Thermodynamic and logical 
reversibilities revisited

Takahiro Sagawa

Department of Basic Science, The University of Tokyo, 3-8-1 Komaba,
Meguro-ku, Tokyo 153-8902, Japan
E-mail: sagawa@noneq.c.u-tokyo.ac.jp

Received 8 November 2013
Accepted for publication 11 February 2014
Published 26 March 2014

Online at stacks.iop.org/JSTAT/2014/P03025
doi:10.1088/1742-5468/2014/03/P03025

Abstract. We review and investigate the general theory of the thermodynamics of computation, and derive the fundamental inequalities that set the lower bounds of the work requirement and the heat emission during a computation. These inequalities constitute the generalized Landauer principle, where the information contents are involved in the second law of thermodynamics. We discuss in detail the relationship between the thermodynamic and logical reversibilities; the former is related to the entropy production in the total system including a heat bath, while the latter is related to the entropy change only in the logical states of the memory. In particular, we clarify that any logically irreversible computation can be performed in a thermodynamically reversible manner in the quasi-static limit, which does not contradict the conventional Landauer principle. Our arguments would serve as the theoretical foundation of the thermodynamics of computation in terms of modern statistical physics.

Keywords: driven diffusive systems (theory)
1. Introduction

The investigation of the thermodynamics of computation and information processing predates the era of computer technology [1]. From the viewpoint of fundamental physics, this research area is closely related to the foundation of the second law of thermodynamics, which dates back to Maxwell’s thought experiment on ‘Maxwell’s demon’ in the nineteenth century [2]. Several decades later, Szilard made the first crucial step toward the quantitative understanding of the relationship between information and thermodynamics [3]. In his thought experiment with a single-particle heat engine, which is called the Szilard engine, the demon can extract $k_B T \ln 2$ of work from a single heat bath through feedback control by using one bit of information, where $k_B$ is the Boltzmann constant and $T$ is the temperature of the bath. By developing Szilard’s observation, Brillouin investigated the relationship between thermodynamic and informational entropies [4, 5]. Recently, a type of Szilard engine has been experimentally
Thermodynamic and logical reversibilities revisited

demonstrated for the first time [6] with a small thermodynamic engine that works at the level of thermal fluctuations.

Another notable observation was made by Landauer [7, 8]. He pointed out that, to erase one bit of information from a memory, at least \( k_B T \ln 2 \) of heat should be emitted into a heat bath and the same amount of work is needed. This has been referred to as the Landauer principle. Bennett also discussed the thermodynamics of computation by considering the concept of logical reversibility and its relationship to thermodynamics [9]–[11]. Further, Zurek discussed the thermodynamics of computation in terms of algorithmic complexity [12, 13]. The Landauer principle has been studied in various aspects [14, 15]; it has been derived in terms of statistical mechanics in some setups [16]–[18], has been demonstrated in concrete systems theoretically [19]–[23] and experimentally [24, 25], and has sparked intense discussion on its validity [26]–[35].

In a different context, the recent advancements in nonequilibrium statistical mechanics have revealed a fundamental aspect of the second law of thermodynamics [36]–[47]. This has enabled us to more quantitatively and comprehensively understand the thermodynamics of computation and information processing. In terms of theory, the second law of thermodynamics and the fundamental nonequilibrium relations (e.g., the fluctuation theorem and the Jarzynski equality) have been generalized to information processing processes such as measurement, information erasure, and feedback control [48]–[71], where information contents and thermodynamic quantities are treated on an equal footing. In particular, the concept of thermodynamic reversibility in the presence of feedback control has been established [57, 61], and several concrete models of thermodynamically reversible information processing have been proposed [53, 57], [59]–[61], [70]. More recently, a variety of autonomous Maxwell demons has attracted much attention [72]–[76]. In terms of experiment, a generalized Jarzynski equality with feedback control has been verified [6]. The Jarzynski equality for information erasure has also been investigated experimentally [25]. In the light of these advancements, we are now in the position to revisit and to further clarify the fundamental concepts in thermodynamics of computation and information processing.

