New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics
Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
6 November 2019

REVISED
21 February 2020

ACCEPTED FOR PUBLICATION
24 March 2020

PUBLISHED
24 June 2020

PAPER

# Thermodynamics of computing with circuits

David H Wolpert[1,2,3] and Artemy Kolchinsky[1]

[1] Santa Fe Institute, Santa Fe, New Mexico
[2] Present address: Also at Complexity Science Hub, Vienna; Arizona State University, Tempe, Arizona.
[3] Author to whom any correspondence should be addressed.

**E-mail:** david.h.wolpert@gmail.com

**Keywords:** stochastic thermodynamics, thermodynamics of computation, information theory, non-equilibrium statistical physics, circuits

## Abstract

Digital computers implement computations using circuits, as do many naturally occurring systems (e.g., gene regulatory networks). The topology of any such circuit restricts which variables may be physically coupled during the operation of the circuit. We investigate how such restrictions on the physical coupling affects the thermodynamic costs of running the circuit. To do this we first calculate the minimal additional entropy production that arises when we run a given gate in a circuit. We then build on this calculation, to analyze how the thermodynamic costs of implementing a computation with a full circuit, comprising multiple connected gates, depends on the topology of that circuit. This analysis provides a rich new set of optimization problems that must be addressed by any designer of a circuit, if they wish to minimize thermodynamic costs.

## 1. Introduction

A long-standing focus of research in the physics community has been how the *energetic* resources required to perform a given computation depend on that computation. This issue is sometimes referred to as the 'thermodynamics of computation' or the 'physics of information' [1–3]. Similarly, a central focus of computer science theory has been how the minimal *computational* resources needed to perform a given computation depend on that computation [4, 5]. (Indeed, some of the most important open issues in computer science, like whether P = NP, concern the relationship between a computation and its resource requirements.) Reflecting this commonality of interests, there was a burst of early research relating the resource concerns of computer science theory with the resource concerns of thermodynamics [6–10].

Starting a few decades after this early research, there was dramatic progress in our understanding of non-equilibrium statistical physics [2, 11–15], which has resulted in new insights into the thermodynamics of computation [2, 3, 13, 16]. In particular, research has derived the '(generalized) Landauer bound' [17–22], which states that the heat generated by a thermodynamically reversible process that sends an initial distribution $p_0(x_0)$ to an ending distribution $p_1(x_1)$ is $kT[S(p_0) - S(p_1)]$ (where $S(p)$ indicates the entropy of distribution $p$, $T$ is the temperature of the single bath, and $k$ is Boltzmann's constant).

Almost all of this work on the Landauer bound assumes that the map taking initial states to final states, $P(x_1|x_0)$, is implemented with a monolithic, 'all-at-once' physical process, jointly evolving all of the variables in the system at once. In contrast, for purely practical reasons modern computers are built out of circuits, i.e., they are built out of networks of 'gates', each of which evolves only a small subset of the variables of the full system [4, 5]. An example of a simple circuit that computes the parity of 3 input bits using two XOR gates, and which we will return to throughout this paper, is illustrated in figure 1.

Similarly, in the natural world, biological cellular regulatory networks carry out complicated computations by decomposing them into circuits of far simpler computations [23–25], as do many other kinds of biological systems [26–29].

As elaborated below, there are two major, unavoidable thermodynamic effects of implementing a given computation with a circuit of gates rather than with an all-at-once process:

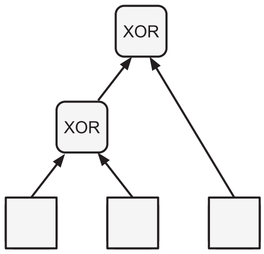

**Figure 1.** A simple circuit that uses two exclusive-OR (XOR) gates to compute the parity of 3 inputs bits. The circuit outputs a 1 if an odd number of input bits are set to 1, and a 0 otherwise.

(I) Suppose we build a circuit out of gates which were manufactured without any specific circuit in mind. Consider such a gate that implements bit erasure, and suppose that it is thermodynamically reversible if $p_0$ is uniform. So by the Landauer bound, it will generate heat $kTS(p_0) = kT\ln 2$ *if run on a uniform distribution*.

Now in general, depending on where such a bit-erasing gate appears in a circuit, the actual initial distribution of its states, $p'_0$, will be non-uniform. This not only changes the Landauer bound for that gate from $kT\ln 2$ to $kTS(p'_0)$; it is now known that since the gate is thermodynamically reversible for $p_0 \neq p'_0$, running that gate on $p'_0$ will *not* be thermodynamically reversible [30]. So the actual heat generated by running that bit will exceed the associated value of the Landauer bound, $kTS(p'_0)$.

(II) Suppose the circuit is built out of two bit-erasing gates, and that each gate is thermodynamically reversible on a uniform input distribution when run separately from the circuit. If the marginal distributions over the initial states of the gates are both uniform, then the heat generated by running each of them is $kT\ln 2$, and therefore the total generated heat is $2kT\ln 2$. Suppose though that there is nonzero statistical coupling between their states under their initial joint distribution. Then as elaborated below, even though each of the gates run separately is thermodynamically reversible, running them in parallel is *not* thermodynamically reversible. So running them generates extra heat beyond the minimum given by applying the Landauer bound to the dynamics of the full joint distribution[4].

These two effects mean that the thermodynamic cost of running a given computation with a circuit will in general vary greatly depending on the precise circuit we use to implement that computation. In the current paper we analyze this dependence.

We make no restriction on the input–output maps computed by each gate in the circuit. They can be either deterministic (i.e., single-valued) or stochastic, logically reversible (i.e., implementing a deterministic permutation of the system's state space, as in Fredkin gates [6]) or not, etc. However, to ground thinking, the reader may imagine that the circuit being considered is a Boolean circuit, where each gate performs one of the usual single-valued Boolean functions, like logical AND gates, XOR gates, etc.

For simplicity, in this paper we focus on circuits whose topology does not contain loops [5, 31], such as the circuit shown in figure 1.

### 1.1. Contributions

We have four primary contributions.

(1) We derive exact expressions for how the entropy flow (EF) and entropy production (EP) produced by a fixed dynamical system vary as one changes the initial distribution of states of that system. These expressions capture effect (I) described above. (These expressions extend an earlier analysis [30]).

(2) We introduce 'solitary processes'. These are a type of physical process that can implement any particular gate in a circuit while respecting the constraints on what variables in the rest of the circuit that gate is coupled with. We can use the thermodynamic properties of solitary processes to analyze effect (II) described above.

3) We combine our first two contributions to analyze the thermodynamic costs of implementing circuits in a 'serial-reinitializing' manner. This means two things: the gates in the circuit are run one at a time, so each gate is run as a solitary process; after a gate is run its input wires are reinitialized, allowing for subsequent reuse of the circuit. In particular, we derive expressions relating the minimal EP generated by running an SR circuit to information-theoretic quantities associated with the wiring diagram of the circuit.

[4] For example, if the initial states of the gates are perfectly correlated, the initial entropy of the two-gate system is $\ln 2$. In this case the running the gates in parallel rather than in a joint system generates extra heat of $2kT\ln 2 - kT\ln 2$, above the minimum possible given by the Landauer bound.

4) Our last contribution is an expression for the extra EP that arises in running an SR circuit if the initial state distributions at its gates differ from the ones that result in minimal EP for each of those gates. This expression involves an information-theoretic function that we call 'multi-divergence' which appears to be new to the literature.

### 1.2. Roadmap

In section 2.1 we introduce general notation, and then provide a minimal summary of the parts of stochastic thermodynamics, information theory and circuit theory that will be used in this paper. We also introduce the definition of the 'islands' of a stochastic matrix in that section, which will play a central role in our analysis. In section 3 we derive an exact expression for how the EF and EP of an arbitrary process depends on its initial state distribution. In section 4 we introduce solitary processes and then analyze their thermodynamics. In section 5 we introduce SR circuits. In section 6 we use the tools developed in the previous sections to analyze the thermodynamic properties of SR circuits. In section 7 we discuss related earlier work. Section 8 concludes and presents some directions for future work. All proofs that are longer than several lines are collected in the appendices.

## 2. Background

Because the analysis of the thermodynamics of circuits involves tools from multiple fields, we review those tools in this section. We also introduce some new mathematical structures that will be central to our analysis, in particular 'islands'. We begin by introducing notation.

### 2.1. General notation

We write a Kronecker delta as $\delta(a, b)$. We write a random variable with an upper case letter (e.g., $X$), and the associated set of possible outcomes with the associated calligraphic letter (e.g., $\mathcal{X}$). A particular outcome of a random variable is written with a lower case letter (e.g., $x$). We also use lower case letters like $p$, $q$, etc to indicate probability distributions.

We use $\Delta_{\mathcal{X}}$ to indicate the set of probability distribution over a set of outcomes $\mathcal{X}$. For any distribution $p \in \Delta_{\mathcal{X}}$, we use $\mathrm{supp}\, p := \{x \in \mathcal{X} : p(x) > 0\}$ to indicate the support of $p$. Given a distribution $p$ over $\mathcal{X}$ and any $\mathcal{Z} \subseteq \mathcal{X}$, we write $p(\mathcal{Z}) = \sum_{x \in \mathcal{Z}} p(x)$ to indicate the probability that the outcome of $X$ is in $\mathcal{Z}$. Given a function $f : \mathcal{X} \to \mathbb{R}$, we write $\mathbb{E}_p[f]$ to indicate $\sum_x p(x) f(x)$, the expectation of $f$ under distribution $p$.

Given any conditional distribution $P(y|x)$ of $y \in \mathcal{Y}$ given $x \in \mathcal{X}$, and some distribution $p$ over $\mathcal{X}$, we write $Pp$ for the distribution over $\mathcal{Y}$ induced by applying $P$ to $p$:

$$[Pp](y) := \sum_{x \in \mathcal{X}} P(y|x) p(x). \tag{1}$$

We will sometimes use the term 'map' to refer to a conditional distribution.

We say that a conditional distribution $P$ is 'logically reversible' if it is deterministic (the entries of $P(y|x)$ are 0/1-valued for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$) and if there do not exist $x, x' \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that $P(y|x) > 0$ and $P(y|x') > 0$. When $\mathcal{Y} = \mathcal{X}$, a logically reversible $P$ is simply a permutation matrix. Given any subset of states $\mathcal{Z} \subseteq \mathcal{X}$, we also say that $P$ is 'logically reversible over $\mathcal{Z}$' if the entries $P(y|x)$ are 0/1-valued for all $x \in \mathcal{Z}$ and $y \in \mathcal{Y}$, and there do not exist $x, x' \in \mathcal{Z}$ and $y \in \mathcal{Y}$ such that $P(y|x) > 0$ and $P(y|x') > 0$.

We write a multivariate random variable with components $V = \{1, 2, \ldots, \}$ as $X_V = (X_1, X_2, \ldots, )$, with outcomes $x_V$. We will also use upper case letters (e.g., $A, V, \ldots,$ ) to indicate sets of variables. For any subset $A \subseteq V$ we use the random variable $X_A$ (and its outcomes $x_A$) to refer to the components of $X_V$ indexed by $A$. Similarly, for a distribution $p_V$ over $X_V$, we write the marginal distribution over $X_A$ as $p_A$. For a singleton set $\{a\}$, we slightly abuse notation and write $X_a$ instead of $X_{\{a\}}$.

### 2.2. Stochastic thermodynamics

We will consider a circuit to be physical system in contact with one or more thermodynamic reservoirs (heat baths, chemical baths, etc). The system evolves over some time interval (sometimes implicitly taken to be $t \in [0, 1]$, where the units of time are arbitrary), possibly while being driven by a work reservoir. We refer to the set of thermodynamic reservoirs and the driving—and, in particular, the stochastic dynamics they induce over the system during $t \in [0, 1]$—as a **physical process**.

We use $\mathcal{X}$ to indicate the finite state space of the system. Physically, the states $x \in \mathcal{X}$ can either be microstates or they can be coarse-grained macrostates under some additional assumptions (e.g., that all macrostates have the same 'internal entropy' [2, 20, 32]).

While much of our analysis applies more broadly, to make things concrete one may imagine that the system undergoes master equation dynamics, also known as a continuous-time Markov chain (CTMC). This kind of dynamics is the basis of stochastic thermodynamics, which is often used to analyze the thermodynamics of discrete-state physical systems. In this subsection we briefly review stochastic thermodynamics, referring the reader to [33, 34] for more details.

Under a CTMC, the probability distribution over $\mathcal{X}$ at time $t$, indicated by $p^t$, evolves according to the master equation

$$\frac{\mathrm{d}}{\mathrm{d}t}p^t(x') = \sum_x p^t(x)K_t(x \to x'), \tag{2}$$

where $K_t$ is the rate matrix at time $t$. For any rate matrix $K_t$, the off-diagonal entries $K_t(x \to x')$ (for $x \neq x'$) indicate the rate at which probability flows from state $x$ to $x'$, while the diagonal entries are fixed by $K_t(x \to x) = -\sum_{x'(\neq x)} K_t(x \to x')$, which guarantees conservation of probability. If the system is connected to multiple thermodynamic reservoirs indexed by $\alpha$, the rate matrix can be further decomposed as $K_t(x \to x') = \sum_\alpha K_t^\alpha(x \to x')$, where $K_t^\alpha$ is the rate matrix at time $t$ corresponding to reservoir $\alpha$.

The term **entropy flow** (EF) refers to the increase of entropy in all coupled reservoirs. The instantaneous rate of EF out of the system at time $t$ is defined as

$$\dot{\mathcal{Q}}(p^t) = \sum_{\alpha,x,x'} p^t(x)K_t^\alpha(x \to x') \ln \frac{K_t^\alpha(x \to x')}{K_t^\alpha(x' \to x)}. \tag{3}$$

The overall EF incurred over the course of the entire process is $\mathcal{Q} = \int_0^1 \dot{\mathcal{Q}}\ \mathrm{d}t$.

The term **entropy production** (EP) refers to the overall increase of entropy, both in the system and in all coupled reservoirs. The instantaneous rate of EP at time $t$ is defined as

$$\dot{\sigma}(p^t) = \frac{\mathrm{d}}{\mathrm{d}t}S(p^t) + \dot{\mathcal{Q}}(p^t). \tag{4}$$

The overall EP incurred over the course of the entire process is $\sigma = \int_0^1 \dot{\sigma}\ \mathrm{d}t$.

Note that we use terms like 'EF' and 'EP' to refer to either the associated rate or the associated integral over a non-infinitesimal time interval; the context should always make the precise meaning clear.

Given some initial distribution $p$, the EF, EP, and the drop in the entropy of the system from the beginning to the end of the process are related according to

$$\mathcal{Q}(p) = \left[S(p) - S(Pp)\right] + \sigma(p). \tag{5}$$

In general, the EF can be written as the expectation $\mathcal{Q}(p) = \sum_x p(x)q(x)$, where $q(x)$ indicates the expected EF arising from trajectories that begin on state $x$. Given that the drop in entropy is a nonlinear function of $p$, while the expectation $\mathcal{Q}(p)$ is a linear function of $p$, equation (5) tells us that EP will generally be a nonlinear function of $p$. Note that if $P$ is logically reversible, then $S(p) = S(Pp)$ and therefore EF and EP will be equal for any $p$.

While the EF can be positive or negative, the log-sum inequality can be used to prove that EP for master equation dynamics is non-negative [15, 35]:

$$\mathcal{Q}(p) \geqslant S(p) - S(Pp). \tag{6}$$

This can be viewed as a derivation of the second law of thermodynamics, given the assumption that our system is evolving forward in time as a CTMC.

All of these results are purely mathematical and hold for any CTMC dynamics, even in contexts having nothing to do with physical systems. However, these results can be interpreted in thermodynamic terms when each $K_t^\alpha$ obeys **local detailed balance** (LDB) with regard to thermodynamic reservoir $\alpha$ [3, 15, 33]. Consider a system with Hamiltonian $H_t(\cdot)$ at time $t$, and let $\alpha$ label a heat bath whose inverse temperature is $\beta_\alpha$. Then, $K_t^\alpha$ will obey LDB when for all $x, x' \in \mathcal{X}$, either $K_t^\alpha(x \to x') = K_t^\alpha(x' \to x) = 0$, or

$$\frac{K_t^\alpha(x \to x')}{K_t^\alpha(x' \to x)} = \mathrm{e}^{\beta_\alpha(H_t(x) - H_t(x'))}. \tag{7}$$

If LDB holds, then EF can be written as [34]

$$\mathcal{Q}(p) = \sum_\alpha \beta_\alpha Q_\alpha(p), \tag{8}$$

where $Q_\alpha$ is the expected amount of heat transferred from the system into bath $\alpha$ during the process.

We end with two caveats concerning the use of stochastic thermodynamics to analyze real-world circuits. First, many of the processes described in this paper require that some transition rates be exactly zero at some moments. In many physical models this implies there are infinite energy barriers at those times. In addition, perfectly carrying out any deterministic map (such as bit erasure) requires the use of infinite energy gaps between some states at some times. Thus, as is conventional (though implicit) in much of the thermodynamics of computation literature, the thermodynamic costs derived in this paper should be understood as limiting values.

Second, there are some conditional distributions that take the system state at time 0 to its state at time 1, $P(x_1|x_0)$, that cannot be implemented by any CTMC [36, 37]. For example, one cannot carry out (or even approximate) a simple bit flip $P(x_1|x_0) = 1 - \delta(x_1, x_0)$ with a CTMC. Now we *can* design a CTMC to implement any given $P(x_1|x_0)$ to arbitrary precision, if the dynamics is expanded to include a set of 'hidden states' in addition to the states in $X$ [21, 22]. However, as we explicitly demonstrate below, SR circuits can be implemented without introducing any such hidden states; this is one of their advantages. (See also example 9 in appendix A).

### 2.3. Information theory

Given two distributions $p$ and $r$ over random variable $X$, we use notation like $S(p)$ for Shannon entropy and $D(p\|r)$ for Kullback–Leibler (KL) divergence. We write $S(Pp)$ to refer to the entropy of the distribution over $Y$ induced by $p(x)$ and the conditional distribution $P$, as defined in equation (5), and similarly for other information-theoretic measures. Given two random variables $X$ and $Y$ with joint distribution $p$, we write $S(p(X|Y))$ for the conditional entropy of $X$ given $Y$, and $I_p(X;Y)$ for the mutual information (we drop the subscript $p$ where the distribution is clear from context). All information-theoretic measures are in nats.

Some of our results below are formulated in terms of an extension of mutual information to more than two random variables that is known as 'total correlation' or **multi-information** [38]. For a random variable $X_A = (X_1, X_2, \ldots, )$, the multi-information is defined as

$$\mathcal{I}(p_A) = \left[\sum_{v\in A} S(p_v)\right] - S(p_A). \tag{9}$$

Some of the other results below are formulated in terms of the **multi-divergence** between two probability distributions over the same multi-dimensional space. This is a recently introduced information-theoretic measure which can be viewed as an extension of multi-information to include a reference distribution. Given two distributions $p_A$ and $r_A$ over $X_A$, the multi-divergence is defined as

$$\mathcal{D}(p_A\|r_A) := D(p_A\|r_A) - \sum_{v\in A} D(p_v\|r_v). \tag{10}$$

Multi-divergence measures how much of the divergence between $p_A$ and $r_A$ arises from the correlations among the variables $X_1, X_2, \ldots,$ rather than from the marginal distributions of each variable considered separately. See appendix A of [3] for a discussion of the elementary properties of multi-divergence and its relation to conventional multi-information. Note that multi-divergence is defined with 'the opposite sign' of multi-information, i.e., by subtracting a sum of terms involving marginal variables from a term involving the joint random variable, rather than vice-versa.

