over the units A, B, C, D, E, F:

1. \(\alpha^1 = (0,0,0,0,0,1)\) is a type-1 distribution;
2. \(\alpha^2 = (0,0,1,0,0,0)\) is a type-1 distribution;
3. \(\alpha^3 = (1/2,0,1/2,0,0,0)\) is a type-2 distribution;
4. \(\alpha^4 = (0,1/2,0,0,1/2,0)\) is a type-2 distribution;

So the distribution \(\pi\) with components

\[
\pi(\alpha^1) = \pi(\alpha^2) = 1/6;
\]

\[
\pi(\alpha^3) = \pi(\alpha^4) = 1/3;
\]

is a centering distribution.

Since \(\pi\) places no probability mass on any type-4 distribution, Proposition \[4\] tells us that the total \(\mathcal{I}^p\) is lower-bounded by

\[
-\Delta \mathcal{I}^N = \frac{1}{3} \Delta \mathcal{I}^{(C,D,E,F)}
\]

Expanding this expression is left as an exercise for the interested reader. (Note that since the dependency structure in Fig. 2 differs from that in Fig. 4, the term \(-\Delta \mathcal{I}^N\) in Eq. (41) is not given by Eq. (40).)
VI. DISCUSSION

In this paper I derive a strengthened version of the conventional Landauer bound, which applies to any composite system whose constituent subsystems obey a given set of constraints on how their dynamics can depend on one another. In contrast to earlier related work, this bound applies even if the subsystems are all evolving in time, with arbitrary overlaps in the dependencies of their dynamics.

To derive this new bound, first I translate any set of dependency constraints into a set of “dependency units”, where each dependency unit $\omega$ is a set of subsystems whose dynamics is self-contained, not depending on the state of any subsystems outside of $\omega$. The precise strengthened form of Landauer’s bound derived here is defined in terms of a “dependency structure”, which is any collection of dependency units which jointly encompass all subsystems in the composite system. The bound is expressed in terms of a novel information-theoretic function, specified by the dependency structure, evaluated for the initial joint distribution over the state of all the subsystems and for the final such joint distribution.

In general, any set of dependency constraints can be represented by more than one dependency structure. For example, this can usually be done by coarsely-graining the units in a dependency structure, i.e., by replacing a set of multiple units $\{\omega, \omega', \ldots\}$ in the dependency structure with a single unit, which encompasses all of the subsystems in $\{\omega, \omega', \ldots\}$. Each of these dependency structures, all representing the same set of dependency constraints, provide a different lower bound on EP. It is left for future work to find general rules for which of a set of candidate dependency structures representing the same set of dependency constraints result in the strongest lower bound on EP (other than simply trying them all).

ACKNOWLEDGMENTS

I would like to thank Jake Hanson and Artemy Kolchinsky for stimulating discussion. 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.

[1] Andre C. Barato, David Hartich, and Udo Seifert, Efficiency of cellular information processing, New Journal of Physics 16 (2014), no. 10, 103024.

[2] Charles H Bennett, Logical reversibility of computation, IBM journal of Research and Development 17 (1973), no. 5, 525–532.

[3] Stefano Bo, Marco Del Giudice, and Antonio Celani, Thermodynamic limits to information harvesting by sensory systems, Journal of Statistical Mechanics: Theory and Experiment 2015 (2015), no. 1, P01014.

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

[6] Massimiliano Esposito and Christian Van den Broeck, Three faces of the second law: 1. master equation formulation, Physical Review E 82 (2010), no. 1, 011143.

[7] D. Hartich, A. C. Barato, and U. Seifert, Stochastic thermodynamics of bipartite systems: transfer entropy inequalities and a Maxwell’s demon interpretation, Journal of Statistical Mechanics: Theory and Experiment 2014 (2014), no. 2, P02016.

[8] David Hartich, Andre C. Barato, and Udo Seifert, Sensory capacity: An information theoretical measure of the performance of a sensor, Physical Review E 93 (2016), no. 2, arXiv:1509.02111.

[9] H-H Hasegawa, J Ishikawa, K Takara, and DJ Driebe, Generalization of the second law for a nonequilibrium initial state, Physics Letters A 374 (2010), no. 8, 1001–1004.

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

[11] Jordan M. Horowitz, Multipartite information flow for multiple Maxwell demons, Journal of Statistical Mechanics: Theory and Experiment 2015 (2015), no. 3, P03006.

