PHYSICAL REVIEW E **79**, 031105 (2009)

Generalizing Landauer's principle

O. J. E. Maroney*

The Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario, Canada N2L 2Y5
and The Centre for Time, The University of Sydney, Sydney, NSW, Australia 2050

(Received 17 August 2007; revised manuscript received 11 December 2008; published 10 March 2009)

In a recent paper [Stud. Hist. Philos. Mod. Phys. **36**, 355 (2005)] it is argued that to properly understand the thermodynamics of Landauer's principle it is necessary to extend the concept of logical operations to include indeterministic operations. Here we examine the thermodynamics of such operations in more detail, extending the work of Landauer to include indeterministic operations and to include logical states with variable entropies, temperatures, and mean energies. We derive the most general statement of Landauer's principle and prove its universality, extending considerably the validity of previous proofs. This confirms conjectures made that all logical operations may, in principle, be performed in a thermodynamically reversible fashion, although logically irreversible operations would require special, practically rather difficult, conditions to do so. We demonstrate a physical process that can perform any computation without work requirements or heat exchange with the environment. Many widespread statements of Landauer's principle are shown to be special cases of our generalized principle.

DOI: 10.1103/PhysRevE.79.031105 PACS number(s): 05.70.−a, 05.30.−d, 89.70.−a

I. INTRODUCTION

Landauer's principle holds a special place in the thermodynamics of computation. It has been described as "the basic principle of the thermodynamics of information processing" [1]. Yet the literature on Landauer's principle is focused almost exclusively on a single, logically irreversible operation and a particular physical procedure by which this operation is performed.[1]

In this paper we seek to analyze the form of Landauer's principle in a more general context, building upon the consideration of the thermodynamics of indeterministic logical operations [2,3]. We will explicitly be considering situations where logical states do not necessarily have uniform mean energies, entropies, or even temperatures and we will work in a framework in which logically reversible and irreversible and logically deterministic and indeterministic operations can be treated on an equal footing. Once we have done this we will have a single framework in which the different aspects of Landauer's principle can be united. Doing so will help to address criticisms [4,5] of the limited validity of previous proofs of Landauer's principle, and criticisms [6] of the conclusions of [3].

This will lead us to the following generalization of Landauer's principle.

Generalized Landauer's principle

A physical implementation of a logical transformation of information has minimal expectation value of the work requirement given by

$$\langle \Delta W \rangle \geqslant \langle \Delta E \rangle - T\Delta S, \tag{1}$$

where $\langle \Delta E \rangle$ is the change in the mean internal energy of the information processing system, ΔS the change in the Gibbs–von Neumann entropy of that system, and T the temperature of the heat bath into which any heat is absorbed. The equality is reachable, in principle, by any logical transformation of information, and if the equality is reached the physical implementation is thermodynamically reversible.

We start by considering what we mean by a logical state, a logical operation, and the requirements for a physical system to be an embodiment of such an operation. We will be considering only the processing of discrete, classical information here, although we will be assuming the fundamental physics is quantum.[2]

We then construct an explicit physical process, based upon the familiar "atom in a box" model, that implements a generic logical operation. Thermodynamic optimization of this model will, in general, require consideration of the probability distribution over the input logical states. As this probability distribution is also required to quantify the Shannon information stored in the system, we will refer to the combination of the logical operation and the probability distribution as a logical transformation of information. The optimal implementation, using the atom in a box physical process, shows that the above limit is reachable in principle. We then demonstrate that this limit cannot be exceeded by any system evolving according to a Hamiltonian evolution.

We then consider in more detail the implications of this limit, including several special cases that correspond to more familiar expressions of Landauer's principle, when the physical implementation conforms to a set of conditions which we refer to as "uniform computing." We will show a less famil-

*o.maroney@usyd.edu.au

[1]Although there is little consensus on the naming of these, we will refer to the logical operation as reset to zero (or RTZ) and the widely used physical process which embodies this operation will be referred to as Landauer erasure (or LE).

[2]The analysis would proceed largely unchanged for classical physics, but it would be unnecessarily cumbersome to attempt both. See [7] for a classical treatment.

 ©2009 The American Physical Society

TABLE I. Logical NOT operation.

NOT	
In	Out
0	1
1	0

TABLE II. Reset to zero operation.

RTZ	
In	Out
0	0
1	0

iar set of conditions, which are nevertheless physically possible, which we call "adiabatic equilibrium computing," and which can embody any logical operation without either exchanging heat with the environment or requiring work to be performed. We conclude that any logical transformation of information can be performed in a thermodynamically reversible manner. As this conclusion may seem surprising, we discuss some of the practical barriers to achieving this and the particular problems presented by logically irreversible operations.

II. LOGICAL STATES AND OPERATIONS

Although Landauer's principle is about the thermodynamics of information processing, very little of the literature surrounding it attempts to define what is meant by a logical operation and what are then the minimal requirements of a physical system for it to be regarded as the embodiment of a logical operation. Without first answering this question, it cannot be certain that the most general relationship between information and thermodynamics has been discovered. In this section the abstract properties of logical states and operations will be considered. This leads to constraints upon a physical system which is required to embody the logical states and operations.

A. Logical states

A logical state simply consists of a variable α, which takes a value from a set $\{1,\ldots,n\}$. If the variable α takes the value x then this means that the logical proposition represented by the statement $\alpha=x$ is true. This paper will consider only classical information processing on finite machines. This produces additional properties, whose assumption is usually implicit.[3]

(1) The set of values is a finite set (and, by implication, discrete).

(2) The values are distinct. In any given instance the variable takes one, and only one, of the possible values.

(3) The values are distinguishable. In any given instance the value taken by the variable can be ascertained.

(4) The values are stable. The value taken by the variable cannot change except as a result of a logical operation.

B. Logical operations

A logical operation LO maps input logical states from the set $\{\alpha\}$ to output logical states from the set $\{\beta\}$:

$$\text{LO:} \quad \alpha \to \beta. \tag{2}$$

The number of input and output states need not be the same. The output states from one logical operation may be used as input states to another logical operation. Tables I and II show the maps for two of the most commonly[4] encountered logical operations that act upon two input states 0 and 1, the NOT operation and the reset to zero (RTZ) operation. These rules can be represented by

$$\text{NOT:} \quad 0 \to 1,$$

$$\text{NOT:} \quad 1 \to 0,$$

$$\text{RTZ:} \quad \{0,1\} \to 0, \tag{3}$$

where use has been made of the fact that the RTZ operation transforms both input states into the 0 output state. The RTZ operation is logically *irreversible*. We shall call a device logically irreversible if the output of a device does not uniquely define the inputs. [8]

If multivalued maps (see [9] Sec. VI A, for example) are to be considered, it is necessary to also define logically indeterministic[5] computation: We shall call a device logically indeterministic if the input of a device does not uniquely define the outputs.

Logically indeterministic operations such as unset from zero (UFZ) and randomize (RND) are given in Tables III and IV; they follow the rules

$$\text{UFZ:} \quad 0 \to \{0,1\},$$

$$\text{RND:} \quad \{0,1\} \to \{0,1\}. \tag{4}$$

UFZ is logically reversible, while RND is logically irreversible.

Logically indeterministic operations are perhaps less commonly encountered than logically deterministic operations, and it has been questioned whether these are really logical operations (the authors of [6], for example, take it as part of

[3]For analog or quantum information processing some of these assumptions can be relaxed. We will not consider the consequences of this here.

[4]There is an even more trivial logical operation: logical do nothing IDN. Including this as a logical operation is not a trivial step, as this is the identity operator. It must also be included as a time delay operator when one considers a sequence of logical operations.

[5]In [3] the term "nondeterministic" was used. Unfortunately, this term has a specific usage in computational complexity classes, which does not quite correspond to the usage here. To attempt to avoid confusion, we have changed our terminology to "indeterministic." In terms of computational complexity classes, this is closest to probabilistic computation. We hope this does not simply introduce more confusion.

TABLE III. Unset from zero operation.

UFZ	
In	Out
0	0
0	1

TABLE IV. Randomization.

RND	
In	Out
0	0
0	1
1	0
1	1

the *definition* of a logical operation that it be a single-valued map). We include them for a number of reasons.

(1) Most importantly, such operations play a significant role in the theory of computational complexity classes for actual computers. The complexity class bounded error probabilistic polynomial time (BPP) represents a class of computational problems for which the inclusion of logically indeterministic operations can produce an accurate answer exponentially faster than any known algorithm consisting only of logically deterministic operations (see [12], Sec. III B 2, for example). Excluding them excludes a genuine class of computational procedures.

(2) By including them we are able to derive a more coherent general framework for the thermodynamics of computation. Excluding them creates an artificial asymmetry, and physical properties ascribed to logically irreversible operations in the literature may be artefacts of the asymmetry caused by this exclusion.

(3) Logically indeterministic transformations of information involve the use of probabilistic inferences. There is a point of view [10,11] that regards probabilistic inferences as a natural generalization of deductive logical inferences;

(4) Finally, there seems no special reason *not* to include them as they form a natural counterpart to the concept of logically irreversible operations. Any conclusion we can draw that applies to the set of all such logical operations must necessarily apply to all logically deterministic operations. Including logically indeterministic operations in our analysis will not invalidate its applicability to logically deterministic operations.

C. Logical transformation of information

To quantify the information being processed by the logical operation, the Shannon information measure will be used. This requires the specification of a probability distribution over the input and output states. If the logical states input to a computation occur with probabilities $P(\alpha)$, then the Shannon information represented by the input states is

$$H_\alpha = -\sum_\alpha P(\alpha)\log_2 P(\alpha). \tag{5}$$

During the logical operation these input states are transformed into output states β. When an input state may be transformed into more than one output state, one must specify the probability $P(\beta|\alpha)$ for each possible output state. For logically deterministic operations, specifying $P(\beta|\alpha)$ is trivial as $\forall\ \alpha\ \exists\ \beta\ P(\beta|\alpha)=1$ {or equivalently $\forall\ \alpha\ \exists\ \beta$ $[\forall\ \beta' \neq \beta P(\beta'|\alpha)=0]$}. Specifying all the nonzero $P(\beta|\alpha)$ completely specifies the rules of the logical operation. We will therefore take the set $\{P(\beta|\alpha)\}$ as the definition of a general logical operation. For logically deterministic operations, this is the list of all combinations of input and output states that have conditional probability 1, which is simply the truth table for the operation.

After the logical operation, the output states β will occur with probability

$$P(\beta) = \sum_\alpha P(\beta|\alpha)P(\alpha), \tag{6}$$

so the Shannon information represented by the output states is

$$H_\beta = -\sum_\beta P(\beta)\log_2 P(\beta). \tag{7}$$

When we refer to a logical transformation of information, we will mean a logical operation, acting upon input states[6] $\{\alpha\}$, that occur with probabilities $P(\alpha)$, which transforms the input states to output states $\{\beta\}$ with conditional probabilities $P(\beta|\alpha)$.

The conditional probability that a given output state β was generated by the input state α is

$$P(\alpha|\beta) = \frac{P(\alpha)P(\beta|\alpha)}{P(\beta)} \tag{8}$$

and the joint probability that there was an input state α and output state β is

$$P(\alpha,\beta) = P(\alpha)P(\beta|\alpha) = P(\beta)P(\alpha|\beta). \tag{9}$$

This gives an equivalent formulation of logical determinism and logical reversibility. A logically deterministic computation is one for which

$$\forall\ \alpha,\beta \quad P(\beta|\alpha) \in \{0,1\}. \tag{10}$$

A logically reversible computation is one for which

$$\forall\ \alpha,\beta \quad P(\alpha|\beta) \in \{0,1\}. \tag{11}$$

This is defined in terms of the set $\{P(\alpha|\beta)\}$. A logical operation has been defined only by the set $\{P(\beta|\alpha)\}$, with the $P(\alpha|\beta)$ dependent upon the input logical state probabilities $P(\alpha)$.

From

$$P(\alpha|\beta) = 0 \Rightarrow P(\alpha,\beta) = P(\beta|\alpha) = 0,$$

[6]For simplicity, input states for which $P(\alpha)=0$, i.e., which are certain to not occur, will not be included in this set.

$$P(\alpha|\beta)=1 \Rightarrow P(\alpha' \neq \alpha|\beta)=0$$

$$\Rightarrow P(\alpha' \neq \alpha,\beta)=P(\beta|\alpha' \neq \alpha)=0, \quad (12)$$

there is an equivalent definition of logically reversible computations. An operation is logically reversible if and only if

$$\forall \ \beta \quad \{P(\beta|\alpha) \neq 0 \Rightarrow [\forall \ \alpha' \neq \alpha P(\beta|\alpha')=0]\}. \quad (13)$$

This definition is now independent of the input probability distribution.

We summarize these properties and some consequences.

1. Logically deterministic operations

$$\forall \ \alpha,\beta \quad P(\beta|\alpha) \in \{0,1\},$$

$$\forall \ \alpha \quad \{P(\alpha|\beta) \neq 0 \Rightarrow [\forall \ \beta' \neq \beta P(\alpha|\beta')=0]\},$$

$$\forall \ \alpha,\beta \quad P(\alpha|\beta) \in \left\{0,\frac{P(\alpha)}{P(\beta)}\right\}. \quad (14)$$

In the case where a particular $\alpha \rightarrow \beta$ transition has $P(\beta|\alpha)=1$, we may refer to this as a logically deterministic *transition*, even if the overall operation is not logically deterministic.

2. Logically reversible operations

$$\forall \ \alpha,\beta \quad P(\alpha|\beta) \in \{0,1\},$$

$$\forall \ \beta \quad \{P(\beta|\alpha) \neq 0 \Rightarrow [\forall \ \alpha' \neq \alpha P(\beta|\alpha')=0]\},$$

$$\forall \ \alpha,\beta \quad P(\beta|\alpha) \in \left\{0,\frac{P(\beta)}{P(\alpha)}\right\}. \quad (15)$$

In the case where a particular $\alpha \rightarrow \beta$ transition has $P(\alpha|\beta)=1$, we may refer to this as a logically reversible *transition*, even if the overall operation is not logically reversible.

D. Physical representation of logical states

We will now consider what the properties above imply for the physical embodiment of logical states and operations upon them. The physical system will have a state space of possible microstates $\{\mu\}$. How can these be used to embody the logical states?

(1) A particular logical state α will be identified with a set of microstates $\{\mu_\alpha\}$ in the state space, in the sense that, when the physical state of the system is one of the microstates $\mu \in \{\mu_\alpha\}$, then the logical state takes the value α.

(2) As logical states are distinct, a given microstate can be identified with one, and only one, input state. Each set of microstates $\{\mu_\alpha\}$ is therefore nonintersecting with any other such set of microstates:

$$\{\mu_\alpha\} \cap \{\mu_{\alpha' \neq \alpha}\} = \varnothing. \quad (16)$$

(3) For the logical states to be distinguishable, it is necessary that it is possible to ascertain the set to which the microstate belongs. We are not considering analog information processing, so the physical interactions must not need to be sensitive to arbitrarily close (using a natural distance measure) states in state space. We replace the point in state space, μ, with the neighborhood of that point, $R(\mu)$. The logical state α is now identified with the region of state space corresponding to the union of all the neighborhoods $\{R(\mu_\alpha)\}$. The neighborhoods corresponding to different logical states must be nonoverlapping.

(4) We can now identify the proposition for the logical state α with the projector K_α onto the region of state space $\{R(\mu_\alpha)\}$:

$$K_\alpha K_{\alpha'} = \delta_{\alpha\alpha'} K_\alpha,$$

$$\sum_\alpha K_\alpha = I,$$

$$K_\alpha[R(\mu_\alpha)] = R(\mu_\alpha),$$

$$K_\alpha[R(\mu_{\alpha' \neq \alpha})] = 0. \quad (17)$$

The proposition α is true if the state μ is in the region of state space $\{R(\mu_\alpha)\}$ projected out by K_α.

(5) For the physical representation of the logical states to be complete, then it must also be the case that, if the state μ is in the region of state space $\{R(\mu_\alpha)\}$ projected out by K_α, then the logical proposition corresponding to the logical state α is true.

(6) For the logical states to be stable, then under the normal evolution of the system, a microstate within the region of state space corresponding to a given logical state must stay within that region of state space.[7] The normal evolution of the system is, trivially, a physical embodiment of the logical do nothing IDN operation.

E. Physical representation of logical operations

During the normal evolution of a system, logical states do not change. To perform nontrivial logical operations new interactions must alter the evolution of the state space. All the essential characteristics of a logical operation are included in the set $\{P(\beta|\alpha)\}$. It follows that a physical process is an embodiment of a logical operation if and only if the evolution of the microstates in the physical process is such that, over an ensemble of microstates in the region $\{R(\mu_\alpha)\}$, the probability that the microstate ends up in the region $\{R(\mu_\beta)\}$ is just $P(\beta|\alpha)$.

(1) We will assume that the laws of physics are Hamiltonian. The evolution of microstates over the state space of the combined system of the logical processing apparatus and the environment must be described by a Hamiltonian evolution operator.

(2) If the interaction of the microstates of the system and the environment is such that any individual microstate μ_α

[7] A weaker condition, acceptable for most practical needs, is that the probability of the microstate leaving the region of a given logical state, during the time scale of the information processing, must be very low.

starting in state α is randomized so that it ends up in the output state β with probability $P(\beta|\alpha)$, then we do not need to be sensitive to the initial probability distribution of the ensemble of microstates within the logical state α. In general, however, we may need to be sensitive to the initial probability distribution ρ_α over microstates corresponding to the logical state α.

