A LETTERS JOURNAL EXPLORING
THE FRONTIERS OF PHYSICS

Società Italiana
di Fisica

LETTER

Marginal and conditional second laws of thermodynamics

To cite this article: Gavin E. Crooks and Susanne Still 2019 *EPL* **125** 40005

EPL, **125** (2019) 40005 www.epljournal.org
doi: 10.1209/0295-5075/125/40005

Marginal and conditional second laws of thermodynamics

Gavin E. Crooks[1] and Susanne Still[2]

[1] *Berkeley Institute for Theoretical Science - Berkeley, CA, USA*
[2] *University of Hawai'i at Mānoa, Department of Information and Computer Sciences and Department of Physics and Astronomy - Honolulu, HI 96822, USA*

received 5 September 2018; accepted in final form 16 January 2019
published online 25 March 2019

PACS 05.70.Ln – Nonequilibrium and irreversible thermodynamics
PACS 05.40.-a – Fluctuation phenomena, random processes, noise, and Brownian motion

Abstract – We show that the total entropy production of a strongly coupled bipartite system can be partitioned into components, which can be used to define local versions of the Second Law that are valid without the usual idealization of weak coupling. The key insight is that causal intervention offers a way to identify those parts of the entropy production that result from feedback between the subsystems. All central relations describing the thermodynamics of strongly coupled systems follow from this observation in a few lines.

Rudolf Clausius' famous statement of the "second fundamental theorem in the mechanical theory of heat" is that "*The entropy of the universe tends to a maximum*" [1]. Although this proclamation has withstood the test of time, in practice measuring the entropy of the entire universe is difficult. As an alternative we can apply the Second Law to any system isolated from outside interactions (a Universe unto itself), as, for example, in Planck's statement of the Second Law: "*Every process occurring in nature ... the sum of the entropies of all bodies taking part in the process is increased*" [2]. Of course, perfectly isolating any system or collection of systems from outside influence is also difficult.

Over the last 150 years thermodynamics has progressed by adopting various idealizations which allow us to isolate and measure that part of the total universal entropy change that is relevant to the behavior of the system at hand. These idealizations include heat reservoirs, work sources, and measurement devices [3,4]. More recently information engines ("Maxwell demons" [5,6]) have been added to the canon to represent idealized computational resources [4,7–10].

In this paper, we demonstrate that we do not need, in principle, to resort to these idealizations. We show how the thermodynamics of strongly coupled systems follow in a straightforward manner from a causal decomposition of their dynamics. This unifying perspective greatly simplifies the treatment, allowing us to assimilate the large recent literature on the thermodynamics of coupled systems in a few short pages. Looking at the problem in the right way then makes it easy to show that conditional and marginalized versions of the Second Law hold locally, even when the system of interest is strongly coupled to other driven, non-equilibrium systems.

Partitions of entropy. – Before considering the partitioning of dissipation, let us recall the partitioning of entropy in information theory. Suppose we have a pair of interacting systems X and Y, whose states are x and y, respectively. The joint entropy of the total system is

$$S_{\mathrm{X,Y}} = -\sum_{x,y} p(x,y) \ln p(x,y). \qquad (1)$$

The marginal entropy S_{X} of system X is the entropy of the marginal distribution, $p(x)$, obtained by summing over the states of the other system,

$$S_{\mathrm{X}} = -\sum_{x} p(x) \ln p(x), \qquad p(x) = \sum_{y} p(x,y). \qquad (2)$$

The conditional entropy,

$$S_{\mathrm{X|Y}} \times = S_{\mathrm{X,Y}} - S_{\mathrm{Y}} = -\sum_{y} p(y) \sum_{x} p(x|y) \ln p(x|y), \qquad (3)$$

is the average entropy of system X given that we observe the state of system Y.

It is also useful to define the pointwise (specific) entropies of individual realizations of the system, whose ensemble averages are the entropies defined previously:

$$s(x,y) = -\ln p(x,y) = s(x|y) + s(y), \tag{4a}$$
$$s(x) = -\ln p(x), \tag{4b}$$
$$s(x|y) = -\ln p(x|y). \tag{4c}$$

The negative log-probability, eq. (4b), has been called "surprisal" in information theory [11].

Dissipation. – Let us now consider dynamical trajectories of a bipartite system. We assume that each side of the system is coupled to idealized constant temperature heat reservoirs, with reciprocal temperature $\beta = 1/k_\mathrm{B}T$, and idealized work sources. We label the controlled parameters of the work source coupled to subsystem X by u, and those corresponding to subsystem Y by v. Although the coupling to the heat baths and work sources are idealized, the two subsystems, X and Y, can be strongly coupled to each other.