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

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

[14] Rolf Landauer, Irreversibility and heat generation in the computing process, IBM journal of research and development 5 (1961), no. 3, 183–191.

[15] William McGill, Multivariate information transmission, Transactions of the IRE Professional Group on Information Theory 4 (1954), no. 4, 93–111.

[16] There is always at least one such $M(\omega)$ for any non-leaf $\omega$, given by the set of subsystems $\cup_{\text{desc}(\omega)}$. Note that in general, any such set of subsystems $M(\omega)$ can be represented by more than one dependency structure $M(\omega)^* \subset \omega^*$. For example, suppose there are leaves in $\text{desc}(\omega)$ that have only one parent, i.e., there are units in $\text{desc}(\omega)$ that are only proper subsets of one other unit in $\text{desc}(\omega)$. Then there is both a dependency structure that contains those leaves that represents $M(\omega)$, and a dependency structure that does not contain those leaves but still represents $M(\omega)$.

[17] For example, because $x_2$ never changes, the minimal EF to implement the dynamics of subsystem 1 with unit
structure $A$ is $S(p(X_1(0)|X_2(0))) - S(p(X_1(t_f),X_2(t_f))) = 0$. The same conclusion holds for the other two subsystems. Moreover, EF is additive over subsystems in this example, by Eq. (12). So the total minimal EF using dependency structure $A'$ to implement the map from $p_0$ to $p_f$ is 0. This is the same as the minimal EF using any dependency structure whatsoever to implement the map from $p_0$ to $p_f$, i.e., it equals the conventional Landauer bound. So the minimal EP to implement that map from $p_0$ to $p_f$ using dependency structure $A'$ is 0.

[18] Manaka Okuyama and Masayuki Ohzeki, Quantum speed limit is not quantum, Physical review letters 120 (2018), no. 7, 070402.

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

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

[21] Takahiro Sagawa and Masahito Ueda, Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics, Physical review letters 109 (2012), no. 18, 180602.

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

[23] U. Seifert, Nonequilibrium thermodynamics of feedback control, Physical Review E 85 (2012), no. 2, 021104.

[24] Tim Schmiedl and Udo Seifert, Optimal finite-time processes in stochastic thermodynamics, Physical review letters 98 (2007), no. 10, 108301.

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

[26] Ito S. Kawaguchi K. Shiraishi, N. and T. Sagawa, Role of measurement-feedback separation in autonomous maxwell’s demons, New Journal of Physics (2015).

[27] Naoto Shiraishi, Ken Funo, and Keiji Saito, Speed Limit for Classical Stochastic Processes, Physical Review Letters 121 (2018), no. 7.

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

[29] K. Takara, H.-H. Hasegawa, and D. J. Driebe, Generalization of the second law for a transition between nonequilibrium states, Phys. Lett. A 375 (2010), no. 2, 88–92.

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

[31] 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.

[32] Gatien Verley, Christian Van den Broeck, and Massimiliano Esposito, Work statistics in stochastically driven systems, New Journal of Physics 16 (2014), no. 9, 095001.

[33] Christopher W Wächtler, Philipp Strasberg, and Tobias Brandes, Stochastic thermodynamics based on incomplete information: generalized jarzynski equality with measurement errors with or without feedback, New Journal of Physics 18 (2016), no. 11, 113042.

[34] David Wolpert and Artemy Kolchinsky, The thermodynamics of computing with circuits, New Journal of Physics (2020).

[35] David Wolpert and Artemy Kolchinsky, The thermodynamics of computing with circuits, New Journal of Physics (2020).

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

[37] David H Wolpert, Minimal entropy production due to constraints on rate matrix dependencies in multipartite processes, arXiv preprint arXiv:2001.02205v3 (2020).

Appendix A: Proof of Eq. (6)

First, note that by Eq. (3), for all conditional distributions $p(x_{\omega_{\alpha}}|x'_{\alpha})$, for all $x_{\alpha},x'_{\alpha}$,

$$K_{x_{\alpha}}^{x'_{\alpha}}(\omega; t) = \sum_{x_{\omega_{\alpha}}=x'_{\omega_{\alpha}}} K_{x_{\omega_{\alpha}}x_{\alpha}}^{x'_{\omega_{\alpha}}x'_{\alpha}}(\omega; t)p(x_{\omega_{\alpha}}|x'_{\omega_{\alpha}})$$  \hspace{1cm} (A1)