(3) The complete statistical state of the logical processing system input to the logical operation is

$$\sum_\alpha P(\alpha)\rho_\alpha \tag{18}$$

where

$$\forall\ \alpha, \quad K_\alpha \rho_\alpha K_\alpha = \rho_\alpha. \tag{19}$$

(4) The complete statistical state of the logical processing system output from the logical operation is required to be

$$\sum_\beta P(\beta)\rho_\beta \tag{20}$$

where

$$\forall\ \beta, \quad K_\beta \rho_\beta K_\beta = \rho_\beta. \tag{21}$$

We have not considered separate systems for the logical input states, the logical processing apparatus or the output states. At first, this seems to assume that the system embodying the logical input states must be the same as the system embodying the logical output states and that the logical processing apparatus cannot have internal states—which would seem to be quite a strong restriction. This is not the case. Let us consider the case where there are three distinct systems: the input state system, with states $\{\rho_\alpha\}$; an output state system, with states $\{\rho_\beta\}$; and an auxiliary system corresponding to all internal and external components of the process, with states $\{\rho^{\mathrm{app}}\}$.

The statistical state is described at the start of the operation by

$$\sum_\alpha P(\alpha)\rho_\alpha \otimes \sum_\beta f_\beta \rho_\beta^{\mathrm{app}} \otimes \rho_\beta, \tag{22}$$

where we have assumed that the input state system is initially uncorrelated to the apparatus but have not assumed the output state system is initially uncorrelated to the internal states $\{\rho_\beta^{\mathrm{app}}\}$ of the apparatus. The effect of the operation would be to evolve the combined system into some new correlated state, combining the three systems:

$$\sum_{\alpha,\beta} P(\alpha,\beta)\rho'_{\alpha,\beta} \otimes \rho_{\alpha,\beta}^{\mathrm{app}} \otimes \rho_\beta. \tag{23}$$

Our approach here is then to consider the state space of the combined system of input, output, and apparatus as a single state space, with input states for α of

$$\rho_\alpha \otimes \left(\sum w_\beta \rho_\beta^{\mathrm{app}} \otimes \rho_\beta\right) \tag{24}$$

and output states for β of

$$\left(\sum_\alpha P(\alpha,\beta)\rho'_{\alpha,\beta} \otimes \rho_{\alpha,\beta}^{\mathrm{app}}\right) \otimes \rho_\beta. \tag{25}$$

We then consider a Hamiltonian evolution on the combined state space to be the operation. This cleanly separates the logical states, embodied by the physical state of the combined state space, from the logical operation, embodied by the Hamiltonian evolution on that state space. We have not restricted ourselves by the assumption of a Hamiltonian evolution on a single state space, as we have the full generality of all possible Hamiltonian interactions allowed between the input state system, the output state system, and the logical processing apparatus. We have avoided, on the other hand, any need to consider the restrictions and complications that would arise if we constructed models based upon specific assumptions as to how the input state, output state, and logical processing apparatus systems are allowed to interact. This completes the physical characterization of logical states and operations.

F. Logical vs microscopic determinism and reversibility

There is one final issue that needs to be stated, for the sake of clarity, regarding the (absence of a) relationship between logical and microscopic indeterminism and irreversibility.

Logical indeterminism does not imply or require the existence of any fundamental indeterminism in the microscopic dynamics of the physical states. Neither is *logical determinism* incompatible with the existence of fundamental indeterministic dynamics.

A specific microstate from the input logical state may evolve deterministically into a specific microstate of an output logical state, while the operation remains logically indeterministic, provided the set of the input microstates corresponding to the same input logical state do not all evolve, with certainty, into microstates of the same output logical state.

A specific microstate from an input logical state may evolve indeterministically into a number of possible microstates, while the operation remains logically deterministic, provided that the set of the input microstates corresponding to the same input logical state can only evolve into microstates from the set corresponding to the same output logical state.

Logical irreversibility does not imply or require the existence of any fundamental irreversibility in the microscopic dynamics of the physical states. Neither is *logical reversibility* incompatible with the existence of fundamental irreversible dynamics.

A specific microstate from the input logical state may evolve reversibly into a specific microstate of an output logical state, while the operation remains logically irreversible, provided the set of the microstates corresponding to the same output logical state have not all evolved, with certainty, from microstates of the same input logical state. A specific microstate from an input logical state may evolve irreversibly into a specific microstate, while the operation remains logically reversible, provided that the set of microstates corresponding to the same output logical state can only have evolved from microstates in the set corresponding to the same input logical state.

III. THERMODYNAMICS OF LOGICAL OPERATIONS

We now undertake the main task of this paper: to determine the limiting thermodynamic cost to a logical operation. We will do this in two steps.

First, we will construct a physical process, capable of implementing any logical operation, as we have defined them, and we will consider the optimum thermodynamic cost to the process. This optimum will be considered in two ways: for individual transitions between specific logical states; and as an expectation value over an ensemble of operations. Both work required to perform the process and heat generated by the process will be calculated, where it is assumed that all heat generated is absorbed by a heat bath at some reference temperature T_R.

To calculate the expectation values, we must consider the probability distribution over the input logical states. For this we will use the probability distribution used to calculate the Shannon information being processed. The optimum process will, of necessity, involve various idealizations (such as frictionless motion and quasistatic processes) that cannot be achieved in practice. The purpose is to demonstrate not that it is possible to build such optimal operations, but rather that there is no physical limitation, in principle, on *how close* one can get to them.

Then we will prove that there cannot exist any physical process that can implement the same logical transformation of information, but with a lower expectation value for either the work requirement or the heat generation. The optimum process, for our particular implementation of a logical transformation of information, is also the optimum for any possible implementation of that transformation.

A. Statistical mechanical assumptions

We will now clearly state the statistical mechanical assumptions that are being made. There are a number of different approaches to the foundations of statistical mechanics and, as the models discussed here involve such idealizations as the treatment of individual atoms, it is important to be clear which approach is being taken. In this paper we will assume the standard structure of Gibbs canonical statistical mechanics: we will be dealing with Hamiltonian flows with probability distributions over a state space, we will assume that a system that has been thermalized can be represented by a canonical distribution over its accessible state space, and will be initially statistically independent of any other system. While these assumptions are clearly open to debate, a full discussion or justification of them lies outside the scope of this paper (although see [13]).

(1) The system consists of the logical processing apparatus (including auxiliary systems as discussed in Sec. II E) and a number of heat baths. A heat bath is simply a system that has been allowed to thermalize at some temperature and is sufficiently large that any energy transfer with the logical processing apparatus will have negligible effect upon the heat baths' internal energy. The Hamiltonian for the combined system is

$$H = H_L + \sum_i (H_i + V_i), \tag{26}$$

where H_L is the internal Hamiltonian for the logical processing apparatus, H_i the internal Hamiltonian of the heat bath i, and V_i the interaction Hamiltonian between the logical processing apparatus and the heat bath i. We assume there is no interaction between heat baths. The density matrix of the combined system is ρ_C, and ρ_L is the marginal density matrix after tracing over the heat bath subsystems.

(2) Work is performed upon the apparatus through the variation of some externally controlled parameter X, which affects the energy eigenvalues and eigenstates,[8]

$$H_L(X) = \sum_n E_n(X)|E_n(X)\rangle\langle E_n(X)|. \tag{27}$$

The mean work performed as the parameter is varied from X_0 to X_1 is given by

$$\Delta W = \int_{X_0}^{X_1} \mathrm{Tr}\left(\frac{\partial H_L(X)}{\partial X}\rho_L(X)\right)dX. \tag{28}$$

Note that the density matrix ρ_L may be varying as X varies. We will assume that neither the internal Hamiltonians of the heat baths nor the interaction Hamiltonians have controllable parameters: work is performed only upon the logical processing apparatus itself.

(3) The mean change in internal energy of the logical processing apparatus is

$$\Delta E = \mathrm{Tr}[H_L(X_1)\rho_L(X_1)] - \mathrm{Tr}[H_L(X_0)\rho_L(X_0)]. \tag{29}$$

(4) If we now assume negligible changes in interaction energies,

$$\forall i \quad \mathrm{Tr}[V_i\rho_C(X_1)] \approx \mathrm{Tr}[V_i\rho_C(X_0)], \tag{30}$$

then

$$\Delta W - \Delta E = \sum_i \Delta Q_i, \tag{31}$$

where

$$\Delta Q_i = \mathrm{Tr}[H_i\rho_i(X_1)] - \mathrm{Tr}[H_i\rho_i(X_0)] \tag{32}$$

is the increase in internal energy of the heat bath i, and ρ_i is the marginal density matrix of the heat bath, after tracing over the logical processing apparatus and all other heat baths.

(5) If the evolution of the density matrix is such that it always remains diagonalized by the energy eigenstate basis, so that

$$\rho_L(X) = \sum_n p_n(X)|E_n(X)\rangle\langle E_n(X)|, \tag{33}$$

then

[8] More generally, one should consider a number of controllable parameters, which are each varying in time.

$$\Delta W = \int_{X_0}^{X_1} \sum_n p_n(X) \frac{\partial E_n(X)}{\partial X} dX, \tag{34}$$

$$\sum_i \Delta Q_i = \int_{X_0}^{X_1} \sum_n \frac{\partial p_n(X)}{\partial X} E_n(X) dX. \tag{35}$$

It is important to note that these five points make no assumption regarding the identification of either thermodynamic entropy or thermal distributions. Neither the canonical distribution nor the Gibbs–von Neumann entropy has been used.

The following results depend upon the assumption that a heat bath is represented by a canonical distribution and that a limiting ideal case exists of thermalization through a succession of brief interactions with small subsystems of a heat bath. The calculations are well known (see [13–15], for example) and the results are stated here purely for clarity. No formal identification of the Gibbs–von Neumann entropy with thermodynamic entropy is required to derive these results.

(6) A system that is brought into contact with an ideal heat bath will, over time periods long with respect to its thermal relaxation time, be well represented by a canonical probability distribution

$$\rho_\alpha = \frac{e^{-H/kT}}{\mathrm{Tr}(e^{-H/kT})} \tag{36}$$

over accessible states of the system, with T being the temperature of the heat bath and H the Hamiltonian of the system over the accessible subspace.

(7) In the limit of isothermal quasistatic processes, the system is in contact with an ideal heat bath at some temperature, and the system stays in thermal equilibrium with the heat bath at all times.

(8) In the limit of adiabatic quasistatic processes (or essentially isolated [14] processes) the system always remains in a (canonically distributed) thermal state but there is zero mean energy flow out of the system ($\Delta W = \Delta E$). The temperature of this state may vary.

(9) We will assume that the only systems with which the information processing system interacts are ideal heat baths at temperatures $\{T_\alpha\}$, $\{T_\beta\}$, and T_R, and a work reservoir, and that there are no initial correlations between the system and the heat baths.

While these assumptions involve significant idealizations, they are the kind of idealizations that are standard in thermodynamics and statistical mechanics. Rather than representing a physically achievable process, they represent the limit of what can be physically achieved. There is no physical reason why one cannot, in principle, get arbitrarily close to these results.

Although the value of the Gibbs–von Neumann entropy $-k\mathrm{Tr}[\rho \ln[\rho]]$ will be calculated for the input and output logical states, all results in this section, in terms of work required and heat generated, are derivable, from the assumptions stated, without needing to identify this property with thermodynamic entropy.[9]

B. Generic logical operation

1. Input logical states

We start the operation with the logical states represented by physical states with the following properties.

(1) An input logical state α to the logical computation is physically embodied by a system confined to some region of state space. The distribution over the microstates of that region gives the density matrix ρ_α.

(2) ρ_α has mean energy $E_\alpha = \mathrm{Tr}(H_L \rho_\alpha) = \mathrm{Tr}(H_\alpha \rho_\alpha)$, where $H_\alpha = K_\alpha H_L K_\alpha$.

(3) For simplicity, in the main section, we will assume that the input logical state α is canonically distributed, as if it has been thermalized with a heat bath at temperature T_α,

$$\rho_\alpha = \frac{e^{-H_\alpha/kT_\alpha}}{\mathrm{Tr}(e^{-H_\alpha/kT_\alpha})}. \tag{37}$$

This assumption is not essential, and can easily be relaxed without affecting any result. If the initial density matrix is not a canonical distribution, then it is possible to construct a unitary operator that acts upon the system in isolation and rotates it into a canonical state, with neither heat nor work requirement.

To give an explicit construction, suppose the initial Hamiltonian and density matrix are $H_\alpha^{(i)}$ and $\rho_\alpha^{(i)}$, such that

$$\rho_\alpha^{(i)} \neq \frac{e^{-H_\alpha^{(i)}/kT_\alpha}}{\mathrm{Tr}(e^{-H_\alpha^{(i)}/kT_\alpha})} \tag{38}$$

for any T_α. Given the diagonal representation

$$\rho_\alpha^{(i)} = \sum_n p_n |\lambda_n\rangle\langle\lambda_n|, \tag{39}$$

then the Hamiltonian, acting in the range $0 < t < \tau$,

$$H_A = \left[\cos^2\left(\frac{\pi t}{2\tau}\right) - \sin^2\left(\frac{\pi t}{\tau}\right)\right] H_\alpha^{(i)} + \left[\sin^2\left(\frac{\pi t}{2\tau}\right) - \sin^2\left(\frac{\pi t}{\tau}\right)\right] H_\alpha - \frac{2\imath\hbar}{\tau}\sin^2\left(\frac{\pi t}{\tau}\right)\ln\left(\sum_n |\gamma_n\rangle\langle\lambda_n|\right) \tag{40}$$

with

$$H_\alpha = \sum_n E_n |\gamma_n\rangle\langle\gamma_n|, \tag{41}$$

$$E_n = E_\alpha - kT_\alpha\left(\ln(p_n) - \sum_m p_m \ln(p_m)\right), \tag{42}$$

varies continuously from $H_\alpha^{(i)}$ to H_α, and has the effect of leaving the system, after time τ, in the stationary canonical state

[9] See [2] and [13] (Chap. 6), where this kind of calculation is carried out in detail for the same kinds of system considered here.

$$\rho_\alpha = \frac{e^{-H_\alpha/kT_\alpha}}{\mathrm{Tr}(e^{-H_\alpha/kT_\alpha})} = \sum_n p_n|\gamma_n\rangle\langle\gamma_n|. \quad (43)$$

ρ_α is unitarily equivalent to $\rho_\alpha^{(i)}$ and $\mathrm{Tr}(H_\alpha\rho_\alpha)=\mathrm{Tr}(H_\alpha^{(i)}\rho_\alpha^{(i)})$. The mean work requirement is zero and no heat is exchanged with the environment. It should be noted that this construction holds even if $\rho_\alpha^{(i)}$ is not diagonalized in the eigenstates of $H_\alpha^{(i)}$.

(4) The Gibbs–von Neumann entropy of the input logical state is: $S_\alpha=-k\,\mathrm{Tr}[\rho_\alpha \ln(\rho_\alpha)]$.

(5) There are M possible input logical states.

2. Output logical states

The output logical states may be similarly characterized as follows.

(1) An output logical state β from the logical computation is physically embodied by a system confined to some region of state space. The distribution over the microstates of that region gives the density matrix ρ_β.

(2) ρ_β has mean energy $E_\beta=\mathrm{Tr}(H_L'\rho_\beta)=\mathrm{Tr}(H_\beta'\rho_\beta)$, where $H_\beta'=K_\beta H_L' K_\beta$.

(3) Again, for convenience we will assume that the output logical state β is canonically distributed as if it has been thermalized at a temperature T_β. This will make the density matrix

$$\rho_\beta = \frac{e^{-H_\beta'/kT_\beta}}{\mathrm{Tr}(e^{-H_\beta'/kT_\beta})}. \quad (44)$$

Again, this assumption is easily dropped. If the final state is required to be a noncanonical density matrix $\rho_\beta^{(f)}$, with Hamiltonian $H_\beta^{(f)}$, then $\rho_\beta^{(f)}$ can be obtained from ρ_β by constructing H_β' and H_B in the same manner as H_α and H_A above.

(4) The Gibbs–von Neumann entropy of the output logical state is $S_\beta=-k\,\mathrm{Tr}[\rho_\beta \ln(\rho_\beta)]$.

(5) There are N possible output logical states.

(6) The output logical state β must occur with probability $P(\beta|\alpha)$ given input logical state α.

We will also note here that probabilities enter the calculation at two levels: as a probability distribution over the microstates within a given logical state, and as a probability distribution over the different logical states. We will attempt to keep these formally separate. From this point onward, the microstate probability will be represented only by the density matrix. Explicitly appearing probabilities and averages will always refer to the probability distribution over the logical states.

C. The transformation of information

If we consider the actual microstate of the system, within the region corresponding to a given logical state, as being some free[10] parameter, then the physical representation of each logical state may initially be regarded as a potential well with some arbitrary shape, such that the free parameter is confined within the well. The potential wells associated with different logical states are in different regions of physical space, separated by high potential barriers, such that there is a very low possibility of transitions between different logical states.

This can be represented as an atom in one of a number of boxes, where the free parameter is the location of the atom in the box. The logical state is represented by the particular box, or potential well, within which the atom is confined. The physical transformation of the information will take place in nine steps. Steps 1–3 will bring the input logical states into standardized physical states at a shared reference temperature. Step 4 is the logically indeterministic implementation of the $P(\beta|\alpha)$ transition. Steps 5 and 6 implement the joining together of the β output states from the different α input states, giving the logically irreversible stage. Steps 6–9 then alter each output logical state to the required final physical state.