In this paper, we review and investigate the long-standing fundamental problems in the thermodynamics of computation, shedding new light on the celebrated Landauer principle from the viewpoint of modern statistical physics. In particular, we clarify the fundamental relationship between the thermodynamic and logical reversibilities. Moreover, on the basis of the second law of thermodynamics, we derive universal thermodynamic inequalities that set the lower bounds of the work requirement and the heat emission during a computation. Our observation does not contradict the conventional Landauer principle, but extends it to a much broader class of memories that perform computation. While we assume that a memory is a classical system, the generalization of our arguments to the quantum cases is straightforward [47].

This paper is organized as follows. In section 2, we discuss the concept of thermodynamic reversibility in terms of the total entropy production in the whole system including a heat bath. In section 3, we discuss the concept of logical reversibility, and illustrate several typical examples of reversible and irreversible computations. In section 4, we discuss the conventional Landauer principle, and clarify the relationship between thermodynamic and logical reversibilities with the standard setup of the Landauer
principle. In section 5, we formulate the general setup of thermodynamic computation, and derive the second law of thermodynamics for computational processes. In section 6, we discuss the situation that there are two memories; this setup enables us to analyze measurement and feedback, which constitute the typical situation of Maxwell’s demon. In section 7, we conclude this paper.

2. Thermodynamic reversibility

In this section, we clarify the concept of thermodynamic reversibility. We consider a time evolution of a thermodynamic system in the presence of heat bath(s). We assume that the system may be driven from equilibrium by changing external parameters such as the volume of the gas. First of all, we roughly characterize the thermodynamic reversibility according to the standard definition in thermodynamics [77, 78]:

A physical process is thermodynamically reversible if and only if its time-reversal is not prohibited by the second law of thermodynamics. Otherwise, the process is thermodynamically irreversible.

To be more precise, we need to consider the ensemble of the system, because, from the microscopic point of view, the dynamics of the phase-space point of the system is stochastic due to thermal fluctuations. The thermodynamic property of the system is described by the probability distribution on the phase space. The thermodynamic reversibility can then be characterized as follows [36]—[42]:

A physical process is thermodynamically reversible if and only if the time evolution of the probability distribution in the process can be time-reversed, where the change of the external parameters is also time-reversed, and the signs of the amounts of the work and the heat are changed.

A process can become thermodynamically reversible in the quasi-static limit, where the change of the external parameters is much slower than the relaxation time of the system, and the state of the system can always be regarded in thermal equilibrium during the process. For example, the quasi-static and isothermal compression and expansion of the gas in a box are both thermodynamically reversible, as they are in time-reversal with each other. We note that all quasi-static processes are not necessarily reversible; there may be quasi-static but irreversible processes such as a weak constant between two baths with different temperatures [77].

To characterize the thermodynamic reversibility in a more quantitative way, we discuss the concept of entropy production. For simplicity, we assume that there is a single heat bath at inverse temperature \( \beta := (k_B T)^{-1} \). The generalization of the following arguments to cases with multiple baths is straightforward. Let \( \mathcal{Y} \) be the phase space of the system, \( y, y' \in \mathcal{Y} \) be the initial and final phase-space points, and \( P[y] \) and \( P'[y'] \) be their probabilities. We consider the Shannon entropy of the probability distribution on the phase space [79, 80]. The initial and final entropies are respectively given by

\[
S := - \int_{y \in \mathcal{Y}} dy \ P[y] \ln P[y], \quad S' := - \int_{y' \in \mathcal{Y}} dy' \ P'[y'] \ln P'[y'].
\] (1)
We denote as $Q$ the average heat absorbed by the system from the heat bath during the dynamics. The total entropy production in the relevant total system (i.e., the whole ‘universe’) including the bath is then given by \[ \Delta S_{\text{tot}} := \Delta S - \beta Q, \] where $\Delta S := S' - S$ is the change in the Shannon entropy of the system. We note that $-\beta Q$ is regarded as the change in the entropy of the heat bath.