A core tenet of non-equilibrium thermodynamics is the *detailed fluctuation theorem* which equates the entropy production (or *dissipation*) Σ to the log ratio of the probabilities of forward and time-reversed trajectories [12–14]:

$$\Sigma_{\mathrm{X,Y}} = \ln \frac{p(\vec{x}, \vec{y}; \vec{u}, \vec{v})}{p(\overleftarrow{x}, \overleftarrow{y}; \overleftarrow{u}, \overleftarrow{v})}. \tag{5}$$

Dissipation is a consequence of breaking time-reversal symmetry. Here, $\vec{x}$ and $\vec{y}$ are trajectories of systems X and Y, respectively, generated while the systems are driven by the external protocols $\vec{u}$ and $\vec{v}$, respectively, which are fixed functions of time. The trajectory $\overleftarrow{x}$ denotes $\vec{x}$ in reverse, running time backwards. Consequently $p(\vec{x}, \vec{y}; \vec{u}, \vec{v})$ is the probability of the forward time trajectories under forward time dynamics, given the forward time protocols, whereas $p(\overleftarrow{x}, \overleftarrow{y}; \overleftarrow{u}, \overleftarrow{v})$ is the probability of the conjugate time-reversed trajectories, given the time-reversed driving. We use a semicolon before the controls to emphasize that the protocols are fixed parameters. For notational simplicity we typically suppress the explicit dependence of the dynamics on the protocols, writing $p(\overleftarrow{x}, \overleftarrow{y})$ for $p(\overleftarrow{x}, \overleftarrow{y}; \overleftarrow{u}, \overleftarrow{v})$, for example.

Suppose we only observe the behavior of one of the subsystems. We can still define marginal trajectory probabilities and the marginal entropy production,

$$\Sigma_{\mathrm{X}} = \ln \frac{p(\vec{x})}{p(\overleftarrow{x})}, \tag{6}$$

and, by similar reasoning, the conditional entropy production,

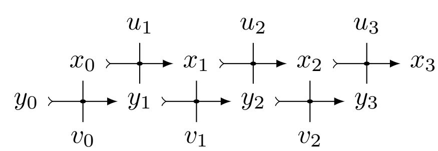
$$\Sigma_{\mathrm{X|Y}} = \ln \frac{p(\vec{x}|\vec{y})}{p(\overleftarrow{x}|\overleftarrow{y})} = \ln \frac{p(\vec{x},\vec{y})}{p(\overleftarrow{x},\overleftarrow{y})} \frac{p(\overleftarrow{y})}{p(\vec{y})}$$
$$= \Sigma_{\mathrm{X,Y}} - \Sigma_{\mathrm{Y}}. \tag{7}$$

Thus, we can partition the total dissipation into local components. However, in order to make these definitions of marginal and conditional dissipation concrete we have to explore their physical meaning.

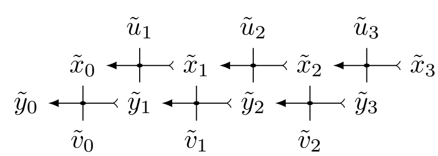
Dynamics. – The first fluctuation theorems for systems driven far from thermodynamic equilibrium were derived under the assumption that a system of interest is driven out of thermodynamic equilibrium by a time-dependent, but fixed, protocol (often implicitly assumed to be applied by an experimenter) [12,13]. Later, two extended scenarios were explored: i) the protocol is a function of the state of the system [7], and ii) the protocol itself is stochastic [15]. In the context of the first extension, the phrase "feedback" was employed, to denote that the system feeds back onto the protocol. Put together, the two extensions result in a strongly coupled bipartite system, as treated in this paper. The situation is now symmetric, both subsystems are feeding back onto each other.

To make the discussion unambiguous, we adopt a specific model of the intersystem dynamics. We could opt for classical mechanics [4], or coupled Langevin dynamics [16], or a continuous time Markov process [17]. But we feel the discussion is most transparent when the dynamics are represented by coupled, discrete time Markov chains. The dynamics of the joint system are assumed to be Markov, and the dynamics of each subsystem are conditionally Markov given the state of the other subsystem. The marginal dynamics are not Markov when we do not know the hidden dynamics of the other subsystem.