Calculations for work requirements, heat generation, and so forth follow the statistical mechanical calculations above. Particularly detailed calculations for atom in a box type systems are considered in references such as [2,16–18]. The key results can be summarized. The Hamiltonian for an infinite square well potential, of width l, holding an atom of mass m is

$$H(l) = \sum_n \frac{\hbar^2\pi^2}{8ml^2} n^2 |E_n\rangle\langle E_n|. \quad (45)$$

Work is performed upon the system by varying the l parameter (width of the box).

In a canonical thermal state at temperature T, the mean energy is

$$E = \frac{\sum_n \frac{\hbar^2\pi^2}{8ml^2} n^2 e^{-(\hbar^2\pi^2/8ml^2kT)n^2}}{\sum_n e^{-(\hbar^2\pi^2/8ml^2kT)n^2}} \approx \frac{1}{2}kT, \quad (46)$$

and the Gibbs–von Neumann entropy is

$$S = \frac{\sum_n \frac{\hbar^2\pi^2}{8ml^2T} n^2 e^{-(\hbar^2\pi^2/8ml^2kT)n^2}}{\sum_n e^{-(\hbar^2\pi^2/8ml^2kT)n^2}} + k\ln\left(\sum_n e^{-(\hbar^2\pi^2/8ml^2kT)n^2}\right)$$
$$\approx \frac{k}{2}\ln\left(\frac{2emkTl^2}{\pi\hbar^2}\right). \quad (47)$$

The approximations hold when the temperature is high with respect to the ground state:

$$kT \gg \frac{\pi\hbar^2}{2em}. \quad (48)$$

(1) The first step is to continuously and slowly deform the potential well of each separate logical state into a square well potential. The square well should be deformed to width $d_\alpha^{(1)}$,

[10] Not externally controlled.

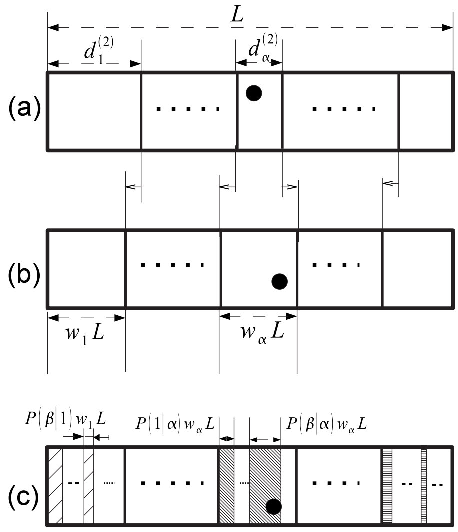

FIG. 1. Arranging input states.

$$d_\alpha^{(1)} = \left(\sqrt{\frac{\pi\hbar^2}{2emkT_\alpha}} \right) e^{S_\alpha/k}, \tag{49}$$

where m is the mass of the atom. This state has mean energy and entropy

$$E_\alpha^{(1)} = \frac{1}{2}kT_\alpha,$$

$$S_\alpha^{(1)} = S_\alpha. \tag{50}$$

If this deformation is carried out sufficiently slowly, the mean heat generation is zero and the work requirement is

$$W_\alpha^{(1)} = \frac{1}{2}kT_\alpha - E_\alpha. \tag{51}$$

This is a mean work requirement for the operation. Fluctuations may occur around this value.

The system may now be pictured as a box, divided with $M-1$ partitions. When the atom is located between the $\alpha-1$ and α partitions, the system is in logical state α. This can be seen in Fig. 1(a).

(2) Remove the system from all contact with heat baths and then, slowly, adiabatically vary the width of each square well to $d_\alpha^{(2)}$:

$$d_\alpha^{(2)} = d_\alpha^{(1)} \sqrt{\frac{T_\alpha}{T_R}}. \tag{52}$$

At the limit of a slow, quasistatic process, this will leave each logical state with a density matrix equal to a canonical thermal system with temperature T_R. The mean energy, entropy, and mean work requirements are

$$E_\alpha^{(2)} = \frac{1}{2}kT_R,$$

$$S_\alpha^{(2)} = k \ln\left[d_\alpha^{(2)} \left(\sqrt{\frac{2emkT_R}{\pi\hbar^2}} \right) \right] = S_\alpha,$$

$$W_\alpha^{(2)} = \frac{1}{2}kT_R - \frac{1}{2}kT_\alpha. \tag{53}$$

As the total width of the box is now

$$L = \sum_{\alpha'} d_{\alpha'}^{(2)} = \left(\sqrt{\frac{\pi\hbar^2}{2emkT_R}} \right) \sum_{\alpha'} e^{S_{\alpha'}/k}, \tag{54}$$

then

$$d_\alpha^{(2)} = L \frac{e^{S_\alpha/k}}{\sum_{\alpha'} e^{S_{\alpha'}/k}}. \tag{55}$$

(3) Now bring the entire system into contact with heat baths at the reference temperature T_R. Slowly and isothermally move the positions of the potential barriers separating the square wells [see Figs. 1(a) and 1(b)]. Move the ith barrier to the position x_i:

$$x_i = L \sum_{\alpha=1}^{\alpha=i} w_\alpha, \tag{56}$$

where $\Sigma_\alpha w_\alpha = 1$. The values of w_α have not been specified. Variation of these will be used to optimize the operation.

Each logical state now has a width $d_\alpha^{(3)} = w_\alpha L$. If $w_\alpha = 0$ for one of the input states α this stage will compress the volume of that state to zero. Clearly this can only be allowed to take place if there is no possibility that the partition is occupied by the atom.

$$E_\alpha^{(3)} = \frac{1}{2}kT_R,$$

$$S_\alpha^{(3)} = k \ln\left[d_\alpha^{(3)} \left(\sqrt{\frac{2emkT_R}{\pi\hbar^2}} \right) \right],$$

$$W_\alpha^{(3)} = kT_R \ln\left(\frac{d_\alpha^{(3)}}{d_\alpha^{(2)}} \right),$$

$$Q_\alpha^{(3)} = kT_R \ln\left(\frac{d_\alpha^{(3)}}{d_\alpha^{(2)}} \right), \tag{57}$$

where $Q_\alpha^{(3)}$ is the heat generated in the heat bath if the atom is in the α partition.

(4) Insert $N-1$ new potential barriers slowly into each partition. Within a given partition α, the barriers should be spaced according to the probabilities $P(\beta|\alpha)$ of the logical operation (see Fig. 2). They have a width

$$d_{\alpha,\beta}^{(4)} = P(\beta|\alpha) w_\alpha L. \tag{58}$$

This is the logically indeterministic step of the computation, in Figs. 1(b) and 1(c). There are M partitions, each with N subpartitions.

If the atom was located in partition α beforehand, and the system has been allowed to thermalize at temperature T_R for a period of time greater than the thermal relaxation time, then the probability of the atom now being located in the (α,β) subpartition is $P(\beta|\alpha)$. For logically deterministic op-

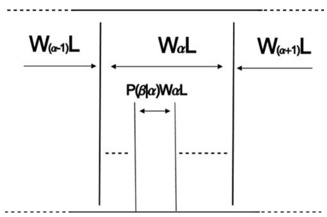

FIG. 2. Inserting subpartitions.

erations, then all nonzero $P(\beta|\alpha)$ are equal to 1 and no partitions need be inserted.

The mean energy and the entropy associated with the atom located within a particular (α,β) subpartition is

$$E^{(4)}_{\alpha,\beta}=\frac{1}{2}kT_R,$$

$$S^{(4)}_{\alpha,\beta}=k\ln\left[d^{(4)}_{\alpha,\beta}\left(\sqrt{\frac{2emkT_R}{\pi\hbar^2}}\right)\right]. \quad (59)$$

(5) Now rearrange the subpartitions so that, for each β, all the β output partitions are adjacent. From each α partition, we gather the first subpartitions, corresponding to output logical state $\beta=1$, and collect them together. Repeat this for each set of β subpartitions, from all the α partitions. Finally this produces a sequence of N β partitions, each with M α subpartitions. This is illustrated in Fig. 3.

(6) Remove the potential barriers within each β output partition and leave the box for a time that is long in comparison to the atoms' thermal relaxation time. The β partition has a width

$$d^{(6)}_{\beta}=w_{\beta}L, \quad (60)$$

where

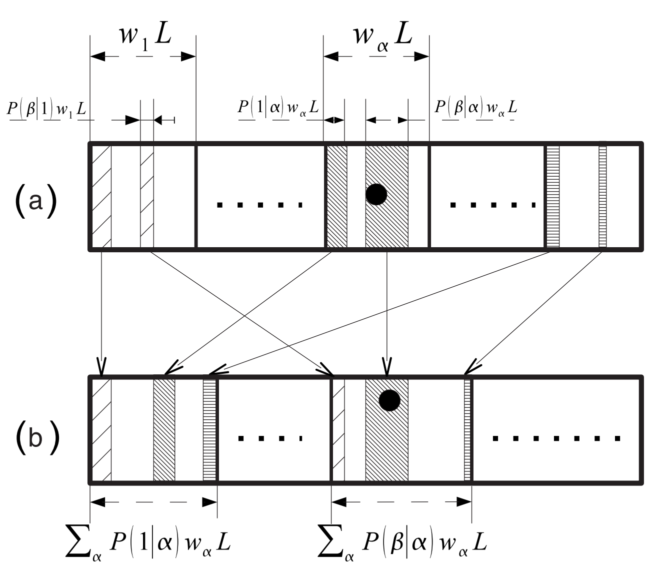

FIG. 3. Rearranging the partitions.

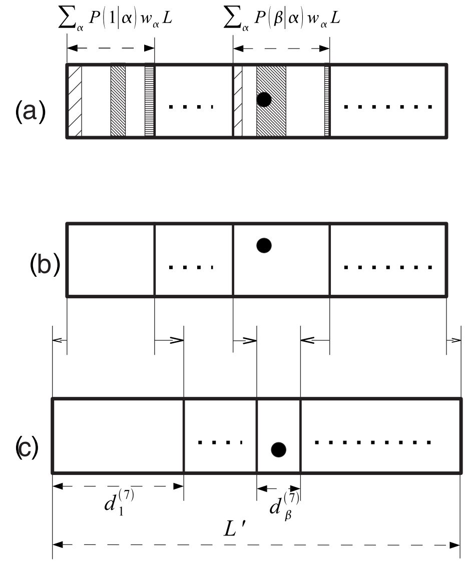

FIG. 4. Arranging the output states

$$w_{\beta}=\sum_{\alpha}w_{\alpha}P(\beta|\alpha), \quad (61)$$

and if the atom is located in the β partition, then

$$E^{(6)}_{\alpha}=\frac{1}{2}kT_R,$$

$$S^{(6)}_{\alpha}=k\ln\left[d^{(6)}_{\beta}\left(\sqrt{\frac{2emkT_R}{\pi\hbar^2}}\right)\right]. \quad (62)$$

This is the logically irreversible stage and is illustrated in Figs. 4(a) and 4(b). This stage is trivial for logically reversible computations, for which each β output partition is composed of only one α subpartition, and so has no internal barriers. Note also that, if $\forall\ \alpha$, $P(\beta|\alpha)=0$, then the atom can never be located in the β partition.

(7) Now slowly and isothermally resize the output partitions. The barriers should be moved until the β partition has width

$$d^{(7)}_{\beta}=\left(\sqrt{\frac{\pi\hbar^2}{2emkT_R}}\right)e^{S_\beta/k}. \quad (63)$$

See Figs. 4(b) and 4(c).

The overall width of the box may be changed by this operation, and is now

$$L'=\sum_{\beta'}d^{(7)}_{\beta'}=L\frac{\sum_{\beta'}\exp^{S_{\beta'}/k}}{\sum_{\alpha'}\exp^{S_{\alpha'}/k}}. \quad (64)$$

so that

$$d^{(7)}_{\beta}=L'\frac{e^{S_\beta/k}}{\sum_{\beta'}\exp^{S_{\beta'}/k}}. \quad (65)$$

For the atom located in the β partition, we have

$$E_\beta^{(7)} = \frac{1}{2}kT_R,$$

$$S_\beta^{(7)} = k \ln\left[d_\beta^{(7)}\left(\sqrt{\frac{2emkT_R}{\pi\hbar^2}}\right)\right] = S_\beta,$$

$$W_\beta^{(7)} = kT_R \ln\left(\frac{d_\beta^{(7)}}{d_\beta^{(6)}}\right),$$

$$Q_\beta^{(7)} = kT_R \ln\left(\frac{d_\beta^{(7)}}{d_\beta^{(6)}}\right), \tag{66}$$

where $Q_\beta^{(7)}$ is the heat generated in the heat bath.

(8) Now remove all contact from the T_R heat baths. With the system thermally isolated, slowly and adiabatic resize the output partitions to the widths

$$d_\beta^{(8)} = d_\beta^{(7)} \sqrt{\frac{T_R}{T_\beta}}. \tag{67}$$

If the atom is in the β partition, the effect of this quasistatic, adiabatic evolution is to leave the atom in a canonical thermal state with temperature T_β,

$$E_\beta^{(8)} = \frac{1}{2}kT_\beta,$$

$$S_\beta^{(8)} = S_\beta,$$

$$W_\beta^{(8)} = \frac{1}{2}kT_\beta - \frac{1}{2}kT_R. \tag{68}$$

(9) The output logical states β are now all at the required temperature and entropy. For completeness, bring each separate β partition into thermal contact with a heat bath at the appropriate temperature T_β and slowly, continuously, and isothermally deform the shape of each square well potential into the final potential for the output logical state:

$$E_\beta^{(9)} = E_\beta,$$

$$S_\beta^{(9)} = S_\beta,$$

$$W_\beta^{(9)} = E_\beta - \frac{1}{2}kT_\beta. \tag{69}$$

This completes the physical implementation of the logical operation.

D. Thermodynamic costs

The procedure detailed in the previous section satisfies the requirements of a generic logical operation. The input logical states are represented by the appropriate physical input states, the output logical states are represented by the appropriate physical output states, and the transitions between them occur with probabilities $P(\beta|\alpha)$.

1. Individual transitions

Addition of the work and heat values across all steps, for a system which starts in logical state α and ends in logical state β, gives

$$\Delta W_{\alpha,\beta} = (E_\beta - T_R S_\beta) - (E_\alpha - T_R S_\alpha) + kT_R \ln\left(\frac{w_\beta}{w_\alpha}\right),$$

$$\Delta Q_{\alpha,\beta} = T_R\left[S_\alpha - S_\beta + k \ln\left(\frac{w_\beta}{w_\alpha}\right)\right]. \tag{70}$$

(1) For a logically reversible transition,

$$\frac{w_\beta}{w_\alpha} = P(\beta|\alpha), \tag{71}$$

and so is independent of the choice of w_α. If the transition is also logically deterministic, $P(\beta|\alpha)=1$ and the logarithmic term is zero. The work requirements are

$$\Delta W_{\alpha,\beta} = (E_\beta - T_R S_\beta) - (E_\alpha - T_R S_\alpha). \tag{72}$$

(2) If the logically reversible transition is *indeterministic*, the work requirement is *reduced* by the quantity $-kT_R \ln[P(\beta|\alpha)]$. If $P(\beta|\alpha)$ is small, this term can be large, even to the extent of making the work requirement negative (i.e., implying work may be *extracted* from the process).

(3) Now consider logically irreversible transitions. When the transition is logically deterministic, w_β is the sum of all the w_α values where the transition is permitted. It is therefore always the case that $w_\beta / w_\alpha \geqslant 1$. This implies an *increased* work requirement compared to a logically reversible, deterministic transition between equivalent (α,β) states.

(4) Finally, logically irreversible, indeterministic transitions may, in principle, take values for w_β / w_α both above and below 1.

Let us consider optimizing the thermodynamic cost of an individual $\alpha \to \beta$ transition. The only free variables are the w_α. For logically reversible transitions, these have no effect and the cost is always

$$\Delta W_{\alpha,\beta} = (E_\beta - T_R S_\beta) - (E_\alpha - T_R S_\alpha) + kT_R \ln[P(\beta|\alpha)],$$

$$\Delta Q_{\alpha,\beta} = T_R\{S_\alpha - S_\beta + k \ln[P(\beta|\alpha)]\}. \tag{73}$$

For logically irreversible transitions, the quantity w_β / w_α should be made as small as possible, subject to the constraint that $\Sigma_\alpha w_\alpha = 1$. From $w_\beta = \Sigma_{\alpha'} w_{\alpha'} P(\beta|\alpha')$ it must be the case that

$$w_\beta \geqslant w_\alpha P(\beta|\alpha). \tag{74}$$

Equality is reached by setting $w_{\alpha'}=0$ for all the input logical states $\alpha' \neq \alpha$ where $P(\beta|\alpha') \neq 0$. This gives $w_\beta = w_\alpha P(\beta|\alpha)$. If the transition is a logically deterministic one, $w_\beta / w_\alpha = 1$, otherwise $w_\beta / w_\alpha < 1$, and the work requirement is reduced (as for a logically reversible, indeterministic transition). The result is similar to that for logically reversible transitions:

$$\Delta W_{\alpha,\beta} \geqslant (E_\beta - T_R S_\beta) - (E_\alpha - T_R S_\alpha) + kT_R \ln[P(\beta|\alpha)],$$

$$\Delta Q_{\alpha,\beta} \geqslant T_R\{S_\alpha - S_\beta + k \ln[P(\beta|\alpha)]\}. \tag{75}$$

2. Expectation values

The problem with optimization for an individual transition is that this can go catastrophically wrong if the operation is performed upon any of the other α' input logical states. For logically irreversible processes, as $w_{\alpha'} \to 0$, then $\Delta W_{\alpha',\beta} \to \infty$.