The second law of thermodynamics can then be expressed as \[ \Delta S_{\text{tot}} \geq 0, \] or equivalently, \[ \Delta S \geq \beta Q. \] We stress that inequality (3) holds for any initial and final nonequilibrium distributions of the system (i.e., for any $P[y]$ and $P'[y']$). Inequality (3) can be derived on the basis of nonequilibrium statistical mechanics [47]. In fact, inequality (3) is a straightforward consequence of the fluctuation theorem [37]–[40].

On the basis of inequality (3), the thermodynamic reversibility can be characterized in terms of the entropy production [77, 78]:

A physical process is thermodynamically reversible if the equality in (3) is achieved (i.e., the total entropy production is zero).

In fact, if $\Delta S_{\text{tot}} > 0$ holds in a process, its time-reversal is impossible because of $\Delta S_{\text{tot}} < 0$. The thermodynamically reversible condition $\Delta S_{\text{tot}} = 0$ can be achieved in the quasi-static limit, where the system is always in thermal equilibrium during the process.

We stress that a process can be thermodynamically reversible even if there is an entropy transfer from the system to the bath or vice versa. If the amount of the increase/decrease in entropy in the system is the same as the amount of the decrease/increase in the bath, the total amount of the entropy increase in the whole universe is zero, where the equality in (3) is achieved and the process is thermodynamically reversible.

As a simple example, we consider a quasi-static isothermal expansion of an ideal gas with $N$ particles. Starting from a thermal equilibrium state, we expand the gas isothermally and quasi-statically, doubling its volume. The entropy of the system is then increased by $N \ln 2$, and the heat emission to the bath is given by $-Q = \beta^{-1}N \ln 2$ (i.e., the entropy of the bath is decreased by $N \ln 2$). Therefore, the total entropy production is given by

$$\Delta S_{\text{tot}} = (-N \ln 2) + N \ln 2 = 0,$$

which implies that the quasi-static isothermal expansion is thermodynamically reversible. Similarly, the quasi-static isothermal compression is also thermodynamically reversible.

We now discuss the relationship between the foregoing thermodynamic perspective and the microscopic perspective based on the reversible Hamiltonian dynamics. One of the crucial advances in modern statistical physics is that it has succeeded in reconciling...
these two perspectives [36, 39, 41]. In fact, the second law (3) can be derived on the basis of the fluctuation theorem, where one needs essentially only two assumptions: that the microscopic dynamics satisfies the Liouville theorem, and the initial distribution of the bath is the canonical distribution [39]. A crucial observation here is that the Shannon entropy on the whole phase space of the system and the bath is different from the thermodynamic entropy in the whole universe; the former is conserved in any Hamiltonian dynamics, while the latter is increased in thermodynamically irreversible processes even if the underlying microscopic Hamiltonian dynamics is reversible.

The change in the thermodynamic entropy is equal to the relative entropy (i.e., the Kullback–Leibler divergence) between the probability distribution in the final state and a reference probability distribution such as the canonical distribution [43]. We again stress that the thermodynamic entropy is not equivalent to the Shannon entropy of the total system including the system and the bath. Correspondingly, we have not considered the Shannon entropy in the bath in the foregoing argument, but regarded the heat transfer \(-\beta Q\) as the entropy change in the bath, which is consistent with the picture that the relative entropy corresponds to the entropy production in the whole universe [39, 43, 47]. The thermodynamic irreversibility (i.e., the positive entropy production in the whole universe) can also be characterized by the gap between the probability distributions of microscopic trajectories in the forward process and the backward one [41]–[43]. We note that the same argument is also valid for quantum systems that obey unitary dynamics [44]–[47], by replacing the Shannon entropy by the von Neumann one.

The concept of thermodynamic reversibility discussed in this paper is consistent with the above-mentioned observation based on the fluctuation theorem, and therefore has a rigid theoretical foundation that is consistent with the microscopic reversible physics. In terms of the fluctuation theorem, \(\Delta S_{\text{tot}} = 0\) holds if and only if the probability distribution of the trajectories in the phase space is the same as that of the time-reversed trajectories, which has been explicitly discussed in, for example, [42, 43].