We can use a causal diagram [18–20] to illustrate the time label conventions for the trajectories of the system and control parameters. First one subsystem updates, then the other, and so on, until time step τ (set to 3 in the following diagram):



Horizontal arrows indicate time evolution, and the other connections indicate causation, where the dynamics of one subsystem are influenced by the external parameters and the current state of the other subsystem. For the corresponding time-reversed trajectory the horizontal arrows flip, but the vertical connections remain unchanged:

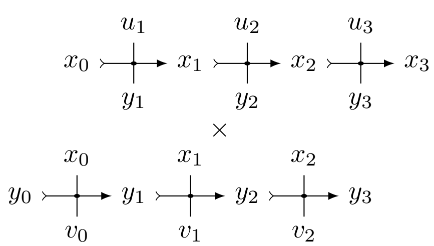


Here the tilde labels the time-reversed configurations of the time-reversed trajectory, $\tilde{x}_t$.

The probability for the joint trajectory is related to total dissipation via the fluctuation theorem, eq. (5). This joint probability naturally splits into two products, each of individual transition probabilities for the respective subsystem. That is in forward time:

$$\begin{aligned}
&p(\vec{x},\vec{y}|x_0,y_0) = \\
&p(y_1|y_0,x_0)p(x_1|x_0,y_1)p(y_2|y_1,x_1)p(x_2|x_1,y_2) \\
&\quad\ldots p(y_\tau|y_{\tau-1},x_{\tau-1})p(x_\tau|x_{\tau-1},y_\tau) \\
&= \prod_{t=0}^{\tau-1} p(y_{t+1}|y_t,x_t) \times \prod_{t=0}^{\tau-1} p(x_{t+1}|x_t,y_{t+1}) \\
&\equiv q(\vec{y};\vec{x},y_0) \times q(\vec{x};\vec{y},x_0). \qquad (8)
\end{aligned}$$

Here, $p(x_{t+1}|x_t,y_{t+1})$ is the probability of jumping from state x_t to x_{t+1} given the current state of the other subsystem (and knowledge of the driving protocol). The expressions $q(\vec{y};\vec{x},y_0)$ and $q(\vec{x};\vec{y},x_0)$ denote the trajectory probabilities of one system given a *fixed* trajectory of the other system. Once again we set off parameters that are fixed (rather than observed and coevolving) with a semicolon. These probabilities are not the same as the conditional distributions $p(\vec{x}|\vec{y},y_0)$, which describe the original process where both systems coevolve and influence each other. We have chosen to use a different symbol (q) for these trajectory probabilities to make this distinction abundantly clear. The dynamics corresponding to the decomposition can be represented by the following pair of diagrams:

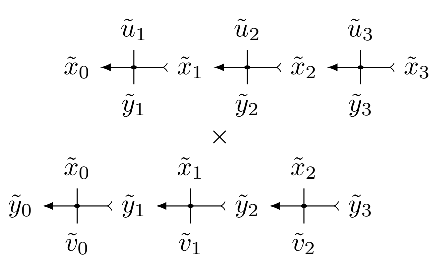


The decomposition of the joint probability in eq. (8) is referred to as a causal intervention [19]. We emphasize that this decomposition is central to disentangling the direct and indirect effects of intersystem coupling.

In reverse time, we have a similar decomposition,

$$\begin{aligned}
&p(\overleftarrow{x},\overleftarrow{y}|\tilde{x}_\tau,\tilde{y}_\tau) = \\
&\prod_{t=0}^{\tau-1} p(\tilde{x}_t|\tilde{x}_{t+1},\tilde{y}_{t+1}) \times \prod_{t=0}^{\tau-1} p(\tilde{y}_t|\tilde{x}_t,\tilde{y}_{t+1}) \\
&\equiv q(\overleftarrow{x};\overleftarrow{y},\tilde{x}_\tau) \times q(\overleftarrow{y};\overleftarrow{x},\tilde{y}_\tau), \qquad (9)
\end{aligned}$$

which is represented by the following pair of diagrams:



This makes it clear that to calculate dissipation in the presence of feedback, we cannot simply replace the coupled reverse time process probability, $p(\overleftarrow{x}|\overleftarrow{y})$ in eq. (7), by its uncoupled counterpart, $q(\overleftarrow{x};\overleftarrow{y}) = q(\overleftarrow{x};\overleftarrow{y},\tilde{x}_\tau)p(\tilde{x}_\tau)$.