We need to consider an optimization over the full set of input logical states, rather than with respect to a single input logical state. For the set of all possible transitions, we will seek to minimize the expectation value, or mean cost, of performing the operation.

This is not the only criterion that could be used. One may seek instead, for example, to optimize by a minimax criterion: minimizing the maximum cost that might be incurred. This would lead to a different set of w_α from those we will calculate here. The maximum cost that might be incurred with such a set would, for certainty, be no higher than the maximum cost we will arrive at here. However, the expectation value for the cost, with the different set, would be at least as high as the expectation value we will find.

To be able to calculate an expectation value, a probability distribution over the input logical states is needed. For this we will use the probabilities that go into the calculation of the Shannon information of the input state: $P(\alpha)$. The probability of the transition $\alpha \to \beta$ occurring is then $P(\beta|\alpha)P(\alpha)$ and the expectation value for the work requirement is

$$\langle \Delta W \rangle = \sum_\beta P(\beta)(E_\beta - T_R S_\beta) - \sum_\alpha P(\alpha)(E_\alpha - T_R S_\alpha) + kT_R \sum_{\alpha,\beta} P(\alpha,\beta) \ln\left(\frac{w_\beta}{w_\alpha}\right), \quad (76)$$

where $P(\alpha,\beta) = P(\beta|\alpha)P(\alpha)$ and $P(\beta) = \Sigma_\alpha P(\alpha,\beta)$.

For logically reversible transformations, this is fixed:

$$\frac{w_\beta}{w_\alpha} = P(\beta|\alpha) = \frac{P(\beta)}{P(\alpha)}. \quad (77)$$

For logically irreversible transformations, we must vary the w_α to minimize the function

$$X = \sum_{\alpha,\beta} P(\alpha,\beta) \ln\left(\frac{w_\beta}{w_\alpha}\right). \quad (78)$$

Consider the similar function

$$Y = \sum_\beta P(\beta) \ln P(\beta) - \sum_\alpha P(\alpha) \ln P(\alpha) = \sum_{\alpha,\beta} P(\alpha,\beta) \ln\left(\frac{P(\beta)}{P(\alpha)}\right), \quad (79)$$

$$X - Y = \sum_{\alpha,\beta} P(\alpha,\beta) \ln\left(\frac{w_\beta P(\alpha)}{P(\beta) w_\alpha}\right) = \sum_{\alpha,\beta} P(\alpha,\beta) \ln\left[\frac{P(\alpha,\beta)}{P(\beta) w(\alpha|\beta)}\right] \geqslant 0, \quad (80)$$

where

$$w(\alpha|\beta) = \frac{P(\beta|\alpha) w_\alpha}{w_\beta} \quad (81)$$

and the equality occurs if and only if $P(\alpha,\beta) = P(\beta) w(\alpha|\beta)$. As Y is independant of the values of w_α, then the minimum value of X is precisely the value of Y. This minimum value of X is reached when $w_\alpha = P(\alpha)$, which leads to $w_\beta = P(\beta)$.

The result can easily be reexpressed as

$$\langle \Delta W \rangle \geqslant \sum_\beta P(\beta)\{E_\beta - T_R[S_\beta - k \ln P(\beta)]\} - \sum_\alpha P(\alpha)\{E_\alpha - T_R[S_\alpha - k \ln P(\alpha)]\}. \quad (82)$$

This is the minimum expectation value of the work requirement for the logical operation, using the physical procedure we have described. The same expression holds for logically reversible, irreversible, deterministic, and indeterministic operations. It is not hard to see that this also minimizes the expectation value of the heat generated:

$$\langle \Delta Q \rangle \geqslant -T_R\left(\sum_\beta P(\beta)[S_\beta - k \ln P(\beta)] - \sum_\alpha P(\alpha)[S_\alpha - k \ln P(\alpha)]\right). \quad (83)$$

As was noted for the case of LE in [3], to achieve the optimal physical implementation of a logically irreversible operation requires the physical process to be designed for the particular probability distribution $P(\alpha)$ over the input logical states.[11] A physical implementation optimized for one input probability distribution will not, in general, be optimized for a different input probability distribution. For logically irreversible operations it is only possible to thermodynamically optimize the logical transformation of information (where the input probability is specified). Without a probability distribution (even a default assumption of equiprobable input states) it does not even make sense to talk about optimizing the expectation value for the work or heat requirements, or about the Shannon information of the input and output states.

3. Multiple heat baths

For completeness, we note that, if there are several heat baths available, at different temperatures, the equations may be easily generalized. Defining

$$\overline{\langle \Delta Q \rangle} = \sum_i \langle \Delta Q_i \rangle, \quad (84)$$

[11] It is worth noting that this is not the same as having a prior knowledge of the input logical states. Having prior knowledge of which input state occurs allows one, trivially, to do rather better than this, by choosing $w'_\alpha = 0$ for all other input states. This optimizes for all individual transitions that come from the known α input state, but requires a different physical implementation each time a different input logical state occurs. That different physical implementation is, in each case, equivalent to a logically reversible operation.

$$\bar{T}=\frac{\sum_i \langle \Delta Q_i\rangle}{\sum_i \frac{\langle \Delta Q_i\rangle}{T_i}}, \quad (85)$$

where $\langle \Delta Q_i\rangle$ is the mean heat generated in a heat bath at temperature T_i, we may simply replace T_R with $\bar{T}$ and $\langle \Delta Q\rangle$ with $\overline{\langle \Delta Q\rangle}$, in Eq. (82), and all subsequent equations. In effect, this is equivalent to the possibility of using reversible Carnot cycles to rearrange heat between any heat baths available, in addition to performing the logical operation with a single heat bath.

The introduction of multiple heat baths has little practical significance though. If

$$\sum_\beta P(\beta)[S_\beta - k \ln P(\beta)] - \sum_\alpha P(\alpha)[S_\alpha - k \ln P(\alpha)] < 0, \quad (86)$$

then the least work is required by generating all the heat in the coolest heat bath available. If

$$\sum_\beta P(\beta)[S_\beta - k \ln P(\beta)] - \sum_\alpha P(\alpha)[S_\alpha - k \ln P(\alpha)] > 0 \quad (87)$$

the opposite is true. The least work involves generating heat only in the hottest heat bath.

E. Optimum physical process

We have shown that a particular physical process can implement a logical operation, with a minimum expectation value for the work required or heat generated. Perhaps other physical processes might exist that can perform the same logical operation at a lower cost? We will now prove that *no* physical process can implement the same logical transformation of information at a lower cost.

The initial statistical state of the logical processing apparatus is

$$\rho_I = \sum_\alpha P(\alpha)\rho_\alpha. \quad (88)$$

The final statistical state is

$$\rho_F = \sum_\beta P(\beta)\rho_\beta. \quad (89)$$

We assume that the environment is initially well described by a canonical thermal state $\rho_E(T_R)$, at temperature T_R, and that it is uncorrelated with the initial state of the logical processing system.

Now consider the initial density matrix of the joint system of the logical processing system and the apparatus

$$\rho = \rho_I \otimes \rho_E(T_R), \quad (90)$$

so

$$\mathrm{Tr}[\rho \ln(\rho)] = \mathrm{Tr}[\rho_I \ln(\rho_I)] + \mathrm{Tr}\{\rho_E(T_R)\ln[\rho_E(T_R)]\}. \quad (91)$$

For any unitary evolution upon the combined system to be a physical representation of the logical state, it must evolve the system to some state ρ' such that the marginal distribution of the information processing apparatus is

$$\rho_F = \mathrm{Tr}_E(\rho'). \quad (92)$$

The marginal distribution of the environment is then

$$\rho_E' = \mathrm{Tr}_F(\rho'). \quad (93)$$

From the well-known [14,19,20] properties of unitary evolutions and density matrices,

$$\mathrm{Tr}[\rho \ln(\rho)] = \mathrm{Tr}[\rho' \ln(\rho')], \quad (94)$$

$$\mathrm{Tr}[\rho' \ln(\rho')] \geq \mathrm{Tr}[\rho_F \ln(\rho_F)] + \mathrm{Tr}[\rho_E' \ln(\rho_E')]. \quad (95)$$

As $\rho_E(T_R)$ is a canonical distribution,

$$\frac{H_E}{kT_R} = -\ln[\rho_E(T_R)] - \ln Z, \quad (96)$$

so

$$\mathrm{Tr}\left[\rho_E'\left(\ln(\rho_E') + \frac{H_E}{kT_R}\right)\right] - \mathrm{Tr}\left[\rho_E(T_R)\left(\ln[\rho_E(T_R)] + \frac{H_E}{kT_R}\right)\right] = \mathrm{Tr}(\rho_E'\{\ln(\rho_E') - \ln[\rho_E(T_R)]\}) \geq 0, \quad (97)$$

where H_E is the internal Hamiltonian of the environment. A simple rearrangement gives

$$\mathrm{Tr}[\rho_I \ln(\rho_I)] - \mathrm{Tr}[\rho_F \ln(\rho_F)] \geq \frac{\mathrm{Tr}[H_E\rho_E(T_R)]}{kT_R} - \frac{\mathrm{Tr}(H_E\rho_E')}{kT_R}. \quad (98)$$

As the physical representations of the logical states are nonoverlapping,

$$-k\mathrm{Tr}[\rho_I \ln(\rho_I)] = \sum_\alpha P(\alpha)[S_\alpha - k \ln P(\alpha)], \quad (99)$$

$$-k\mathrm{Tr}[\rho_F \ln(\rho_F)] = \sum_\beta P(\beta)[S_\beta - k \ln P(\beta)]. \quad (100)$$

The expectation value for the work performed upon the system must equal[12] the expectation value for the change in the internal energy of the system plus the expectation value for the change in the internal energy of the environment:

$$\langle \Delta W\rangle = \sum_\beta P(\beta)E_\beta + \mathrm{Tr}(H_E\rho_E') - \sum_\alpha P(\alpha)E_\alpha - \mathrm{Tr}[H_E\rho_E(T_R)]. \quad (101)$$

[12] We assume that the interaction energy between system and environment is negligible at the start and end of the operation. Both this assumption and the assumption that the environment is initially an uncorrelated Gibbs state do not appear to hold in [21,22].

From this we conclude that, for *any* physical process that takes input logical states $\{\alpha\}$ with probabilities $P(\alpha)$ and produces output logical states $\{\beta\}$ with probabilities $P(\beta)$ then the expectation value of the work requirement for this process cannot be less than

$$\langle\Delta W\rangle \geq \sum_\beta P(\beta)\{E_\beta - T_R[S_\beta - k\ln P(\beta)]\} - \sum_\alpha P(\alpha)\{E_\alpha - T_R[S_\alpha - k\ln P(\alpha)]\}. \quad (102)$$

There is no physical process that can do better, in terms of an expectation value for the work requirement, or for the heat generation, than the process developed in Sec. III B.

We emphasize that the relationships we have derived in this section do not depend upon the results of the specific process we examined in the previous section. No assumptions are made regarding the details of the physical process that represents the logical operation, beyond the requirements that it is a unitary evolution of the combined state space of system and environment and does, in fact, faithfully represent the operation. No assumptions are required about the physical representation of the input and output logical states, except those made in Sec. II D. It is not assumed that the environment is an ideal heat bath, is in some thermodynamic limit, or is in thermal equilibrium after the operation. The results require only that the environment be a canonically distributed and uncorrelated system at the start of the operation. Given these assumptions, the result follows: there is no physical representation of the logical operation that has a lower expectation value for the work requirements or heat generation.

IV. GENERALIZED LANDAUER PRINCIPLE

There are several different, but formally equivalent, ways of expressing the generalized Landauer principle (GLP). It will be convenient to use the notation

$$\langle\Delta E\rangle = \sum_\beta P(\beta)E_\beta - \sum_\alpha P(\alpha)E_\alpha,$$

$$\Delta S = \sum_\beta P(\beta)[S_\beta - k\ln P(\beta)] - \sum_\alpha P(\alpha)[S_\alpha - k\ln P(\alpha)],$$

$$\Delta H = -\sum_\beta P(\beta)\log_2 P(\beta) + \sum_\alpha P(\alpha)\log_2 P(\alpha) \quad (103)$$

for: the change in the expectation value for the internal energy of the information processing apparatus; the change in the Gibbs–von Neumann entropy of the statistical ensemble describing the information processing system; and the change in the Shannon information of the logical states over the course of the operation.

A. Work requirements

GLP1: Work. A logical transformation of information has a minimal expectation value for the work requirement given by

$$\langle\Delta W\rangle \geq \langle\Delta E\rangle - T_R\Delta S. \quad (104)$$

B. Heat generation

We note that

$$\langle\Delta Q\rangle = \langle\Delta W\rangle - \langle\Delta E\rangle \quad (105)$$

is equal to the expectation value of the heat generated in the heat bath.

GLP2: Heat. A logical transformation of information has a minimal expectation value for the heat generated in the environment of

$$\langle\Delta Q\rangle \geq -T_R\Delta S. \quad (106)$$

It is important to remember that the term ΔS appearing in GLP1 and GLP2 is *not* the change in Shannon information ΔH between the input and output states. It is the change in the Gibbs–von Neumann entropy of the logical system, taking into account any changes in the entropies of the subensembles that represent the input and output logical states. It can be related to the change in the Shannon information by

$$\Delta S = \sum_\beta P(\beta)S_\beta - \sum_\alpha P(\alpha)S_\alpha + k\Delta H\ln 2. \quad (107)$$

C. Entropic cost

The change in the Gibbs–von Neumann entropy of the environmental heat bath is given by

$$\Delta S_{\mathrm{HB}} = -k\,\mathrm{Tr}[\rho_E'\ln(\rho_E')] + k\mathrm{Tr}\{\rho_E(T_R)\ln[\rho_E(T_R)]\}, \quad (108)$$

which gives the entropic form of the generalized Landauer principle.

GLP3: Entropy. A logical transformation of information requires a minimal change in the Gibbs–von Neumann entropies of the marginal statistical states of an information processing apparatus ΔS and its environment ΔS_{HB} of

$$\Delta S_{\mathrm{HB}} + \Delta S \geq 0. \quad (109)$$

This is a trivial consequence of the requirements that the evolution be unitary and that the statistical states of the logical processing system and the environment be initially uncorrelated. The expectation value of the heat generated in the environment is at least equal to the increase in the Gibbs–von Neumann entropy of the marginal state of the heat bath:

$$\langle\Delta Q\rangle \geq T_R\Delta S_{\mathrm{HB}}. \quad (110)$$

This allows us to deduce GLP1 or GLP2 from GLP3, but not the reverse.[13]

D. Information

If we *define* the term

[13]In the limiting case of an *ideal* heat bath and quasistatic processes, the equality is reached and the deduction can then go in both directions.

$$\Delta S_L = \sum_\beta P(\beta)S_\beta - \sum_\alpha P(\alpha)S_\alpha, \tag{111}$$

we get

$$\Delta S_{\rm HB} + \Delta S_L \geq -k\Delta H \ln 2. \tag{112}$$

This expression seems suggestive. If we regard the terms $\Delta S_{\rm HB}$ and ΔS_L as changes in the entropies of the non-information-bearing degrees of freedom of the environment and the apparatus, respectively, then we appear to have provided a quantitative version of Bennet's statement that "any logically irreversible manipulation of information [ΔH]… must be accompanied by a corresponding [$k \ln 2$] entropy increase in the non-information-bearing degrees of freedom of the information processing apparatus [ΔS_L] or its environment [$\Delta S_{\rm HB}$]" [1] although, unlike Bennett, we do not restrict this to irreversible transformations of data.

This produces what may be taken as the information form of the GLP.

GLP4: Information. A logical transformation of information requires an increase of entropy of the non-information-bearing degrees of freedom of the information processing apparatus and its environment of at least $-k \ln 2$ times the change in the total quantity of Shannon information over the course of the operation:

$$\Delta S_{\rm NIBDF} \geq -k\Delta H \ln 2, \tag{113}$$

where $\Delta S_{\rm NIBDF}=\Delta S_{\rm HB}+\Delta S_L$. This is quite generally true and follows directly from GLP3 and the definition of ΔS_L.

V. MODELS OF COMPUTING

We will now discuss some of the consequences that can be drawn from the generalized Landauer principle by varying the thermodynamic properties of the input and output states. This allows us to consider the effects of having different energies and entropies for the physical states that embody the logical states and has some surprising consequences.

A. Uniform computing

Assumption 1. When we make the assumption that the computation takes place at the same temperature throughout, such that

$$\forall \quad \alpha,\beta \quad T_R = T_\alpha = T_\beta, \tag{114}$$

then we shall call this *isothermal* computing.

In the most commonly encountered set of assumptions for the thermodynamics of computation, we have, in addition to the assumption of isothermal computing, the assumption of uniform states.