We now assume that the initial and final distributions are the canonical distributions

\[
P[y] = e^{\beta(F - E[y])}, \quad P'[y'] = e^{\beta(F' - E'[y'])},
\]

where \(E[y] (E'[y'])\) and \(F (F')\) are the initial (final) energy and the initial (final) free energy of the system, respectively. In this case, we obtain

\[
S = \beta(E - F), \quad S' = \beta(E' - F'),
\]

where \(E (E')\) is the ensemble average of the initial (final) energy. Therefore,

\[
\Delta S = \beta(\Delta E - \Delta F),
\]

where \(\Delta E := E' - E\) and \(\Delta F := F' - F\). Let \(W\) be the average work performed on the system during the process. The total entropy production is then given by

\[
\Delta S_{\text{tot}} = \beta(W - \Delta F),
\]

where we used the first law of thermodynamics:

\[
W = \Delta E - Q.
\]
Therefore, the second law (3) reduces to
\[ W \geq \Delta F, \]  
(11)
where the equality can be achieved in the quasi-static limit.

If the final distribution is different from the canonical distribution, we can show that
\[ \Delta S_{\text{tot}} \leq \beta (W - \Delta F), \]  
(12)
where the final free energy is given by the equilibrium one corresponding to the final Hamiltonian. In this case, we again obtain inequality (11). We note that the equality in (12) is achieved if the final distribution is given by the canonical distribution.

3. Logical reversibility

In this section, we discuss the concept of logical reversibility, and introduce the logical entropy that is decreased by logically irreversible computation.

3.1. Definition and examples

We consider a memory that has several logical states, in which information is stored. A computational process is performed by changing an input logical state into an output one with a certain algorithm. Let \( \mathcal{M} \) and \( \mathcal{M}' \) be the sets of input and output logical states, respectively. We assume that these are finite sets. For example, if the input consists of \( n \) bits, we set \( \mathcal{M} = \{0,1\}^n \). We note that the logical states do not have one-to-one correspondence to the physical phase space, as discussed in section 4 in detail.

We next formulate (deterministic) computational processes. In this paper, a computational process \( \hat{C} \) is defined as a map
\[ \hat{C} : \mathcal{M} \rightarrow \mathcal{M}'. \]  
(13)
We also call \( \hat{C} \) as a gate. We note that \( \hat{C} \) is not necessarily a surjection.

In order to discuss the logical reversibility and its relationship to thermodynamics, we do not need the detailed characterization of computable functions in terms of computability theory [81]; the following argument is applicable to any map \( \hat{C} : \mathcal{M} \rightarrow \mathcal{M}' \).

We now define the logical reversibility [8]–[11]:

A computational process \( \hat{C} \) is logically reversible if and only if it is an injection. In other words, \( \hat{C} \) is logically reversible if and only if, for any output logical state, there is a unique input logical state. Otherwise, \( \hat{C} \) is logically irreversible.

We note that, if \( \mathcal{M}' = \mathcal{M} \), a computational process is reversible if and only if it is a bijection (i.e., an injection and a surjection). In general, a computational process is reversible if and only if we can precisely estimate the input state from the output state. In fact, if a computational process is reversible and \( \hat{C} \) is an injection, we can define the reversed computational process
\[ \hat{C}^{-1} : \mathcal{M}' \rightarrow \mathcal{M}, \]  
(14)
where the domain of \( \hat{C}^{-1} \) is given by the image of \( \hat{C} \) denoted as \( \hat{C}(\mathcal{M}) \subset \mathcal{M}' \).
We next discuss several simple examples of computation. We first consider the case that both the input and output are one bit so that $\mathcal{M} = \mathcal{M}' = \{0, 1\}$. A simple example of reversible gate is NOT, which is defined as

$$0 \mapsto 1, \quad 1 \mapsto 0.$$  \hspace{1cm} (15)

This is clearly a bijection, and the reversal of NOT is also NOT. A simple example of irreversible gate is the information erasure that is referred to as ERASE:

$$0 \mapsto 0, \quad 1 \mapsto 0.$$  \hspace{1cm} (16)

This is not a bijection, as the logical state is always 0 after the computation; we cannot estimate the input state from the output state.