Detailed fluctuation theorem. – For the complete system, the total pathwise entropy production consists of the change in the entropy of the environment (due to the flow of heat from the baths) and a boundary term $\Delta s_{\mathrm{X,Y}}$ [12–14],

$$\Sigma_{\mathrm{X,Y}} = \Delta s_{\mathrm{X,Y}} - \beta Q_{\mathrm{X,Y}}. \qquad (10)$$

This boundary term is the difference in pointwise entropy between the initial configurations of the forward and reverse trajectories,

$$\Delta s_{\mathrm{X,Y}} = -\ln p(\tilde{x}_\tau,\tilde{y}_\tau) + \ln p(x_0,y_0). \qquad (11)$$

Typically, we either assume that the system is initially in thermodynamic equilibrium for both the forward and reversed processes (as we do for the Jarzynski equality [21]), or we assume that the final ensemble of the forward process is the same as the initial, time-reversed probabilities of the reversed process, $p(\tilde{x}_\tau,\tilde{y}_\tau) = p(x_\tau,y_\tau)$ [14,22]. In general, the initial ensembles need not have any simple relationship: for instance, we might be observing a short segment of a much longer driven process.

The energy E of the total system can be written as the two subsystem Hamiltonians plus an interaction term,

$$E_{\mathrm{X,Y}}(x,y;u,v) = E_{\mathrm{X}}(x;u) + E_{\mathrm{Y}}(y;v) + E^{\mathrm{int}}_{\mathrm{X:Y}}(x,y). \quad (12)$$

The external baths and control parameters couple to the internal states of each system separately and do not couple directly to the interaction energy, which ensures that the source of energy flowing into the system is unambiguous.

The heat is the flow of energy into the system due to interactions with the bath [12,21,23–25]. We can split the total heat into the heat flow for each of the two subsystems, $Q_{\mathrm{X,Y}} = Q_{\mathrm{X}} + Q_{\mathrm{Y}}$,

$$\begin{aligned}
Q_{\mathrm{X}} &= \sum_{t=0}^{\tau-1}\Big[E_{\mathrm{X}}(x_{t+1},u_{t+1}) + E^{\mathrm{int}}_{\mathrm{X:Y}}(x_{t+1},y_{t+1}) \\
&\quad -E_{\mathrm{X}}(x_t,u_{t+1}) - E^{\mathrm{int}}_{\mathrm{X:Y}}(x_t,y_{t+1})\Big], \qquad (13a)\\
Q_{\mathrm{Y}} &= \sum_{t=0}^{\tau-1}\Big[E_{\mathrm{Y}}(y_{t+1},v_t) + E^{\mathrm{int}}_{\mathrm{X:Y}}(x_t,y_{t+1}) \\
&\quad -E_{\mathrm{Y}}(y_t,v_t) - E^{\mathrm{int}}_{\mathrm{X:Y}}(x_t,y_t)\Big]. \qquad (13b)
\end{aligned}$$

Heat flow is the change in energy when the state of a system updates with fixed configurations of the connected systems. Thus, the heat flow into subsystem X involves both a flow of energy from the bath, and a flow of energy mediated by the interacting subsystem Y. Conversely for system Y.

Local detailed fluctuation theorems. – If the trajectory of system Y is fixed, then its dynamics act as an idealized work source to system X, and we can write down a standard fluctuation theorem for system X alone,

$$\ln \frac{q(\vec{x};\vec{y},x_0)}{q(\overleftarrow{x};\overleftarrow{y},\tilde{x}_\tau)}\frac{p(x_0)}{p(\tilde{x}_\tau)} = \Delta s_{\mathrm{X}} - \beta Q_{\mathrm{X}}. \tag{14}$$

This is the fluctuation theorem we would obtain were there no feedback from X to Y[1]. With feedback, eq. (14) no longer correctly describes the entropy production of subsystem X. Assuming that it does, leads to apparent contradictions that would imply that the Second Law has to be modified [7].

What is the quantity that correctly describes the entropy production of subsystem X while coupled to the coevolving subsystem Y? The answer depends on what information the observer has at hand. If the coevolving state of subsystem Y can be observed at all times, then the conditional entropy production, $\Sigma_{\mathrm{X|Y}}$, eq. (7), best describes the dissipation encountered by system X alone. In the absence of this knowledge, we have to integrate out the state-space trajectories of Y, and therefore we can make statements only about the marginal dissipation, Σ_{X}, eq. (6). While we have no guarantee that the average of eq. (14) is positive when there is feedback between the systems, we do know that both the marginal and the conditional entropy production obey the Second Law, because their averages can be written as Kullback-Leibler divergences [26–29], which means that they are non-negative quantities.