Assumption 2. The physical states that represent the logical states all have the same entropy and mean energy, so that

$$\forall \quad \alpha,\beta \quad E_R = E_\alpha = E_\beta,$$

$$\forall \quad \alpha,\beta \quad S_R = S_\alpha = S_\beta. \tag{115}$$

This reduces the generalized Landauer principle to the form of

$$\langle \Delta W\rangle \geq -kT_R\Delta H \ln 2,$$

$$\langle \Delta Q\rangle \geq -kT_R\Delta H \ln 2, \tag{116}$$

where ΔH is the change in Shannon information over the course of the transformation. This is the usual form in which Landauer's principle is encountered.

The necessary and sufficient conditions for Eqs. (116) to hold is a weaker condition

Assumption 3. Uniform computing:

$$\sum_\alpha P(\alpha)E_\alpha = \sum_\beta P(\beta)E_\beta,$$

$$\sum_\alpha P(\alpha)S_\alpha = \sum_\beta P(\beta)S_\beta. \tag{117}$$

B. Equilibrium computing

The simplifying assumption of uniform computing is made so universally that it might be questioned whether there is any value to considering nonuniform computing. To answer this, consider the following assumption.

Assumption 4. In equilibrium computing, the input and output states are constructed to be canonical thermal systems, at temperature T_R, with the properties

$$E_\alpha - T_R S_\alpha + kT_R \ln P(\alpha) = C_A,$$

$$E_\beta - T_R S_\beta + kT_R \ln P(\beta) = C_A, \tag{118}$$

where C_A is a constant related to the overall size of the logical processing apparatus. This yields the relationships

$$\langle \Delta E\rangle - T_R\Delta S = 0, \tag{119}$$

and reduces the generalized Landauer principle to

$$\langle \Delta W\rangle \geq 0, \tag{120}$$

although

$$\langle \Delta Q\rangle \geq -T_R\Delta S \tag{121}$$

still. The equality can, of course, only be reached in the limit of slow processes.

Assumption 5. The necessary and sufficient assumption for Eq. (120) to hold is that of zero mean work,

$$\sum_\alpha P(\alpha)\{E_\alpha - T_R[S_\alpha + k \ln P(\alpha)]\}$$

$$= \sum_\beta P(\beta)\{E_\beta - T_R[S_\beta + k \ln P(\beta)]\} \tag{122}$$

Assumption 5 implies only that the average work requirement can approach zero, over all the possible transitions between logical states. Assumption 4 ensures that there is a zero mean work requirement $\Delta W_{\alpha,\beta}$ for all individual (α,β) transitions.

C. Adiabatic computing

Assumption 6. To eliminate mean heat generation in the ideal limit, the necessary and sufficient condition is that of zero mean heat generation,

$$\sum_{\alpha} P(\alpha)[S_\alpha + k \ln P(\alpha)] = \sum_{\beta} P(\beta)[S_\beta + k \ln P(\beta)], \quad (123)$$

leading to

$$\langle \Delta Q \rangle \geq 0, \quad (124)$$

although this does not eliminate mean work requirements

$$\langle \Delta W \rangle = \langle \Delta E \rangle. \quad (125)$$

Again, this is only the expectation value over all transitions. To ensure that the mean heat generated $\Delta Q_{\alpha,\beta}$ is zero for each individual (α,β) transition, requires a further assumption

Assumption 7. Adiabatic computing is represented by

$$S_\alpha + k \ln P(\alpha) = C_B,$$

$$S_\beta + k \ln P(\beta) = C_B, \quad (126)$$

where C_B is an apparatus-related constant.

D. Adiabatic equilibrium computing

Assumption 8. Combining the assumptions of adiabatic and equilibrium computing gives the requirement of adiabatic equilibrium computing,

$$E_\alpha = E_\beta = E_R,$$

$$S_\alpha + k \ln P(\alpha) = C_C,$$

$$S_\beta + k \ln P(\beta) = C_C, \quad (127)$$

which yields, $\forall\ \alpha,\beta$

$$\Delta W_{\alpha,\beta} \geq 0,$$

$$\Delta Q_{\alpha,\beta} \geq 0, \quad (128)$$

with equality being reachable as a limiting case, and C_C again a machine-dependent constant.

This result may seem surprising. It suggests that it is possible to design a computer to perform any combination of logical operations, with no exchange of heat with the environment and requires no work to be performed upon it. This must be as true for logically irreversible operations as for logically reversible operations, and as true for logically indeterministic operations as for logically deterministic operations.

To understand this better, let us consider what happens in adiabatic equilibrium computing. We can use the square well potential as the physical model of the logical states, as the internal energy of these states is $\frac{1}{2}kT$. Varying the width of the square well potential for each input and output logical state satisfies the remaining conditions.

Implementing the model of adiabatic equilibrium computing on the processes of Sec. III simplifies the procedure significantly.

(1) There is no need to resize the input states, as these will already be canonically distributed. Steps 1–3 are redundant.

(2) Potential barriers are inserted into the α states, corresponding to the conditional probabilities $P(\beta|\alpha)$, as in step 4.

(3) The separate portions of the β output states are brought into adjacent positions as in step 5.

(4) The potential barriers within each β output state are removed, as in step 6.

(5) These output states are *already* canonically distributed. There is therefore no need for a resizing of the output states and steps 7–9 are unnecessary.

None of these stages requires any work to be performed upon the system or exchange of heat with the environment. The computation is reduced to a process of rearranging a canonical ensemble from one set of canonically distributed orthogonal subensembles into a different set of canonically distributed orthogonal subensembles, in accordance with the computational probabilities $P(\beta|\alpha)$.

As the probabilities of the different output states cannot change between logical operations[14] then the canonically distributed output states can be used as canonically distributed input states to any new logical operation. This thermodynamic model may therefore proceed indefinitely without generating any heat or requiring any work.

Before leaving this subject, let us just note one feature of equilibrium computing. Logically deterministic, irreversible computations are able to avoid generating heat, in this model, by increasing the size of the physical states representing the logical states. This does *not* mean that the logical processing apparatus itself needs to be increasing in size. Although the size of the individual states has increased, the number of logical states has decreased (by the definition of a logically deterministic, irreversible computation!). Whenever the equality in assumption 6 holds, the two effects cancel out and the overall size of the logical processing apparatus can remain constant.

VI. THERMODYNAMIC REVERSIBILITY

We have not yet examined the question of whether these operations are thermodynamically reversible. This is a subtle question and depends upon what one takes to be the statistical mechanical generalization of thermodynamic entropy and thermodynamic reversibility. We will first discuss how this appears from the perspective of three different approaches to entropy, and then from a definition based on thermodynamic cycles that is not directly based upon any definition of entropy.

It is worth remembering that a net increase in entropy is considered to be taken as a sign of *irreversibility* because net decreases in entropy cannot occur (or are unlikely). An "entropy" that can be *systematically* decreased may be a useful indicator of some properties, but its increase cannot automatically be regarded as an indicator of irreversibility, whether thermodynamic or of some other kind.

We will consider three possible conditions for thermodynamic reversibility and irreversibility.

[14]By definition anything that changes the probabilities of a state must be a logical transformation of the data.

(1) The thermodynamic entropy is the entropy of the individual state. If the system is in logical state α, then the thermodynamic entropy is S_α. The net entropy change for a particular logical transition, from logical state α to logical state β is

$$S_\beta - S_\alpha + \frac{\Delta Q_{\alpha,\beta}}{T_R}. \tag{129}$$

A transition is thermodynamically reversible if the decrease in individual state entropy from the input to output logical states is equal to the heat generated in the heat bath, divided by the temperature of the heat bath. A transition is thermodynamically irreversible if the decrease in individual state entropy is less than this. Decreases in individual state entropy greater than this cannot occur.

(2) The thermodynamic entropy is the entropy of the individual state, but is only nondecreasing *on average*. If the system is in logical state α, then the thermodynamic entropy is S_α, but this may decrease provided it does not decrease on average. The average change is

$$\sum_\beta P(\beta)S_\beta - \sum_\alpha P(\alpha)S_\alpha + \frac{\langle \Delta Q_{\alpha,\beta}\rangle}{T_R}. \tag{130}$$

A logical transformation of information is thermodynamically reversible if the average decrease in individual state entropy over all the transitions from input to output logical states is equal to the average heat generated in the heat bath, divided by the temperature of the heat bath. The transformation is thermodynamically irreversible if the average decrease in individual state entropy is less than this. Average decreases in individual state entropy greater than this cannot occur.

(3) The thermodynamic entropy is the Gibbs–von Neumann entropy of the *marginal* statistical states. If the statistical state of the system is $\rho=\Sigma_\alpha P(\alpha)\rho_\alpha$, the thermodynamic entropy is $-k\,\mathrm{Tr}[\rho\ln(\rho)]$. A logical transformation of information is thermodynamically reversible if the decrease in Gibbs–von Neumann entropy from the input to output statistical states is equal to the average heat generated in the heat bath, divided by the temperature of the heat bath. The transformation is thermodynamically irreversible if the decrease in Gibbs–von Neumann entropy is less than this. Decreases in Gibbs–von Neumann entropy greater than this cannot occur.

The first two conditions imply thermodynamic irreversibility for logically deterministic, irreversible operations. Unfortunately, it will be shown that neither condition can consistently account for logically indeterministic operations, which can systematically decrease the relevant entropy measure by quantities greater than should be permitted.

The third condition gives an entropy that is consistently nondecreasing (provided there are no spontaneous or preexisting correlations with heat baths). Logically indeterministic operations do not decrease this entropy. On the other hand, logically irreversible operations no longer necessarily increase this entropy measure either. According to the Gibbs–von Neumann measure, all logical operations may be implemented in a thermodynamically reversible manner.

We will be making the standard assumptions that all processes can take place with ideal heat baths and sufficiently slowly that equalities are reached as the limiting cases. Without these assumptions no process can be thermodynamically reversible. We will therefore replace the appropriate inequalities with equalities.

A. Individual logical state entropy

The net individual state entropy change, for a particular logical transition, gives

$$S_\beta - S_\alpha + \frac{\Delta Q_{\alpha,\beta}}{T_R} \geqslant k\ln[P(\beta|\alpha)]. \tag{131}$$

Allowed logically deterministic transitions require $P(\beta|\alpha)=1$. The equality is automatically reached for logically deterministic, reversible transitions, and which are therefore thermodynamically reversible. For logically deterministic, irreversible transitions, the equality requires $w_\alpha=1$. This is possible only if no other input logical states are allowed. Such an operation would be trivially logically reversible as there is only one permissible input logical state. So, according to this entropy measure, logically deterministic irreversible transitions must be thermodynamically irreversible.

As $\ln[P(\beta|\alpha)]\leqslant 0$ it is possible that

$$S_\beta - S_\alpha + \frac{\Delta Q_{\alpha,\beta}}{T_R} < 0. \tag{132}$$

This gives a net decrease in individual state entropy. For this to happen, the transition must be logically indeterministic. Optimally implemented, logically indeterministic, reversible transitions will *always* decrease individual state entropy.

If an entropy increase is indicative of thermodynamic irreversibility because entropy decreases are impossible, this measure of entropy cannot be seen as a good indicator of thermodynamic irreversibility. Any apparent irreversibility can actually be reversed.

B. Average state entropy

In statistical mechanics fluctuations occur. Perhaps the demand for a *strictly* nondecreasing entropy might be the problem. What of the *average* change in entropy? Does this give a good indicator of thermodynamic irreversibility?

This gives

$$\sum_\beta P(\beta)S_\beta - \sum_\alpha P(\alpha)S_\alpha + \frac{\langle \Delta Q_{\alpha,\beta}\rangle}{T_R} = -k\Delta H\ln 2. \tag{133}$$

What we have here is the ideal limit case of GLP4, with $\Delta S_{\mathrm{NIBDF}}$ now representing the average change in entropy:

$$\Delta S_{\mathrm{NIBDF}} = \sum_\beta P(\beta)S_\beta - \sum_\alpha P(\alpha)S_\alpha + \frac{\langle \Delta Q_{\alpha,\beta}\rangle}{T_R}. \tag{134}$$

Logically deterministic, irreversible operations have $\Delta H < 0$, so

$$\Delta S_{\mathrm{NIBDF}} > 0, \tag{135}$$

and the net mean change in individual state entropy of the system and environment is strictly increasing. Again, logically deterministic, irreversible operations must be, on average, individual state entropy increasing.

The problem with the argument should be immediately apparent: for logically reversible, indeterministic operations $\Delta H > 0$ and by the same reasoning and arguments it is possible that

$$\Delta S_{\mathrm{NIBDF}} = -k\Delta H \ln 2 < 0. \tag{136}$$

Not only can logically indeterministic operations reduce individual state entropy on individual transitions, they can even reduce this entropy *on average*.

C. Gibbs–von Neumann entropy

The entropy measure which includes the effects of the statistical mixture over the states, the Gibbs–von Neumann entropy over the ensemble, gives the initial entropy of the logical processing system,

$$S_I = -k\,\mathrm{Tr}[\rho_I \ln(\rho_I)], \tag{137}$$

where

$$\rho_I = \sum_\alpha P(\alpha)\rho_\alpha, \tag{138}$$

and the final entropy

$$S_F = -k\,\mathrm{Tr}[\rho_F \ln(\rho_F)], \tag{139}$$

where

$$\rho_F = \sum_\beta P(\beta)\rho_\beta. \tag{140}$$

In [13] it is argued that the Gibbs–von Neumann entropy is indeed the correct statistical mechanical generalization of thermodynamic entropy, although this identification has not been assumed anywhere within this paper.[15]

When we consider the Gibbs–von Neumann entropy, the most appropriate form of the GLP is GLP3. In this case, the limiting behavior gives

$$\Delta S_{\mathrm{HB}} + \Delta S = 0. \tag{141}$$

As any logical operation may reach this limit, the Gibbs–von Neumann entropy regards all logical operations as being possible in a thermodynamically reversible manner.

D. Discussion

As some of these results may seem surprising or counterintuitive, and appear to contradict widely stated expressions of the implications of the thermodynamics of logically irreversible operations, let us examine them in more detail.

1. The RLE-LE cycle

First, let us take the examples of the logically deterministic, irreversible, reset to zero (RTZ) operation and the logically reversible, indeterministic unset from zero (UFZ) operation (see the Appendix). If the argument is accepted that the optimal procedure to implement RTZ is entropy increasing, then it must also be accepted that the optimal procedure for UFZ can be entropy *decreasing*.

That this must be the case can be seen by considering the reverse Landauer erasure (RLE) operation immediately followed by the Landauer erasure (LE) operation. If these two procedures are matched in terms of the probabilities and input and output states, then the result is to leave both the logical system and the environment in their initial states. The total entropy must be the same at the end of such a procedure, as at the start, and it follows if it *increases* during LE, then it *must* decrease during RLE.

As a simple example, using the assumptions of uniform computing, and an initial input state of 0, the process of RLE extracts $kT \ln 2$ heat from the environment, and converts it into work. The output state of RLE is an equiprobable distribution of logical states 0 and 1, each of which has the same entropy as the initial 0 state.

This is input to the LE procedure, which requires $kT \ln 2$ heat to be generated in the environment and leaves the output state as 0. The system and environment are left in the same logical and thermodynamic states as at the beginning of the process. There is a zero net work requirement and a zero net heat generation. The combination of RLE followed by LE is clearly a thermodynamically reversible cycle.

It follows that the net change in entropy over the course of the two operations must be zero for both system and environment. To argue that the net change in entropy for the LE procedure is $k \ln 2$, requires, for the overall change in entropy to be zero, the change in entropy during the RLE operation to be $-k \ln 2$.

Both the individual state entropies and the average state entropy do indeed decrease by $k \ln 2$ during the RLE operation. The Gibbs–von Neumann entropy remains constant, as the mixing entropy increases by $k \ln 2$ to compensate. During the course of the LE operation, the individual state and average state entropies increase by $k \ln 2$. In the conventional operation of the LE process, this is associated with heat generated in the environment, and is often considered to be the source of an irreversible entropy increase. However, we can clearly see that from the point of view of the Gibbs–von Neumann entropy, there is a compensating reduction of $k \ln 2$ associated with the reduction in the mixing entropy.

2. Uniform computing

We can easily generalize this to situations where the quantity of information erased is less than 1 bit,[16] and in doing so will see more clearly the need to optimize the operation to

[15] We have calculated the Gibbs–von Neumann entropies for *individual* states, in canonical distributions, but even here the calculation of the mean work requirements and mean heat generated did not depend upon any identification of this as a *thermodynamic* entropy.

[16] This cycle was detailed in [3].

the probability distribution. We simply need to implement an UFZ(p) operation, followed by a RTZ(p) operation.

We start with a standard atom in a box, and the partition divides the box exactly in half. The atom is on the left-hand side (which represents logical state 0) with certainty.

a. RLE(p). The RLE(p) operation consists of the following steps.

(1) Isothermally move the partition to the right hand side, extracting $W_1=kT\ln 2$ heat as work.

(2) Insert the partition at location $x=pL$ in the box, where the width of the box is L. The atom is, with probability p, on the left-hand side of the partition.

(3) Isothermally move the partition to the center of the box ($x=\frac{1}{2}L$). If the atom is on the left-hand side, the work requirement is $kT\ln(2p)$ while if it is on the right, the work requirement is $kT\ln[2(1-p)]$. The mean work required in this stage is

$$W_2 = kT[p\ln p + (1-p)\ln(1-p) + \ln 2] \tag{142}$$

so the net work for the operation is

$$W_1 + W_2 = kT[p\ln p + (1-p)\ln(1-p)], \tag{143}$$

which is negative, representing a net extraction of work.