Next, use Eq. (6) to expand

$$\langle \omega(t) \rangle = \sum_{x_{\alpha},x'_{\alpha}} p_{x_{\alpha}}(t) \sum_{x_{\omega_{\alpha}}=x'_{\omega_{\alpha}}} K_{x_{\omega_{\alpha}}x_{\alpha}}^{x'_{\omega_{\alpha}}x'_{\alpha}}(\omega; t)p_{x'_{\omega_{\alpha}}|x'_{\alpha}) \left( \ln \frac{K_{x_{\omega_{\alpha}}x_{\alpha}}^{x'_{\omega_{\alpha}}x'_{\alpha}}(\omega; t)p_{x'_{\omega_{\alpha}}|x'_{\alpha})}{K_{x_{\omega_{\alpha}}x_{\alpha}}^{x'_{\omega_{\alpha}}x'_{\alpha}}(\omega; t)p_{x_{\omega_{\alpha}}|x_{\alpha}}) + \ln \frac{\ln K_{x_{\omega_{\alpha}}x_{\alpha}}^{x'_{\omega_{\alpha}}x'_{\alpha}}(\omega; t)p_{x'_{\omega_{\alpha}}|x'_{\alpha})}{\ln K_{x_{\omega_{\alpha}}x_{\alpha}}^{x'_{\omega_{\alpha}}x'_{\alpha}}(\omega; t)p_{x_{\omega_{\alpha}}|x_{\alpha}}) \right)$$  \hspace{1cm} (A2)

Using the log sum inequality [5] shows that for each $x_{\alpha},x'_{\alpha}$, the resultant value of the first term in the inner sum on the RHS of Eq. (A2) is bounded by

$$\sum_{x_{\omega_{\alpha}}=x'_{\omega_{\alpha}}} K_{x_{\omega_{\alpha}}x_{\alpha}}^{x'_{\omega_{\alpha}}x'_{\alpha}}(\omega; t)p_{x'_{\omega_{\alpha}}|x'_{\alpha}) \ln \frac{K_{x_{\omega_{\alpha}}x_{\alpha}}^{x'_{\omega_{\alpha}}x'_{\alpha}}(\omega; t)p_{x'_{\omega_{\alpha}}|x'_{\alpha})}{K_{x_{\omega_{\alpha}}x_{\alpha}}^{x'_{\omega_{\alpha}}x'_{\alpha}}(\omega; t)p_{x_{\omega_{\alpha}}|x_{\alpha}}) \geq \sum_{x_{\omega_{\alpha}}=x'_{\omega_{\alpha}}} K_{x_{\omega_{\alpha}}x_{\alpha}}^{x'_{\omega_{\alpha}}x'_{\alpha}}(\omega; t)p_{x'_{\omega_{\alpha}}|x'_{\alpha}) \ln \frac{\sum_{x_{\omega_{\alpha}}=x'_{\omega_{\alpha}}} K_{x_{\omega_{\alpha}}x_{\alpha}}^{x'_{\omega_{\alpha}}x'_{\alpha}}(\omega; t)p_{x'_{\omega_{\alpha}}|x'_{\alpha})}{\sum_{x_{\omega_{\alpha}}=x'_{\omega_{\alpha}}} K_{x_{\omega_{\alpha}}x_{\alpha}}^{x'_{\omega_{\alpha}}x'_{\alpha}}(\omega; t)p_{x_{\omega_{\alpha}}|x_{\alpha}}$$
where the second line uses Eq. (A1).

Combining and using Eq. (A1) again establishes that

$$\langle a^\omega(t) \rangle \geq \sum_{x_a,x_b} K_{x_a}^a(\omega; t) p_{x_a,x_b}(t) \ln \left[ \frac{K_{x_b}^a(\omega; t)}{K_{x_a}^a(\omega; t)} \right]$$

(A4)

Again plugging into Eq. (6), this time for the EP rate of unit \( \alpha \), completes the proof.

\[\text{Appendix B: Proof of Eq. (12)}\]

Fix the time \( t \) and make it implicit from now on. Define \( \mathcal{T} \) as the set of all possible state transitions in \( X \), involving an arbitrary number of the subsystems in \( \mathcal{N} \). (So there are \(|X||\{X\}-1|\) elements of \( \mathcal{T} \), where \(|X| = \prod_{x \in \mathcal{N}} |X_x| \) is the number of joint states of the composite system.) For all \( \tau \in \mathcal{T} \), define \( f(\tau) \) to be the expected heat flow rate into the bath for all transitions in \( \tau \). Finally, define \( \tau_\omega \) as the set of all possible state transitions in \( X \) in which no subsystem outside of \( \omega \) changes state. So the set \( \mathcal{T}_{\mathcal{N}^*} := \{ \tau_\omega : \omega \in \mathcal{N}^* \} \) is a poset, ordered by the set inclusion relation over the \( \alpha \in \mathcal{N}^* \).