We can now write the marginal dissipation using the causal decompositions, eqs. (8) and (9):

$$\begin{aligned}
\Sigma_{\mathrm{X}} &= \ln \frac{p(\vec{x})}{p(\overleftarrow{x})} = \ln \frac{p(\vec{x},\vec{y})}{p(\overleftarrow{x},\overleftarrow{y})}\frac{p(\overleftarrow{y}|\overleftarrow{x})}{p(\vec{y}|\vec{x})} && (15a)\\
&= \ln \frac{p(\vec{x},\vec{y}|x_0,y_0)}{p(\overleftarrow{x},\overleftarrow{y}|\tilde{x}_\tau,\tilde{y}_\tau)}\frac{p(x_0,y_0)}{p(\tilde{x}_\tau,\tilde{y}_\tau)}\frac{p(\overleftarrow{y}|\overleftarrow{x},\tilde{y}_\tau)}{p(\vec{y}|\vec{x},y_0)}\frac{p(\tilde{y}_\tau|\tilde{x}_\tau)}{p(y_0|x_0)} && (15b)\\
&= \ln \frac{q(\vec{y};\vec{x},y_0)}{q(\overleftarrow{y};\overleftarrow{x},\tilde{y}_\tau)}\frac{q(\vec{x};\vec{y},x_0)}{q(\overleftarrow{x};\overleftarrow{y},\tilde{x}_\tau)}\frac{p(\overleftarrow{y}|\overleftarrow{x},\tilde{y}_\tau)}{p(\vec{y}|\vec{x},y_0)}\frac{p(x_0)}{p(\tilde{x}_\tau)} && (15c)\\
&= \ln \frac{p(x_0)}{p(\tilde{x}_\tau)} + \ln \frac{q(\vec{x};\vec{y},x_0)}{q(\overleftarrow{x};\overleftarrow{y},\tilde{x}_\tau)} && (15d)\\
&\quad - \left(\ln \frac{p(\vec{y}|\vec{x},y_0)}{p(\overleftarrow{y}|\overleftarrow{x},\tilde{y}_\tau)} - \ln \frac{q(\vec{y};\vec{x},y_0)}{q(\overleftarrow{y};\overleftarrow{x},\tilde{y}_\tau)}\right)\\
&= \Delta s_{\mathrm{X}} - \beta Q_{\mathrm{X}} - \Sigma^{\mathrm{trn}}_{\mathrm{X}}. && (15e)
\end{aligned}$$

Here, we have written down in eq. (15a) the definition of the marginal dissipation (see eq. (6)), and expanded it using the definition of conditional probability $p(a|b) = p(a,b)/p(b)$. Then we split out the initial state probabilities in eq. (15b), and in eq. (15c) split the probability of the joint trajectory into components without feedback (see (8) and (9)). In eq. (15d) we gather the entropy production into 3 terms. The first two terms describe the entropy production of subsystem X if there were no feedback from X to Y (eq. (14)). The last term characterizes the thermodynamic effects of feedback. We identify this term as the *transferred dissipation*,

$$\Sigma^{\mathrm{trn}}_{\mathrm{X}} = \ln \frac{p(\vec{y}|\vec{x},y_0)}{p(\overleftarrow{y}|\overleftarrow{x},\tilde{y}_\tau)} - \ln \frac{q(\vec{y};\vec{x},y_0)}{q(\overleftarrow{y};\overleftarrow{x},\tilde{y}_\tau)}. \tag{16}$$

It consists of the difference between subsystem Y's conditional entropy production with feedback, compared to without feedback. The subscript indicates the energy sink, the subsystem into which energy is flowing. If system Y did not influence the behavior of system X then $q(\vec{y};\vec{x},y_0)$ would equal $p(\vec{y}|\vec{x},y_0)$ (and similarly for the time-reversed components), and the transferred dissipation would be zero.

With this insight, we can now appreciate the fact that causal intervention, and decomposition of the joint probability of the system as a whole into disconnected parts in both forward and reverse time (eqs. (8) and (9)) is crucial in thinking about feedback systems, because it allows us to decompose the entropy production, revealing the contribution due to feedback.