We next consider the case that both the input and output are two bits so that $\mathcal{M} = \mathcal{M}' = \{0, 1\}^2 = \{00, 01, 10, 11\}$. A simple example of reversible gate is CNOT (controlled-NOT), which is defined as

$$00 \mapsto 00, \quad 01 \mapsto 01, \quad 10 \mapsto 11, \quad 11 \mapsto 10,$$  \hspace{1cm} (17)

where the first (left) bit is the control bit and the second (right) bit is the target bit. If the control gate is 1, CNOT behaves as NOT on the target bit. Otherwise, CNOT is just identity. CNOT is a bijection and its reversal is also CNOT.

CNOT can be used for a measurement (i.e., the copy of information); if the input of the target bit is 0, the input of the control bit is copied to the output of the target bit by CNOT:

$$00 \mapsto 00, \quad 10 \mapsto 11.$$  \hspace{1cm} (18)

CNOT can also be used for feedback control, where the control bit is the feedback controller and the target bit is to be controlled. In the case of feedback control, we exchange the roles of the bits from the case of measurement (18); we regard the first (left) bit as the target bit and the second (right) bit as the control bit for feedback control. Before the feedback, the input states of the two bits are assumed to be the same, which is realized after the measurement. The target bit is then flipped if the control bit is 1:

$$00 \mapsto 00, \quad 11 \mapsto 01,$$  \hspace{1cm} (19)

where the output of the target bit is 0 irrespective of its input.

We next consider the case that the input is two bit and the output is one bit, where $\mathcal{M} = \{0, 1\}^2$ and $\mathcal{M}' = \{0, 1\}$. In this case, any computational process is irreversible. In fact, the numbers of the elements in $\mathcal{M}$ and $\mathcal{M}'$ are 4 and 2, respectively, and therefore any map from $\mathcal{M}$ to $\mathcal{M}'$ cannot be an injection. An example of such an irreversible gate is XOR, which is defined as

$$00 \mapsto 0, \quad 01 \mapsto 1, \quad 10 \mapsto 1, \quad 11 \mapsto 0.$$  \hspace{1cm} (20)
3.2. Reversible extension

We next show that any irreversible computation can be extended to a reversible computation, which has been discussed by Bennett in detail [9]. In fact, we can show the following theorem:

For any \( \hat{C} : \mathcal{M} \rightarrow \mathcal{M}' \), there exist a finite set \( \mathcal{M}'' \) and a map \( \hat{C}_{\text{ex}} : \mathcal{M} \rightarrow \mathcal{M}' \times \mathcal{M}'' \), such that \( \hat{C}_{\text{ex}} \) is logically reversible and the restriction of \( \hat{C}_{\text{ex}} \) on \( \mathcal{M} \rightarrow \mathcal{M}' \) is equivalent to \( \hat{C} \).

Here, \( \mathcal{M}'' \) can be regarded as an ancilla or an environment. We call \( \hat{C}_{\text{ex}} \) the reversible extension of \( \hat{C} \). We note that the reversible extension is not unique.

Before the proof of the theorem, we illustrate the reversible extension of ERASE. Let \( \mathcal{M}'' = \{0, 1\} \). We then define \( \hat{C}_{\text{ex}} : \{0, 1\} \rightarrow \{0, 1\}^2 \) by

\[
0 \mapsto 00, \quad 1 \mapsto 01,
\]

where the first (left) bit of the output is in \( \mathcal{M} \). \( \hat{C}_{\text{ex}} \) is clearly an injection and therefore reversible. Moreover, its restriction on \( \mathcal{M} \rightarrow \mathcal{M}' \) is equivalent to ERASE. Intuitively, this extension describes that the erased information can be kept in ancilla \( \mathcal{M}'' \) (i.e., can remain in the environment).

We now show a simple constructive proof of the theorem: we set \( \mathcal{M}'' = \mathcal{M} \) and define \( \hat{C}_{\text{ex}} \) as

\[
\hat{C}_{\text{ex}}(m) = (\hat{C}(m), m) \in \mathcal{M}' \times \mathcal{M}'',
\]

which satisfies the condition of the theorem.