Now, we find that the individual state entropy and average state entropy remain the same as at the start of the operation, despite the fact that $kT[p\ln p+(1-p)\ln(1-p)]$ work has been extracted from the heat bath. From the point of view of the Gibbs–von Neumann entropy, this is compensated by the increase in mixing entropy between the two logical states.

b. LE(p). If we follow this with an LE(p) operation, we have the following steps.

(1) Isothermally move the partition to the position $x=pL$. If the atom is on the left-hand side, the work requirement is $-kT\ln(2p)$ while if it is on the right, the work requirement is $-kT\ln[2(1-p)]$. The mean work required in this stage is

$$W_3 = -kT[p\ln p + (1-p)\ln(1-p) + \ln 2]. \tag{144}$$

(2) Remove the partition from the box.

(3) Insert the partition in the right-hand side of the box and isothermally move it to the center. This requires $W_4 = kT\ln 2$ work, so the net work is

$$W_3 + W_4 = -kT[p\ln p + (1-p)\ln(1-p)]. \tag{145}$$

Again, both the individual and average state entropy are unchanged, while work is converted to heat in the environment. The Gibbs–von Neumann entropy, however, shows a compensating decrease in mixing entropy.

The net work and net heat generated, over the course of the cycle, is zero:

$$W_1 + W_2 + W_3 + W_4 = 0. \tag{146}$$

If the heat generated in the environment during the LE(p) operation is an indicator of an irreversible entropy increase, we have to explain a corresponding systematic *reduction* in entropy during the RLE(p) operation. As we noted, entropy increases are associated with irreversibility precisely because corresponding systematic entropy decreases are supposed to be impossible.

c. LE(p'). Let us now consider following the RLE(p) operation with LE(p'), where the erasure operation has been optimized for a different probability distribution.

(1) Isothermally move the partition to the position $x=p'L$. If the atom is on the left-hand side, the work requirement is $-kT\ln(2p')$, while if it is on the right, the work requirement is $-kT\ln[2(1-p')]$. The mean work required in this stage is

$$W_5 = -kT[p\ln p' + (1-p)\ln(1-p') + \ln 2]. \tag{147}$$

(2) Remove the partition from the box.

(3) Insert the partition in the right-hand side of the box and isothermally move it to the center. This requires $W_6 = kT\ln 2$ work, so the net work is

$$W_5 + W_6 = -kT[p\ln p' + (1-p)\ln(1-p')]. \tag{148}$$

The net work required over the RLE(p)−LE(p') cycle is

$$W_1 + W_2 + W_5 + W_6 = kT\left[p\ln\left(\frac{p}{p'}\right) + (1-p)\ln\left(\frac{1-p}{1-p'}\right)\right] \geq 0, \tag{149}$$

with equality occurring if, and only if, $p=p'$.

Once again, both the individual and average state entropy are unchanged. In this case, however, the cycle generates a net heat in the environment, unless $p=p'$. This cycle is, in general, thermodynamically irreversible.

From the point of view of the Gibbs–von Neumann entropy, it is the removal of the partition from the location $x=p'L$, when the probability is p, that is associated with an uncompensated entropy increase. We can see this by noting that, if we reinsert the partition at $x=p'L$, we do not recover the previous statistical state, as the probability of the atom being on the left-hand side would then be p'. To recover the statistical state we need to reinsert the partition at $x=pL$ and then move it isothermally to $x=p'L$. This isothermal movement of the partition requires, on average,

$$kT\left[p\ln\left(\frac{p}{p'}\right) + (1-p)\ln\left(\frac{1-p}{1-p'}\right)\right] \geq 0$$

work to be performed.

Even so, let us note that had LE(p') in fact followed the RLE(p') operation, it would have been thermodynamically reversible. The physical process involved in performing the LE(p) [or LE(p')] operation cannot be said to be intrinsically thermodynamically reversible (or irreversible) in itself. Whether it is thermodynamically reversible or not depends upon the statistical state upon which it acts.

3. Adiabatic equilibrium computing

Let us look at the same logical cycle, but with a different computing model: adiabatic equilibrium. Again we start with a standard atom in a box. As the atom is in logical state 0 with certainty, the conditions of Eq. (127) require that logical state 0 occupies the entire box.

a. RLE(p). The RLE(p) operation now consists of a single step:

(1) Insert the partition at location $x=pL$ in the box, where the width of the box is L. The atom is, with probability p, on the left-hand side of the partition.

No work is required or heat generated. The individual and average state entropies have decreased, with the average state entropy decreasing by $k[p\ln p+(1-p)\ln(1-p)]$. The Gibbs–von Neumann entropy remains the same, as the mixing entropy compensates for this.

b. LE(p). If we follow this with an LE(p) operation, we have the following step.

(1) Remove the partition from the box.

Both the individual and average state entropies are increased, with the average state entropy increasing by $k[p\ln p+(1-p)\ln(1-p)]$. The Gibbs–von Neumann entropy, however, shows a compensating decrease in mixing entropy.

We see how, in the case of adiabatic equilibrium computing, the generation of heat in the environment is replaced by changes in the entropies of the individual states (or, as [1] refers to it, the non-information-bearing degrees of freedom of the apparatus). Although there is an increase in such entropies during the LE(p) process, there is an exactly equivalent decrease during the RLE(p) process. Again, if we take the increase during LE(p) to be indicative of a thermodynamic irreversibility, we are left with the challenge of accounting for the systematic decrease during the RLE(p) operation.

c. LE(p'). Following RLE(p) with an LE(p') operation under the assumptions of adiabatic equilibrium does not entirely make sense, as adiabatic equilibrium requires the physical representation of the logical states to be tailored to the probability of the state occurring. However, we may consider the optimum implementation of RTZ(p'), on the assumption that the probability of the logical state 0 is p', with the partition initially located at $x=pL$ and the process leaving the system in a state compatible with adiabatic equilibrium computation.

(1) Isothermally move the partition to the position $x=p'L$. If the atom is on the left-hand side, the work requirement is $-kT\ln(p'/p)$ while if it is on the right, the work requirement is $-kT\ln[(1-p')/(1-p)]$. The mean work required[17] in this stage is

$$kT\left[p\ln\left(\frac{p}{p'}\right)+(1-p)\ln\left(\frac{1-p}{1-p'}\right)\right]\geqslant 0 \tag{151}$$

with equality occurring if, and only if, $p=p'$.

(2) Remove the partition from the box.

The net work required over the RLE(p)$-LE(p')$ cycle is again

$$kT\left[p\ln\left(\frac{p}{p'}\right)+(1-p)\ln\left(\frac{1-p}{1-p'}\right)\right]\geqslant 0. \tag{152}$$

Once again, from the point of view of the Gibbs–von Neumann entropy, it is the removal of the partition from the location $x=p'L$, when the probability is p, that is associated with an uncompensated entropy increase.

4. Generic logical operations

Now let us consider a generic logical transformation of information. Start with input logical states α, physically represented by states with energies and entropies E_α and S_α, and define a logical operation by the transition probabilities $P(\beta|\alpha)$ to the output logical states β with physical state energies and entropies E_β and S_β.

To thermodynamically optimize the physical process, we need a probability distribution $P(\alpha)$. The β output states will then occur with probabilities

$$P(\beta)=\sum_\alpha P(\beta|\alpha)P(\alpha). \tag{153}$$

Writing

$$\rho_I=\sum_\alpha P(\alpha)\rho_\alpha,$$

$$\rho_F=\sum_\beta P(\beta)\rho_\beta,$$

then the optimal thermodynamic cost of this is

$$\begin{aligned}\langle\Delta W\rangle &= [\mathrm{Tr}(H\rho_I)-T_RS(\rho_I)]-[\mathrm{Tr}(H\rho_F)-T_RS(\rho_F)]\\ &=\sum_\beta P(\beta)(E_\beta-T_RS_\beta)-\sum_\alpha P(\alpha)(E_\alpha-T_RS_\alpha)\\ &\quad+kT_R\sum_{\alpha,\beta}P(\alpha,\beta)\ln\left(\frac{P(\beta)}{P(\alpha)}\right),\end{aligned}$$

$$\langle\Delta Q\rangle=-T_RS(\rho_I)+T_RS(\rho_F). \tag{154}$$

We can now define a physical process that acts upon the physical states $\{\beta\}$, and evolves them into the physical states $\{\alpha\}$, with probabilities[18] given by

[17] Note that, had the probability of logical state 0 *actually* been p', the work required would have been

$$kT\left[p'\ln\left(\frac{p}{p'}\right)+(1-p')\ln\left(\frac{1-p}{1-p'}\right)\right]\leqslant 0, \tag{150}$$

so work would have been extracted in the process.

[18] For logical operations taking as input states $\{\beta\}$ and producing output states $\{\alpha\}$, we will use the notation Π for the corresponding probabilities.

$$\Pi(\alpha|\beta) = \frac{P(\beta|\alpha)P(\alpha)}{\sum_\alpha P(\beta|\alpha)P(\alpha)}. \quad (155)$$

It is straightforward to see that if this acts upon states $\{\beta\}$, occurring with probabilities $P(\beta)$, then it produces the states $\{\alpha\}$ with probabilities $P(\alpha)$. If the physical process is optimized for these probabilities, then the thermodynamic cost is

$$\langle \Delta W_\Pi \rangle = -\langle \Delta W \rangle,$$

$$\langle \Delta Q_\Pi \rangle = -\langle \Delta Q \rangle. \quad (156)$$

So for any logical transformation of information, optimally implemented, there exists a second operation, which when optimally implemented restores the original statistical state, and for which the total expectation value of the work requirement and the total expectation value of the energy generated in the environment are zero. This is true regardless of whether the original operation is logical reversible, irreversible, deterministic, or indeterministic.

As we have noted before, however, to achieve this optimum for logically irreversible operations, the physical process must take into account the probability distribution $P(\alpha)$ over the input logical states. One cannot create a physical process that implements a logically irreversible operation which will be thermodynamically optimal for every probability distribution over the input logical states. This differs from logically reversible operations, which may be represented by a physical process that is thermodynamically optimal for any probability distribution over the input logical states.

We will now look at the effect of an operation that is not optimized for the right set of probabilities. Suppose we have an operation with the same transition probabilities $\Pi(\alpha|\beta)$ above, but where the physical process has been optimized for the input probability distribution $\Pi(\beta)$. The output states are expected to occur with probabilities

$$\Pi(\alpha) = \sum_\beta \Pi(\alpha|\beta)\Pi(\beta) \quad (157)$$

and the expected thermodynamic cost, from Eq. (76), is

$$\langle \Delta W_\Pi \rangle = \sum_\alpha \Pi(\alpha)(E_\alpha - T_R S_\alpha) - \sum_\beta \Pi(\beta)(E_\beta - T_R S_\beta) + kT_R \sum_{\alpha,\beta} \Pi(\alpha,\beta) \ln\left(\frac{w_\alpha}{w_\beta}\right) \quad (158)$$

with $w_\alpha = \Pi(\alpha)$, $w_\beta = \Pi(\beta)$, and $\Pi(\alpha,\beta) = \Pi(\alpha|\beta)\Pi(\beta)$.

The input states do not occur with $\Pi(\beta)$ but with $P(\beta)$. The actual thermodynamic cost incurred is

$$\langle \Delta W'_\Pi \rangle = \sum_\alpha P(\alpha)(E_\alpha - T_R S_\alpha) - \sum_\beta P(\beta)(E_\beta - T_R S_\beta) + kT_R \sum_{\alpha,\beta} P(\alpha,\beta) \ln\left(\frac{w_\alpha}{w_\beta}\right). \quad (159)$$

The combined cycle now has a cost

$$\langle \Delta W'_\Pi \rangle + \langle \Delta W \rangle = kT_R \sum_{\alpha,\beta} P(\alpha,\beta) \ln\left(\frac{w_\alpha P(\beta)}{w_\beta P(\alpha)}\right), \quad (160)$$

which can be rearranged to give

$$\langle \Delta W'_\Pi \rangle + \langle \Delta W \rangle = kT_R \sum_{\alpha,\beta} P(\alpha,\beta) \ln\left(\frac{P(\alpha,\beta)}{\Pi(\beta|\alpha)P(\alpha)}\right) \geqslant 0, \quad (161)$$

where $\Pi(\beta|\alpha)\Pi(\alpha) = \Pi(\alpha|\beta)\Pi(\beta)$.

Equality can occur in two ways. First, and most simply, if $\Pi(\beta) = P(\beta)$. The input states to the $\Pi(\alpha|\beta)$ operation occur with the optimal probabilities.

Second, if the second operation is a logically reversible operation, then

$$\forall \ \beta \quad [\Pi(\alpha|\beta) \neq 0 \Rightarrow \forall \ \ \alpha' \neq \alpha \Pi(\alpha'|\beta) = 0]. \quad (162)$$

As $\Pi(\alpha|\beta) = P(\alpha|\beta)$, it follows that the first operation must have been logically deterministic:

$$\forall \ \beta \quad [P(\alpha|\beta) \neq 0 \Rightarrow \forall \ \ \alpha' \neq \alpha P(\alpha'|\beta) = 0]. \quad (163)$$

Together this means that

$$P(\alpha|\beta) = \frac{P(\alpha)}{P(\beta)} = \Pi(\alpha|\beta) = \frac{w_\alpha}{w_\beta} \quad (164)$$

and $\langle \Delta W'_\Pi \rangle + \langle \Delta W \rangle = 0$, regardless of the value of w_β. This shows, once more, that logically reversible operations may be thermodynamically optimized without reference to the probability distribution over their input states.

A corollary to this is worth noting. While the second logical operation, if logically reversible, may be implemented and optimized without reference to the probability distribution over the input states, its very definition depends upon the probability distribution over the input states of the first operation. The first operation is defined by the set of transition probabilities $\{P(\beta|\alpha)\}$, while the second is *defined* by

$$\Pi(\alpha|\beta) = P(\alpha|\beta) = \frac{P(\beta|\alpha)P(\alpha)}{\sum_\alpha P(\beta|\alpha)P(\alpha)}. \quad (165)$$

There is, in general, only one way to make this independent of $\{P(\alpha)\}$: if the first operation is logically reversible, then $P(\alpha|\beta) \in \{0,1\}$. The second operation is now logically deterministic and $\Pi(\alpha|\beta) \in \{0,1\}$ does not require the $\{P(\alpha)\}$.

We can summarize this, as follows. If an operation $\{P(\beta|\alpha)\}$ is logically reversible, then it is possible to calculate a (logically deterministic) reverse operation $\{\Pi(\alpha|\beta)\}$, independently of the first input probability distribution $\{P(\alpha)\}$. However, if $\{P(\beta|\alpha)\}$ is logically indeterministic, then optimizing the reverse operation requires the output probability distribution $\{P(\beta)\}$.

Conversely, if an operation $\{P(\beta|\alpha)\}$ is logically deterministic, then it is possible to thermodynamically optimize a (logically reversible) reverse operation $\{\Pi(\alpha|\beta)\}$, independently of the first output probability distribution $\{P(\beta)\}$.

However, if $\{P(\beta|\alpha)\}$ is logically irreversible, then the very calculation of the probabilities $\{\Pi(\alpha|\beta)\}$ require the first input probability distribution $\{P(\alpha)\}$.

In general, it is only for logically deterministic, reversible operations (which are permutations) that one can construct optimal reverse operations independently of the probability distributions.

E. Thermodynamic irreversibility

The reverse operations considered in the preceding discussion have the property of restoring the original statistical state of the logical system. They do not, in general, restore the original logical state. The question of what is the "correct" thermodynamic entropy to use in such situations is not uncontroversial and can depend upon differing physical interpretations of the probabilities of the initial and final logical states. It will therefore be helpful to consider an approach to thermodynamic reversibility that does not depend upon such definitions.

We will use this to discuss further that the thermodynamic optimization of logically irreversible operations is not possible without specifying the probability distribution over the input states. Then we consider two additional sources of thermodynamic irreversibility that occur in the practical construction of information processing systems.

1. Thermodynamic cycles

In phenomenological thermodynamics, in any closed cycle, where a system returns to its initial state, the total heat generated in heat baths in the process must satisfy

$$\sum_i \frac{Q_i}{T_i} \geqslant 0. \tag{166}$$

As is well known, in statistical mechanics this can no longer be relied upon. There is some probability for the equality being violated. However, provided the system does return to its initial macroscopic state with certainty, then

$$\sum_i \frac{\langle Q_i \rangle}{T_i} \geqslant 0 \tag{167}$$

still holds. We will regard such a cycle, for which the equality holds, to be a thermodynamically reversible cycle, and use the following definition[19] of a thermodynamically reversible process: If a given physical process can, in principle, be included in at least one thermodynamically reversible cycle, then it is a thermodynamically reversible process.

To say otherwise would require one either to say that the overall cycle is thermodynamically reversible, although one of the steps in the cycle is not (which challenges what it could possibly mean to refer to that step as thermodynamically irreversible) or to say that the overall cycle is thermodynamically irreversible, despite the fact that it restores the original state with certainty and generates no net heat in any heat bath (and which means that the entropy of the universe must be the same at the end as the start of the cycle).

Conversely, if a given physical process cannot, even in principle, be included in *any* thermodynamically reversible cycles, then it is a thermodynamically irreversible process.