### 2.4. 'Island' decomposition of a conditional distribution

A central part of our analysis will involve the equivalence relation,

$$x \sim x' \Leftrightarrow \exists y : P(y|x) > 0, \quad P(y|x') > 0. \tag{11}$$

In words, $x \sim x'$ if there is a non-zero probability of transitioning to some state $y$ from both $x$ and $x'$ under the conditional distribution $P(y|x)$. We define an **island** of the conditional distribution $P(y|x)$ as any connected subset of $\mathcal{X}$ given by the transitive closure of this equivalence relation. The set of islands of any $P(\cdot|\cdot)$ form a partition of $\mathcal{X}$, which we write as $L(P)$.

We will also use the notion of the islands of the conditional distribution $P$ restricted to some subset of states $\mathcal{Z} \subseteq \mathcal{X}$. We write $L_{\mathcal{Z}}(P)$ to indicate the partition of $\mathcal{Z}$ generated by the transitive closure of the relation given by equation (11) for $x, x' \in \mathcal{Z}$. Note that in this notation, $L(P) = L_{\mathcal{X}}(P)$.

As an example, if $P(y|x) > 0$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ (i.e., any final state $y$ can be reached from any initial state $x$ with non-zero probability), then $L(P)$ contains only a single island. As another example, if $P(y|x)$ implements a deterministic function $f : \mathcal{X} \to \mathcal{Y}$, then $L(P)$ is the partition of $\mathcal{X}$ given by the

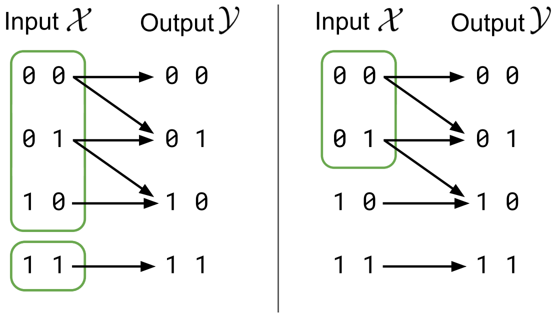

**Figure 2.** Left: the island decomposition for the conditional distribution in equation (13). The two islands are indicated by the two rounded green boxes. Right: the island decomposition for this map with $\mathcal{X}$ restricted to the subset of states $\mathcal{Z} = \{00, 01\}$. For this subset of states, there is only one island, indicated by the round green box.

pre-images of $f$, $L(P) = \{f^{-1}(y) : y \in \mathcal{Y}\}$. For example, the conditional distribution that implements the logical AND operation of two binary variables,

$$P(c|a, b) = \delta(c, a\,b) \tag{12}$$

has two islands, corresponding to $(a, b) \in \{(0, 0), (0, 1), (1, 0)\}$ and $(a, b) \in \{(1, 1)\}$, respectively. As a final example, let $P$ be the following conditional distribution:

$$P(y|x) = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{13}$$

where the rows and columns corresponds to the ordered states $\mathcal{X} = \mathcal{Y} = \{00, 01, 10, 11\}$. The island decomposition for this map is illustrated in figure 2 (left). We also show the island decomposition for this map restricted to subset of states $\mathcal{Z} = \{00, 01\}$ in figure 2 (right).

For any distribution $p$ over $\mathcal{X}$, any $\mathcal{Z} \subseteq \mathcal{X}$, and any $c \in L_{\mathcal{Z}}(P)$, $p(c) = \sum_{x \in c} p(x)$ is the probability that the state of the system is contained in island $c$. It will be helpful to use the unusual notation $p^c(x)$ to indicate the conditional probability of $x$ within island $c$. Formally, $p^c(x) = p(x)/p(c)$ if $x \in c$, and $p^c(x) = 0$ otherwise.

Intuitively, the islands of a conditional distribution are ‘firewalled’ subsystems, both computationally and thermodynamically isolated from one another for the duration of the process implementing that conditional distribution. In particular, we will show below that the EP of running $P(y|x)$ on an initial distribution $p$ can be written as a weighted sum of the EPs involved in running $P$ on each separate island $c \in L(P)$, where the weight for island $c$ is given by $p(c)$.

### 2.5. Circuit theory

For the purposes of this paper, a **(logical) circuit** is a special type of Bayes net [39–41]. Specifically, we define any circuit $\Phi$ as a tuple $(V, E, F, \mathcal{X}_V)$. The pair $(V, E)$ specifies the vertices and edges of a directed acyclic graph (DAG). (We sometimes call this DAG the **wiring diagram** of the circuit.) $\mathcal{X}_V$ is a Cartesian product $\prod_v \mathcal{X}_v$, where each $\mathcal{X}_v$ is the set of possible states associated with node $v$. $F$ is a set of conditional distributions, indicating the logical maps implemented at the non-root nodes of the DAG.

Following the convention in the Bayes nets literature, we orient edges in the direction of information flow. Thus, the inputs to the circuit are the roots of the associated DAG and the outputs are the leaves of the DAG [5]. Without loss of generality, we assume that each node $v$ has a special ‘initialized state’, indicated as $\emptyset$.

We use the term **gate** to refer to any non-root node, **input node** to refer to any root node, and **output node** or **output gate** to refer to a leaf node. For simplicity, we assume that all output nodes are gates, i.e., there is no root node which is also a leaf node. We write IN and $x_{\mathrm{IN}}$ to indicate the set of input nodes and their joint state, and similarly write OUT and $x_{\mathrm{OUT}}$ for the output nodes.

We write the set of all gates in a given circuit as $G \subseteq V$, and use $g \in G$ to indicate a particular gate. We indicate the set of all nodes that are parents of gate $g$ as $\mathrm{pa}(g)$. We indicate the set of nodes that includes gate $g$ and all parents of $g$ as $n(g) := \{g\} \cup \mathrm{pa}(g)$.

[5] The reader should be warned that much of the computer science literature adopts the opposite convention.

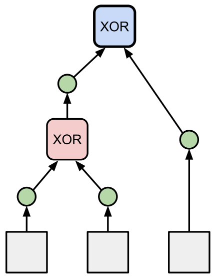

**Figure 3.** The 3-bit parity circuit of figure 1 represented as a wired circuit. Squares represent input nodes, rounded boxes represent non-wire gates, and smaller green circles represent wire gates. The output XOR gate is in blue, while the other (non-output) XOR gate is in red.

As mentioned, $F$ is a set of conditional distributions, indicating the logical maps implemented by each gate of the circuit. The element of $F$ corresponding to gate $g$ is written as $\pi_g(x_g|x_{\mathrm{pa}(g)})$. In conventional circuit theory, each $\pi_g$ is required to be deterministic (i.e., 0/1-valued). However, we make no such restriction in this paper. We write the overall conditional distribution of output gates given input nodes implemented by the circuit $\Phi$ as

$$\pi_\Phi(x_{\mathrm{OUT}}|x_{\mathrm{IN}}) = \sum_{x_{G\setminus\mathrm{OUT}}} \prod_{g\in G} \pi_g(x_g|x_{\mathrm{pa}(g)}). \tag{14}$$

We can illustrate this formalism using the parity circuit shown in figure 1. Here, $V$ has 5 nodes, corresponding to the 3 input nodes and the two gates. The circuit operates over bits, so $\mathcal{X}_v = \{0, 1\}$ for each $v \in V$. Both gates carry out the $\mathsf{XOR}$ operation, so both elements of $F$ are given by $\pi_g(x_g|x_{\mathrm{pa}(g)}) = \delta(x_g, \mathsf{XOR}(x_{\mathrm{pa}(g)}))$ (where $\mathsf{XOR}(x_{\mathrm{pa}(g)}) = 1$ when the two parents of gate $g$ are in different states, and $\mathsf{XOR}(x_{\mathrm{pa}(g)}) = 0$ otherwise). Finally, $E$ has four elements representing the edges connecting the nodes in $V$, which are shown as arrows in figure 1.

In the conventional representation of a physical circuit as a (Bayes net) DAG, the wires in the physical circuit are identified with edges in the DAG. However, in order to account for the thermodynamic costs of communication between gates along physical wires, it will be useful to represent the wires themselves as a special kind of gate. This means that the DAG $(V, E)$ we use to represent a particular physical circuit is not the same as the DAG $(V', E')$ that would be used in the conventional computer science representation of that circuit. Rather $(V, E)$ is constructed from $(V', E')$ as follows.

To begin, $V = V'$ and $E = E'$. Then, for each edge $(v \to \tilde{v}) \in E'$, we first add a **wire gate** $w$ to $V$, and then add two edges to $E$: an edge from $v$ to $w$ and an edge from $w$ to $\tilde{v}$. So a wire gate $w$ has a single parent and a single child, and implements the identity map, $\pi_w(x_w|x_{\mathrm{pa}(w)}) = \delta(x_w, x_{\mathrm{pa}(w)})$. (This is an idealization of the real world, in which wires have nonzero probability of introducing errors.) We sometimes call $(V, E)$ the **wired circuit**, to distinguish it from the original **logical circuit** defined as in computer science theory, $(V', E')$. We use $W \subset G$ to indicate the set of wire gates in a wired circuit.

Every edge in a wired circuit either connects a wire gate to a non-wire gate or vice versa. Physically, the edges of the DAG of a wired circuit do not represent interconnects (e.g., copper wires), as they do in a logical circuit. Rather they only indicate physical identity: an edge $e \in E$ going into a wire gate $w$ from a non-wire node $v$ indicates that the same physical variable will be written as either $X_v$ or $X_{\mathrm{pa}(w)}$. Similarly, an edge $e \in E$ going into a non-wire gate $g$ from a wire gate $w$ indicates that $X_w$ is the same physical variable (and so always has the same state) as the corresponding component of $X_{\mathrm{pa}(g)}$. However, despite this modified meaning of the nodes in a wired circuit, equation (14) still applies to any wired circuit, as well as applying to the corresponding logical circuit. In figure 3, we demonstrate how to represent the 3-bit parity circuit from figure 1 as a wired circuit.

We use the word ‘circuit’ to refer to either an abstract wired (or logical) circuit, or to a physical system that implements that abstraction. Note that there are many details of the physical system that are not specified in the associated abstract circuit. When we need to distinguish the abstraction from its physical implementation, we will refer to the latter as a **physical circuit**, with the former being the corresponding wired circuit. The context will always make clear whether we are using terms like ‘gate’, ‘circuit’, etc, to refer to physical systems or to their formal abstractions.

Even if one fully specifies the distinct physical subsystems of a physical circuit that will be used to implement each gate in a wired circuit, we still do not have enough information concerning the physical circuit to analyze the thermodynamic costs of running it. We still need to specify the initial states of those subsystems (before the circuit begins running), the precise sequence of operations of the gates in the circuit, etc. However, before considering these issues, we need to analyze the general form of the thermodynamic costs of running individual gates in a circuit, isolated from the rest of the circuit. We do that in the next section.

## 3. Decomposition of EF

Suppose we have a fixed physical system whose dynamics over some time interval is specified by a conditional distribution $P$, and let $p$ be its initial state distribution, which we can vary. We decompose the EF of running that system into a sum of three functions of $p$. Applied to any specific gate in a circuit (the 'fixed physical system'), this decomposition tells us how the thermodynamic costs of that gate would change if the distribution of inputs to the gate were changed.

First, equation (6) tells us that the minimal possible EF, across all physical processes that transform $p$ into $P' := Pp$, is given by the drop in system entropy. We refer to this drop as the **Landauer cost** of computing $P$ on $p$, and write it as

$$\mathcal{L}(p) := S(p) - S(Pp). \tag{15}$$

Since EF is just Landauer cost plus EP, our next task is to calculate how the EP incurred by a fixed physical process depends on the initial distribution $p$ of that process. To that end, in the rest of this section we show that EP can be decomposed into a sum of two non-negative functions of $p$. Roughly speaking, the first of those two functions reflects the deviation of the initial distribution $p$ from an 'optimal' initial distribution, while the second term reflects the remaining EP that would occur even if the process were run on that optimal initial distribution.

To derive this decomposition, we make use of a mathematical result provided by the following theorem. The theorem considers any function of the initial distribution $p$ which can be written in the form $S(Pp) - S(p) + \mathbb{E}_p[f]$ (i.e., the increase of Shannon entropy plus an expectation of some quantity with respect to $p$). The EP incurred by a physical process can be written in this form (by equation (5), where $\mathbb{E}_p[f]$ refers to the EF). Further below, we will also consider other functions, which are closely related to EP, that can be written in this special form. The theorem shows that any function with this special form can be decomposed into a sum of the two terms described above: the first term reflecting deviation of $p$ from the optimal initial distribution (relative to all distributions with support in some restricted set of states, which we indicate as $\mathcal{Z}$), and a remainder term.

**Theorem 1.** Consider any function $\Gamma : \Delta_{\mathcal{X}} \to \mathbb{R}$ of the form

$$\Gamma(p) := S(Pp) - S(p) + \mathbb{E}_p[f]$$

where $P(y|x)$ is some conditional distribution of $y \in \mathcal{Y}$ given $x \in \mathcal{X}$ and $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ is some function. Let $\mathcal{Z}$ be any subset of $\mathcal{X}$ such that $f(x) < \infty$ for $x \in \mathcal{Z}$, and let $q \in \Delta_{\mathcal{Z}}$ be any distribution that obeys

$$q^c \in \underset{r:\mathrm{supp}\, r \subseteq c}{\arg\min}\, \Gamma(r) \quad \text{for all } c \in L_{\mathcal{Z}}(P).$$

Then, each $q^c$ will be unique, and for any $p$ with $\mathrm{supp}\, p \subseteq \mathcal{Z}$,

$$\Gamma(p) = D(p\|q) - D(Pp\|Pq) + \sum_{c \in L_{\mathcal{Z}}(P)} p(c)\Gamma(q^c).$$

We emphasize that $P$ and $f$ are implicit in the definition of $\Gamma$. We remind the reader that the definition of $L_{\mathcal{Z}}$ and $q^c$ is provided in section 2.4. The proof is provided in appendix A.

Note that theorem 1 does not suppose that $q$ is unique, only that the conditional distributions within each island, $\{q^c\}_c$, are. Moreover, as implied by the statement of the theorem, the overall probability weights assigned to the separate islands, $\{q(c)\}_c$, has no effect on the value of $\Gamma$.

Consider some conditional distribution $P(y|x)$, with $\mathcal{Y} = \mathcal{X}$, implemented by a physical process. Then if we take $\mathcal{Z} = \mathcal{X}$ and $\mathbb{E}_p[f] = \mathcal{Q}$ in theorem 1, the function $\Gamma$ is just the EP of running the conditional distribution $P(y|x)$. This establishes the following decomposition of EP:

$$\sigma(p) = D(p\|q) - D(Pp\|Pq) + \sum_{c \in L(P)} p(c)\sigma(q^c). \tag{16}$$

We emphasize that equation (16) holds without any restrictions on the process, e.g., we do not require that the process obey LDB. In fact, equation (16) even holds if the process does not evolve according to a CTMC (as long as EP can be defined via equation (5)).

We refer to the first term in equation (16), the drop in KL divergence between $p$ and $q$ as both evolve under $P$, as **mismatch cost**[6]. Mismatch cost is non-negative by the data-processing inequality for KL divergence [42]. It equals zero in the special case that $p^c = q^c$ for each island $c \in L_{\mathcal{Z}}(P)$. We refer to any such initial distribution $p$ that results in zero mismatch cost as a **prior distribution** of the physical process that implements the conditional distribution $P$ (the term 'prior' reflects a Bayesian interpretation of $q$; see [20, 30].) If there is more than one island in $L_{\mathcal{Z}}(P)$, the prior distribution is not unique.

We call the second term in our decomposition of EP in equation (16), $\sum_{c\in L_{\mathcal{Z}}(P)} p(c)\sigma(q^c)$, the **residual EP**. In contrast to mismatch cost, residual EP does not involve information-theoretic quantities, and depends linearly on $p$. When $L_{\mathcal{Z}}(P)$ contains a single island, this 'linear' term reduces to an additive constant, independent of the initial distribution. The residual EP terms $\{\sigma(q^c)\}_c$ are all non-negative, since EP is non-negative.

Concretely, the conditional distributions $\{q^c\}_c$ and the corresponding set of real numbers $\{\sigma(q^c)\}_c$ depend on the precise physical details of the process, beyond the fact that the process implements $P$. Indeed, by appropriate design of the 'nitty gritty' details of the physical process, it is possible to have $\sigma(q^c) = 0$ for all $c \in L_{\mathcal{Z}}(P)$, in which case the residual EP would equal zero for all $p$. (For example, this will be the case if the process is an appropriate quasi-static transformation; see [21, 43].)

Imagine that the conditional distribution $P$ is logically reversible over some set of states $\mathcal{Z} \subseteq \mathcal{X}$, and that $\operatorname{supp} p \subseteq \mathcal{Z}$. Then both mismatch cost and Landauer cost must equal zero, and EF must equal EP, which in turn must equal residual EP[7]. Conversely, if $P$ is not logically reversible over $\mathcal{Z}$, then mismatch cost cannot be zero for all initial distributions $p$ with $\operatorname{supp} p \subseteq \mathcal{Z}$ (for such a $P$, regardless of what $q$ is, there will be some $p$ with $\operatorname{supp} p \subseteq \mathcal{Z}$ such that the KL divergence between $p$ and $q$ will shrink under the mapping $P$). Thus, for any fixed process that implements a logically irreversible map, there will be some initial distributions $p$ that result in unavoidable EP.

To provide some intuition into these results, the following example reformulates the EP of a very commonly considered scenario as a special case of equation (16):

**Example 1.** Consider a physical system evolving according to an irreducible master equation, while coupled to a single thermodynamic reservoir and without external driving. Because there is no external driving, the master equation is time-homogeneous with some unique equilibrium distribution $p_{\text{eq}}$. So the system is relaxing toward that equilibrium as it undergoes the conditional distribution $P$ over the interval $t \in [0, 1]$.

For this kind of relaxation process, it is well known that the EP can be written as [34, 44, 45]:

$$\sigma(p) = D(p\|p_{\text{eq}}) - D(Pp\|p_{\text{eq}}). \tag{17}$$

Equation (17) can also be derived from our result, equation (16), since

(a) Taking $\mathcal{Z} = \mathcal{X}$, $P$ has a single island (because the master equation is irreducible, and therefore any state is reachable from any other over $t \in [0, 1]$);

(b) The prior distribution within this single island is $q = p_{\text{eq}}$ (since the EP would be exactly zero if the system were started at this equilibrium, which is a fixed point of $P$);

(c) The residual EP is $\sigma(q) = 0$ (again using fact that EP is exactly zero for $p = p_{\text{eq}}$, and that there is a single island);

(d) $Pq = p_{\text{eq}}$ (since there is no driving, and the equilibrium distribution is a fixed point of $P$).

Thus, equation (16) can be seen as a generalization of the well-known relation given by equation (17), which is defined for simple relaxation processes, to processes that are driven and possibly connected to multiple reservoirs.