The extension of ERASE (21) is a special case of extension (22). We note that extension (22) may be redundant in general. For example, in the case of XOR, the extension (22) is given by

\[
00 \mapsto 000, \quad 01 \mapsto 101, \quad 10 \mapsto 110, \quad 11 \mapsto 011,
\]

where \( \mathcal{M}'' = \{0, 1\}^2 \). However, there is a simpler reversible extension of XOR:

\[
00 \mapsto 00, \quad 01 \mapsto 10, \quad 10 \mapsto 11, \quad 11 \mapsto 01,
\]

where \( \mathcal{M}'' = \{0, 1\} \).

3.3. Entropy change in computation

We next consider the concept of logical entropy, which is defined by the Shannon entropy of the logical states.

We consider the probability distribution on \( \mathcal{M} \) (i.e., \( P[m] \) for \( m \in \mathcal{M} \)) with \( \sum_{m \in \mathcal{M}} P[m] = 1 \). After computation \( \hat{C} \), the probability distribution on \( \mathcal{M}' \) (i.e., \( P'[m'] \) for \( m' \in \mathcal{M}' \)) is given by

\[
P'[m'] = \sum_{m : \hat{C}(m) = m'} P[m],
\]

where \( \hat{C} : \mathcal{M} \rightarrow \mathcal{M}' \) is a map.
where the sum in the right-hand side is taken over $m$ satisfying $\hat{C}(m) = m'$. If the computation is reversible, equation (25) reduces to

$$P'[m'] = P[\hat{C}^{-1}(m')] \quad \text{(if } m' \in \hat{C}(\mathcal{M})),$$

$$P'[m'] = 0 \text{ (otherwise)},$$

which describes the conservation of probability.

We now define the initial and final logical entropies by [79, 80]

$$H(\mathcal{M}) := - \sum_{m \in \mathcal{M}} P[m] \ln P[m],$$

$$H'(\mathcal{M}') := - \sum_{m' \in \mathcal{M}'} P'[m'] \ln P'[m'].$$  \hspace{1cm} (27)

If the computation is logically reversible, the logical entropy states do not change. In fact, by using equation (26), we obtain

$$H'(\mathcal{M}') = - \sum_{m' \in \hat{C}(\mathcal{M})} P[\hat{C}^{-1}(m')] \ln P[\hat{C}^{-1}(m')]$$

$$= - \sum_{m \in \mathcal{M}} P[m] \ln P[m] = H(\mathcal{M}).$$  \hspace{1cm} (29)

In contrast, if the computation is logically irreversible, the logical entropy states is decreased:

$$H(\mathcal{M}) \geq H'(\mathcal{M}').$$  \hspace{1cm} (30)

In fact, by noting that $P'[m'] = \sum_{m \colon \hat{C}(m) = m'} P[m]$ (i.e., $P[m]/P'[m']$ is a probability distribution over the $m$ that satisfy $\hat{C}(m) = m'$), we obtain

$$H(\mathcal{M}) - H'(\mathcal{M}') = - \sum_{m'} P'[m'] \sum_{m \colon \hat{C}(m) = m'} \frac{P[m]}{P'[m']} \ln \frac{P[m]}{P'[m']} \geq 0.$$  \hspace{1cm} (31)

We note that inequality (30) is a special case of the data processing inequality in information theory [80].