Next, by our supposition that no state transition can occur that simultaneously changes the state of all subsystems a unless \( \alpha \in \mathcal{N}^* \), every element of \( \mathcal{T} \) that can occur is contained in at least one element of \( \mathcal{T}_{\mathcal{N}^*} \). So \( \mathcal{T}_{\mathcal{N}^*} \) is a cover of the set of all state transitions that can occur. Accordingly, by Rota’s extension of the inclusion-exclusion principle,

$$f(\mathcal{T}_{\mathcal{N}^*}) = \sum_{\omega \in \mathcal{N}^*} f(\tau_\omega)$$

(B1)

(Note that unlike in the conventional inclusion-exclusion principle, in this extension we don’t need to have \( f(i) \) defined for individual subsystems \( i \).)

Identifying \( f(\mathcal{T}_{\mathcal{N}^*}) \) with \( \langle \hat{Q}(t) \rangle \) and identifying \( f(\tau_\omega) \) with \( \langle \hat{Q}^\omega(t) \rangle \) for each \( \omega \in \mathcal{N}^* \) completes the proof.

\[\text{Appendix C: Proof of Proposition 1}\]

The proof has two parts. First, I construct a function \( f : \mathcal{N} \to \mathbb{R} \) such that for all units \( \omega \in \mathcal{N}^* \), \( \sum_{i \in \mathcal{N}^*} f_i = \sigma^\omega \). (Note that in general, any subsystem \( i \) will be in more than one unit \( \omega \), and so this function \( f \) is the solution to a set of coupled equations.) I will then apply the inclusion-exclusion principle with this \( f \) in order to replace the first in-ex sums over unit \( \omega \) on the RHS of Eq. (15) with a conventional sum over subsystems, \( \sum_i f_i \).

In the second part of the proof I use the hypothesized existence of a centering distribution to provide a lower bound on \( \sum_i f_i \), expressed purely in terms of \( p_0 \) and \( p_{\mathcal{L}} \). Plugging in to Eq. (15) then completes the proof.

To begin, for each unit \( \omega \), define \( \overline{\omega} \) as the set of all subsystems in \( \omega \) that are not in any of the units in \( \text{desc}(\omega) \). (As an example, in Fig. 1, \( \overline{\omega} \) is the pair of subsystems 1 and 2.) Because \( \mathcal{N}^* \) is closed under intersections and covers \( \mathcal{N} \), every subsystem is in \( \overline{\omega} \) for exactly one unit \( \omega \in \mathcal{N}^* \). Moreover, because there are no orphan units allowed, \( \overline{\omega} \) is nonempty for every unit \( \omega \). Note that if \( \omega \subset \omega' \) for two units \( \omega, \omega' \), then the subsystems in \( \omega \) must evolve independently of the states of any subsystems in \( \omega' \), but the reverse need not be true. In other words, if there is an edge from \( \omega' \) to \( \omega \), then there may be subsystems in \( \omega' \) whose dynamics depends on the state of subsystems in \( \omega \), but not vice-versa.

For all \( j \in \mathbb{N} \), let \( \Omega_j \) be the set of all nodes in \( \Gamma_{\mathcal{N}^*} \), with height \( j \). So in particular, \( \Omega_1 \) is the the set all units with no subunits. For every \( \omega \in \Omega_1 \), for all subsystems \( i \in \omega \), set

$$f_i := \frac{\sigma^\omega}{|\omega|} \quad \text{(C1)}$$

So by construction, for all \( \omega \in \Omega_1 \),

$$\sum_{i \in \omega} f_i = \sigma^\omega \quad \text{(C2)}$$

Since \( \sigma^\omega \geq 0 \) for all \( \omega \in \mathcal{N}^* \), this means that \( \sum_{i \in \omega} f_i \geq 0 \) for all \( \omega \in \Omega_1 \).