The following example addresses the effect of possible discontinuities in the island decomposition of $P$ on our decomposition of thermodynamic costs:

**Example 2.** Mismatch cost and residual EP are both defined in terms of the island decomposition of the conditional distributions $P$ over some set of states $\mathcal{Z}$. That decomposition in turn depends on which (if any) entries in the conditional probability distribution $P$ are exactly 0. This suggests that the decomposition of equation (16) can depend discontinuously on very small variations in $P$ which replace strictly zero entries in $P$ with infinitesimal values, since such variations will change the island decomposition of $P$.

[6] [30] derived equation (16) for the special case where $P$ has a single island within $\mathcal{X}$, and only provided a lower bound for more general cases. In that paper, mismatch cost is called the 'dissipation due to incorrect priors', due to a particular Bayesian interpretation of $q$.

[7] When $P$ is logically reversible over initial states $\mathcal{Z}$, each state in $\mathcal{Z}$ is a separate island, which means that EF, EP, and residual EP can be written as $\sum_{x\in\mathcal{Z}} p(x)\sigma(u^x)$, where $u^x$ indicates a distribution which is a delta function over state $x$.

To address this concern, first note that if $P \simeq P'$, then the EP of the real-world process that implements $P'$ can be approximated as

$$\begin{aligned}\sigma'(p) &= S(P'p) - S(p) + \mathcal{Q}'(p)\\ &\simeq S(Pp) - S(p) + \mathcal{Q}'(p),\end{aligned} \tag{18}$$

where $\mathcal{Q}'(p)$ is the EF function of the real-world process, with the approximation becoming exact as $P \to P'$[8]. If we now apply theorem 1 to the right-hand side of equation (18), we see that so long as $P'$ is close enough to $P$, we can approximate $\sigma'(p)$ as a sum of mismatch cost and residual EP using the islands of the idealized map $P$, instead of the actual map $P'$.

## 4. Solitary processes

Implicit in the definition of a physical circuit is that it is 'modular', in the sense that when a gate in the circuit runs, it is physically coupled to the gates that are its direct inputs, and those that directly get its output, but is *not* physically coupled to any other gates in the circuit. This restriction on the allowed physical coupling is a constraint on the possible processes that implement each gate in the circuit. It has major thermodynamic consequences, which we analyze in this section.

To begin, suppose we have a system that can be decomposed into two separate subsystems, $A$ and $B$, so that the system's overall state space $\mathcal{X}$ can be written as $\mathcal{X} = \mathcal{X}_A \times \mathcal{X}_B$, with states $(x_A, x_B)$. For example, $A$ might contain a particular gate and its inputs, while $B$ might consist of all other nodes in the circuit. We use the term **solitary process** to refer to a physical process over state space $\mathcal{X}_A \times \mathcal{X}_B$ that takes place during $t \in [0, 1]$ where:

(a) $A$ evolves independently of $B$, and $B$ is held fixed:

$$P(x'_A, x'_B|x_A, x_B) = P_A(x'_A|x_A)\delta(x'_B, x_B). \tag{19}$$

(b) The EF of the process depends only on the initial distribution over $\mathcal{X}_A$, which we indicate with the following notation:

$$\mathcal{Q}(p) = \mathcal{Q}_A(p_A). \tag{20}$$

(c) The EF is lower bounded by the change in the marginal entropy of subsystem $A$,

$$\mathcal{Q}_A(p_A) \geqslant S(p_A) - S(P_A p_A). \tag{21}$$

Note that it may be that some subset $A'$ of the variables in subsystem $A$ do not change their state during the solitary process. In that sense such variables would be like the variables in $B$. However, if the dynamics of those variables in $A$ that *do* change state depends on the values of the variables in $A'$, then in general the variables in $A'$ cannot be assigned to $B$; they have to be included in subsystem $A$ in order for condition (b) to be met.

**Example 3.** A concrete example of a solitary process is a CTMC where at all times, the rate matrix $K_t$ has the decoupled structure

$$K_t(x_V \to x'_V) = \delta(x_B, x'_B)\sum_{\alpha} K_t^{A,\alpha}(x_A \to x'_A) \tag{22}$$

for $x_V \neq x'_V$, where $K_t^{A,\alpha}$ indicates the rate matrix for subsystem $A$ and thermodynamic reservoir $\alpha$ at time $t$[9].

To verify that this CTMC is a solitary process, first plug the rate matrix in equation (22) into equation (2) and simplify, giving

$$\frac{\mathrm{d}}{\mathrm{d}t}p^t(x'_A, x'_B) = p^t(x'_B)\sum_{x_A} p^t(x_A|x'_B)\sum_{\alpha} K_t^{A,\alpha}(x_A \to x'_A).$$

Marginalizing the above equation, we see that the distribution over the states of $A$ evolves independently of the state of $x_B$, according to

$$\frac{\mathrm{d}}{\mathrm{d}t}p^t(x'_A) = \sum_{x_A} p^t(x_A)\sum_{\alpha} K_t^{A,\alpha}(x_A \to x'_A).$$

[8] The fact that $S(P\prime p) \to S(Pp)$ as $P \to P\prime$ follows from [52, theorem 17.3.3], and the assumption of a finite state space.

[9] In fact, in [3] solitary processes are defined as CTMCs with this form.

Note also that given the form of equation (22), the state of $B$ does not change. Thus, the conditional distribution carried out by this CTMC over any time interval must have the form of equation (19). (See also appendix B in [3].)

Next, plug equation (22) into equation (3) and simplify to get

$$\dot{\mathcal{Q}}(p^t) = \sum_{\alpha, x_A, x'_A} p^t(x_A) K_t^{A,\alpha}(x_A \to x'_A) \ln \frac{K_t^{A,\alpha}(x_A \to x'_A)}{K_t^{A,\alpha}(x'_A \to x_A)}. \tag{23}$$

Thus, the EF incurred by the process evolves exactly as if $A$ were an independent system connected to a set of thermodynamic reservoirs. Therefore, a joint system evolving according to equation (22) will satisfy equations (20) and (21).

We refer to the lower bound on the EF of subsystem $A$, as given in equation (21), as the **subsystem Landauer cost** for the solitary process. We make the associated definition that the **subsystem EP** for the solitary process is

$$\hat{\sigma}_A(p_A) := \mathcal{Q}_A(p_A) - \left[S(p_A) - S(P_A p_A)\right], \tag{24}$$

which by equation (21) is non-negative. Note that if $P_A$ is a logically reversible conditional distribution, then subsystem EP is equal to the EF incurred by the solitary process.

In general, $S(p_A) - S(P_A p_A)$, the subsystem Landauer cost, will not equal $S(p_{AB}) - S(Pp_{AB})$, the Landauer cost of the entire joint system. Loosely speaking, an observer examining the entire system would ascribe a different value to its entropy change during the solitary process than would an observer examining just subsystem $A$—even though subsystem $B$ does not change its state. We use the term **Landauer loss** to refer to this difference in Landauer costs,

$$\mathcal{L}^{\text{loss}}(p) := \left[S(p_A) - S(P_A p_A)\right] - \left[S(p_{AB}) - S(Pp_{AB})\right]. \tag{25}$$

Assuming that the lower bound in equation (21) can be saturated, since the bound in equation (6) can be saturated, the Landauer loss is the increase in the minimal EF that must be incurred by any process that carries out $P$ if that process is required to be a solitary process.

By using the fact that subsystem $B$ remains fixed throughout a solitary process, the Landauer loss can be rewritten as the drop in the mutual information between $A$ and $B$, from the beginning to the end of the solitary process,

$$\mathcal{L}^{\text{loss}}(p) = I_p(A;B) - I_{Pp}(A;B). \tag{26}$$

Applying the data processing inequality establishes that Landauer loss is non-negative [46]. (See section 7 for a discussion of the relation between solitary processes and other processes that have been considered in the literature.)

If $P_A$ (and thus also $P$) is logically reversible, then the Landauer loss will always be zero. However, for other conditional distributions, there is always some $p$ that results in strictly positive Landauer loss. Moreover, we can rewrite it as

$$\mathcal{L}^{\text{loss}}(p) = \sigma(p_{AB}) - \hat{\sigma}_A(p_A). \tag{27}$$

So in general the subsystem EP will be less than the overall EP of the entire system[10].

Finally, note that $\mathcal{Q}_A(p_A)$ is a linear function of the distribution $p_A$ (since EF functions are linear). Combining this fact with theorem 1, while taking $\mathcal{Z} = \mathcal{X}_A$, allows us to expand the subsystem EP as

$$\hat{\sigma}_A(p_A) = D(p_A \| q_A) - D(P_A p_A \| P_A q_A) + \sum_{c \in L(P_A)} p_A(c) \hat{\sigma}_A(q_A), \tag{28}$$

where $q_A$ is a distribution over $\mathcal{X}_A$ that satisfies $\hat{\sigma}_a(q_A^c) = \min_{r:\text{supp}\, r \subseteq c} \hat{\sigma}_A(r)$ for all $c \in L(P_A)$. As before, both the drop in KL divergence and the term linear in $p_A(c)$ are non-negative. We will sometimes refer to that drop in KL divergence as **subsystem mismatch cost**, with $q_A$ the **subsystem prior**, and refer to the linear term as **subsystem residual EP**. Intuitively, subsystem Landauer cost, subsystem EP, subsystem mismatch cost, and subsystem residual EP are simply the values of those quantities that an observer would ascribe to subsystem $A$ if they observed it independently of $B$.

[10] See [3] for an example explicitly illustrating how the rate matrices change if we go from an unconstrained process that implements $P_A$ to a solitary process that does so, and how that change increases the total EP. That example also considers the special case where the prior of the full $A \times B$ system is required to factor into a product of a distribution over the initial value of $\mathcal{X}_A$ times a distribution over the initial value of $\mathcal{X}_B$. In particular, it shows that the Landauer loss is the minimal value of the mismatch cost in this special case.

## 5. Serial-reinitialized circuits

As mentioned at the end of section 2.5, specifying a wired circuit does not specify the initial distributions of the gates in the physical circuit, the sequence in which the gates in the physical circuit are run, etc. So it does not fully specify the dynamics of a physical system that implements that wired circuit. In this section we introduce one relatively simple way of mapping a wired circuit to such a full specification. In this specification, the gates are run serially, one after the other. Moreover, the gates reinitialize the states of their parent gates after they run, so that the entire circuit can be repeatedly run, incurring the same expected thermodynamic costs each time. We call such physical systems **serial reinitialized implementations** of a given wired circuit, or just SR circuits for short.

For simplicity, in the main text of this paper we focus on the special case in which all non-output nodes have out-degree 1, i.e., where each non-output node is the parent of exactly one gate. See appendix C for a discussion of how to extend the current analysis to relax this requirement, allowing some nodes to have out-degree larger than 1.

There are several properties that jointly define the SR circuit implementation of a given wired circuit.

First, just before the physical circuit starts to run, all of its nodes have a special initialized value with probability 1, i.e., $x_v = \emptyset$ for all $v \in V$ at time $t = 0$. Then the joint state of the input nodes $x_{\mathrm{IN}}$ is set by sampling $p_{\mathrm{IN}}(x_{\mathrm{IN}})$[11]. Typically this setting of the state of the input nodes is done by some offboard system, e.g., the user of the digital device containing the circuit. We do not include the details of this offboard system in our model of the physical circuit. Accordingly, we do not include the thermodynamic costs of setting the joint state of the input nodes in our calculation of the thermodynamic costs of running the circuit[12].

After $x_{\mathrm{IN}}$ is set this way, the SR circuit implementation begins. It works by carrying out a sequence of solitary processes, one for each gate of the circuit, including wire gates. At all times that a gate $g$ is 'running', the combination of that gate and its parents (which we indicate as $n(g)$) is the subsystem $A$ in the definition of solitary processes. The set of all other nodes of the wired circuit ($V\backslash n(g)$) constitute the subsystem $B$ of the solitary process. The temporal ordering of the solitary processes must be a topological ordering consistent with the wiring diagram of the circuit: if gate $g$ is an ancestor of gate $g'$, then the solitary process for gate $g$ completes before the solitary process for gate $g'$ begins.

When the solitary process corresponding to any gate $g \in G$ begins running, $x_g$ is still set to its initialized state, $\emptyset$, while all of the parent nodes of $g$ are either input nodes, or other gates that have completed running and are set to their output values. By the end of the solitary process for gate $g$, $x_g$ is set to a random sample of the conditional distribution $\pi_g(x_g|x_{\mathrm{pa}(g)})$, while its parents are reinitialized to state $\emptyset$. More formally, under the solitary process for gate $g$, nodes $n(g)$ evolve according to

$$P_g(x'_{n(g)}|x_{n(g)}) := \pi_g(x'_g|x_{\mathrm{pa}(g)}) \prod_{v\in\mathrm{pa}(g)} \delta(x'_v, \emptyset) \tag{29}$$

while all nodes $V\backslash n(g)$ do not change their states. (Recall notation from section 2.5). Note that this means that the input nodes are reinitialized as soon as their child gates have run.

**Example 4.** In this example we demonstrate how to implement an XOR gate $g$ in an SR circuit with a CTMC, i.e., how to carry out the following logical map on the state of gate $g$,

$$\pi_g(x_g|x_{\mathrm{pa}(g)}) = \delta(x_g, \mathsf{XOR}(x_{\mathrm{pa}(g)})),$$

and then reset the gate's parents. The CTMC involves a sequence of two solitary processes over $n(g)$. The time-dependent rate matrix for both solitary processes has the form

$$K_t(x_V \to x'_V) = \delta(x_{V\backslash n(g)}, x'_{V\backslash n(g)}) K_t^{n(g)}(x_{n(g)} \to x'_{n(g)})$$

for all $x_V \neq x'_V$ (compare to equation (22), where for simplicity we assume there is a single thermodynamic reservoir). The two solitary processes differ in their associated subsystem rate matrices $K_t^{n(g)}$.

[11] Strictly speaking, if the circuit is a Bayes net, then $p_{\mathrm{IN}}$ should be a product distribution over the root nodes. Here we relax this requirement of Bayes nets, and let $p_{\mathrm{IN}}$ have arbitrary correlations.

[12] For example, it could be that at some $t < 0$, the joint state of the input nodes is some special initialized state $\vec{\emptyset}$ with probability 1, and that the initialized joint state is then overwritten with the values copied in from some variables in an offboard system, just before the circuit starts. The joint entropy of the offboard system and the circuit would not change in this overwriting operation, and so it is theoretically possible to perform that operation with zero EF [2]. However, to be able to run the circuit again after it finishes, with new values at the input nodes set this way, we need to reinitialize those input nodes to the joint state $\vec{\emptyset}$ . As elaborated below, we do include the thermodynamic costs of reinitializing those input nodes in preparation of the next run of the circuit. This is consistent with modern analyses of Maxwell's demon, which account for the costs of reinitializing the demon's memory in preparation for its next run [2, 3].

In the first solitary process, the state of the gate's parents is held fixed, while the gate's output is changed from the initialized state to the correct XOR value. For $t \in [0, 1]$ (the units of time are arbitrary), the subsystem rate matrix that implements this solitary process is

$$K_t^{n(g)}\left(x_{n(g)} \to x'_{n(g)}\right) = \delta(x_{\mathrm{pa}(g)}, x'_{\mathrm{pa}(g)})\eta\left[(1-t)\delta(x'_g, \emptyset)/4 + t\pi_g(x'_g|x'_{\mathrm{pa}(g)})/4\right], \tag{30}$$

for $x_{n(g)} \neq x'_{n(g)}$, where $\eta > 0$ is the relaxation speed. Note that the term $\delta(x'_{n(g)}, \emptyset)$ inside the square brackets encodes the assumption that the initial state of the gate is $\emptyset$ with probability 1, while the factor of $1/4$ encodes the assumption that the initial distribution over the four possible states of the gate's parents is uniform.

From the beginning to the end of the first solitary process, the nodes $n(g)$ are updated according to the conditional probability distribution $P_g^{(1)}$, given by the time-ordered exponential of the rate matrix in equation (30) over $t \in [0, 1]$. In the quasi-static limit $\eta \to \infty$, this conditional distribution becomes

$$P_g^{(1)}(x'_{n(g)}|x_{n(g)}) = \delta(x_{\mathrm{pa}(g)}, x'_{\mathrm{pa}(g)})\pi_g(x'_g|x_{\mathrm{pa}(g)}).$$

In the second solitary process, the gate's output is held fixed while the gate's parents are reinitialized. Redefining the time coordinate so that this second process also transpires in $t \in [0, 1]$, its subsystem rate matrix is

$$K_t^{n(g)}\left(x_{n(g)} \to x'_{n(g)}\right) = \delta(x_g, x'_g)\eta\left[(1-t)\pi_g(x'_g|x'_{\mathrm{pa}(g)})/4 + t\prod_{v \in \mathrm{pa}(g)} \delta(x'_v, \emptyset)/2\right], \tag{31}$$

for $x_{n(g)} \neq x'_{n(g)}$, where $\eta$ is again the relaxation speed. Note that $\pi_g(x'_g|x'_{\mathrm{pa}(g)})/4$ is what the distribution over nodes $n(g)$ would be at the beginning of the second solitary process, if the distribution at the beginning of the first solitary process was $\delta(x'_g, \emptyset)/4$. From the beginning to the end of the second solitary process, the nodes $n(g)$ are updated according to the conditional probability distribution $P_g^{(2)}$, which is given by the time-ordered exponential of the rate matrix equation (31). In the quasi-static limit $\eta \to \infty$, this conditional distribution is

$$P_g^{(2)}(x'_{n(g)}|x_{n(g)}) = \delta(x_g, x'_g)\prod_{v \in \mathrm{pa}(g)} \delta(x'_v, \emptyset).$$

The sequence of two solitary processes causes the nodes in $n(g)$ to be updated according to the conditional distribution $P_g = P_g^{(1)}P_g^{(2)}$. In the quasi-static limit, this is

$$P_g(x'_{n(g)}|x_{n(g)}) = \pi_g(x'_g|x_{\mathrm{pa}(g)})\prod_{v \in \mathrm{pa}(g)} \delta(x'_v, \emptyset), \tag{32}$$

which recovers equation (29), as desired.

We now compute thermodynamic costs for the XOR gate. Let $\mathcal{Q}(p_{\mathrm{pa}(g)})$ be the total EF incurred by running the sequence of two solitary process, given some initial distribution $p_{\mathrm{pa}(g)}$ over the parents of gate $g$. Using results from section 4, write this EF as

$$\mathcal{Q}(p_{\mathrm{pa}(g)}) = S(p_{\mathrm{pa}(g)}) - S(\pi_g p_{\mathrm{pa}(g)}) + D(p_{\mathrm{pa}(g)}\|q_{\mathrm{pa}(g)}) - D(\pi_g p_{\mathrm{pa}(g)}\|\pi_g q_{\mathrm{pa}(g)}) + \sum_{c \in L(\pi_g)} p_{\mathrm{pa}(g)}(c)\hat{\sigma}_{n(g)}(q_{\mathrm{pa}(g)}), \tag{33}$$

where the three lines correspond to subsystem Landauer cost, subsystem mismatch cost, and subsystem residual EP, respectively. To derive this decomposition, we applied theorem 1, while considering the subset of states $\mathcal{Z} = \{x_{n(g)} \in \mathcal{X}_{n(g)} : x_g = \emptyset\}$ (note that for this $\mathcal{Z}$, $L_{\mathcal{Z}}(P) = L(\pi_g)$).