In the case of the reversible extension (22), the logical entropy of the extended logical states does not change: $H(\mathcal{M}) = H'(\mathcal{M}' \times \mathcal{M}'')$. From the subadditivity of the Shannon entropy, we have

$$H'(\mathcal{M}' \times \mathcal{M}'') \leq H'(\mathcal{M}') + H'(\mathcal{M}''),$$  \hspace{1cm} (32)

and therefore, together with inequality (30),

$$H'(\mathcal{M}') \leq H(\mathcal{M}) \leq H'(\mathcal{M}') + H'(\mathcal{M}'').$$  \hspace{1cm} (33)

We discuss simple examples with $\mathcal{M} = \mathcal{M}' = \{0, 1\}$. Let $p := P[0]$ and $p' := P'[0]$ be the probabilities of the input 0 and output 0, respectively. In the case of NOT, $p' = 1 - p$ and therefore $H(\mathcal{M}) = -p \ln p - (1-p) \ln(1-p) = H'(\mathcal{M})$. In the case of ERASE, $p' = 1$ for any $p$. Therefore, $H(\mathcal{M}) = -p \ln p - (1-p) \ln(1-p)$ and $H'(\mathcal{M}') = 0$, which implies that the logical entropy is decreased by $H(\mathcal{M})$ by information erasure. In the case of the reversible extension of erasure (21), the equality in (32) is achieved with $H'(\mathcal{M}') = 0$ and $H'(\mathcal{M}'') = -p \ln p - (1-p) \ln(1-p)$. Therefore, the decrease in the entropy of the memory is compensated for by the increase in the entropy of the ancilla or environment, such that the total entropy is conserved.
4. Conventional Landauer principle

In this section, we discuss the conventional Landauer principle [7], and clarify the relationship between the thermodynamic and logical reversibilities in the standard setup of information erasure.

We consider the information erasure of one bit of information from a memory in the presence of a single heat bath at inverse temperature $\beta$. As a simple and conventional setup, we consider a binary symmetric potential as a physical model of the memory (figure 1(a)) [7, 8, 10], where the height of the barrier is assumed to be much larger than the thermal fluctuation. If the particle is in the left (right) well, the logical state is ‘0’ (‘1’). An idealized model of the double-well memory is shown in figure 1(b), where the left (right) box corresponds to the left (right) well, and the wall at the center of the box corresponds to the barrier of the binary potential. We note that these two models are not completely equivalent. In fact, the width of the barrier is finite in the double-well memory but infinitely small in the two-box memory. We also note that the height of the barrier in the double-well memory can be very large but is still finite theoretically, while the barrier in the two-box memory can be regarded as perfectly impenetrable.

Before the erasure, the probability of ‘0’ and ‘1’ are assumed to be equally $1/2$. After the erasure, the particle is in ‘0’ with unit probability. The logical entropy is $\ln 2$ before the erasure, while it is 0 after the erasure. Therefore, the logical entropy changes by $\Delta H = -\ln 2$ during the erasure. We assume that the initial probability distribution in the memory is thermal equilibrium, and that the final probability distribution is in conditional thermal equilibrium only in the well of ‘0’.

The crucial observation in the Landauer principle [7]–[11] is that the logical entropy must be included as a part of the total entropy, and be treated on an equal footing with the thermodynamic entropy. The change in the total entropy of the memory is then given by

$$\Delta S = \Delta H = -\ln 2,$$

and therefore, the second law (3) reduces to

$$-Q \geq \beta^{-1} \ln 2,$$

Figure 1. (a) Schematic of a memory with a binary symmetric potential. If the particle is in the left (right) well, the logical state is 0 (1). (b) A model of the memory with two boxes, which is an idealization of the binary-potential memory (a). If the particle is in the left (right) box, the logical state is 0 (1).
where $-Q$ is the heat that is emitted to the heat bath during the erasure. Inequality (35) is the conventional Landauer principle and the right-hand side is called the Landauer bound. Inequality (35) implies that the decrease in the entropy by $\ln 2$ in the memory must be compensated for by the increase in the entropy of the bath by at least $\ln 2$, which is accompanied by the inevitable heat emission of $-Q = \beta^{-1} \ln 2$. We note that the increase in the entropy of the bath corresponds to the increase in the entropy of ancilla $M''$ in terms of the reversible erasure (21) in section 3.2. The Landauer principle can then be summarized as follows:

*A positive amount of the heat emission is inevitable during the logically irreversible information erasure.*

We next consider the work needed for the erasure. Since the internal energy of the memory does not change during the erasure, we have $-Q = W$ from the first law of thermodynamics. Therefore, we obtain the minimal work needed for the information erasure:

$$W \geq \beta^{-1} \ln 2. \quad (36)$$

Inequality (36) is also referred to as the Landauer principle.