To avoid interpretational problems over probability, we will require that the thermodynamically reversible cycle starts, and ends, with the system in a physical state that represents a fixed logical state a, with certainty.

2. Optimal implementations

Take any logical operation, defined by the set $\{P(\beta|\alpha)\}$, and construct a physical implementation of that operation, optimized for the values w_α and $w_\beta = \Sigma_\alpha P(\beta|\alpha) w_\alpha$. This physical process will implement the $\{P(\beta|\alpha)\}$ operation regardless of the input state probabilities.

We now also construct two further operations: a logically reversible, indeterministic operation, generalizing the UFZ operation, that acts on a as the sole possible logical input state, and outputs state α with probability $P(\alpha) = w_\alpha$; and a logically irreversible, deterministic operation, generalizing RTZ, that acts on the logical states $\{\beta\}$, and always outputs logical state a. The physical implementation of this second operation is optimized for probabilities $P(\beta) = w_\beta$. Both these are well-defined physical processes.

It is clear that the sequence of these three operations forms a closed cycle, starting and ending in logical state a, with certainty. It is trivial to show that the optimal implementation of these operations produces a net thermodynamic cost of zero, over the course of the cycle. The cycle is, unquestionably, a thermodynamically reversible cycle. The given physical process that implements the logical operation must, then, be regarded as a thermodynamically reversible process.

3. Suboptimal implementations

If we had used a different initial operation, generating the logical state α with probability $P'(\alpha)$, and a final operation optimized for probabilities $P'(\beta) = \Sigma_\alpha P(\beta|\alpha) P'(\alpha)$, then it is straightforward to show that the cost would be

$$W = kT \sum_{\alpha,\beta} P'(\alpha,\beta) \ln\left(\frac{P'(\alpha) w_\beta}{P'(\beta) w_\alpha}\right) \geqslant 0. \tag{168}$$

Equality is in general reachable if either $w_\alpha = P'(\alpha)$ or $\forall \alpha, \beta$, $P(\beta|\alpha) w_\alpha \in (0, w_\beta)$, the latter being possible only if the logical operation $\{P(\beta|\alpha)\}$ is logically reversible.

This leaves us with the following conclusions.

(1) For any logical operation $\{P(\beta|\alpha)\}$, there exist physical implementations of that operation which can be included in thermodynamically reversible cycles.

(2) For any logical operation $\{P(\beta|\alpha)\}$, and for any given probability distribution $P(\alpha)$ over the input logical states, there exist physical implementations of that operation which can be included in thermodynamically reversible cycles.

(3) A given physical implementation of logical operation $\{P(\beta|\alpha)\}$ cannot be included in thermodynamically revers-

[19] We define the condition in this way to take into account the fact that for any physical process, it is always trivially possible to find some closed cycle incorporating that process for which inequality is strictly positive.

ible cycles for generic probability distributions over the input states unless it is a logically reversible operation.

It is not possible to characterize a particular physical process that implements a logically irreversible operation as thermodynamically reversible, independently of the specification of the statistical state on which it acts. Does this mean that we cannot characterize the physical process as thermodynamically reversible at all?

This situation is not unknown in statistical mechanics, or even phenomenological thermodynamics. Let us consider a large container, divided in half by a removable partition, and in the container is a macroscopic gas. The pressure on both sides of the partition is initially equal, and the gas is always kept in isothermal contact with a single heat bath. Removal and reinsertion of the partition is clearly thermodynamically reversible.

If we slowly, isothermally, slide the partition to the left, compressing half the gas and expanding the other half, until the compressed gas occupies only one-third of the container, the pressure on the left side is double the pressure on the right side (net work is required). Removal and reinsertion of the partition at this off-center position is not thermodynamically reversible.

This thermodynamic irreversibility is not simply due to the off-center position of the partition. Start with the partition in the center, but now with gas initially prepared to be at twice the pressure on the right-hand side of the container as on the left-hand side. Simply removing the partition from the center of the box is now thermodynamically irreversible. Isothermally moving the partition to the left until the left-hand side holds only one-third of the container's volume equalizes the pressure (and extracts work). Now the off-center removal and reinsertion of the partition become thermodynamically reversible.

The parallel to the model used for logical operations should be clear.[20] A given sequence of actions cannot, in general, be regarded as thermodynamically reversible independently of the state on which they act. To describe a phenomenological thermodynamic process as thermodynamically reversible, it is necessary to specify both the sequence of actions *and the state on which they act* in the definition of the physical process. This carries over into statistical mechanics and, as we have seen above, into the thermodynamics of computation.

The situation also bears some similarity to data compression from a signal source. A given coding scheme will be optimal only for a *particular* distribution of probabilities of signals from the source. Should the signals, in fact, be generated with a different probability distribution, then the mean length of the encoded signals will be greater than the Shannon information of the source. That Shannon's coding theorem is of practical utility indicates that it is not inconceivable that there may be information processing problems where the probability distribution over the logical states may be available when designing optimal physical implementations.

[20]Indeed, if we are considering a statistical mechanical N-atom gas, with $N=1$, it is exactly the same model.

4. Uncertain operations

If the logical operation acts upon a set of statistical states, but it is uncertain which operations have acted upon the system in the past, an additional source of thermodynamic irreversibility may occur. As an example of this, let us consider a bit that has been deterministically set to either zero or one, from a standard state a, and now needs to be reset to the standard state.

If the first operation set the bit to zero, the operation is UFZ(1), and the work required was

$$\Delta W_0 = (E_0 - T_R S_0) - (E_a - T_R S_a), \tag{169}$$

and if set to one, UFZ(0) gives

$$\Delta W_1 = (E_1 - T_R S_1) - (E_a - T_R S_a). \tag{170}$$

If the reset operation is optimized with values $w_0+w_1=1$, then it is RTZ(w_0),

$$\Delta W_{R0} = (E_a - T_R S_a) - (E_0 - T_R S_0) - kT_R \ln w_0,$$

$$\Delta W_{R1} = (E_a - T_R S_a) - (E_1 - T_R S_1) - kT_R \ln w_1, \tag{171}$$

giving total costs

$$\Delta W_{T0} = \Delta W_{R0} + \Delta W_0 = -kT_R \ln w_0 \geq 0,$$

$$\Delta W_{T1} = \Delta W_{R1} + \Delta W_1 = -kT_R \ln w_1 \geq 0. \tag{172}$$

The equalities can be reached by setting $w_0=1$ or $w_1=1$, respectively, but this is only possible if the other is zero—which would require an infinite amount of work if the wrong operation had taken place.

If we assign nonzero probabilities to the set operations of p_0 and p_1, then the expected cost for the cycle is

$$\Delta W_T = -kT_R \sum_{i=0,1} p_i \ln w_i \geq -kT_R \sum_i p_i \ln p_i > 0, \tag{173}$$

with the equality occurring if $w_i=p_i$. Clearly this is a thermodynamically irreversible cycle, despite the fact that each of the three logical operations [UFZ(1), UFZ(0), RTZ(w_0)] can be individually incorporated in a thermodynamically reversible cycle. What is the source of the irreversibility?

There are a number of ways one can regard this. Both the deterministic set operations are, in themselves, thermodynamically reversible. It could be argued that the irreversibility in whichever of the ΔW_{T0} or the ΔW_{T1} cycles actually took place is then through the reset operation, which was designed for the possibility of either deterministic set operation.

A different way to perceive the situation is to regard it as being either ΔW_{T0}, which may be thermodynamically optimized by setting $w_0=1$, or ΔW_{T1}, which may be optimized by $w_1=1$. In either case the cycle becomes thermodynamically reversible. The source of thermodynamic irreversibility would then be that the reset operation was not optimized for the correct probabilities (which must now be regarded as either $p_0=1$ or $p_1=1$, corresponding to which operation actually did take place).

Yet another way would be to consider a new class of operation: an "uncertain" operation, where there is an uncertainty as to which actual operation took place. In this case we have an uncertain set operation, which could be defined as $p_0\text{UFZ}(1)+p_1\text{UFZ}(0)$. This operation has a work requirement

$$\Delta W_U = \sum_{i=0,1} p_i(E_i - T_R S_i) - (E_a - T_R S_a). \tag{174}$$

Viewed as a logical operation, this would take as input logical state 0 with probability 1, and output states 0 and 1 with probabilities p_0 and p_1. The optimal implementation of such a logical transformation of information would be $\text{UFZ}(p_0)$, which has cost

$$\Delta W = \sum_{i=0,1} p_i(E_i - T_R S_i) - (E_a - T_R S_a) + kT_R \sum_i p_i \ln p_i. \tag{175}$$

As a logical transformation of information, the uncertain set operation is clearly sub-optimal. It is thermodynamically irreversible, as it cannot be included in any thermodynamically reversible cycle.

What is the "correct" way to view this? We are not sure this is a well-posed question. However, what all three explanations have in common is that the thermodynamic irreversibility is a consequence of the uncertainty over which the logical operation took place. It is this that prevents the construction of a thermodynamically reversible cycle.

Suppose we have a number of different processes, labeled with γ, and each implements a logical operation $\{P(\beta|\alpha,\gamma)\}$, optimized for input state probabilities $P(\alpha)$. The optimal cost for operation γ is

$$\Delta W_\gamma = \sum_\beta P(\beta|\gamma)[E_\beta - T_R S_\beta + kT_R \ln P(\beta|\gamma)] - \sum_\alpha P(\alpha)[E_\alpha - T_R S_\alpha + kT_R \ln P(\alpha)], \tag{176}$$

where $P(\beta|\gamma)=\Sigma_\alpha P(\beta|\alpha,\gamma)P(\alpha)$.

We now assign a probability $P(\gamma)$ to each logical operation occurring [and take for granted $P(\alpha,\gamma)=P(\alpha)P(\gamma)$]. The cost of this generic uncertain operation is

$$\langle \Delta W_\gamma \rangle = \sum_{\alpha,\beta,\gamma} P(\alpha,\beta,\gamma) \times \left[E_\beta - E_\alpha - T_R\left(S_\beta - S_\alpha - k \ln \frac{P(\beta|\gamma)}{P(\alpha)} \right) \right]. \tag{177}$$

This produces the output states $\{\beta\}$ with probabilities $P(\beta) = \Sigma_{\alpha,\gamma} P(\beta|\alpha,\gamma)P(\alpha)P(\gamma)$.

Now, to complete the cycle, we consider an optimized reset operation, which acts upon states $\{\beta\}$ to produce the standard state a, and an optimal operation that acts upon a and produces the logical states $\{\alpha\}$ with probability $P(\alpha)$. Combining these two has the cost

$$\Delta W_R = \sum_\alpha P(\alpha)[E_\alpha - T_R S_\alpha + kT_R \ln P(\alpha)] - \sum_\beta P(\beta)[E_\beta - T_R S_\beta + kT_R \ln P(\beta)], \tag{178}$$

giving a total cost for the cycle of

$$\langle \Delta W_\gamma \rangle + \Delta W_R = kT_R \sum_{\beta,\gamma} P(\beta,\gamma) \ln \frac{P(\beta,\gamma)}{P(\beta)P(\gamma)} \geqslant 0. \tag{179}$$

Equality is reached only if $P(\beta,\gamma)=P(\beta)P(\gamma)$, i.e., there is no correlation between the occurrence of the β output states and which γ operation actually took place. The thermodynamic irreversibility that occurs if $P(\beta|\gamma) \neq P(\beta)$ does not depend upon whether the operation required to restore the original statistical state is logically reversible or logically irreversible.

In the familiar case of the uncertain set–reset cycle there is a compression of the logical state space during the reset operation and the compensating increase in the non-information-bearing degrees of freedom of system or environment may give the impression that the source of the thermodynamic irreversibility is the logical irreversibility of the reset operation. The generic uncertain operation shows that this is not the case. In fact an optimal operation that restores the $P(\alpha)$ distribution from the $P(\beta)$ distribution could be logically reversible and the cycle still be thermodynamically irreversible provided $P(\beta|\gamma) \neq P(\beta)$. It is the uncertainty over which γ operation took place that is the source of the thermodynamic irreversibility.

As before, this situation has well-known parallels in standard statistical mechanics. The spread of gas molecules into a box, shielded from any outside interference, can in principle be reversed. (Spin-echo experiments have even demonstrated similar reversals to this in the laboratory.) However, this reversal is very sensitive to uncertainty in the outside forces that act upon the gas. In a famous calculation, Borel showed that the gravitational influence of remote stars could change the microscopic state of an expanding macroscopic gas within seconds. Reversing that expansion would be possible, in principle, if there was highly detailed knowledge of the gravitational influence of the remote bodies on the gas (or if microscopic state of the expanded gas molecules turned out to be independant of that influence) but becomes impossible when the gravitational influence is uncertain.

5. Partial operations

A third reason for the occurrence of thermodynamic irreversibility is that the physical implementation of the logical operation is not able to take into account the existence of correlations between systems, and can only act upon part of the total logical state.[21] We will show that, in this case, logically reversible operations are able to avoid the thermodynamic irreversibility, although logically irreversible operations are still not *always* thermodynamically irreversible.

[21] See also [34].

Suppose the input logical states factorize into the product of two subsystems, with the logical states of the first system in the set $\{\alpha\}$ and the second system in $\{\gamma\}$, so the joint system is described by the logical states $\{(\alpha,\gamma)\}$. Now consider a logical operation that acts only on the α states, with probabilities $P(\beta|\alpha)$. If the physical implementation of this logical operation has no access to the γ system, then the physical implementation can only be optimized with respect to the marginal probabilities

$$P(\alpha) = \sum_{\gamma} P(\alpha,\gamma). \tag{180}$$

The system ends up in output states from the product of the states of the $\{\gamma\}$ and $\{\beta\}$ systems, $\{(\beta,\gamma)\}$, with probabilities

$$P(\beta,\gamma) = \sum_{\alpha} P(\beta|\alpha)P(\alpha,\gamma). \tag{181}$$

The resulting thermodynamic cost of the partially optimized operation is

$$\Delta W_P = \sum_{\beta} P(\beta)\{E_\beta - T_R[S_\beta - k\ln P(\beta)]\} - \sum_{\alpha} P(\alpha)\{E_\alpha - T_R[S_\alpha - k\ln P(\alpha)]\} \tag{182}$$

where $P(\beta)=\Sigma_\gamma P(\beta,\gamma)$ and we have assumed

$$E_{\alpha,\gamma} = E_\alpha + E_\gamma,$$

$$E_{\beta,\gamma} = E_\beta + E_\gamma,$$

$$S_{\alpha,\gamma} = S_\alpha + S_\gamma,$$

$$S_{\beta,\gamma} = S_\beta + S_\gamma. \tag{183}$$

An optimal operation for restoring the states (α,γ), with probabilities $P(\alpha,\gamma)$ has a thermodynamic cost of

$$\Delta W_R = \sum_{\alpha,\gamma} P(\alpha,\gamma)\{E_\alpha - T_R[S_\alpha - k\ln P(\alpha,\gamma)]\} - \sum_{\beta,\gamma} P(\beta,\gamma)\{E_\beta - T_R[S_\beta - k\ln P(\beta,\gamma)]\} \tag{184}$$

so the net cost for the cycle is

$$\frac{\Delta W_R + \Delta W_P}{kT_R} = kT_R\left(\sum_{\alpha,\gamma} P(\alpha,\gamma)\ln\frac{P(\alpha,\gamma)}{P(\alpha)} - \sum_{\beta,\gamma} P(\beta,\gamma)\ln\frac{P(\beta,\gamma)}{P(\beta)}\right). \tag{185}$$

This can be expressed as changes in conditional or correlation information:

$$\frac{\Delta W_R + \Delta W_P}{kT_R} = -\sum_{\alpha,\beta,\gamma} P(\alpha,\beta,\gamma)[\ln P(\gamma|\beta) - \ln P(\gamma|\alpha)]$$
$$= -\sum_{\alpha,\beta,\gamma} P(\alpha,\beta,\gamma)\left(\ln\frac{P(\beta,\gamma)}{P(\beta)P(\gamma)} - \ln\frac{P(\alpha,\gamma)}{P(\alpha)P(\gamma)}\right). \tag{186}$$

Use of the identity

$$P(\alpha,\gamma|\beta)P(\beta|\alpha) = P(\beta,\gamma|\alpha)P(\alpha|\beta) \tag{187}$$

gives the form of the conditional correlations:

$$\frac{\Delta W_R + \Delta W_P}{kT_R} = -\sum_{\alpha,\beta,\gamma} P(\alpha,\beta,\gamma)\left(\ln\frac{P(\beta,\gamma|\alpha)}{P(\beta|\alpha)P(\gamma|\alpha)} - \ln\frac{P(\alpha,\gamma|\beta)}{P(\alpha|\beta)P(\gamma|\beta)}\right). \tag{188}$$

As $P(\beta|\alpha,\gamma)=P(\beta|\alpha)$, then

$$P(\gamma,\beta|\alpha) = P(\beta|\alpha,\gamma)P(\gamma|\alpha) = P(\beta|\alpha)P(\gamma|\alpha). \tag{189}$$

The α states screen off any correlation between the β and γ states and the first term is zero, so

$$\frac{\Delta W_R + \Delta W_P}{kT_R} = \sum_{\alpha,\beta,\gamma} P(\alpha,\beta,\gamma)\ln\frac{P(\alpha,\gamma|\beta)}{P(\alpha|\beta)P(\gamma|\beta)} \geqslant 0. \tag{190}$$

Equality occurs if and only if β screens off any correlations between α and γ:

$$P(\alpha,\gamma|\beta) = P(\alpha|\beta)P(\gamma|\beta). \tag{191}$$

This can happen directly if there is no initial correlation between the α and γ systems, so that $P(\alpha,\gamma)=P(\alpha)P(\gamma)$. With $P(\beta|\alpha,\gamma)=P(\beta|\alpha)$ it follows that $P(\alpha,\beta,\gamma) = P(\alpha,\beta)P(\gamma)$ and from that Eq. (191) holds, as might be expected.