To compute the second and third of those terms, note that in the quasi-static limit, the prior distribution is uniform:

$$q_{\mathrm{pa}(g)}(x_{\mathrm{pa}(g)}) = 1/4. \tag{34}$$

To see this, suppose that the distribution over $n(g)$ when the sequence of processes begins is given by $p_{n(g)}(x_{n(g)}) = \delta(x_g, \emptyset)q_{\mathrm{pa}(g)}(x_{\mathrm{pa}(g)})$. Then,

(a) The system will remain in equilibrium during the first solitary process, thereby incurring zero EP. At the end of the first solitary process, it will have distribution

$$[P_g^{(1)}p_{n(g)}](x_{n(g)}) = q_{\mathrm{pa}(g)}(x_{\mathrm{pa}(g)})\pi_g(x'_g|x_{\mathrm{pa}(g)}). \tag{35}$$

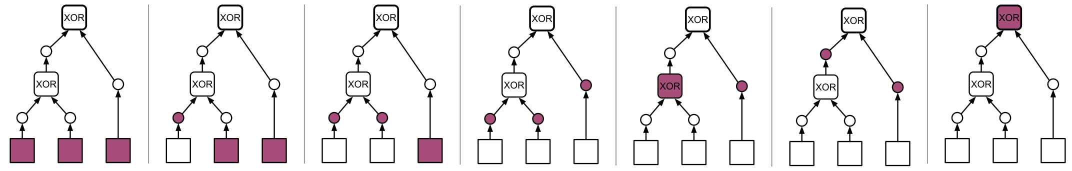

**Figure 4.** An SR implementation of the wired circuit shown in figure 3. Each diagram represents one step of the SR implementation, with white shapes indicating nodes set to their initialized value ($\emptyset$) and maroon shapes indicating nodes that can have non-initialized values. The implementation starts with only the input nodes set to non-initialized values (left-most diagram) and ends with only the output gates set to non-initialized values (right-most diagram).

(b) Given that the system starts the second solitary process with this distribution $P_g^{(1)}p_{n(g)}$, it will remain in equilibrium throughout the second solitary process, thereby again incurring zero EP.

So that sequence of processes will incur zero EP—the minimum possible—if the initial distribution is $q_{\mathrm{pa}(g)}$ over $\mathrm{pa}(g)$ (and $x_g = \emptyset$), as claimed. In addition, the fact that the minimal EP that can be generated for any initial distribution is strictly zero means that the subsystem residual EP vanishes. This fully specifies all terms in equation (33) as a function of $p_{\mathrm{pa}(g)}$.

As a concrete example of this analysis, consider the initial distribution which is uniform over states $\{00, 01, 10\}$:

$$p_{\mathrm{pa}(g)}(x_{\mathrm{pa}(g)}) = [1 - \delta(x_{\mathrm{pa}(g)}, 11)]/3.$$

For this distribution, the subsystem Landauer cost is

$$S(p_{\mathrm{pa}(g)}) - S(\pi_g p_{\mathrm{pa}(g)}) = \ln\, 3 + [(1/3)\ln(1/3) + (2/3)\ln(2/3)] \approx 0.46.$$

The subsystem EP is

$$D(p_{\mathrm{pa}(g)}\|q_{\mathrm{pa}(g)}) - D(\pi_g p_{\mathrm{pa}(g)}\|\pi_g q_{\mathrm{pa}(g)}) = [\ln\, 4 - \ln\, 2] - [S(p_{\mathrm{pa}(g)}) - S(\pi_g p_{\mathrm{pa}(g)})] \approx 0.23.$$

We end by noting that the XOR gate may also incur some EP which is not accounted for by these calculations, due to loss of correlations between the nodes $n(g)$ and the rest of the circuit as the gate runs. This is quantified by the Landauer loss, which can be evaluated using equations (25) and (26) or equation (27).

Given the requirement that the solitary processes are run accordingly to a topological ordering, equation (29) ensures that once all the gates of the circuit have run, the state of the output gates of the circuit have been set to a random sample of $\pi_\Phi(x_{\mathrm{OUT}}|x_{\mathrm{IN}})$, while all non-output nodes are back in their initialized states, i.e., $x_v = \emptyset$ for $v \in V\backslash\mathrm{OUT}$.

**Example 5.** Consider the 3-bit parity circuit shown in figure 3. An SR implementation of this wired circuit would run its 6 gates in topological order, such that each gate computes its output and then reinitializes its parents. One such sequence of steps is shown in figure 4 (note that some other topological orderings are also possible). Each XOR gate could be implemented by the kind of CTMC described in example 4. Each wired gate could be run by a similar kind CTMC, but which carries out the identity map $\pi_g(x'_g|x_{\mathrm{pa}(g)}) = \delta(x'_g, x_{\mathrm{pa}(g)})$, instead of the XOR map.

After the SR circuit has run, some ‘offboard system’ may make a copy of the state of the output gate for subsequent use, e.g., by copying it into some of the input bits of some downstream circuit(s), onto an external disk, etc. Regardless, we assume that after the circuit finishes, but before the circuit is run again, the state of the output nodes have also been reinitialized to $\emptyset$. Just as we do not model the physical mechanism by which new inputs are sampled for the next run of the circuit, we also do not model the physical mechanism by which the output of the circuit is reinitialized. Accordingly, in our calculation of the thermodynamic costs of running the circuit, we do not account for any possible cost of reinitializing the output[13]

This kind of cyclic procedure for running the circuit allows the circuit to be re-used an arbitrary number of times, while ensuring that each time it will have the same expected thermodynamic behavior (Landauer cost, mismatch cost, etc), and will carry out the same map $\pi_\Phi$ from input nodes to output gates.

[13] Suppose that the outputs of circuit $\Phi$ were the inputs of some subsequent circuit $\Phi'$. That would mean that when $\Phi'$ reinitializes its inputs, it would reinitialize the outputs of $\Phi$. Since we ascribe the thermodynamic costs of that reinitialization to $\Phi'$, it would result in double-counting to also ascribe the costs of reinitializing $\Phi$'s outputs to $\Phi$.

## 6. Thermodynamic costs of SR circuits

In general, there are multiple decompositions of the EF and EP incurred by running any given SR circuit. They differ in how much of the detailed structure of the circuit they incorporate. In this section we present some of these decompositions. (See appendix B for all proofs of the results in this section).

### 6.1. General decomposition of EF and EP

Let $p$ refer to the initial distribution over the joint state of all nodes in the circuit. By equation (5), the total EF incurred by implementing some overall map $P$ which takes the initial joint state of *all* nodes in the full circuit to the final joint state is

$$Q(p) = \mathcal{L}(p) + \sigma(p). \tag{36}$$

where

$$\mathcal{L}(p) := S(p) - S(Pp) \tag{37}$$

is the Landauer cost of computing $P$ for initial distribution $p$.

The first term in equation (37), $\mathcal{L}$, is the minimal EF that must be incurred by any process over $\mathcal{X}_{\mathrm{IN}} \times \mathcal{X}_{\mathrm{OUT}}$ that implements $P$, without any constraints on how the variables in $\mathcal{X}_{\mathrm{IN}} \times \mathcal{X}_{\mathrm{OUT}}$ are coupled, and without any reference to a set of intermediate subsystems (e.g., gates) that may connect the input and output variables[14].

The second term in equation (44) is the EP, which reflects the thermodynamic irreversibility of the SR circuit. Using equation (16), the EP can be further decomposed as

$$\sigma(p) = \left[D(p\|q) - D(Pp\|Pq)\right] + \sum_{c\in L(P)} p(c)\sigma(q^c). \tag{38}$$

The decrease in KL reflects the mismatch cost, arising from the discrepancy between $p(x_V)$, the actual initial distribution over all nodes of the circuit defined in equation (39), and $q(x_V)$, the optimal prior distribution over the joint state of all the nodes of the circuit which would result in the least EP. The last sum in equation (38) reflects the residual EP, reflecting EP that remains even when the circuit is initialized with the optimal prior distribution.

Suppose we know that the dynamics is actually implemented with an SR circuit, but do not know the precise wiring diagram. Then we know that the initial joint distribution over all the nodes is

$$p(x_V) = p_{\mathrm{IN}}(x_{\mathrm{IN}}) \prod_{v\in V\setminus \mathrm{IN}} \delta(x_v, \emptyset), \tag{39}$$

and the ending joint distribution is

$$[Pp](x_V) = p_{\mathrm{OUT}}(x_{\mathrm{OUT}}) \prod_{v\in V\setminus \mathrm{OUT}} \delta(x_v, \emptyset), \tag{40}$$

So $S(p) = S(p_{\mathrm{IN}})$ and $S(Pp) = S(p_{\mathrm{OUT}}) = S(\pi_\Phi p_{\mathrm{IN}})$, where $\pi_\Phi$ is the conditional distribution of the final joint state of the output gates given the initial joint state of the input nodes, defined in equation (14). Combining gives

$$\mathcal{L}(p) = S(p_{\mathrm{IN}}) - S(\pi_\Phi p_{\mathrm{IN}}) \tag{41}$$

Similarly, the EP becomes

$$\sigma(p) = \left[D(p_{\mathrm{IN}}\|q_{\mathrm{IN}}) - D(\pi_\Phi p_{\mathrm{IN}}\|\pi_\Phi q_{\mathrm{IN}})\right] + \sum_{c\in L(\pi_\Phi)} p(c)\sigma(q^c_{\mathrm{IN}}). \tag{42}$$

While the expressions in equations (38) and (42) for EP must be equal, how they decompose that EP among a mismatch cost term and a residual EP differ. The two decompositions differ because they define the ‘optimal initial distribution’ relative to differ sets of possible distributions, resulting in different prior distributions (which are also defined over different sets of outcomes). Also note that the residual EP terms in equation (42) are defined in terms of a more constrained minimization problem than the residual EP terms in equation (38). Thus, given the same initial distribution $p$, the residual EP in equation (42) will generally be larger than the residual EP in equation (38), while the mismatch cost in equation (38) will generally be larger than the mismatch cost in equation (42). We also emphasize that the island decompositions appearing in the two expressions are different.

[14] See [3] for a discussion of how this bound applies in the case of logically reversible circuits.

### 6.2. Circuit-based decompositions of EF and EP

The decompositions of EF and EP given in (equations (36), (38) and (42)) do not involve the wiring diagram of the SR circuit. As an alternative, we can exploit that wiring diagram to formulate a decomposition of EF and EP which separates the contributions from different gates. In general, such circuit-based decompositions allow for a finer-grained analysis of the EP in SR circuits than do the decompositions proposed in the last section. In particular, they allow us to derive some novel connections between nonequilibrium statistical physics, computer science theory, and information theory, as discussed in the next two subsections.

Before discussing these circuit-based decompositions, we introduce some new notation. We write $p_{\mathrm{pa}(g)}(x_{\mathrm{pa}(g)})$ and $p_{n(g)}(x_{n(g)}) = p_{\mathrm{pa}(g)}(x_{\mathrm{pa}(g)})\delta(x_g, \emptyset)$ for the distributions over $x_{\mathrm{pa}(g)}$ and $x_{n(g)}$, respectively, at the beginning of the solitary process that implements gate $g$. We write the EF function of the solitary process of gate $g$ as $\mathcal{Q}_g(p_{n(g)})$, and its subsystem EP as

$$\hat{\sigma}_g(p_{n(g)}) := \mathcal{Q}_g(p_{n(g)}) - [S(p_{n(g)}) - S(P_g p_{n(g)})]. \tag{43}$$

We also write $p^{\mathrm{beg}(g)}$ and $p^{\mathrm{end}(g)}$ to indicate the joint distribution over *all* circuit nodes at the beginning and end, respectively, of the solitary process that runs gate $g$. As an illustration of this notation, $p^{\mathrm{beg}(g)}(x_{\mathrm{pa}(g)}) = p_{\mathrm{pa}(g)}(x_{\mathrm{pa}(g)})$. On the other hand, $p^{\mathrm{end}(g)}(x_g) = (\pi_g p_{\mathrm{pa}(g)})(x_g)$ is the distribution over $x_g$ after gate $g$ runs, and $p^{\mathrm{end}(g)}(x_{\mathrm{pa}(g)})$ is a delta function about the joint state of the parents of $g$ in which they are all initialized, by equation (29). Note that since we are considering solitary processes, $p^{\mathrm{beg}(g)}(x_{V\setminus n(g)}) = p^{\mathrm{end}(g)}(x_{V\setminus n(g)})$.

We now present our first circuit-based decomposition, and then we explain what its terms mean in detail:

**Theorem 2.** *The total EF incurred by running an SR circuit where p is the initial distribution over the joint state of all nodes in the circuit is*

$$\mathcal{Q}(p) = \mathcal{L}(p) + \underbrace{\mathcal{L}^{\mathrm{loss}}(p) + \mathcal{M}(p) + \mathcal{R}(p)}_{\sigma(p)}. \tag{44}$$

(1) The first term in equation (44), $\mathcal{L}(p)$, is the Landauer cost of the circuit, as described in equation (37). This Landauer cost can be further decomposed into contributions from the individual gates. Specifically, write $\mathcal{L}_g(p)$ for the drop in the entropy of the entire circuit during the time that the solitary process for gate $g$ runs, given that the input distribution over the entire circuit is $p$:

$$\mathcal{L}_g(p) := S(p^{\mathrm{beg}(g)}) - S(p^{\mathrm{end}(g)}). \tag{45}$$

Note that the distribution over the states of the entire physical circuit at the end of the running of any gate is the same as the distribution at the beginning of the running of the next gate. So by canceling terms, and using the fact that entropy does not change when a wire gate runs, we can expand $\mathcal{L}$ as

$$\mathcal{L}(p) = \sum_{g\in G\setminus W} \mathcal{L}_g(p). \tag{46}$$

(Recall from section 2.5 that $W$ is the set of wire gates in the circuit.) This decomposition will be useful below.

(2) The second term in equation (44), $\mathcal{L}^{\mathrm{loss}}(p)$, is the unavoidable additional EF that is incurred by any SR implementation of the SR circuit on initial distribution $p$, above and beyond $\mathcal{L}$, the Landauer cost of running the map $\pi_\Phi$ on initial distribution $p$. We refer to this unavoidable extra EF as the **circuit Landauer loss**. It equals the sum of the subsystem Landauer losses incurred by each non-wire gate's solitary process,

$$\mathcal{L}^{\mathrm{loss}}(p) = \sum_{g\in G\setminus W} \mathcal{L}_g^{\mathrm{loss}}(p), \tag{47}$$

where $\mathcal{L}_g^{\mathrm{loss}}(p) = S(p_{\mathrm{pa}(g)}) - S(\pi_g p_{\mathrm{pa}(g)}) - \mathcal{L}_g(p)$.

Each term $\mathcal{L}_g^{\mathrm{loss}}(p)$ in this sum is non-negative (see end of section 4), and so $\mathcal{L}^{\mathrm{loss}}(p) \geqslant 0$. Note that we can omit wires from the sum in equation (47) because $\pi_g$ is logically reversible for any wire gate $g$, which means that $\mathcal{L}_g^{\mathrm{loss}}(p) = 0$ for such gates.

We define **circuit Landauer cost** to be the minimal EF incurred by running any SR implementation of the circuit, i.e.,

$$\mathcal{L}^{\mathrm{circ}}(p) = \mathcal{L}(p) + \mathcal{L}^{\mathrm{loss}}(p) \tag{48}$$

$$= \sum_{g \in G \setminus W} \left[ S(p_{\mathrm{pa}(g)}) - S(\pi_g p_{\mathrm{pa}(g)}) \right]. \tag{49}$$

Recall that $\mathcal{L}(p)$ is the minimal EF that must be generated by any physical process that carries out the map $P$ on initial distribution $p$. So by equation (48), $\mathcal{L}^{\mathrm{loss}}(p)$ is the minimal additional EF that must be generated if we use an SR circuit to carry out $P$ on $p$, no matter how efficient the gates in the circuit are. In this sense, equation (48) can be viewed as an extension of the generalized Landauer bound, to concern SR circuits.

(3) The third term in equation (44), $\mathcal{M}$, reflects the EF incurred because the actual initial distribution of each gate $g$ is not the optimal one for that gate (i.e., not one that minimizes subsystem EP within each island of the conditional distribution $P_g$, defined in equation (29)). We refer to this cost as the **circuit mismatch cost**, and write it as

$$\mathcal{M}(p) = \sum_{g \in G} \left[ D\left(p_{n(g)} \| q_{n(g)}\right) - D\left(P_g p_{n(g)} \| P_g q_{n(g)}\right) \right] \tag{50}$$

where the **prior** $q_{n(g)}$ is a distribution over $\mathcal{X}_{n(g)}$ whose conditional distributions over the islands $c \in L(P_g)$ all obey $\hat{\sigma}_g(q^c_{n(g)}) = \min_{r:\mathrm{supp}\, r \subseteq c} \hat{\sigma}_g(r)$. Note that we must include wire gates $g$ in the sum in equation (50) even though $\pi_g$ for a wire gate is logically reversible. This is because the associated overall map over $n(g)$, equation (29), is not logically reversible over $n(g)$[15].

$\mathcal{M}$ is non-negative, since each gate's subsystem mismatch cost is non-negative. Moreover, $\mathcal{M}$ achieves its minimum value of 0 when $p^{c_g}_{n(g)} = q^{c_g}_{n(g)}$ for all $g \in G$ and all islands $c_g \in L(P_g)$. (Recall that subsystem priors like $q^{c_g}_{n(g)}$ reflect the specific details of the underlying physical process that implements the gate $g$, such as how its energy spectrum evolves as it runs.)

Suppose that one wishes to construct a physical system to implement some circuit, and can vary the associated subsystem priors $q^{c_g}_{n(g)}$ arbitrarily. Then in order to minimize mismatch cost one should choose priors $q^{c_g}_{n(g)}$ that equal the actual associated initial distributions $p^c_{n(g)}$. Moreover, those actual initial distributions $p^{c_g}_{n(g)}$ can be calculated from the circuit's wiring diagram, together with the input distribution of the entire circuit, $p_{\mathrm{IN}}$, by 'propagating' $p_{\mathrm{IN}}$ through the transformations specified by the wiring diagram. As a result, given knowledge of the wiring diagram and the input distribution of the entire circuit, in principle the priors can be set so that mismatch cost vanishes.

(4) The fourth term in equation (44), $\mathcal{R}$, reflects the remaining EF incurred by running the SR circuit, and so we call it **circuit residual EP**. Concretely, it equals the subsystem EP that would be incurred even if the initial distribution within each island of each gate were optimal:

$$\mathcal{R}(p) = \sum_{g \in G} \sum_{c \in L(P_g)} p_{n(g)}(c)\, \hat{\sigma}_g(q^c_{n(g)}). \tag{51}$$

Circuit residual EP is non-negative, since each $\hat{\sigma}_g$ is non-negative. Since for every gate $g$, $p_{n(g)}(c)$ is a linear function of the initial distribution to the circuit as a whole, circuit residual EP also depends linearly on the initial distribution. Like the priors of the gates, the residual EP terms $\{\hat{\sigma}_g(q^c_{n(g)})\}$ reflect the 'nitty-gritty' details of how the gates run.

To summarize, the EF incurred by a circuit can be decomposed into the Landauer cost (the contribution to the EF that would arise even in a thermodynamically reversible process) plus the EP (the contribution to that EF which is thermodynamically irreversible). In turn, there are three contributions to that EP:

(*a*) Circuit Landauer loss, which is independent of how the circuit is physically implemented, but does depend on the wiring diagram of the circuit, the conditional distributions implemented by the gates, and the initial distribution over inputs. It is a nonlinear function of the distribution over inputs, $p_{\mathrm{IN}}$.