To summarize, the menagerie of local detailed fluctuation theorems (eqs. (5)–(7)) can be expressed solely in terms of differences in local pointwise entropies, eq. (4), the heat flow, eq. (13), and the transferred dissipation,

$$\begin{aligned}
\Sigma_{\mathrm{X,Y}} &= \Delta s_{\mathrm{X,Y}} - \beta Q_{\mathrm{X}} - \beta Q_{\mathrm{Y}}, && (17a)\\
\Sigma_{\mathrm{X}} &= \Delta s_{\mathrm{X}} - \beta Q_{\mathrm{X}} - \Sigma^{\mathrm{trn}}_{\mathrm{X}}, && (17b)\\
\Sigma_{\mathrm{X|Y}} &= \Delta s_{\mathrm{X|Y}} - \beta Q_{\mathrm{X}} + \Sigma^{\mathrm{trn}}_{\mathrm{Y}}. && (17c)
\end{aligned}$$

The joint entropy production $\Sigma_{\mathrm{X,Y}}$ is the usual dissipation of the entire system. The marginal entropy production Σ_{X} is the appropriate dissipation to consider if we cannot observe the dynamics of system Y, while the conditional entropy production $\Sigma_{\mathrm{X|Y}}$ should be considered when the dynamics of Y can be observed. This conditional entropy production contains a transferred dissipation from subsystem X to subsystem Y.

Transferred dissipation. – Transferred dissipation can be further decomposed into time-forward and time-reversed components (analogous to the decomposition of the entropy production rate into the Shannon entropy rate and a time-reversed entropy rate [26]),

$$\begin{aligned}
\Sigma^{\mathrm{trn}}_{\mathrm{Y}} &= \ln \frac{p(\vec{x}|\vec{y},x_0)}{q(\vec{x};\vec{y},x_0)} - \ln \frac{p(\overleftarrow{x}|\overleftarrow{y},\tilde{x}_\tau)}{q(\overleftarrow{x};\overleftarrow{y},\tilde{x}_\tau)}\\
&= \ln \frac{p(\vec{x}|\vec{y},x_0)q(\vec{y};\vec{x},y_0)}{p(\vec{y},\vec{x}|x_0,y_0)} - \ln \frac{q(\overleftarrow{y};\overleftarrow{x},\tilde{y}_\tau)p(\overleftarrow{x}|\overleftarrow{y},\tilde{x}_\tau)}{p(\overleftarrow{y},\overleftarrow{x}|\tilde{x}_\tau,\tilde{y}_\tau)}\\
&= \ln \frac{q(\vec{y};\vec{x},y_0)}{p(\vec{y}|y_0)} - \ln \frac{q(\overleftarrow{y};\overleftarrow{x},\tilde{y}_\tau)}{p(\overleftarrow{y}|\tilde{y}_\tau)}\\
&= T_{\vec{\mathrm{Y}}} - T_{\overleftarrow{\mathrm{Y}}}. && (18a)
\end{aligned}$$

[1]The boundary term $\Delta s_{\mathrm{X}} = -\ln p(\tilde{x}_\tau) + \ln p(x_0)$ is the pointwise entropy difference between the initial configurations of the forward and reverse trajectories of system X alone.

After some additional manipulation we recognize that the first term is the sum of the pointwise *transfer entropies* [30],

$$\begin{aligned} T_{\vec{\mathrm{Y}}} &= \ln\left(\prod_{t=0}^{\tau-1} p(y_{t+1}|y_t, x_t)\right) - \ln\left(\prod_{t=0}^{\tau-1} p(y_{t+1}|y_{0:t})\right) \\ &= \sum_{t=0}^{\tau-1} \ln\frac{p(y_{t+1}|y_{0:t}, x_{0:t})}{p(y_{t+1}|y_{0:t})} = \sum_{t=0}^{\tau-1} i(y_{t+1} : x_{0:t}|y_{0:t}). \end{aligned} \tag{19}$$

Here $i(a : b|c) = \ln p(a, b|c) - \ln p(a|c)p(b|c)$ is the pointwise conditional mutual information, and the slice notation $x_{a:b}$ is shorthand for the sequence $x_a, x_{a+1}, \ldots, x_b$. Thus, $T_{\vec{\mathrm{Y}}}$ is the total pointwise transfer entropy from X to Y for the forward time trajectory.