To see the effect of logical reversibility, rewrite Eq. (190) as

$$\frac{\Delta W_R + \Delta W_P}{kT_R} = \sum_{\alpha,\beta,\gamma} P(\alpha,\beta,\gamma)[\ln P(\alpha|\gamma,\beta) - \ln P(\alpha|\beta)]. \tag{192}$$

If the operation is logically reversible $P(\alpha|\beta)\in\{0,1\}$. This gives

$$P(\alpha|\beta) = 0 \Rightarrow P(\alpha,\beta,\gamma) = 0,$$

$$P(\alpha|\beta) = 1 \Rightarrow P(\alpha|\gamma,\beta) = 1, \tag{193}$$

and the summation is identically zero. Logically reversible operations avoid the thermodynamically irreversible cost.[22]

[22]However, one should note that it is still possible for some logically *irreversible* operation to satisfy the conditions for thermodynamic reversibility, Eq. (191), for particular correlations between the α and γ systems.

VII. CONCLUSIONS

The focus on the process of Landauer *erasure* can give the impression that Landauer's *principle* should be exclusively about the thermodynamics of logically irreversible processes and further that the heat generation of such processes implies thermodynamic irreversibility: "To erase a bit of information in an environment at temperature T requires dissipation of energy $\geqslant kT \ln 2$" [23,24]. "In erasing one bit …of information one dissipates, on average, at least $k_B T \ln 2$ of energy into the environment" [25]. "A logically irreversible operation must be implemented by a physically irreversible device, which dissipates heat into the environment" [26]. "Erasure of one bit of information increases the entropy of the environment by at least $k \ln 2$" ([27], p. 27). "Any logically irreversible manipulation of data … must be accompanied by a corresponding entropy increase in the non-information-bearing degrees of freedom of the information processing apparatus or its environment. Conversely, it is generally accepted that any logically reversible transformation of information can in principle be accomplished by an appropriate physical mechanism operating in a thermodynamically reversible fashion." [1].

However, it should be noted that not all advocates of Landauer's principle regard the process of erasure as *necessarily* thermodynamically irreversible: "A logically irreversible operation …may be thermodynamically reversible or not depending on the data to which it is applied. If it is applied to random data …it is thermodynamically reversible, because it decreases the entropy of the data while increasing the entropy of the environment by the same amount" [1].

In [3] it was argued that there exists a valid thermodynamically reverse process to Landauer erasure, but which needs to be classified as logically indeterministic, which we called reverse Landauer erasure. Consideration of the thermodynamic consequences of the existence of this process led us to conclude that there was no convincing evidence that logically irreversible operations had special thermodynamic characteristics. Instead, we hypothesized that a generalized form of Landauer's principle should be possible that made no reference to irreversibility, whether logical or thermodynamic. This was expressed in two conjectures ([3], p. 362):

Conjecture E. Any logically irreversible transformation of information can in principle be accomplished by an appropriate physical mechanism operating in a thermodynamically reversible fashion.

Conjecture F. A logical operation needs to generate heat equal to at least $-kT \ln 2$ times the change in the total quantity of Shannon information over the operation, or

$$\Delta W \geqslant kT \ln 2(H_i - H_f).$$

In this paper we have both proved and generalized these conjectures. Our approach has been to take the widest definition of logical operations available and the most general procedure for physically implementing these operations that we can. This requires us to consider logically indeterministic operations as well as deterministic ones, logically reversible operations as well as irreversible ones.

Other papers have made some consideration of Landauer erasure in the context of non-uniform temperatures [28], entropy [29,30], and energy [25], while [31] combines varying entropy and energy. Nonuniform input probabilities are considered in the proofs of [25,29]. The thermodynamics of logically indeterministic operations does not seem to have been considered before [2,3], although [32] (Chap. VI) is close, and it is noticeable that [1] refers throughout to deterministic computation. Recently, a paper has appeared by Turgut [7] deriving similar results using classical phase space arguments.

General proofs of Landauer's principle seem hard to come by (as pointed out in [5]) although [25] derives similar results to those of Sec. III E, but restricted to the RTZ operation, and under the assumption that logical states are represented by pure quantum states (an assumption shared with [33]). Here we allow logical states to be represented by density matrices and consider any logical operation. We have considered the most general setting for physically implementing classical logical operations, covering and extending these earlier results. We derive the most general statement of Landauer's principle, prove it cannot be exceeded and give a limiting process which can achieve it.

The general statement of Landauer's principle we arrived at is the following.

Generalized Landauer principle. A physical implementation of a logical transformation of information has minimal expectation value of the work requirement given by

$$\langle \Delta W \rangle \geqslant \langle \Delta E \rangle - T\Delta S, \tag{194}$$

where $\langle \Delta E \rangle$ is the change in the mean internal energy of the information processing system, ΔS the change in the Gibbs–von Neumann entropy of that system, and T the temperature of the heat bath into which any heat is absorbed. The equality is reachable, in principle, by any logical operation, and if the equality is reached the physical implementation is thermodynamically reversible.

We have then shown how various additional assumptions and simplifications can lead to more familiar versions of Landauer's principle that can be found in the literature and these are special cases of the GLP. Generalizations about the relationship between information processing and thermodynamic entropy based upon these special cases can be misleading.

In particular, we have argued, counter to a widespread version of Landauer's principle, that there is nothing in principle that prevents a logically irreversible operation from being implemented in a thermodynamically reversible manner. What differs between logically irreversible operations and logically reversible operations is that to thermodynamically optimize physical implementations of the former it is necessary to take into account the probability distribution over the complete set of input logical states. A physical implementation of a logically irreversible operation, optimized for a particular input probability distribution, will not be thermodynamically irreversible for a different input probability distribution. If the physical implementation cannot access a correlated system, then logically irreversible operations may incur additional costs.

As the practical business of actually building physical devices to implement logical operations will typically not be able to make such optimizations, it is natural to assume an equiprobable distribution over a subsystem, and expect thermodynamic irreversibility. Nevertheless the point remains: *in principle* it is always possible to physically implement logically irreversible transformations of information in thermodynamically reversible ways. There are many practical reasons why a logically irreversible operation may not be thermodynamically optimized, and it is clearly important and useful to explore such problems. In this paper, however, we are primarily concerned with the question: What is the *fundamental limit* for thermodynamical optimization of the physical implementation of a given logical operation?

We have demonstrated that, under the same conditions of uniform computing that imply that logically deterministic, irreversible operations generate heat, logically indeterministic, reversible operations extract heat from the environment which can be converted into work. At the same time we have demonstrated that, under other conditions, adiabatic equilibrium computing, information processing is able to progress without any exchange of work or heat, regardless of the type of logical operation.

The thermodynamic reversibility of all logical operations is, of course, based upon the definition of thermodynamic reversibility given in Secs. VI C and VI E. Other approaches to thermodynamics (such as [4–6]) use different concepts of entropy and correspondingly different definitions of thermodynamic irreversibility from those in this paper. Ultimately the most important question is not what particular quantity one chooses to label as the "thermodynamic" entropy. The GLP we have derived here is valid, whether one chooses to regard the Gibbs–von Neumann entropy as the true thermodynamic entropy, or not. What is important is the actual work required to drive a system, the actual heat generated by that system. As there is no disagreement over the fundamental microscopic dynamics, it would be surprising if we were unable to be able to agree on these values, regardless of the definition of entropy to which we choose to adhere.

ACKNOWLEDGMENTS

The author would like to thank Basil Hiley, Keith Bowden, Chris Timpson, John Barrett, James Ladyman, and Tony Short for various discussions on subjects related to computation, information, and entropy and Steve Weinstein for his comments and advice. The author would also like to thank an anonymous referee for raising the question of the uncertain set operation. Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI.

APPENDIX: ONE-BIT LOGICAL OPERATIONS

1. Do nothing (IDN)

The simplest operation is

$$P(\beta=0|\alpha=0)=1,$$
$$P(\beta=0|\alpha=1)=0,$$
$$P(\beta=1|\alpha=0)=0,$$
$$P(\beta=1|\alpha=1)=1, \tag{A1}$$

giving

$$P(\alpha=0|\beta=0)=1,$$
$$P(\alpha=0|\beta=1)=0,$$
$$P(\alpha=1|\beta=0)=0,$$
$$P(\alpha=1|\beta=1)=1, \tag{A2}$$

and is logically deterministic and reversible.

2. Logical NOT (NOT)

Logical NOT, acting upon an input bit with probability p of being in state 0, is very simple:

$$P(\beta=0|\alpha=0)=0,$$
$$P(\beta=0|\alpha=1)=1,$$
$$P(\beta=1|\alpha=0)=1,$$
$$P(\beta=1|\alpha=1)=0, \tag{A3}$$

giving

$$P(\alpha=0|\beta=0)=0,$$
$$P(\alpha=0|\beta=1)=1,$$
$$P(\alpha=1|\beta=0)=1,$$
$$P(\alpha=1|\beta=1)=0. \tag{A4}$$

This is logically deterministic and reversible.

3. Reset to zero: [RTZ(p)]

If the input state 0 occurs with probability p, then the RTZ(p) operation has the properties

$$P(\beta=0|\alpha=0)=1,$$
$$P(\beta=0|\alpha=1)=1, \tag{A5}$$

giving

$$P(\alpha=0|\beta=0)=p,$$
$$P(\alpha=1|\beta=0)=1-p. \tag{A6}$$

This is logically deterministic and irreversible. As $\forall\ \alpha\ P(\beta=1|\alpha)=0$, the state $\beta=1$ is not an output state of the operation and we leave it out of the table.

4. Unset from zero [UFZ(*p*)]

The reverse operation to RTZ, where the state 0 is taken to state 0 with probability p, will be called here the unset from zero operation. In [3] this operation was described in terms of the physical process that reverses LE, so was called reverse Landauer erasure or RLE. In this paper we will refer to the logical operation as UFZ, and to the specific physical process that can be used to embody it as RLE. This operation may also be characterized as a random number generator:

$$P(\beta=0|\alpha=0)=p,$$

$$P(\beta=1|\alpha=0)=1-p, \tag{A7}$$

giving

$$P(\alpha=0|\beta=0)=1,$$

$$P(\alpha=0|\beta=1)=1. \tag{A8}$$

This is indeterministic but reversible. As $\forall\ \beta\ P(\alpha=1|\beta)=0$ the state $\alpha=1$ is not an input state of the operation and we leave it out of the table.

5. Randomize [RND(*p*,*p*′)]

This operation takes an input probability of p of the state being 0 and produces 0 with an output probability of p', regardless of input state:

$$P(\beta=0|\alpha=0)=p',$$

$$P(\beta=0|\alpha=1)=p',$$

$$P(\beta=1|\alpha=0)=1-p',$$

$$P(\beta=1|\alpha=1)=1-p', \tag{A9}$$

giving

$$P(\alpha=0|\beta=0)=p,$$

$$P(\alpha=0|\beta=1)=p,$$

$$P(\alpha=1|\beta=0)=1-p,$$

$$P(\alpha=1|\beta=1)=1-p. \tag{A10}$$

This is indeterministic and irreversible.

We note that $\text{RTZ}(p)\equiv\text{RND}(p,1)$ and $\text{UFZ}(p)\equiv\text{RND}(1,p)$.

6. General one bit [GOB(p,p_{00},p_{11})]

Finally, we consider the most generic operation possible for one input bit and one output bit. The operation can be wholly defined by one input probability p and two conditional probabilities p_{00} and p_{11}:

$$P(\alpha=0)=p,$$

$$P(\alpha=1)=1-p,$$

$$P(\beta=0|\alpha=0)=p_{00},$$

$$P(\beta=0|\alpha=1)=1-p_{11},$$

$$P(\beta=1|\alpha=0)=1-p_{00},$$

$$P(\beta=1|\alpha=1)=p_{11}, \tag{A11}$$

giving

$$P(\alpha=0,\beta=0)=pp_{00},$$

$$P(\alpha=0,\beta=1)=p(1-p_{00}),$$

$$P(\alpha=1,\beta=0)=(1-p)(1-p_{11}),$$

$$P(\alpha=1,\beta=1)=(1-p)p_{11}, \tag{A12}$$

and

$$P(\beta=0)=pp_{00}+(1-p)(1-p_{11}),$$

$$P(\beta=1)=p(1-p_{00})+(1-p)p_{11}, \tag{A13}$$

so

$$P(\alpha=0|\beta=0)=\frac{pp_{00}}{pp_{00}+(1-p)(1-p_{11})},$$

$$P(\alpha=0|\beta=1)=\frac{p(1-p_{00})}{p(1-p_{00})+(1-p)p_{11}},$$

$$P(\alpha=1|\beta=0)=\frac{(1-p)(1-p_{11})}{pp_{00}+(1-p)(1-p_{11})},$$

$$P(\alpha=1|\beta=1)=\frac{(1-p)p_{11}}{p(1-p_{00})+(1-p)p_{11}}. \tag{A14}$$

In general, this is logically indeterministic and irreversible, but can become logically reversible or deterministic under the right limits:

$$\text{IDN}\equiv\text{GOB}(p,1,1),$$

$$\text{NOT}\equiv\text{GOB}(p,0,0),$$

$$\text{RTZ}(p)\equiv\text{GOB}(p,1,0),$$

$$\text{UFZ}(p)\equiv\text{GOB}(1,p,-),$$

$$\text{RND}(p,p')\equiv\text{GOB}(p,p',1-p'). \tag{A15}$$

[1] C. H. Bennett, Stud. Hist. Philos. Mod. Phys. **34**, 501 (2003).
[2] O. J. E. Maroney, Ph. D. thesis, Birkbeck College, University of London, 2002, e-print arXiv:quant-ph/0411172.
[3] O. J. E. Maroney, Stud. Hist. Philos. Mod. Phys. **36**, 355 (2005).
[4] O. R. Shenker, http://philsci-archive.pitt.edu/archive/00000115/
[5] J. D. Norton, Stud. Hist. Philos. Mod. Phys. **36**, 375 (2005).
[6] J. Ladyman, S. Presnell, Anthony J. Short, and B. Groisman, Stud. Hist. Philos. Mod. Phys. **38**, 58 (2007).
[7] S. Turgut, Phys. Rev. E. (to be published), e-print arXiv:cond-mat/0609198.
[8] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961), reprinted in [27].
[9] G. E. Revesz, *Introduction to Formal Languages* (Dover, New York, 1991).
[10] G. Boole, *An Investigation of the Laws of Thought* (Dover, New York, 1958).
[11] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, U.K., 2003).
[12] I. L. Chuang and M. A. Nielsen, *Quantum Computation and Quantum Information* (Cambridge, Cambridge, U.K., 2001).
[13] O. J. E. Maroney, eprint arXiv:quant-ph/0701127.
[14] R. C. Tolman, *The Principles of Statistical Mechanics* (Dover, New York, 1979).
[15] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Ox Bow Press, Woodbridge, CT, 1981).
[16] W. H. Zurek, in *Frontiers of Non-Equilibrium Statistical Physics*, edited by G. T. Moore and M. O. Scully (Plenum Press, New York, 1984), pp. 151–161, reprinted in [27].
[17] H. S. Leff, Am. J. Phys. **63**, 895 (1995).
[18] C. M. Bender, D. C. Brody, and B. K. Meister, J. Phys. A **33**, 4427 (2000).
[19] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).
[20] M. H. Partovi, Phys. Lett. A **137**, 440 (1989).
[21] A. E. Allahverdyan and T. M. Nieuwenhuizen, Phys. Rev. E **64**, 056117 (2001).
[22] A. E. Allahverdyan and T. M. Nieuwenhuizen, e-print arXiv:cond-mat/0206052.
[23] C. M. Caves, Phys. Rev. Lett. **64**, 2111 (1990).
[24] C. M. Caves, Phys. Rev. E **47**, 4010 (1993), reprinted in [27].
[25] B. Piechocinska, Phys. Rev. A **61**, 062314 (2000), reprinted in [27].
[26] J. Bub, Stud. Hist. Philos. Mod. Phys. **32**, 569 (2001).
[27] *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, edited by H. S. Leff and A. F. Rex (IOP, Bristol, 2003).
[28] B. W. Schumacher, in *Physical Origins of Time Asymmetry*, edited by J. J. Halliwell, J. Perez-Mercader, and W. H. Zurek (Cambridge University Press, Cambridge, U.K., 1994), pp. 90–98.
[29] K. Shizume, Phys. Rev. E **52**, 3495 (1995), reprinted in [27].
[30] P. N. Fahn, Found. Phys. **26**, 71 (1996).
[31] M. M. Barkeshli, e-print arXiv:cond-mat/0504323.
[32] O. Penrose, *Foundations of Statistical Mechanics* (Pergamon, Oxford, 1970).
[33] K. Jacobs, e-print arXiv:quant-ph/0512105.
[34] N. G. Anderson, Phys. Lett. A **372**, 5552 (2008).