(*b*) Circuit mismatch cost, which does depend on how the circuit is physical implemented (via the priors), as well as the wiring diagram. It is also a nonlinear function of $p_{\mathrm{IN}}$.

(*c*) Circuit residual EP, which also depends on how the circuit is physical implemented. It is a linear function of $p_{\mathrm{IN}}$. However, no matter what the wiring diagram of the circuit is, if we implement each of

[15] There are several ways that we could manufacture wires that would allow us to exclude them from the sum in equation (50). One way would be to modify the conditional distribution $\mathrm{P_g}$ of the wire gates, replacing the logically irreversible equation (29) with the logically reversible $P_{\mathrm{g}}(x'_{n(g)}|x_{n(g)}) = \delta(xg, x_{\mathrm{pa}(g)})\delta(x'_{\mathrm{pa}(g)}, x_g)$ (thus, each wire gate would basically 'flip' its input and output). Since in an SR circuit an initial state $\mathrm{x_g} \neq \emptyset$ should not arise for any gate g, such modified wire gates would always end up performing the same logical operation as would wire gates that obey equation (29). Another way that wire gates could be excluded from the sum in equation (50) is if their priors had the form $q_n(g)(x_{\mathrm{pa}}(g), x_g) = p_{\mathrm{pa}}(g)(x_{\mathrm{pa}}(g))\delta(x_g, \emptyset)$ (see appendix A for details).

the gates in a circuit with a quasistatic process, then the associated circuit residual EP is identically zero, independent of $p_{\mathrm{IN}}$[16].

There are other useful decompositions of the EP incurred by an SR circuit that incorporate the wiring diagram. One such alternative decomposition, which is our second main result, leaves the circuit Landauer loss term in equation (44) unchanged, but modifies the circuit mismatch cost and the circuit residual EP terms.

**Theorem 3.** *The total EF incurred by running an SR circuit where p is the initial distribution over the joint state of all nodes in the circuit is*

$$\mathcal{Q}(p) = \mathcal{L}(p) + \underbrace{\mathcal{L}^{\text{loss}}(p) + \mathcal{M}'(p) + \mathcal{R}'(p)}_{\sigma(p)} . \tag{52}$$

To present this decomposition, recall from equation (29) that for any gate $g$, the distribution over $\mathcal{X}_{n(g)}$ has partial support at the beginning of the solitary process that implements $P_g$, since there is 0 probability that $x_g \neq \emptyset$. We use this fact to apply theorem 1 to equation (18), while taking $\mathcal{Z} = \{x_{n(g)} \in \mathcal{X}_{n(g)} : x_g = \emptyset\}$. This allows us to express the modified circuit mismatch cost by replacing all of the map $P_g(x'_{n(g)}|x_{n(g)})$ in the summand in equation (50) with $P_g(x'_g|x_{\mathrm{pa}(g)}) = \pi_g$:

$$\mathcal{M}'(p) = \sum_{g \in G \setminus W} \left[ D(p_{\mathrm{pa}(g)} \| q_{\mathrm{pa}(g)}) - D(\pi_g p_{\mathrm{pa}(g)} \| \pi_g q_{\mathrm{pa}(g)}) \right] \tag{53}$$

where the priors $q_{\mathrm{pa}(g)}$ are defined in terms of the island decompositions of the associated conditional distributions $\pi_g$, rather than in terms of the island decompositions of the conditional distributions $P_g$. Note that can exclude the wire gates from the sum in equation (53) because each wire gate's $\pi_g$ is logically reversible, and so the associated drop in KL divergence are zero. Then, the modified circuit residual EP is

$$\mathcal{R}'(p) = \sum_{g \in G} \sum_{c \in L(\pi_g)} p_{\mathrm{pa}(g)}\, \hat{\sigma}_g(q^c_{\mathrm{pa}(g)}), \tag{54}$$

where each term $\hat{\sigma}_g(q^c_{\mathrm{pa}(g)})$ is given by appropriately modifying the arguments in equation (43). In deriving equation (54), we used the fact that $L_{\mathcal{Z}}(P_g) = L(\pi_g)$ (for $\mathcal{Z}$ as defined above for each gate g).

As with the analogous results in the previous section, theorems 2 and 3 differ, because they define 'optimal initial distribution' relative to different sets of possibilities. In particular, the decomposition in theorem 2 will generally have a larger mismatch cost and smaller residual EP term than the decomposition in theorem 3.

For the rest of this section, we will use the term 'circuit mismatch cost' to refer to the expression in equation (53) rather than the expression in equation (50), and similarly will use the term 'circuit residual EP' to refer to the expression in equation (54) rather than the expression in equation (51).

### 6.3. Information theory and circuit Landauer loss

By combining equations (47) and (26), we can write circuit Landauer loss as

$$\mathcal{L}^{\text{loss}}(p) = \sum_{g \in G \setminus W} \left[ I(X_{\mathrm{pa}(g)}; X_{V \setminus \mathrm{pa}(g)}) - I(X_g; X_{V \setminus g}) \right] \tag{55}$$

Any nodes that belong to $V \setminus n(g)$ and that are in their initialized state when gate $g$ starts to run will not contribute to the drop in mutual information terms in equation (55). Keeping track of such nodes and simplifying establishes the following:

**Corollary 4.** *The circuit Landauer loss is*

$$\mathcal{L}^{\text{loss}}(p) = \mathcal{I}(p_{\mathrm{IN}}) - \mathcal{I}(\pi_\Phi p_{\mathrm{IN}}) - \sum_{g \in G \setminus W} \mathcal{I}(p_{\mathrm{pa}(g)}).$$

(We remind the reader that $\mathcal{I}(p_A)$ refers to the multi-information between the variables indexed by $A$.) Corollary 4 suggest a set of novel optimization problems for how to design SR circuits: given some desired computation $\pi_\Phi$ and some given initial distribution $p_{\mathrm{IN}}$, find the circuit wiring diagram that carries out $\pi_\Phi$ while minimizing the circuit Landauer loss. Presuming we have a fixed input distribution $p$ and map $\pi_\Phi$, the term $\mathcal{I}(p_{\mathrm{IN}}) - \mathcal{I}(\pi_\Phi p_{\mathrm{IN}})$ in corollary 4 is an additive constant that does not depend on the particular

[16] To see this, simply consider equation (23) in example 3, and recall that it is possible to choose a time-varying rate matrix to implement any desired map over any space in such a way that the resultant EF equals the drop in entropies [21].

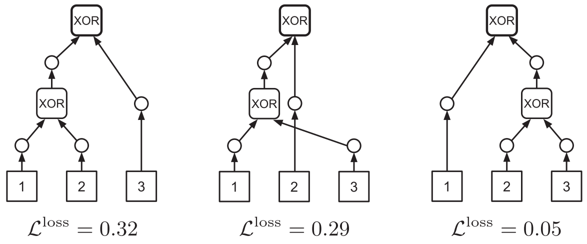
XOR
XOR
1
2
3
$\mathcal{L}^{\text{loss}} = 0.32$
XOR
XOR
1
2
3
$\mathcal{L}^{\text{loss}} = 0.29$
XOR
XOR
1
2
3
$\mathcal{L}^{\text{loss}} = 0.05$

**Figure 5.** Three different circuit wiring diagrams for computing 3-bit parity using XOR gates. The circuit Landauer loss for each wiring diagram is shown below, given the input distribution of equation (56).

choice of circuit wiring diagram. So the optimization problem can be reduced to finding which wiring diagram results in a minimal value of $\sum_{g\in G\setminus W}\mathcal{I}(p_{\text{pa}(g)})$. In other words, for fixed $p$ and map $\pi_\Phi$, to minimize the Landauer loss we should choose the wiring diagram for which the parents of each gate are as strongly correlated among themselves as possible. Intuitively, this ensures that the 'loss of correlations', as information is propagated down the circuit, is as small as possible.

In general, the distributions over the outputs of the gates in any particular layer of the circuit will affect the distribution of the inputs of all of the downstream gates, in the subsequent layers of the circuit. This means that the sum of multi-informations in corollary 4 is an inherently global property of the wiring diagram of a circuit; it cannot be reduced to a sum of properties of each gate considered in isolation, independently of the other gates. This makes the optimization problem particularly challenging. We illustrate this optimization problem in the following example.

**Example 6.** Consider again the case where we want our circuit to compute the 3-bit parity function using 2-input XOR gates, i.e., it want it to implement the map

$$\pi_\Phi(x_g|x_1,x_2,x_3) = \delta(x_g,\delta(x_1+x_2+x_3,1)+\delta(x_1+x_2+x_3,3)).$$

Suppose we happen to know that the input distribution to the circuit will be

$$p_{\text{IN}}(x_1,x_2,x_3) = \frac{1}{Z}\mathrm{e}^{-[\phi(x_1)\phi(x_2)/4+\phi(x_1)\phi(x_3)/2+\phi(x_2)\phi(x_3)]}, \tag{56}$$

where $Z$ is a normalization constant and $\phi(x) := 2x-1$ is a function that maps bit values $x\in\{0,1\}$ to spin values, $\phi(x)\in\{-1,1\}$. The distribution equation (56) is a Boltzmann distribution for a pairwise spin model, where inputs 2 and 3 have the strongest coupling strength (1), inputs 1 and 3 have intermediate-strength coupling strength (1/2), and inputs 1 and 2 have the weakest coupling strength (1/4).

We wish to find the wiring diagram connecting our XOR gates that has minimal circuit Landauer cost for this distribution over its three input bits. It turns out that we can restrict our search to three possible wiring diagrams, which are shown in figure 5. We indicate the circuit Landauer loss for the input distribution of equation (56) for each of those three wiring diagrams. So for this input distribution, the right-most wiring diagram results in minimal circuit Landauer loss. Note that this wiring diagram aligns with the correlational structure of the input distribution (given that inputs 2 and 3 have the strongest statistical correlation).

An interesting variant of the optimization problem described above arises if we model the residual EP terms for the wire gates. In any SR circuit, wire gates carry out a logically reversible operation on their inputs. Thus, by equation (16), *all* of the EF generated by any wire gates is residual EP. If we allow the physical lengths of wires to vary, then as a simple model we could presume that the residual EP of any wire is proportional to its length. This would allow us to incorporate into our analysis the thermodynamic effect of the geometry with which a circuit is laid out on a two-dimensional circuit boards, in addition to the thermodynamic effect of the topology of that circuit.

Finally, note that for any set of nodes $A$, multi-information can be bounded as

$$0\leqslant\mathcal{I}(p_A)\leqslant\textstyle\sum_{v\in A}S(p_v)\leqslant\sum_{v\in A}|\mathcal{X}_v|.$$

Given this, corollary 4 implies

$$\mathcal{L}^{\text{loss}}(p)\leqslant\mathcal{I}(p_{\text{IN}})\leqslant\ln|\mathcal{X}_{\text{IN}}|. \tag{57}$$

This means that for a fixed input state space, the circuit Landauer loss cannot grow without bound as we vary the wiring diagram. Interestingly, this bound on Landauer loss only holds for SR circuits that have out-degree 1. If we consider SR circuits that have out-degree greater than 1, then the circuit Landauer cost can be arbitrarily large. This is formalized as the following proposition, which is proved in appendix C.

**Proposition 1.** *For any $\pi_\Phi$, non-delta function input distribution $p_{\mathrm{IN}}$, and $\kappa \geqslant 0$, there exist an SR circuit with out-degree greater than 1 that implements $\pi_\Phi$ for which $\mathcal{L}^{\mathrm{loss}}(p) \geqslant \kappa$.*

### 6.4. Information theory and circuit mismatch loss

Landauer loss captures the gain in minimal EF due to using an SR circuit, which becomes equal to the gain in actual EF if there is no mismatch cost or residual EP. It is harder to make general statements about the gain in actual EF due to using an SR circuit, i.e., when the mismatch cost and/or residual EP is nonzero. In this subsection we make some preliminary remarks about this issue.

Imagine that we wish to build a physical process that implements some computation $\pi_\Phi(x_{\mathrm{OUT}}|x_{\mathrm{IN}})$ over a space $\mathcal{X}_{\mathrm{IN}} \times \mathcal{X}_{\mathrm{OUT}}$. Suppose we want this process to achieve minimal EP when run with inputs generated by $q_{\mathrm{IN}}$ (e.g., if we expect future inputs to the process to be generated by sampling $q_{\mathrm{IN}}$), and as usual assume the initial value of $x_{\mathrm{OUT}}$ will be $\emptyset$ whenever it is run. Using the decomposition of equation (42) and assuming that the residual EP of the process is zero, the EF that such a process would generate if it is actually run with an input distribution $p$ (initialized like SR circuits are, so that it has the form of equation (39)) is given by the sum of the Landauer cost and the mismatch cost, with no Landauer loss term. We write this as

$$\mathcal{Q}_{\mathrm{AO}}(p) = \mathcal{L}(p) + D(p_{\mathrm{IN}}\|q_{\mathrm{IN}}) - D(\pi_\Phi p_{\mathrm{IN}}\|\pi_\Phi q_{\mathrm{IN}}), \tag{58}$$

Note that in order for the EF generated by an actual physical process to be given by equation (58), the prior of that process must be $q_{\mathrm{IN}}$, and in general this may require that the process couple together arbitrary sets of variables. This means that the EF generated by implementing $\pi_\Phi(x_{\mathrm{OUT}}|x_{\mathrm{IN}})$ with an SR circuit cannot obey equation (58) in general, due to restrictions on what variables can be coupled in such a circuit. (One can verify, for example, that the prior distribution $q_{\mathrm{IN}}$ of a circuit consisting of two disconnected bit erasing gates must be a product distribution over the two input bits.) To emphasize this distinction, we will refer to a process whose EF is given by equation (58)) as an 'all-at-once' (AO) process (indicated by the subscript 'AO' in equation (58)).

For practical reasons, it may be quite difficult to construct an AO process that implements $\pi_\Phi$, and we must use a circuit implementation instead. In particular, even though the circuit as a whole cannot have prior $q_{\mathrm{IN}}$, suppose we can set the priors $q_{\mathrm{pa}(g)}$ at its gates by propagating $q_{\mathrm{IN}}$ through the wiring diagram of the circuit. Assuming again that there is zero EP, the EF that must be incurred by any such SR circuit implementation of $\pi_\Phi$ on input distribution $p$, assuming some particular wiring topology and gate priors, is given by the decomposition of theorem 3,

$$\mathcal{Q}_{\mathrm{circ}}(p) = \mathcal{L}(p) + \mathcal{L}^{\mathrm{loss}}(p) + \mathcal{M}'(p). \tag{59}$$

We now ask: how much larger is this EF incurred by the SR circuit implementation, compared to that of the original AO process? Subtracting equation (58) from equation (59) gives

$$\begin{aligned}\Delta\mathcal{Q} &= \mathcal{Q}_{\mathrm{circ}}(p_{\mathrm{IN}}) - \mathcal{Q}_{\mathrm{AO}}(p_{\mathrm{IN}})\\ &= \mathcal{L}^{\mathrm{loss}}(p_{\mathrm{IN}}) + \mathcal{M}^{\mathrm{loss}}(p_{\mathrm{IN}}\|q_{\mathrm{IN}})\end{aligned} \tag{60}$$

where we have defined $\mathcal{M}^{\mathrm{loss}}$ as the difference between the circuit mismatch cost, $\mathcal{M}'(p)$, and the mismatch cost of the AO process. We refer to that difference in mismatch costs as the **circuit mismatch loss**, and use equation (53) to express it as

$$\mathcal{M}^{\mathrm{loss}}(p_{\mathrm{IN}}\|q_{\mathrm{IN}}) = \mathcal{D}(\pi_\Phi p_{\mathrm{IN}}\|\pi_\Phi q_{\mathrm{IN}}) - \mathcal{D}(p_{\mathrm{IN}}\|q_{\mathrm{IN}}) + \sum_{g\in G\setminus W} \mathcal{D}(p_{\mathrm{pa}(g)}\|q_{\mathrm{pa}(g)}), \tag{61}$$

where $\mathcal{D}$ refers to the multi-divergence, defined in equation (10). Equation (61) can be compared to corollary 4, which expresses the circuit Landauer cost rather than circuit mismatch cost, and involves multi-informations rather than multi-divergences.

Interestingly, while circuit Landauer loss is non-negative, circuit mismatch loss can either be positive or negative. In fact, depending on the wiring diagram, $p_{\mathrm{IN}}$ and $q_{\mathrm{IN}}$, the sum of circuit mismatch loss and circuit Landauer loss can be negative. This means that when the actual input distribution $p_{\mathrm{IN}}$ is different from the prior distribution of the AO process, the 'closest equivalent circuit' to the AO process may actually incur less EF than the corresponding AO process. This occurs because an SR circuit cannot implement some of the prior distributions that an AO process can implement, so the two implementations end up having different priors. This is illustrated in the following example.

**Example 7.** Assume the desired computation is the erasure of two bits, $\pi_\Phi(x_3, x_4|x_1, x_2) = \delta(x_3, 0)\delta(x_4, 0)$, where $x_1$ and $x_2$ refers to the input bits, and $x_3$ and $x_4$ refers to the output bits. The prior distribution implemented by an AO process is given by $q_{\mathrm{IN}}(0, 0) = q(1, 1) = \epsilon < 1/2$, and $q_{\mathrm{IN}}(0, 1) = q(1, 0) = 1/2 - \epsilon$. The actual input distribution is given by a delta function distribution, $p_{\mathrm{IN}}(x_1, x_2) = \delta(x_1, 0)\delta(x_2, 0)$.

We now implement this computation using an SR circuit which consists of two disconnected erasure gates. The closest equivalent SR circuit has gate priors given by the uniform marginal distributions, $q(x_1) = 1/2$ and $q(x_2) = 1/2$. Then the difference between the EF of the AO process and the SR circuit is

$$\begin{aligned}\Delta \mathcal{Q} &= -\left[S(p_{\mathrm{IN}}) + D(p_{\mathrm{IN}}\|q_{\mathrm{IN}})\right] + \left[\sum_g S(p_{\mathrm{pa}(g)}) + D(p_{\mathrm{pa}(g)}\|q_{\mathrm{pa}(g)})\right]\\ &= \sum_{x_1,x_2} p_{\mathrm{IN}}(x_1, x_2) \ln q_{\mathrm{IN}}(x_1, x_2) - \sum_{x_1} p(x_1) \ln q(x_1) - \sum_{x_2} p(x_2) \ln q(x_2)\\ &= \ln \epsilon + 2 \ln 2\end{aligned}$$

This can be made arbitrarily negative by taking $\epsilon$ sufficiently close to zero. Thus, the EF of the AO process may be arbitrarily larger than the EF of the closest equivalent SR circuit.

## 7. Related work

The issue of how the thermodynamic costs of a circuit depend on the constraints inherent in the topology of the circuit has not previously been addressed using the tools of modern nonequilibrium statistical physics. Indeed, this precise issue has received very little attention in *any* of the statistical physics literature. A notable exception was a 1996 paper by Gernshenfeld [47], which pointed out that all of the thermodynamic analyses of conventional (irreversible) computing architectures at the time were concerned with properties of individual gates, rather than entire circuits. That paper works through some elementary examples of the thermodynamics of circuits, and analyzes how the global structure of circuits (i.e., their wiring diagram) affects their thermodynamic properties. Gernshenfeld concludes that the 'next step will be to extend the analysis from these simple examples to more complex systems'[17].