Transfer entropy has been investigated as a measure of Granger statistical causality [30–32], and has recently been recognized as a component of the thermodynamic dissipation [20,33–37]. However, transfer entropy can only equal the total transferred dissipation if we construe a process with feedback only in the time-forward dynamics, but no feedback in the time-reversed dynamics. Such time-reverse feedback-free reference systems, first studied in refs. [7,8], do not have a clear physical interpretation, because we cannot turn off the coupling between subsystems *only* during the time-reversed process. Therefore, we must include the time-reversed component of the transferred dissipation in order to fully appreciate the thermodynamic costs associated with interactions between subsystems.

Information engines. – An interesting idealized limit to consider is when the interaction energy is zero $E^{\mathrm{int}}_{\mathrm{X:Y}} = 0$, but the dynamics are still coupled. Removing the energetic component of the interaction forces us to carefully consider the role of information flow and computation in thermodynamics, and the relationship between information and entropy [7–10,33,35,38–42]. Although no energy flows between the systems, the transferred dissipation can still be non-zero. From the point of view of system X, system Y becomes a purely computational resource (a "demon"). This resource has an irreducible thermodynamic cost which is captured by the transferred dissipation. Neglecting this cost leads to Maxwell's demon paradoxes where the Second Law appears to be violated.

Local Second Law. – We are now in a position to express the averages of the total, marginal, and conditional dissipation by averages of the quantities we found in our decomposition (eq. (17)). Remember that we can also write them as Kullback-Leibler divergences, and hence each obeys a Second Law like inequality:

$$\begin{aligned} \langle \Sigma_{\mathrm{X,Y}} \rangle &= \Delta S_{\mathrm{X,Y}} - \beta\langle Q_{\mathrm{X}}\rangle - \beta\langle Q_{\mathrm{Y}}\rangle \qquad (20a) \\ &= \sum_{\vec{x},\vec{y}} p(\vec{x},\vec{y}) \ln\frac{p(\vec{x},\vec{y})}{p(\overleftarrow{x},\overleftarrow{y})} \geq 0, \end{aligned}$$

$$\begin{aligned} \langle \Sigma_{\mathrm{X}} \rangle &= \Delta S_{\mathrm{X}} - \beta\langle Q_{\mathrm{X}}\rangle - \langle \Sigma^{\mathrm{trn}}_{\mathrm{X}}\rangle \qquad (20b) \\ &= \sum_{\vec{x}} p(\vec{x}) \ln\frac{p(\vec{x})}{p(\overleftarrow{x})} \geq 0, \\ \langle \Sigma_{\mathrm{X|Y}} \rangle &= \Delta S_{\mathrm{X|Y}} - \beta\langle Q_{\mathrm{X}}\rangle + \langle \Sigma^{\mathrm{trn}}_{\mathrm{Y}}\rangle \qquad (20c) \\ &= \sum_{\vec{y}} p(\vec{y}) \sum_{\vec{x}} p(\vec{x}|\vec{y}) \ln\frac{p(\vec{x}|\vec{y})}{p(\overleftarrow{x}|\overleftarrow{y})} \geq 0. \end{aligned}$$

Since the total dissipation is the sum of a conditional and marginal dissipation, it follows that, on average, each of the marginal and conditional dissipations are less than the total dissipation, and we summarize:

$$\langle \Sigma_{\mathrm{X,Y}} \rangle \geq \left\{ \langle \Sigma_{\mathrm{X}} \rangle, \langle \Sigma_{\mathrm{Y}} \rangle, \langle \Sigma_{\mathrm{X|Y}} \rangle, \langle \Sigma_{\mathrm{Y|X}} \rangle \right\} \geq 0. \tag{21}$$

This analysis shows that strongly coupled systems obey a local Second Law of thermodynamics. Proper treatment has to either consider the dynamics of one system alone, and study the marginal dissipation, or account for the behavior of other systems directly coupled to the system of interest, and study the conditional dissipation. In either case the system's average dissipation is non-negative and less than the total dissipation. Thus, when studying small parts of the entire Universe, we are allowed to neglect the dissipation occurring elsewhere that is irrelevant to the behavior of the system at hand.

* * *

We would like to acknowledge enlightening discussions with David Sivak, Rob Shaw, Tony Bell, Jordan Horowitz, Sosuke Ito, Takahiro Sagawa and Jim Crutchfield. This work was supported by the Foundational Questions Institute (Grant No. FQXi-RFP3-1345).