Next, note that any for any \( \omega \in \mathcal{N}^* \), the set of units in \( \text{desc}(\omega) \) is closed under intersections. This allows us to define

$$f_i := \frac{1}{|\omega|} \left( \sigma^\omega - \sum_{\omega'' \in \text{desc}(\omega)} \sigma^{\omega''} \right) \quad \text{(C3)}$$

for all \( j \in \mathcal{N} \), all \( \omega \in \Omega_j \), and all \( i \in \overline{\omega} \). (Note that any such \( i \) will be assigned a value \( f_i \) exactly once in this procedure.)

In general, it could be that \( f_j \) is negative. Note though that since no two units in \( \Omega_1 \) have any overlap, for all \( \omega \in \Omega_2 \),

$$\sum_{\omega' \in \text{desc}(\omega)} \sigma^{\omega'} = \sum_{\omega' \in \text{desc}(\omega)} \sum_{i \in \omega'} f_i \quad \text{(C4)}$$
Therefore by Eq. (C3), for all \( \omega \in \Omega_2 \),
\[
\sum_{i \in \omega} f_i = \sum_{i \in \text{desc}(\omega)} f_i + \sum_{i \in \omega} f_i = \sum_{\omega' \in \text{desc}(\omega)} \sigma^{\omega'} + \sum_{i \in \omega} f_i \tag{C7}
\]
\[
\omega \in \text{desc}(\omega) \wedge \sum_{i \in \omega} f_i = \sigma^{\omega} \tag{C8}
\]
i.e., Eq. (C2) holds for unit \( \omega \).

Next, assume that Eq. (C2) holds for all \( \omega \in \Omega_{k-1} \) for some integer \( k > 2 \). Then by Eq. (C3) and the inclusion-exclusion principle, for any \( \nu \in \Omega_k \),
\[
\sum_{i \in \nu} f_i = \sum_{i \in \text{desc}(\nu)} f_i + \sum_{i \in \nu} f_i = \sum_{\nu' \in \text{desc}(\nu)} \sigma^{\nu'} + \sum_{i \in \nu} f_i \tag{C9}
\]
\[
\sum_{\nu' \in \text{desc}(\nu)} \omega \in \text{desc}(\nu') \sum_{i \in \nu} f_i = \sigma^{\nu} \tag{C10}
\]
\[
\omega \in \text{desc}(\nu) \sum_{i \in \nu} f_i = \sigma^{\nu} \tag{C11}
\]
\[
\omega \in \text{desc}(\nu) \omega \in \text{desc}(\nu) \sum_{i \in \nu} f_i = \sigma^{\nu} \tag{C12}
\]
(Note that the inclusion-exclusion principle holds for arbitrary functions \( f \), not just for nowhere-negative functions.) Since \( \sigma^{\nu} \geq 0 \), this means that we are guaranteed that \( \sum_{i \in \nu} f_i \geq 0 \).

Iterate this procedure going from nodes in \( \Omega_{k-1} \) to those in \( \Omega_k \) until all units have been considered, so that values \( f_i \) have been assigned to all subsystems \( i \in \mathcal{N} \). By induction, at the end of this procedure, for all units \( \omega \in \mathcal{N}^* \), Eq. (C2) will hold and \( \sum_{i \in \omega} f_i \geq 0 \). In addition, by the inclusion-exclusion principle,
\[
\sum_{\omega \in \mathcal{N}^*} \omega \in \text{desc}(\omega) \sum_{i \in \omega} f_i = \sum_{i \in \mathcal{N}} f_i \tag{C13}
\]
\[
\omega \in \mathcal{N}^* \sum_{\omega \in \text{desc}(\omega)} \sum_{i \in \omega} f_i = \sum_{i \in \mathcal{N}} f_i \tag{C14}
\]
(Note that since \( \mathcal{N} \notin \mathcal{N}^* \), Eq. (C12) does not imply that \( \sum_{i \in \mathcal{N}} f_i = \sigma^{\mathcal{N}} \). So we cannot combine Eq. (C14) with the fact that global EP is \( \geq 0 \) to establish that \( \sum_{\omega \in \mathcal{N}^*} \sigma^{\omega} \) is also non-negative.)