There are also several papers that do not address circuits, but focus on tangentially related topics, using modern nonequilibrium statistical physics. Ito and Sagawa [40, 41] considered the thermodynamics of (time-extended) Bayesian networks [48, 49]. They divided the variables in the Bayes net into two sets: the sequence of states of a particular system through time, which they write as $X$, and all external variables that interact with the system as it evolves, which they write as $\mathcal{C}$. They then derive and investigate an integral fluctuation theorem [12, 15, 33, 50] that relates the EP generated by $X$ and the EP flowing between $X$ and $\mathcal{C}$. (See also [51]).

Note that [40] focuses on the EP generated by a proper subset of the nodes in the entire network. In contrast, our results below concern the EP generated by all nodes. In addition, while [40] concentrates on an integral fluctuation theorem involving EP, we give an exact expression for the expected EP.

Otsubo and Sagawa [52] considered the thermodynamics of stochastic Boolean network models of gene regulatory networks. They focused in particular on characterizing the information-theoretic and dissipative properties of 3-node motifs. While their study does concern dynamics over networks, it has little in common with the analysis in the current paper, in particular due to its restriction to 3-node systems.

Solitary processes are similar to 'feedback control' processes, which have attracted much attention in the thermodynamics of information literature [2, 53, 54]. In feedback control processes, there is a subsystem $A$ that evolves while coupled to another subsystem $B$, which is held fixed. (This joint evolution is often used to represent either $A$ making a measurement of the state of $B$, or the state of $B$ being used to determine which control protocol to apply to $B$.) It has been shown for feedback control processes that the total EP incurred by the joint $A \times B$ system is the 'subsystem EP' of $A$, plus the drop in the mutual information between $A$ and $B$ [53]. Formally, this is identical to equation (26).

Crucially however, in feedback control processes there is no assumption that $A$ and $B$ are physically decoupled. (Formally, equation (20) is not assumed.) Therefore the change in mutual information can either be negative or positive in those processes (the latter occurs, for instance, when $A$ performs a

[17] This prescient article even contains a cursory discussion of the thermodynamic consequences of providing a sequence of non-IID rather than IID inputs to a computational device, a topic that has recently received renewed scrutiny [55, 56, 66, 67].

measurement of the state of $B$). In addition, the 'subsystem EP' in these processes can be negative. For this reason, in feedback control processes there is no simple relationship between subsystem EP and the total EP incurred by the joint $A \times B$ system. In contrast, in solitary processes $A$ and $B$ are physically decoupled (see equations (20) and (21)). For this reason, in solitary processes subsystem EP is non-negative, as is the drop in mutual information, equation (26), and so each of them is a lower bound on the total EP incurred by the joint $A \times B$ system.

Boyd *et al* [55] considered the thermodynamics of 'modular' systems, which in our terminology are a special type of solitary processes, with extra constraints imposed. In particular, to derive their results, [55] assumes there is exactly one thermodynamic reservoir (in their case, a heat bath). That restricts the applicability of their results. Nonetheless, individual gates in a circuit are run by solitary processes, and one could require that they in fact be run by modular systems, in order to analyze the thermodynamics of (appropriately constrained) circuits. However, instead of focusing on this issues, [55] focuses on the thermodynamics of 'information ratchets' [56], modeling them as a sequence of iterations of a single solitary process, successively processing the symbols on a semi-infinite tape. In contrast, we extend the analysis of single solitary processes operating in isolation to analyze full circuits that comprise multiple interacting solitary processes.

Riechers [57] also contains work related to solitary processes, assuming a single heat bath, like [55]. [57] exploits the decomposition of EP into 'mismatch cost' plus 'residual EP' introduced in [30], in order to analyze thermodynamic attributes of a special kind of circuit. The analysis in that paper is not as complicated as either the analysis in the current paper or the analysis in [55]. That is because [57] does not focus on how the thermodynamic costs of running a system are affected if we impose a constraint on how the system is allowed to operate (e.g., if we require that it use solitary processes). In addition, the system considered in that paper is a very special kind of circuit: a set of $N$ disconnected gates, working in parallel, with the outputs of those gates never combined.

[3] is a survey article relating many papers in the thermodynamics of computation. To clarify some of those relationships, it introduces a type of process related to solitary processes, called 'subsystem processes'. (See also [58].) For the purposes of the current paper though, we need to understand the thermodynamics specifically of solitary processes. In addition, being a summary paper, [3] presents some results from the arXiv preprint version of the current paper, [58]. Specifically, [3, 58] summarize some of the thermodynamics of straight-line circuits subject to the extra restriction (not made in the current paper) that there only be a single output node.

There is a fairly extensive literature on 'logically reversible circuits' and their thermodynamic properties [3, 6, 59–61]. This work is based on the early analysis in [62], and so it is not grounded in modern nonequilibrium statistical physics. Indeed, modern nonequilibrium statistical physics reveals some important subtleties and caveats with the thermodynamic properties of logically reversible circuits [3]. Also see [16] for important clarifications of the relationship between thermodynamic and logical reversibility, not appreciated in some of the research community working on logically reversible circuits.

Finally, another related paper is [63]. This paper starts by taking a distilled version of the decomposition of EP in [30] as given. It then discusses some of the many new problems in computer science theory that this decomposition leads to, both involving circuits and involving many other kinds of computational system.

## 8. Discussion and future work

It is important to emphasize that SR circuits are somewhat unrealistic models of real digital circuits. For example, many real digital circuits have multiple gates running at the same time, and often do not reinitialize their gates after they are run. In addition, many real digital circuits have characteristics like loops and branching. This makes them challenging to model at all using simple solitary processes. Extending our analysis to these more general models of circuits is an important direction for future work. Nonetheless, it is worth mentioning that all of the thermodynamic costs discussed above—including Landauer loss, mismatch cost, and residual EP—are intrinsic to *any* physical process, as described in section 3. So versions of them arise in those other kinds of circuits, only in modified form.

An interesting set of issues to investigate in future work is the scaling properties of the thermodynamic costs of SR circuits. In conventional circuit complexity theory [4, 5] one first specifies a 'circuit family' which comprises an infinite set of circuits that have different size input spaces but that are all (by definition) viewed as 'performing the same computation'. For example, one circuit family is given by an infinite set of circuits each of which has a different size input space, and outputs the single bit of whether the number of 1's in its input string is odd or even. Circuit complexity theory is concerned with how various resource costs in making a given circuit (e.g., the number of gates in the circuit) scales with the size of the circuit as one goes through the members of a circuit family. For example, it may analyze how the number of gates in a set of circuits, each of which determines whether its input string contains an odd number of 1's, scales with the size of those input strings. One interesting set of issues for future research is to perform these kinds of scaling analyses when the 'resource costs' are thermodynamic costs of running the circuit rather than conventional costs like the number of gates. In particular, it is interesting to consider classes of circuit families defined in terms of such costs, in analogy to the complexity classes considered in computer science theory, like P/poly, or P/log.

Other interesting issues arise if we formulate a cellular automaton (CA) as a circuit with an infinite number of nodes in each layer, and an infinite number of layers, each layer of the circuit corresponding to another timestep of the CA. For example, suppose we are given a particular CA rule (i.e., a particular map taking the state of each layer $i$ to the state of layer $i+1$) and a particular distribution over its initial infinite bit pattern. These uniquely specify the 'thermodynamic EP rate', given by the total EP generated by running the CA for $n$ iterations (i.e., for reaching the $n$th layer in the circuit), divided by $n$. It would be interesting to see how this EP rate depends on the CA rule and initial distribution over bit patterns.

Finally, another important direction for future work arises if we broaden our scope beyond digital circuits designed by human engineers, to include naturally occurring circuits such as brains and gene regulatory networks. The 'gates' in such circuits are quite noisy—but all of our results hold independent of the noise levels of the gates. On the other hand, like real digital circuits, these naturally occurring circuits have loops, branching, concurrency, etc, and so might best be modeled with some extension of the models introduced in this paper. Again though, the important point is that whatever model is used, the EP generated by running a physical system governed by that model would include Landauer loss, mismatch cost, and residual EP.

## Acknowledgments

We would like to thank Josh Grochow for helpful discussion, and thank the Santa Fe Institute for helping to support this research. This paper was made possible through Grant No. CHE-1648973 from the U.S. National Science Foundation, Grant No. FQXi-RFP-1622 from the Foundational Questions Institute, and Grant No. FQXi-RFP-IPW-1912 from the Foundational Questions Institute and Fetzer Franklin Fund, a donor advised fund of Silicon Valley Community Foundation.

## Appendix A. Proof of theorem 1 and related results

### A.1. Preliminaries

Consider a conditional distribution $P(y|x)$ that specifies the probability of 'output' $y \in \mathcal{Y}$ given 'input' $x \in \mathcal{X}$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Given some $\mathcal{Z} \subseteq \mathcal{X}$, the island decomposition $L_{\mathcal{Z}}(P)$ of P, and any $p \in \Delta_{\mathcal{X}}$, let $p(c) = \sum_{x \in c} p(x)$ indicate the total probability within island $c$, and

$$p^c(x) := \begin{cases} \dfrac{p(x)}{p(c)} & \text{if } x \in c \text{ and } p(c) > 0 \\ 0 & \text{otherwise} \end{cases}$$

indicate the conditional probability of state $x$ within island $c$.

In our proofs below, we will make use of the notion of relative interior. Given a linear space $V$, the relative interior of a subset $A \subseteq V$ is defined as [68]

$$\mathrm{relint}\, A := \{x \in A : \forall y \in A, \exists \epsilon > 0\, \text{s.t.}\, x + \epsilon(x - y) \in A\}.$$

Finally, for any $g$, we use notation

$$\partial_x^+ g|_{x=a} := \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \left(g(a+\epsilon) - g(a)\right)$$

to indicate the right-handed derivative of $g(x)$ at $x = a$.

### A.2. Proofs

Given some conditional distribution $P(y|x)$ and function $f : \mathcal{X} \to \mathbb{R}$, we consider the function $\Gamma : \Delta_{\mathcal{X}} \to \mathbb{R}$ as

$$\Gamma(p) := S(Pp) - S(p) + \mathbb{E}_p[f].$$

Note that $\Gamma$ is continuous on the relative interior of $\Delta_{\mathcal{X}}$.

**Lemma A1.** *For any $a, b \in \Delta_{\mathcal{X}}$, the directional derivative of $\Gamma$ at a toward* b *is given by*

$$\partial_\epsilon^+ \Gamma(a + \epsilon(b - a))|_{\epsilon=0} = D(Pb\|Pa) - D(b\|a) + \Gamma(b) - \Gamma(a).$$

**Proof.** Let $a^\epsilon := a + \epsilon(b - a)$. Using the definition of $\Gamma$, write

$$\partial_\epsilon^+ \Gamma(a^\epsilon) = \partial_\epsilon^+ [S(Pa^\epsilon) - S(a^\epsilon)] + \partial_\epsilon^+ \mathbb{E}_{a^\epsilon}[f]. \tag{A1}$$

Then, consider the first term on the right-hand side,

$$\begin{aligned}\partial_\epsilon^+ [S(Pa^\epsilon) - S(a^\epsilon)] &= -\sum_{y\in\mathcal{Y}} \left[\partial_\epsilon^+ a^\epsilon(y) \ln a^\epsilon(y) + \partial_\epsilon^+ [Pa^\epsilon](y)\right] + \sum_{x\in\mathcal{X}} \left[\partial_\epsilon^+ a^\epsilon(x) \ln a^\epsilon(x) + \partial_\epsilon^+ a^\epsilon(x)\right] \\ &= -\sum_{y\in\mathcal{Y}} (b(y) - a(y)) \ln a^\epsilon(y) + \sum_{x\in\mathcal{X}} (b(x) - a(x)) \ln a^\epsilon(x)\end{aligned}$$

Evaluated at $\epsilon = 0$, the last line can be written as

$$\begin{aligned}&-\sum_{y\in\mathcal{Y}} (b(y) - a(y)) \ln a(y) + \sum_{x\in\mathcal{X}} (b(x) - a(x)) \ln a(x) \\ &\quad = D(Pb\|Pa) + S(Pb) - S(Pa) - D(b\|a) - S(b) + S(a)\end{aligned}$$

We next consider the $\partial_\epsilon^+ \mathbb{E}_{a^\epsilon}[f]$ term,

$$\begin{aligned}\partial_\epsilon^+ \mathbb{E}_{a^\epsilon}[f] &= \partial_\epsilon^+ \left[\sum_{x\in\mathcal{X}} (a(x) + \epsilon(b(x) - a(x))) f(x)\right] \\ &= \mathbb{E}_b[f] - \mathbb{E}_a[f].\end{aligned}$$

Combining the above gives

$$\begin{aligned}\partial_\epsilon^+ \Gamma(a^\epsilon)|_{\epsilon=0} &= D(Pb\|Pa) - D(b\|a) + S(Pb) - S(b) - (S(Pa) - S(a)) + \mathbb{E}_b[f] - \mathbb{E}_a[f] \\ &= D(Pb\|Pa) - D(b\|a) + \Gamma(b) - \Gamma(a).\end{aligned}$$

□

**Theorem A1.** *Let* V *be a convex subset of $\Delta$. Then for any* q *∈ arg min*$_{s\in V}$ $\Gamma$(s) *and any* p ∈ V,

$$\Gamma(p) - \Gamma(q) \geqslant D(p\|q) - D(Pp\|Pq). \tag{A2}$$

*Equality holds if* q *is in the relative interior of* V.

**Proof.** Define the convex mixture $q^\epsilon := q + \epsilon(p - q)$. By lemma A1, the directional derivative of $\Gamma$ at $q$ in the direction $p - q$ is

$$\partial_\epsilon^+ \Gamma(q^\epsilon)|_{\epsilon=0} = D(Pp\|Pq) - D(p\|q) + \Gamma(p) - \Gamma(q).$$

At the same time, $\partial_\epsilon^+ \Gamma(q^\epsilon)|_{\epsilon=0} \geqslant 0$, since $q$ is a minimizer within a convex set. equation (A2) then follows by rearranging.

When $q$ is in the relative interior of $V$, $q - \epsilon(p - q) \in V$ for sufficiently small $\epsilon > 0$. Then,

$$\begin{aligned}0 &\leqslant \lim_{\epsilon\to 0^+} \frac{1}{\epsilon} \left(\Gamma(q - \epsilon(p - q)) - \Gamma(q)\right) \\ &= -\lim_{\epsilon\to 0^-} \frac{1}{\epsilon} \left(\Gamma(q + \epsilon(p - q)) - \Gamma(q)\right) \\ &= -\lim_{\epsilon\to 0^+} \frac{1}{\epsilon} \left(\Gamma(q + \epsilon(p - q)) - \Gamma(q)\right) \\ &= -\partial_\epsilon^+ \Gamma(q^\epsilon)|_{\epsilon=0}.\end{aligned}$$

where in the first inequality comes from the fact that $q$ is a minimizer, in the second line we change variables as $\epsilon \mapsto -\epsilon$, and the third line we use the continuity of $\Gamma$ on interior of the simplex. Combining with the above implies

$$\partial_\epsilon^+ \Gamma(q^\epsilon) = D(Pp\|Pq) - D(p\|q) + \Gamma(p) - \Gamma(q) = 0.$$

□

**Lemma A2.** *For any* $c \in L(P)$ *and* $q \in \underset{s:\text{supp } s \subseteq c}{\arg\min}\, \Gamma(s)$,

$$\text{supp}\, q = \{x \in c : f(x) < \infty\}.$$

**Proof.** We prove the claim by contradiction. Assume that $q$ is a minimizer with $\text{supp}\, q \subset \{x \in c : f(x) < \infty\}$. Note there cannot be any $x \in \text{supp}\, q$ and $y \in \mathcal{Y} \backslash \text{supp}\, Pq$ such that $P(y|x) > 0$ (if there were such an $x$, $y$, then $q(y) = \sum_{x'} P(y|x')q(x') \geqslant P(y|x)q(x) > 0$, contradicting the statement that $y \in \mathcal{Y}\backslash\text{supp}\, Pq$). Thus, by definition of islands, there must be an $\hat{x} \in c\backslash\text{supp}\, q$, $\hat{y} \in \text{supp}\, Pq$ such that $f(\hat{x}) < \infty$ and $P(\hat{y}|\hat{x}) > 0$.

Define the delta-function distribution $u(x) := \delta(x, \hat{x})$ and the convex mixture $q^\epsilon(x) = (1-\epsilon)q(x) + \epsilon u(x)$ for $\epsilon \in [0, 1]$. We will also use the notation $q^\epsilon(y) = \sum_x P(y|x)q(x)$.

Since $q$ is a minimizer of $\Gamma$, $\partial_\epsilon \Gamma(q^\epsilon)|_{\epsilon=0} \geqslant 0$. Since $\Gamma$ is convex, the second derivative $\partial_\epsilon^2 \Gamma(q^\epsilon) \geqslant 0$ and therefore $\partial_\epsilon \mathit{\Gamma}(q^\epsilon) \geqslant 0$ for all $\epsilon \geqslant 0$. Taking $a = q^\epsilon$ and $b = u$ in lemma A1 and rearranging, we then have

$$\begin{aligned}\Gamma(u) &\geqslant D(u\|q^\epsilon) - D(Pu\|Pq^\epsilon) + \Gamma(q^\epsilon)\\ &\geqslant D(u\|q^\epsilon) - D(Pu\|Pq^\epsilon) + \Gamma(q),\end{aligned} \tag{A3}$$

where the second inequality uses that $q$ is a minimizer of $\Gamma$. At the same time,

$$\begin{aligned}D(u\|q^\epsilon) - D(Pu\|Pq^\epsilon) &= \sum_y P(y|\hat{x}) \ln \frac{q^\epsilon(y)}{q^\epsilon(\hat{x})P(y|\hat{x})}\\ &= P(\hat{y}|\hat{x}) \ln \frac{q^\epsilon(\hat{y})}{\epsilon P(\hat{y}|\hat{x})} + \sum_{y \neq \hat{j}} P(y\widehat{|i}) \ln \frac{q^\epsilon(y)}{\epsilon P(y|\hat{x})}\\ &\geqslant P(\hat{y}|\hat{x}) \ln \frac{(1-\epsilon)q(\hat{y})}{\epsilon P(\hat{y}|\hat{x})} + \sum_{y \neq \hat{j}} P(y\widehat{|i}) \ln \frac{\epsilon P(y|\hat{x})}{\epsilon P(y|\hat{x})}\\ &= P(\hat{y}|\hat{x}) \ln \frac{(1-\epsilon)}{\epsilon} \frac{q(\hat{y})}{P(\hat{y}|\hat{x})},\end{aligned} \tag{A4}$$

where in the secondline we have used that $q^\epsilon(\hat{x}) = \epsilon$, and in the third that $q^\epsilon(y) = (1-\epsilon)q(y) + \epsilon P(y|\hat{x})$, so $q^\epsilon(y) \geqslant (1-\epsilon)q(y)$ and $q^\epsilon(y) \geqslant \epsilon P(y|\hat{x})$.