REFERENCES

[1] Clausius R., *Philos. Mag. Ser. 4*, **35** (1868) 405.
[2] Planck M., *Treatise on Thermodynamics* (Longmans, Green, and Co., London) 1897.
[3] Callen H. B., *Thermodynamics and an Introduction to Thermostatistics*, 2nd edition (Wiley, New York) 1985.
[4] Deffner S. and Jarzynski C., *Phys. Rev. X*, **3** (2013) 041003.
[5] Maxwell J. C., *Theory of Heat* (Appleton, London) 1871.
[6] Szilard L., *Z. Phys.*, **53** (1929) 840.
[7] Sagawa T. and Ueda M., *Phys. Rev. Lett.*, **104** (2010) 090602.
[8] Horowitz J. M. and Vaikuntanathan S., *Phys. Rev. E*, **82** (2010) 061120.
[9] Parrondo J. M. R., Horowitz J. M. and Sagawa T., *Nat. Phys.*, **11** (2015) 131.
[10] Boyd A. B., Mandal D. and Crutchfield J. P., *Phys. Rev. E*, **95** (2017) 012152.
[11] Samson E. W., *ETC: Rev. Gen. Semant.*, **10** (1953) 283.
[12] Crooks G. E., *J. Stat. Phys.*, **90** (1998) 1481.

[13] Jarzynski C., *J. Stat. Phys.*, **98** (2000) 77.
[14] Seifert U., *Phys. Rev. Lett.*, **95** (2005) 040602.
[15] Still S., Sivak D. A., Bell A. J. and Crooks G. E., *Phys. Rev. Lett.*, **109** (2012) 120604.
[16] Munakata T. and Rosinberg M. L., *Phys. Rev. Lett.*, **112** (2014) 180601.
[17] Horowitz J. M. and Esposito M., *Phys. Rev. X*, **4** (2014) 031015.
[18] Davis J. A., *The Logic Of Causal Order* (SAGE Publications, London) 1985.
[19] Pearl J., *Causality: Models, Reasoning and Inference*, 2nd edition (Cambridge University Press, New York) 2009.
[20] Ito S. and Sagawa T., *Phys. Rev. Lett.*, **111** (2013) 180603.
[21] Jarzynski C., *Phys. Rev. Lett.*, **78** (1997) 2690.
[22] Crooks G. E., *Phys. Rev. E*, **60** (1999) 2721.
[23] Hunter J. E. III, Reinhardt W. P. and Davis T. F., *J. Chem. Phys.*, **99** (1993) 6856.
[24] Sekimoto K., *Prog. Theor. Phys. Suppl.*, **130** (1998) 17.
[25] Peliti L., *Phys. Rev. Lett.*, **101** (2008) 098903.
[26] Gaspard P., *J. Chem. Phys.*, **120** (2004) 8898.
[27] Jarzynski C., *Phys. Rev. E*, **73** (2006) 046105.
[28] Kawai R., Parrondo J. M. R. and Van den Broeck C., *Phys. Rev. Lett.*, **98** (2007) 080602.
[29] Gomez-Marin A., Parrondo J. M. R. and Van den Broeck C., *Phys. Rev. E*, **78** (2008) 011107.
[30] Schreiber T., *Phys. Rev. Lett.*, **85** (2000) 461.
[31] Granger C. W. J., *Econometrica*, **37** (1969) 424.
[32] James R. G., Barnett N. and Crutchfield J. P., *Phys. Rev. Lett.*, **116** (2016) 238701.
[33] Sagawa T. and Ueda M., *Phys. Rev. E*, **85** (2012) 021104.
[34] Prokopenko M., Lizier J. T. and Price D. C., *Entropy*, **15** (2013) 524.
[35] Hartich D., Barato A. C. and Seifert U., *J. Stat. Mech.: Theory Exp.*, **2014** (2014) P02016.
[36] Horowitz J. M. and Sandberg H., *New J. Phys.*, **16** (2014) 125007.
[37] Spinney R. E., Lizier J. T. and Prokopenko M., *Phys. Rev. E*, **94** (2016) 022135.
[38] Toyabe S., Sagawa T., Ueda M., Muneyuki E. and Sano M., *Nat. Phys.*, **6** (2010) 988.
[39] Sagawa T., *Prog. Theor. Phys.*, **127** (2012) 1.
[40] Koski J. V., Maisi V. F., Sagawa T. and Pekola J. P., *Phys. Rev. Lett.*, **113** (2014) 030601.
[41] Horowitz J. M., *J. Stat. Mech.: Theory Exp.*, **2015** (2015) P03006.
[42] Bartolotta A., Carroll S. M., Leichenauer S. and Pollack J., *Phys. Rev. E*, **94** (2016) 022102.