It will be convenient to define new variables that equal sums of \( f_i \) over small sets of subsystems \( i \). For all \( \omega \in \mathcal{N}^* \), define
\[
\omega \in \mathcal{N}^* g_\omega := \sum_{i \in \omega} f_i \tag{C15}
\]
\[
\omega \in \mathcal{N}^* \omega \in \text{desc}(\omega) \sigma^{\omega} - \sum_{\omega' \in \text{desc}(\omega)} \sigma^{\omega'} \tag{C16}
\]
where the second line follows from Eq. (C3). Since each subsystem \( i \) is in \( \text{desc}(\omega) \) for exactly one unit \( \omega \), for any \( \nu \in \mathcal{N}^* \),
\[
g_\nu + \sum_{\omega \in \text{desc}(\nu)} g_\omega = \sum_{i \in \nu} f_i + \sum_{\omega \in \text{desc}(\nu)} \sum_{i \in \omega} f_i \tag{C17}
\]
\[
= \sum_{i \in \nu} f_i \tag{C18}
\]
\[
\omega \in \mathcal{N}^* \omega \in \text{desc}(\omega) g_\omega = \sum_{i \in \nu} f_i \tag{C19}
\]
where the last line uses Eq. (C12). Using similar reasoning shows that \( \sum_{\omega \in \mathcal{N}^*} g_\omega = \sum_{i \in \mathcal{N}} f_i \). Combining this with Eq. (C14) and Eq. (C15) gives
\[
\omega \in \mathcal{N}^* \omega \in \text{desc}(\omega) g_\omega = \sum_{i \in \mathcal{N}} f_i \tag{C20}
\]
This completes the first part of the proof. In the second part I derive a lower bound on the RHS of Eq. (C20). First, to reduce the complexity of the equations, I will translate all distributions \( \alpha \) into binary-valued vectors:

1. \( \hat{\alpha} = \alpha \) for any type-1 distribution \( \alpha \);
2. \( \hat{\alpha} = \alpha \sum_{\omega \in \text{desc}(\omega)} \omega \in \mathcal{N}^* f_\omega \) for any type-2 distribution \( \alpha \);
3. \( \hat{\alpha} = \alpha \sum_{\omega \in \mathcal{N} \setminus \mathcal{N}^*} f_\omega \) for any type-3 distribution \( \alpha \);
4. \( \hat{\alpha} = \alpha \sum_{\omega \in \mathcal{N} \setminus \mathcal{N}^*} f_\omega \) for any type-4 distribution \( \alpha \);

(Note that every component of every vector \( \hat{\alpha} \) is either a 0 or a 1. I will refer to any vectors that obey condition (1) as type-1 vectors, and similarly for vectors obeying conditions (2), (3) and / or (4).)

Now make three suppositions. First, suppose that for all vectors \( \hat{\alpha} \) of types 1, 2 or 3,
\[
\omega \in \mathcal{N}^* \omega \in \text{desc}(\omega) g_\omega \hat{\alpha} \geq 0 \tag{C21}
\]
Next, suppose that for all vectors \( \hat{\alpha} \) of type-4, and all associated dependency structures \( \hat{\alpha}^* \),
\[
\omega \in \mathcal{N}^* \omega \in \text{desc}(\omega) g_\omega \hat{\alpha} \geq -\Delta \hat{\alpha}^* \tag{C22}
\]
Now, hypothesize that there is a centering vector \( \gamma \) all of whose components are non-negative such that
\[
\omega \in \mathcal{N}^* \omega \in \text{desc}(\omega) \gamma \hat{\alpha} \hat{\alpha} \omega + \sum_{\omega \in \mathcal{N}^*} \gamma \hat{\alpha} \hat{\alpha} \omega = 1 \tag{C23}
\]
for all \( \omega \in \mathcal{N}^* \). Applying \( \sum_{\omega \in \mathcal{N}^*} g_\omega \) to both sides of Eq. (C23) and then plugging in Eqs. (C21) and (C22) establishes that if those three equations hold,
\[
\omega \in \mathcal{N}^* \omega \in \text{desc}(\omega) g_\omega \geq -\sum_{\omega \in \mathcal{N}^*} \gamma \hat{\alpha} \hat{\alpha} \omega \tag{C24}
\]
Plugging this into Eq. (C20) shows that if we can prove that the suppositions Eqs. (C21) and (C22) always hold, then we will have proven that for any centering vector \( \gamma \),

\[
\sigma \geq -\Delta I^{N_*} - \sum_{d \in V^2(N_\gamma)} \gamma_d \Delta I^{d_*} \tag{C25}
\]

To begin, use Eq. (C16) to conclude that \( g_\omega = \sigma_\omega \) for all \( \omega \in \Omega_1 \) (which have no descendants) and so \( g_\omega \geq 0 \) for all \( \omega \in \Omega_1 \). Next, combine this fact that \( g_\omega = \sigma_\omega \) for all leaf nodes \( \omega \) with Eqs. (9) and (C19) to also conclude that \( g_\omega \geq 0 \) for all \( \omega \in \Omega_2 \). Combining these two results means that Eq. (C21) holds for all type-1 vectors \( \hat{\alpha} \).