Note that the right-hand side of equation (A4) goes to $\infty$ as $\epsilon \to 0$. Combined with equation (A3) and that $\Gamma(q)$ is finite implies that $\Gamma(u) = \infty$. However, $\Gamma(u) = S(P(Y|\hat{x})) + f(\hat{x}) \leqslant |\mathcal{Y}| + f(\hat{x})$, which is finite. We thus have a contradiction, so $q$ cannot be the minimizer. □

**Lemma A3.** *For any island* $c \in L(P)$, $q \in \underset{s:\text{supp } s \subseteq c}{\arg\min}\, \Gamma(p)$ *is unique.*

**Proof.** Consider any two distributions $p, q \in \underset{s:\text{supp } s \subseteq c}{\arg\min}\, \Gamma(s)$, and let $P' = Pp$, $q' = Pq$. We will prove that $p = q$.

First, note that by lemma A2, $\text{supp}\, q = \text{supp}\, p = c$. By theorem A1,

$$\begin{aligned}\Gamma(p) - \Gamma(q) &= D(p\|q) - D(p'\|q')\\ &= \sum_{x,y} p(x)P(y|x) \ln \frac{p(x)q'(y)}{q(x)p'(y)}\\ &= \sum_{x,y} p(x)P(y|x) \ln \frac{p(x)P(y|x)}{q(x)p'(y)P(y|x)/q'(y)}\\ &\geqslant 0\end{aligned}$$

where the last line uses the log-sum inequality. If the inequality is strict, then $p$ and $q$ cannot both be minimizers, i.e., the minimizer must be unique, as claimed.

If instead the inequality is not strict, i.e., $\Gamma(p) - \Gamma(q) = 0$, then there is some constant $\alpha$ such that for all $x$, $y$ with $P(y|x) > 0$,

$$\frac{p(x)P(y|x)}{q(x)p'(y)P(y|x)/q'(y)} = \alpha \tag{A5}$$

which is the same as

$$\frac{p(x)}{q(x)} = \alpha\frac{p'(y)}{q'(y)}. \tag{A6}$$

Now consider any two different states $x, x' \in c$ such that $P(y|x) > 0$ and $P(y|x') > 0$ for some $y$ (such states must exist by the definition of islands). For equation (A6) to hold for both $x$, $x'$ with that same, shared $y$, it must be that $p(x)/q(x) = p(x')/q(x')$. Take another state $x'' \in c$ such that $\mathrm{P}(y'|x'') > 0$ and $P(y'|x') > 0$ for some $y'$. Since this must be true for all pairs $x, x' \in c$, $p(x)/q(x) = \text{const}$ for all $x \in c$, and $p = q$, as claimed. □

**Lemma A4.** $\Gamma(p) = \sum_{c\in L(P)} p(c)\Gamma(p^c)$.

**Proof.** First, for any island $c \in L(P)$, define

$$\phi(c) = \{y \in \mathcal{Y} : \exists x \in c \,\text{s.t.}\, P(y|x) > 0\}.$$

In words, $\phi(c)$ is the subset of output states in $\mathcal{Y}$ that receive probability from input states in $c$. By the definition of the island decomposition, for any $y \in \phi(c)$, $P(y|x) > 0$ only if $y \in c$. Thus, for any $p$ and any $y \in \phi(c)$, we can write

$$\frac{p(y)}{p(c)} = \frac{\sum_x P(y|x)p(x)}{p(c)} = \sum_{x\in\mathcal{X}} P(y|x)p^c(x). \tag{A7}$$

Using $p = \sum_{c\in L(P)} p(c)p^c$ and linearity of expectation, write $\mathbb{E}_p[f] = \sum_{c\in L(P)} p(c)\mathbb{E}_{p^c}[f]$. Then,

$$\begin{aligned} S(Pp) - S(p) &= -\sum_y p(y)\ln\, p(y) + \sum_x p(x)\ln\, p(x) \\ &= \sum_{c\in L(P)} p(c)\left[-\sum_{y\in\phi(c)} \frac{p(y)}{p(c)}\ln\frac{p(y)}{p(c)} + \sum_{x\in c}\frac{p(x)}{p(c)}\ln\frac{p(x)}{p(c)}\right] \\ &= \sum_{c\in L(P)} p(c)\left[S(Pp^c) - S(p^c)\right], \end{aligned}$$

where in the last line we have used equation (A7). Combining gives

$$\begin{aligned} \Gamma(p) &= \sum_{c\in L(P)} p(c)\left[S(Pp^c) - S(p^c) + \mathbb{E}_{p^c}[f]\right] \\ &= \sum_{c\in L(P)} p(c)\Gamma(p^c). \end{aligned}$$

□

We are now ready to prove the main result of this appendix.

**Theorem 1** *Consider any function $\Gamma : \Delta_{\mathcal{X}} \to \mathbb{R}$ of the form*

$$\Gamma(p) := S(Pp) - S(p) + \mathbb{E}_p[f]$$

*where $P(y|x)$ is some conditional distribution of $y \in \mathcal{Y}$ given $x \in \mathcal{X}$ and $f : \mathcal{X} \to \mathbb{R}\cup\{\infty\}$ is some function. Let $\mathcal{Z}$ be any subset of $\mathcal{X}$ such that $f(x) < \infty$ for $x \in \mathcal{Z}$, and let $q \in \Delta_{\mathcal{Z}}$ be any distribution that obeys*

$$q^c \in \underset{r:\text{supp}\, r\subseteq c}{\textit{argmin}}\, \Gamma(r) \qquad \textit{for all } c \in L_{\mathcal{Z}}(P).$$

*Then, each $q^c$ will be unique, and for any $p$ with $\text{supp} p \subseteq \mathcal{Z}$,*

$$\Gamma(p) = D(p\|q) - D(Pp\|Pq) + \sum_{c\in L_{\mathcal{Z}}(P)} p(c)\Gamma(q^c).$$

**Proof.** We prove the theorem by considering two cases separately.

**Case 1:** $\mathcal{Z} = \mathcal{X}$. This case can be assumed when $f(x) < \infty$ for all $x$, so that $L_{\mathcal{Z}}(P) = L(P)$. Then, by lemma A4, we have $\Gamma(p) = \sum_{c \in L(P)} p(c)\Gamma(p^c)$. By lemma A2 and theorem A1,

$$\Gamma(p^c) - \Gamma(q^c) = D(p^c \| q^c) - D(Pp^c \| Pq^c),$$

where we have used that if some supp $q^c = c$, then $q^c$ is in the relative interior of the set $\{s \in \Delta_{\mathcal{X}} : \operatorname{supp} s \subseteq c\}$. $q^c$ is unique by lemma A3.

At the same time, observe that for any $p, r \in \Delta_{\mathcal{X}}$,

$$\begin{aligned} D(p\|r) - D(Pp\|Pr) &= \sum_x p(x) \ln \frac{p(x)}{r(x)} - \sum_y p(y) \ln \frac{p(y)}{r(y)} \\ &= \sum_{c \in L(P)} p(c) \left[ \sum_{x \in c} \frac{p(x)}{p(c)} \ln \frac{p(x)/p(c)}{r(x)/r(c)} - \sum_{y \in \phi(c)} \frac{p(y)}{p(c)} \ln \frac{p(y)/p(c)}{r(y)/r(c)} \right] \\ &= \sum_{c \in L(P)} p(c) \left[ D(p^c \| r^c) - D(Pp^c \| Pr^c) \right]. \end{aligned}$$

The theorem follows by combining.

**Case 2:** $\mathcal{Z} \subset \mathcal{X}$. In this case, define a 'restriction' of $f$ and P to domain $\mathcal{Z}$ as follows:

(a) Define $\tilde{f} : \mathcal{Z} \to \mathbb{R}$ via $\tilde{f}(x) = f(x)$ for $x \in \mathcal{Z}$.

(b) Define the conditional distribution $\tilde{P}(y|x)$ for $y \in \mathcal{Y}, x \in \mathcal{Z}$ via $\tilde{P}(y|x) = P(y|x)$ for all $y \in \mathcal{Y}, x \in \mathcal{Z}$.

In addition, for any distribution $p \in \Delta_{\mathcal{X}}$ with supp $p \subseteq \mathcal{Z}$, let $\tilde{p}$ be a distribution over $\mathcal{Z}$ defined via $\tilde{p}(x) = p(x)$ for $x \in \mathcal{Z}$. Now, by inspection, it can be verified that for any $p \in \Delta_{\mathcal{X}}$ with supp $p \subseteq \mathcal{Z}$,

$$\Gamma(p) = S(\tilde{P}\tilde{p}) - S(\tilde{p}) + \mathbb{E}_{\tilde{p}}[\tilde{f}] := \tilde{\Gamma}(\tilde{p}) \tag{A8}$$

We can now apply case 1 of the theorem to the function $\tilde{\Gamma} : \Delta_{\mathcal{Z}} \to \mathbb{R}$, as defined in terms of the tuple $(\mathcal{Z}, \tilde{f}, \tilde{P})$ (rather than the function $\Gamma : \Delta_{\mathcal{X}} \to \mathbb{R}$, as defined in terms of the tuple $(\mathcal{X}, f, P)$). This gives

$$\tilde{\Gamma}(\tilde{p}) = D(\tilde{p}\|\tilde{q}) - D(\tilde{P}\tilde{p}\|\tilde{P}\tilde{q}) + \sum_{c \in L(\tilde{P})} \tilde{p}(c)\tilde{\Gamma}(\tilde{q}^c), \tag{A9}$$

where, for all $c \in L(\tilde{P})$, $\tilde{q}^c$ is the unique distribution that satisfies $\tilde{q}^c \in \arg\min_{r \in \Delta_{\mathcal{Z}} : \operatorname{supp} r \subseteq c} \tilde{\Gamma}(r)$.

Now, let $q$ be the natural extension of $\tilde{q}$ from $\Delta_{\mathcal{Z}}$ to $\Delta_{\mathcal{X}}$. Clearly, for all $c \in L(\tilde{P})$, $\Gamma(q^c) = \tilde{\Gamma}(\tilde{q}^c)$ by equation (A8). In addition, each $q^c$ is the unique distribution that satisfies $q^c \in \arg\min_{r \in \Delta_{\mathcal{X}} : \operatorname{supp} r \subseteq c} \Gamma(r)$. Finally, it is easy to verify that $D(\tilde{p}\|\tilde{q}) = D(p\|q)$, $D(\tilde{P}\tilde{p}\|\tilde{P}\tilde{q}) = D(Pp\|Pq)$, $L(\tilde{P}) = L_{\mathcal{Z}}(P)$ (recall the definition of $L_{\mathcal{Z}}$ from section 2.4). Combining the above results with equation (A8) gives

$$\Gamma(p) = \tilde{\Gamma}(\tilde{p}) = D(p\|q) - D(Pp\|Pq) + \sum_{c \in L_{\mathcal{Z}}(P)} p(c)\Gamma(q^c).$$

□

**Example 8.** Suppose we are interested in thermodynamic costs associated with functions $f$ whose image contains the value infinity, i.e., $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$. For such functions, $\Gamma(p) = \infty$ for any $p$ which has support over an $x \in \mathcal{X}$ such that $f(x) = \infty$. In such a case it is not meaningful to consider a prior distribution $q$ (as in theorem 1) which has support over any $x$ with $f(x) = \infty$. For such functions we also are no longer able to presume that the optimal distribution has full support within each island of $c \in L(P)$, because in general the proof of lemma A2 no longer holds when $f$ can take infinite values.

Nonetheless, by equation (A9), for the purposes of analyzing the thermodynamic costs of actual initial distributions $p$ that have finite $\Gamma(p)$ (and so have zero mass on any $x$ such that $f(x) = \infty$), we can always carry out our usual analysis if we first reduce the problem to an appropriate 'restriction' of $f$.

**Example 9.** Suppose we wish to implement a (discrete-time) dynamics $P(x\prime|x)$ over $\mathcal{X}$ using a CTMC. Recall from the end of section 2.2 that by appropriately expanding the state space $\mathcal{X}$ to include a set of 'hidden states' $\mathcal{Z}$ in addition to $\mathcal{X}$, and appropriately designing the rate matrices over that expanded state space $\mathcal{X} \cup \mathcal{Z}$, we can ensure that the resultant evolution over $\mathcal{X}$ is arbitrarily close to the desired conditional distribution P. Indeed, one can even design those rates matrices over $\mathcal{X} \cup \mathcal{Z}$ so that not only is the dynamics over $\mathcal{X}$ arbitrarily close to the desired $P$, but in addition the EF generated in running that CTMC over $\mathcal{X} \cup \mathcal{Z}$ is arbitrarily close to the lower bound of equation (21) [21].

However, in any real-world system that implements some $P$ with a CTMC over an expanded space $\mathcal{X} \cup \mathcal{Z}$, that lower bound will not be achieved, and nonzero EP will be generated. In general, to analyze the EP of such real-world systems one has to consider the mismatch cost and residual EP of the full CTMC over the expanded space $\mathcal{X} \cup \mathcal{Z}$. Fortunately though, we can design the CTMC over $\mathcal{X} \cup \mathcal{Z}$ so that when it begins the implementation of $P$, there is zero probability mass on any of the states in $\mathcal{Z}$ [21, 22]. If we do that, then we can apply equation (A9), and so restrict our calculations of mismatch cost and residual EP to only involve the dynamics over $\mathcal{X}$, without any concern for the dynamics over $\mathcal{Z}$.

**Example 10.** Our last example is to derive the alternative decomposition of the EP of an SR circuit which is discussed in section 6.2. Recall that due to equation (29), the initial distribution over any gate in an SR circuit has partial support. This means we can apply equation (29) to decompose the EF, in direct analogy to the use of theorem 1 to derive theorem 2—only with the modification that the spaces $X$ and $Y$ are set to $\mathcal{X}_{\mathrm{pa}(g)}$ and $\mathcal{X}_g$, respectively, rather than both set to $\mathcal{X}_{n(g)}$, as was done in deriving theorem 2. (Note that the islands also change when we apply equation (29) rather than theorem 1, from the islands of $P_g$ to the islands of $\pi_g$). The end result is a decomposition of EF just like that in theorem 2, in which we have the same circuit Landauer cost and circuit Landauer loss expressions as in that theorem, but now have the modified forms of circuit mismatch cost and of circuit residual EP introduced in section 6.2.

## Appendix B. Thermodynamics costs for SR circuits

To begin, we will make use of the fact that there is no overlap in time among the solitary processes in an SR circuit, so the total EF incurred can be written as

$$\mathcal{Q}(p) = \sum_{g\in G} \mathcal{Q}_g(p_{n(g)}). \tag{B1}$$

Moreover, for each gate $g$, the solitary process that updates the variables in $n(g)$ starts with $x_g$ in its initialized state with probability 1. So we can overload notation and write $\mathcal{Q}_g(p_{\mathrm{pa}(g)})$ instead of $\mathcal{Q}_g(p_{n(g)})$ for each gate $g$.

### B.1. Derivation of theorem 2 and equation (49)

Apply theorem 1 to equation (18) to give

$$\hat{\sigma}_g(p_{n(g)}) = D\left(p_{n(g)} \| q_{n(g)}\right) - D\left(P_g p_{n(g)} \| P_g q_{n(g)}\right) + \sum_{c\in L(P_g)} p_{n(g)}(c)\hat{\sigma}_g(q^c_{n(g)}), \tag{B2}$$

where $q_n(g)$ is a distribution that satisfies

$$q^c_{n(g)} \in \mathop{\arg\min}_{r:\mathrm{supp}\, r\subseteq c} \hat{\sigma}_g(r)$$

for all islands $c \in L(P_g)$. Next, for convenience use equation (B1) to write $\mathcal{Q}(p)$ as

$$\mathcal{Q}(p) = \mathcal{L}(p) + \left(\left[\sum_{g\in G} \mathcal{Q}_g(p_{\mathrm{pa}(g)})\right] - \mathcal{L}(p)\right) \tag{B3}$$

The basic decomposition of $\mathcal{Q}(p)$ given in theorem 2 into a sum of $\mathcal{L}$ (defined in equation (37)), $\mathcal{L}^{\mathrm{loss}}$ (defined in equation (47)), $\mathcal{M}$ (defined in equation (50)), and $\mathcal{R}$ (defined in equation (51)) comes from combining equations (43), (B2) and (B3) and then grouping and redefining terms.

Next, again use the fact that the solitary processes have no overlap in time to establish that the minimal value of the sum of the EPs of the gates is the sum of the minimal EPs of the gates considered separately of one another. As a result, we can jointly take $\hat{\sigma}_g(p_{n(g)}) \to 0$ for all gates $g$ in the circuit [21]. We can then use equation (B1) to establish that the minimal EF of the circuit is simply the sum of the minimal EFs of running each of the gates in the circuit, i.e., the sum of the subsystem Landauer costs of running the gates. In other words, the circuit Landauer cost is

$$\mathcal{L}^{\mathrm{circ}}(p) = \sum_{g\in G} \left[S(p_{n(g)}) - S(P_g p_{n(g)})\right] \tag{B4}$$

$$= \sum_{g\in G} \left[S(p_{\mathrm{pa}(g)}) - S(\pi_g p_{\mathrm{pa}(g)})\right] \tag{B5}$$

$$= \sum_{g\in G\setminus W} \left[S(p_{\mathrm{pa}(g)}) - S(\pi_g p_{\mathrm{pa}(g)})\right] \tag{B6}$$

To derive the second line, we have used the fact that in an SR circuit, each gate is set to its initialized value at the beginning of its solitary process with probability 1, and that its parents are set to their initialized states with probability 1 at the end of the process. Then to derive the third line we have used the fact that wire gates implement the identity map, and so $S(p_{\mathrm{pa}}(g)) - S(\pi_g p_{\mathrm{pa}}(g)) = 0$ for all $g \in W$.

This establishes equation (49).

### B.2. Derivation of equation (55)

To derive equation (55), we first write Landauer cost as

$$\mathcal{L}(p) = S(p) - S(Pp) = \sum_{g\in G} [S(p_V^{\mathrm{beg}(g)}) - S(p_V^{\mathrm{end}(g)})].$$

We then write circuit Landauer loss as

$$\begin{aligned}\mathcal{L}^{\mathrm{loss}}(p) &= \sum_{g\in G} \left[S(p_{n(g)}) - S(P_g p_{n(g)})\right] - \mathcal{L}(p)\\ &= \sum_{g\in G} \left[(S(p_{n(g)}) - S(p_V^{\mathrm{beg}(g)})) - (S(P_g p_{n(g)}) - S(p_V^{\mathrm{end}(g)}))\right]\\ &= \sum_{g\in G} \left[(S(p_{n(g)}) + S(p_{V\setminus n(g)}^{\mathrm{beg}(g)}) - S(p_V^{\mathrm{beg}(g)})) - (S(P_g p_{n(g)}) + S(p_{V\setminus n(g)}^{\mathrm{end}(g)}) - S(p_V^{\mathrm{end}(g)}))\right] && \text{(B7)}\\ &= \sum_{g\in G} \left[I_{p^{\mathrm{beg}(g)}}(X_{n(g)}; X_{V\setminus n(g)}) - I_{p^{\mathrm{end}(g)}}(X_{n(g)}; X_{V\setminus n(g)})\right], && \text{(B8)}\end{aligned}$$

In equation (B7), we used the fact that a solitary process over $n(g)$ leaves the nodes in $V\setminus n(g)$ unmodified, thus $S(p_{V\setminus n(g)}^{\mathrm{beg}(g)}) = S(p_{V\setminus n(g)}^{\mathrm{end}(g)})$.