Next, note that Eq. (C19) means that for any \( \omega \in N^* \),

\[
\sum_{\omega' \subseteq \omega} g_{\omega'} = \sigma_\omega \tag{C26}
\]

So by the non-negativity of local EP, for all \( \omega \in N^* \),

\[
\sum_{\omega} \sum_{\omega' \subseteq \omega} \delta(\omega', \omega) g_{\omega'} \geq 0 \tag{C27}
\]

This means that Eq. (C21) holds for all type-2 vectors \( \hat{\alpha} \).

Now consider any pair of nodes \( \omega \in N^*, v \subset \omega \). Using Eq. (C19) for both \( \omega \) and \( v \) and then applying Eq. (8) establishes that

\[
0 \leq \sum_{\omega' \in \text{fa}(\omega)} g_{\omega'} - \sum_{\omega' \in \text{fa}(v)} g_{\omega'}
= \sum_{\omega' \in \text{fa}(\omega) \setminus \text{fa}(v)} g_{\omega'} \tag{C28}
\]

This means that Eq. (C21) also holds for all type-3 vectors \( \hat{\alpha} \). Combining establishes our first goal, of showing that Eq. (C21) holds for all vectors \( \hat{\alpha} \in V^1(N^*) \), of types 1, 2 or 3.

Next, consider any pair of a unit \( \omega \) and set of subsystems \( M(\omega) \subset \omega \) such that there is a dependency structure \( M(\omega)^* \subset \omega^* \). Use Eq. (C12), the inclusion-exclusion principle, Eq. (C15), and then Eq. (C19) to expand

\[
\sigma^\omega = \sum_{\omega' \in \text{M}(\omega)^*} \sigma^\omega' - \sum_{\omega' \in \text{M}(\omega)^*} \left[ \sum_{i \in \text{fa}(\omega')} f_i \right] \tag{C30}
\]

\[
= \sigma^\omega - \sum_{i \in \text{M}(\omega)} f_i \tag{C31}
\]

\[
= \sigma^\omega - \sum_{\omega' \in \text{M}(\omega)^*} \sum_{i \in \omega} f_i \tag{C32}
\]

\[
= \sigma^\omega - \sum_{\omega' \in \text{M}(\omega)^*} g_{\omega'} \tag{C33}
\]

\[
= g_\omega + \sum_{\omega' \in \text{desc}(\omega)} g_{\omega'} - \sum_{\omega' \in \text{M}(\omega)^*} g_{\omega'} \tag{C34}
\]

Eq. (16) then establishes that

\[
\sum_{\omega' \in \text{M}(\omega)^*} g_{\omega'} \geq -\Delta I^{M(\omega)^*} \tag{C37}
\]

Plugging this into the definition of type-4 vectors for \( \hat{\alpha} = M(\omega) \) and \( \hat{\alpha} = M(\omega)^* \) confirms that Eq. (C21) holds.

Combining establishes that for any centering vector \( \gamma \),

\[
\sigma \geq -\Delta I^{N_*} - \sum_{d \in V^2(N_\gamma)} \gamma_d \Delta I^{d_*} \tag{C38}
\]

Finally, normalize each vector \( \hat{\alpha} \) to recover the distributions \( \alpha \), and define the distribution \( \pi(\alpha) \) by normalizing \( \gamma(\alpha) \). It follows that \( \pi(\alpha) \) is the uniform distribution, so that \( \pi \) is a centering distribution.

### Appendix D: Proof that any dependency structure is centered

To begin, choose \( V^1(N^*) \) to be the set of all type-1 distributions, i.e., \( V^1(N^*) \) is the set of all distributions \( \delta^\omega \) for any \( \omega \) of height \( \leq 2 \).

Next, for each \( \omega \) of height greater than 2, plug \( M(\omega) = \text{desc}(\omega) \) and any arbitrary single one of the possible dependency structures \( M(\omega)^* \) into Eq. (C23) to define a distribution

\[
\alpha(\omega) = \sum_{\omega' \in \text{M}(\omega)^*} \delta^\omega' \tag{D1}
\]

and associated dependency structure \( \alpha(\omega)^* \). Choose \( V^2(N^*) \) to be the set of all such \( \alpha(\omega) \), one per \( \omega \), as one ranges over all \( \omega \) of height greater than 2.

By construction, \( V(N^*) \) is exactly the set of all vectors \( \delta^\omega \) as one ranges over all \( \omega \in N^* \). Accordingly, the sum of all distributions in \( V(N^*) \) is 1, and the average of those distributions is the uniform distribution, \( (1/|N^*|, 1/|N^*|, \ldots) \). Therefore the set of those vectors is centered, where the centering distribution \( \pi(\alpha) = 1/|N^*| \) for all \( \alpha \in V(N^*) \). This completes the proof.

### Appendix E: Proof of Proposition 2

First, note that for a uniform distribution over the states of the composite system, the entropy of every subsystem \( i \) with \( |X_i| \) states is \( \ln |X_i| \). Furthermore, no matter
what the unit structure is, one can create a process consistent with that structure that results in this uniform distribution as the final distribution. (Just choose the rate matrix of every subsystem $i$ to uniformly randomize $x_i$ by the end of the process.)

Now assign values $f_i = \ln |X_i|$ to all subsystems. By construction, for all units $\omega$, $\sum_{i \in \omega} f_i = S^\omega$. In addition,

$$\sum_{\omega} \left[ \sum_{i \in \omega} f_i \right] = \sum f_i$$

(E1)

by the inclusion-exclusion principle. But the sum on the RHS just equals $S^N$, the entropy of the full system, since the subsystems are statistically independent under the final distribution. Therefore $I^{N^*} = 0$ for this ending distribution. Similarly, for this ending distribution, $I^\alpha = 0$ for every $\alpha \in V^2(N^*)$.

So to find a situation where Proposition 1 holds, it suffices to find an initial distribution $p$ such that $I^{N^*}(p) \geq 0$ while $I^\alpha(p) = 0$ for all $\alpha \in V^2(N^*)$. To do that, label the states of each subsystem $i$ by the first $|X_i|$ counting numbers. Define $M := \min_{i \in N}|X_i|$, and define $\Gamma^R_{N^*}$ as the (units corresponding to the) root nodes of the dependency graph $\Gamma_{N^*}$. Note that since by hypothesis $N \notin N^*$, there must be at least two distinct root nodes in $\Gamma^R_{N^*}$. Furthermore, since there are no orphan units, all of (the units corresponding to) those root nodes contain subsystems that are not in any other units, i.e., for all $\omega \in \Gamma^R_{N^*}$, $\overline{\omega} \neq \emptyset$.

Next, define $T := \cup_{\omega \in \Gamma^R_{N^*}} \overline{\omega}$, and fix the state of each subsystem $j \notin T$, i.e., set the distribution over the state of that subsystem to a delta function. Set the joint distribution over the remaining subsystems to

$$p(x_T) = \frac{1}{M} \sum_{k=1}^M \prod_{i \in T} \delta(x_i, k)$$

(E2)

So all subsystems that only occur in a single root unit are perfectly coupled with one another, with a uniform distribution over the set of $M$ possible joint states they can adopt.

The entropy of the full joint distribution defined this way is $S(p) = \ln M$. So to prove that $I^{N^*}(p) \geq 0$ we need to show that

$$\sum_{\omega} S(p_\omega) \geq \ln M$$

(E3)

To do that, assign the value $f_i = \ln |X_i|$ to all subsystems $j \notin T$. For each subsystem $j \in T$, where $\omega(j)$ is the unique unit containing $j$, assign the value

$$f_j = \ln \frac{M}{|\overline{\omega(j)}|}$$

(E4)

where $|\overline{\omega(j)}|$ is the number of elements in $\overline{\omega(j)}$.

By construction, for all units $\omega \in N^*$,

$$S(p_\omega) = \sum_{j \in \omega} f_j$$

(E5)

Accordingly, by the inclusion-exclusion principle

$$\sum_{\omega} S(p_\omega) = \sum_{i \in N} f_i$$

(E6)

$$= \sum_{\omega \in \Gamma^R_{N^*}} \ln M$$

(E7)

Since $\Gamma^R_{N^*}$ contains at least two units, this means that

$$\sum_{\omega} S(p_\omega) > \ln M = S(p)$$

(E8)

In addition, the entropy of every unit $\omega \notin \Gamma^R_{N^*}$ equals 0. Accordingly, $I^\alpha(p) = 0$ for all $\alpha \in V^2(N^*)$.

This completes the proof.