Given the assumption that $x_g = \emptyset$ at the beginning of the solitary process for gate $g$, we can rewrite

$$I_{p^{\mathrm{beg}(g)}}(X_{n(g)}; X_{V\setminus n(g)}) = I_{p^{\mathrm{beg}(g)}}(X_{\mathrm{pa}(g)}; X_{V\setminus \mathrm{pa}(g)}). \tag{B9}$$

Similarly, because $X_v = \emptyset$ for all $v \in \mathrm{pa}(g)$ at the end of the solitary process for gate $g$, we can rewrite

$$I_{p^{\mathrm{end}(g)}}(X_{n(g)}; X_{V\setminus n(g)}) = I_{p^{\mathrm{end}(g)}}(X_g; X_{V\setminus g}). \tag{B10}$$

Finally, for any wire gate $g \in W$, given the assumption that $X_g = \emptyset$ at the beginning of the solitary process, we can write

$$I_{p^{\mathrm{beg}(g)}}(X_{\mathrm{pa}(g)}; X_{V\setminus \mathrm{pa}(g)}) = I_{p^{\mathrm{end}(g)}}(X_g; X_{V\setminus g}). \tag{B11}$$

Equation (55) then follows from combining equations (B8)–(B11) and simplifying.

### B.3. Derivation of corollary 4 and equation (57)

First, write circuit Landauer loss as

$$\mathcal{L}^{\mathrm{loss}}(p) = \sum_{g\in G\setminus W} \left[S(p_{\mathrm{pa}(g)}) - S(p_g)\right] - \mathcal{L}(p). \tag{B12}$$

Then, rewrite the sum in equation (B12) as

$$\begin{aligned}\sum_{g\in G\setminus W} \left[S(p_{\mathrm{pa}(g)}) - S(p_g)\right] &= \sum_{g\in G\setminus W} S(p_{\mathrm{pa}(g)}) - \sum_{g\in G\setminus W} S(p_g)\\ &= \sum_{g\in G\setminus W} S(p_{\mathrm{pa}(g)}) - \sum_{g\in G\setminus (W\cup \mathrm{OUT})} S(p_g) - \sum_{g\in \mathrm{OUT}} S(p_g)\\ &= \sum_{g\in G\setminus W} S(p_{\mathrm{pa}(g)}) - \sum_{v\in V\setminus (W\cup \mathrm{OUT})} S(p_v) + \sum_{v\in \mathrm{IN}} S(p_v) - \sum_{g\in \mathrm{OUT}} S(p_g). && \text{(B13)}\end{aligned}$$

Now, notice that for every $v \in V\setminus(W \cup \mathrm{OUT})$ (i.e., every node which is not a wire and not an output), there is a corresponding wire $w$ which transmits $v$ to its child, and which has $S(p_w) = S(p_v)$. This lets us rewrite equation (B13) as

$$\begin{aligned}
&\sum_{g\in G\setminus W} S(p_{\mathrm{pa}(g)}) - \sum_{w\in W} S(p_w) + \sum_{v\in \mathrm{IN}} S(p_v) - \sum_{g\in \mathrm{OUT}} S(p_g)\\
&= \sum_{g\in G\setminus W}\left[S(p_{\mathrm{pa}(g)}) - \sum_{v\in \mathrm{pa}(g)} S(p_v)\right] + \sum_{v\in \mathrm{IN}} S(p_v) - \sum_{g\in \mathrm{OUT}} S(p_g)\\
&= -\sum_{g\in G}\mathcal{I}(p_{\mathrm{pa}(g)}) + \sum_{v\in \mathrm{IN}} S(p_v) - \sum_{g\in \mathrm{OUT}} S(p_g)
\end{aligned} \tag{B14}$$

where in the second line we have used the fact that every wire belongs to exactly one set pa($g$) for a non-wire gate $g$, and in the last line we used the definition of multi-information. Then, using the definition $\mathcal{L}(p) = S(p_{\mathrm{IN}}) - S(p_{\mathrm{OUT}})$, the definition of multi-information, and by combining equations (B12)–(B14), we have

$$\mathcal{L}^{\mathrm{loss}}(p) = \mathcal{I}(p_{\mathrm{IN}}) - \mathcal{I}(\pi_\Phi p_{\mathrm{IN}}) - \sum_{g\in G}\mathcal{I}(p_{\mathrm{pa}(g)}). \tag{B15}$$

To derive equation (57), note that

$$\mathcal{I}(p_{\mathrm{IN}}) = \left[\sum_{v\in \mathrm{IN}} S(p_v)\right] - S(p_{\mathrm{IN}}) \leqslant \sum_{v\in \mathrm{IN}} S(p_v) \leqslant \sum_{v\in \mathrm{IN}} \ln|\mathcal{X}_v| = \ln\left|\prod_{v\in \mathrm{IN}}\mathcal{X}_v\right| = \ln|\mathcal{X}_{\mathrm{IN}}|. \tag{B16}$$

## Appendix C. SR circuits with out-degree greater than 1

In this appendix, we consider a more general version of SR circuits, in which non-output gates can have out-degree greater than 1.

First, we need to modify the definition of an SR circuit in section 5. This is because in SR circuits, the subsystem corresponding to a given gate $g$ reinitializes all of the parents of that gate to their initialized state, $\emptyset$. If, however, there is some node $v$ that has out-degree greater than 1—i.e., has more than one child—then we must guarantee that no such $v$ is reinitialized by one its children gates before all of its children gates have run. To do so, we require that each non-output node $v$ in the circuit is reinitialized only by the last of its children gates to run, while the earlier children (if any) apply the identity map to $v$.

Note that this rule could result in different thermodynamic costs of an overall circuit, depending on the precise topological order we use to determine which of the children of a given $v$ reinitialized $v$. This would mean that the entropic costs of running a circuit would depend on the (arbitrary) choice we make for the topological order of the gates in the circuit. This issue will not arise in this paper however. To see why, recall that we model the wires in the circuit themselves as gates, which have both in-degree and out-degree equal to 1. As a result, if $v$ has out-degree greater than 1, then $v$ is not a wire gate, and therefore all of its children must be wire gates—and therefore none of those children has multiple parents. So the problem is automatically avoided.

We now prove that for SR circuits with out-degree greater than 1, circuit Landauer loss can be arbitrarily large.

**Proposition 1** *For any $\pi_\Phi$, non-delta function input distribution $p_{\mathrm{IN}}$, and $\kappa \geqslant 0$, there exist an SR circuit with out-degree greater than 1 that implements $\pi_\Phi$ for which $\mathcal{L}^{\mathrm{loss}}(p) \geqslant \kappa$.*

**Proof.** Let $\Phi = (V, E, F, \mathcal{X})$ be such a circuit that implements $\pi_\Phi$. Given that $p$ is not a delta function, there must be an input node, which we call $v$, such that $S(p_v) > 0$. Take $g \in \mathrm{OUT}$ to be any output gate of $\Phi$, and let $\pi_g \in F$ be its update map.

Construct a new circuit $\Phi' = (V', E', F', \mathcal{X}')$, as follows:

(a) $V' = V \cup \{w', g', w''\}$;

(b) $E' = E \cup \{(v, w'), (w', g'), (g', w''), (w'', g)\}$;

(c) $F' = (F \setminus \pi_g) \cup \{\pi_{w'}, \pi_{w''}, \pi_{g'}, \pi'_g\}$ where

$$\begin{aligned}
\pi_{w'}(x_{w'}|x_v) &= \delta(x_{w'}, x_v)\\
\pi_{w''}(x_{w''}|x_{g'}) &= \delta(x_{w''}, x_{g'})\\
\pi_{g'}(x_{g'}|x_w) &= \delta(x_{g'}, \emptyset)\\
\pi'_g(x_g|x_{\mathrm{pa}(g)}, x_{w''}) &= \pi_g(x_g|x_{\mathrm{pa}(g)}).
\end{aligned}$$

(d) $\mathcal{X}_{w'} = \mathcal{X}_{g'} = \mathcal{X}_{w''} = \mathcal{X}_v$.

In words, $\Phi'$ is the same as $\Phi$ except that: (a) we have added an 'erasure gate' $g'$ which takes $v$ as input (through a new wire gate $w'$), and (b) this erasure gate is provided as an additional input, which is completely ignored, to one of the existing output gates $g$ (through a new wire gate $w''$).

It is straightforward to see that $\pi_{\Phi}^{\prime} = \pi_{\Phi}$. At the same time, $S(p_{\mathrm{pa}(g')}) - S(\Pi_{g'} p_{\mathrm{pa}(g')}) = S(p_v)$, thus

$$\mathcal{L}^{\mathrm{loss}}_{\Phi'}(p) = \mathcal{L}^{\mathrm{loss}}_{\Phi}(p) + S(p_v), \tag{C1}$$

where $\mathcal{L}^{\mathrm{loss}}_{\Phi'}$ and $\mathcal{L}^{\mathrm{loss}}_{\Phi}$ indicate the circuit Landauer loss of $\Phi'$ and $\Phi$ respectively. This procedure can be carried out again to create a new circuit $\Phi''$ from $\Phi'$, which also implements $\pi_{\Phi}$ but which now has Landauer loss $\mathcal{L}^{\mathrm{loss}}_{\Phi''}(p) = \mathcal{L}^{\mathrm{loss}}_{\Phi}(p) + 2S(p_v)$. Iterating, we can construct a circuit with an arbitrarily large Landauer loss which implements $\pi_{\Phi}$. □

## ORCID iDs

David H Wolpert https://orcid.org/0000-0003-3105-2869
Artemy Kolchinsky https://orcid.org/0000-0002-3518-9208

## References

[1] Bennett C H 1982 The thermodynamics of computation—a review *Int. J. Theor. Phys.* **21** 905–40

[2] Parrondo J M, Horowitz J M and Sagawa T 2015 Thermodynamics of information *Nat. Phys.* **11** 131–9

[3] Wolpert D 20419 The stochastic thermodynamics of computation *J. Phys. A: Math. Theor.* **52** 193001

[4] Arora S and Barak B 2009 *Computational Complexity: A Modern Approach* (Cambridge: Cambridge University Press)

[5] Savage J E 1998 *Models of Computation* (Reading, MA: Addison-Wesley) vol 136

[6] Fredkin E and Toffoli T 1982 Conservative logic *Int. J. Theor. Phys.* **21** 219–53

[7] Lloyd S 1989 Use of mutual information to decrease entropy: implications for the second law of thermodynamics *Phys. Rev.* A **39** 5378

[8] Caves C M 1993 Information and entropy *Phys. Rev.* E **47** 4010

[9] Lloyd S 2000 Ultimate physical limits to computation *Nature* **406** 1047–54

[10] Like this early work, in this paper we focus on computation by classical systems; see [78, 79] and associated articles for recent work on the resources required to perform quantum computations.

[11] Jarzynski C 1997 Nonequilibrium equality for free energy differences *Phys. Rev. Lett.* **78** 2690

[12] Crooks G E 1998 Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems *J. Stat. Phys.* **90** 1481–7

[13] Hasegawa H-H, Ishikawa J, Takara K and Driebe D 2010 Generalization of the second law for a nonequilibrium initial state *Phys. Lett.* A **374** 1001–4

[14] Takara K, Hasegawa H-H and Driebe D Dec. 2010 Generalization of the second law for a transition between nonequilibrium states *Phys. Lett.* A **375** 88–92

[15] Seifert U 2012 Stochastic thermodynamics, fluctuation theorems and molecular machines *Rep. Prog. Phys.* **75** 126001

[16] Sagawa T 2014 Thermodynamic and logical reversibilities revisited *J. Stat. Mech.* P03025

[17] Maroney O 2009 Generalizing landauer's principle *Phys. Rev.* E **79** 031105

[18] Turgut S 2009 Relations between entropies produced in nondeterministic thermodynamic processes *Phys. Rev.* E **79** 041102

[19] Faist P, Dupuis F, Oppenheim J and Renner R 2015 The minimal work cost of information processing *Nat. Commun.* **6**

[20] Wolpert D H 2015 Extending Landauer's bound from bit erasure to arbitrary computation (arXiv:1508.05319 [cond-mat.stat-mech])

[21] Owen J A, Kolchinsky A and Wolpert D H 2018 Number of hidden states needed to physically implement a given conditional distribution *New J. Phys.* **21** 013022

[22] Wolpert D H, Kolchinsky A and Owen J A 2019 A space/time tradeoff for implementing a function with master equation dynamics *Nat. Commun.* **10** 1727

[23] Wang R-S, Saadatpour A and Albert R 2012 Boolean modeling in systems biology: an overview of methodology and applications *Phys. Biol.* **9** 055001

[24] Yokobayashi Y, Weiss R and Arnold F H 2002 Directed evolution of a genetic circuit *Proc. Natl Acad. Sci.* **99** 16587–91

[25] Brophy J A and Voigt C A 2014 Principles of genetic circuit design *Nat. Methods* **11** 508

[26] Qian L and Winfree E 2011 Scaling up digital circuit computation with dna strand displacement cascades *Science* **332** 1196–201

[27] Clune J, Mouret J-B and Lipson H 2013 The evolutionary origins of modularity *Proc. R. Soc.* B **280** 20122863

[28] Melo D, Porto A, Cheverud J M and Marroig G 2016 Modularity: genes, development, and evolution *Annu. Rev. Ecol. Evol. Syst.* **47** 463–86

[29] Deem M W 2013 Statistical mechanics of modularity and horizontal gene transfer *Annu. Rev. Condens. Matter Phys.* **4** 287–311

[30] Kolchinsky A and Wolpert D H 2017 Dependence of dissipation on the initial distribution over states *J. Stat. Mech.* 083202

[31] Wegener I 1991 *The Complexity of Boolean Functions* (New York: Wiley)

[32] Wolpert D H 2016 The free energy requirements of biological organisms; implications for evolution *Entropy* **18** 138

[33] Van den Broeck C and Esposito M 2015 Ensemble and trajectory thermodynamics: a brief introduction *Physica* A **418** 6–16

[34] Esposito M and Van den Broeck C 2010 Three faces of the second law. I. Master equation formulation *Phys. Rev.* E **82** 011143

[35] Esposito M and Van den Broeck C 2011 Second law and Landauer principle far from equilibrium *Europhys. Lett.* **95** 40004

[36] Lencastre P, Raischel F, Rogers T and Lind P G 2016 From empirical data to time-inhomogeneous continuous Markov processes *Phys. Rev.* E **93** 032135

[37] Kingman J F C Jan. 1962 The imbedding problem for finite Markov chains *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* **1** 14–24

[38] Watanabe S 1960 Information theoretical analysis of multivariate correlation *IBM J. Res. Dev.* **4** 66–82

[39] Koller D and Friedman N 2009 *Probabilistic Graphical Models* (Cambridge, MA: MIT Press)

[40] Ito S and Sagawa T 2013 Information thermodynamics on causal networks *Phys. Rev. Lett.* **111** 180603

[41] Ito S and Sagawa T 2016 *Information Flow and Entropy Production on Bayesian Networks* (Mathematical Foundations and Applications of Graph Entropy) (New York: Wiley) vol 3 2

[42] Csiszar I and Körner J 2011 *Information Theory* (Coding Theorems for Discrete Memoryless Systems) (Cambridge: Cambridge University Press)

[43] Esposito M, Kawai R, Lindenberg K and Van den Broeck C 2010 Finite-time thermodynamics for a single-level quantum dot *Europhys. Lett.* **89** 20003

[44] Schlögl F 1971 On stability of steady states *Z. für Physik A Hadrons Nucl.* **243** 303–10

[45] Schnakenberg J 1976 Network theory of microscopic and macroscopic behavior of master equation systems *Rev. Mod. Phys.* **48** 571

[46] Cover T M and Thomas J A 2012 *Elements of Information Theory* (New York: Wiley)

[47] Gershenfeld N 1996 Signal entropy and the thermodynamics of computation *IBM Syst. J.* **35** 577–86

[48] Koller D and Friedman N 2009 *Probabilistic Graphical Models: Principles and Techniques* (Cambridge, MA: MIT press)

[49] Neapolitan R E *et al* 2004 *Learning Bayesian Networks* (Saddle River, NJ: Pearson) vol 38

[50] Rao R and Esposito M 2019 *Detailed Fluctuation Theorems: A Unifying Perspective* (Santa Fe, NM: Santa Fe Institute Press) 2018 Detailed fluctuation theorems: a unifying perspective *The Energetics of Computing in Life and Machines* (Santa Fe, NM: Santa Fe Institute Press)

[51] Ito S 2016 *Inform ation Thermodynamics on Causal Networks and its Application to Biochemical Signal Transduction* (Berlin: Springer)

[52] Otsubo S and Sagawa T 2018 Information-thermodynamic characterization of stochastic boolean networks (arXiv:1803.04217)

[53] Sagawa T and Ueda M 2008 Second law of thermodynamics with discrete quantum feedback control *Phys. Rev. Lett.* **100** 080403

[54] Sagawa T and Ueda M 2012 Fluctuation theorem with information exchange: role of correlations in stochastic thermodynamics *Phys. Rev. Lett.* **109** 180602

[55] Boyd A B, Mandal D and Crutchfield J P 2018 Thermodynamics of modularity: structural costs beyond the landauer bound *Phys. Rev.* X **8** 031036

[56] Mandal D and Jarzynski C 2012 Work and information processing in a solvable model of Maxwell's demon *Proc. Natl Acad. Sci.* **109** 11641–5

[57] Riechers P 2018 Transforming metastable memories: The nonequilibrium thermodynamics of computation *Energetics of Computing in Life and Machines* ed D H Wolpert, C P Kempes, P Stadler and J Grochow (Santa Fe, NM: Santa Fe Institute Press)

[58] Wolpert D H and Kolchinsky A 2018 Exact, complete expressions for the thermodynamic costs of circuits (arXiv:1806.04103v2)

[59] Drechsler R and Wille R 2012 Reversible circuits: recent accomplishments and future challenges for an emerging technology *Progress in VLSI Design and Test* (Berlin: Springer) pp 383–92

[60] Perumalla K S 2013 *Introduction to Reversible Computing* (Boca Raton, FL: CRC Press)

[61] Frank M P 2005 Introduction to reversible computing: motivation, progress, and challenges *Proc. of the 2nd Conf. on Computing Frontiers* (ACM) pp 385–90

[62] Landauer R 1961 Irreversibility and heat generation in the computing process *IBM J. Res. Dev.* **5** 183–91

[63] Grochow J and Wolpert D 2018 Beyond number of bit erasures: new complexity questions raised by recently discovered thermodynamic costs of computation *SIGACT News* **49** 54

[64] Gour G, Müller M P, Narasimhachar V, Spekkens R W and Halpern N Y 2015 The resource theory of informational nonequilibrium in thermodynamics *Phys. Rep.* **583** 1–58

[65] Brandao F G, Horodecki M, Oppenheim J, Renes J M and Spekkens R W 2013 Resource theory of quantum states out of thermal equilibrium *Phys. Rev. Lett.* **111** 250404

[66] Boyd A B, Mandal D and Crutchfield J P 2016 Identifying functional thermodynamics in autonomous Maxwellian ratchets *New J. Phys.* **18** 023049

[67] Strasberg P, Schaller G, Brandes T and Esposito M 2017 Quantum and information thermodynamics: a unifying framework based on repeated interactions *Phys. Rev.* X **7** 021003

[68] Borwein J and Goebel R 2003 Notions of relative interior in banach spaces *J. Math. Sci.* **115** 2542–53