If the information erasure is quasi-static, the equalities in (35) and (36) can be achieved as

$$-Q = \beta^{-1} \ln 2, \quad W = \beta^{-1} \ln 2. \quad (37)$$

A concrete protocol of such a quasi-static erasure will be discussed below. Therefore, the total entropy production is given by

$$\Delta S_{\text{tot}} = (-\ln 2) - (-\ln 2) = 0, \quad (38)$$

which implies that the quasi-static erasure is thermodynamically reversible, while it is logically irreversible. We now conclude that:

*The logically irreversible erasure can be performed in a thermodynamically reversible manner in the quasi-static limit.*

This does not contradict the conventional Landauer principle. In fact, the logical reversibility is defined only by the reversibility of the logical states, which is related only to the logical entropy. In contrast, the thermodynamic reversibility is related to the reversibility of the relevant total system (i.e., the whole universe) including the heat bath, and to the total entropy production as discussed in section 2. Therefore, these two reversibilities are not equivalent in general. We note that, if the erasure is not quasi-static but is performed with a finite velocity, the erasure becomes thermodynamically irreversible. We summarize the thermodynamic and logical reversibilities for the information erasure in table 1.

Let us discuss the above point in more detail. During the information erasure, the entropy is transferred from the logical (i.e., accessible) degrees of freedom in the memory into the environmental (i.e., inaccessible or microscopic) degrees of freedom in
Thermodynamic and logical reversibilities revisited

Table 1. Summary for the conventional setup of information erasure.

|                        | Quasi-static | Finite velocity |
|------------------------|--------------|-----------------|
| Thermodynamically      | Reversible   | Irreversible    |
| Logically              | Irreversible | Irreversible    |
| Heat emission          | $=\beta^{-1}\ln 2$ | $>\beta^{-1}\ln 2$ |

The bath. Even if the erased information may still be kept in the degrees of freedom of the environment in principle, one cannot access or recognize in practice such dissipated information in the bath in a real computation. This implies that the erasure is logically irreversible (i.e., the accessibility to the stored information is lost). We note that the logical reversibility depends on the boundary between the logical and environmental degrees of freedom. For example, in terms of the reversible extension of the erasure (21), the erased information is kept in ancilla $\mathcal{M}'$. If all of the microscopic physical states in the whole universe were regarded as ‘logical states’, the information erasure could be logically reversible.

In contrast, the above-mentioned entropy transfer from the accessible to inaccessible degrees of freedom does not imply the thermodynamic irreversibility. The second law of thermodynamics is relevant only to the reversibility and irreversibility of the whole universe from the macroscopic point of view, as discussed in section 2. We stress that the thermodynamic reversibility does not depend on the boundary between the logical and environmental degrees of freedom. This shows the fundamental difference between the logical and thermodynamic reversibilities.

A simple model of the information erasure that achieves the Landauer bound is illustrated as follows. We consider the information erasure with the two-box memory shown in figure 2(a), which is an idealization of the information erasure with the binary-potential model as shown in figure 2(b). Let $V/2$ be the volume of each box. Before the erasure, the particle is in the left or right box with equal probability $1/2$. We then remove the wall without performing any work on the memory, and next compress the box from right to left quasi-statically and isothermally. The particle is then in the left box with unit probability. The work that is performed on the memory during the compression process is given by

$$W = \int_{V/2}^{V} \frac{\beta^{-1}}{V'} \, dV' = \beta^{-1} \ln 2,$$

where we have used the equation of states of the ideal gas with a single particle. Therefore, the Landauer bound (36) is achieved in this erasure protocol. By noting that $Q = -W$ holds in this case, the equality in (35) is also achieved.

We show in figure 2(c) the time-reversal of the foregoing protocol of the information erasure with the two-box memory. The logical state of the memory is initially 0 with unit probability, which is the same as that after the information erasure. We then expand the left box quasi-statically and isothermally, so that the volume of the box is doubled. During this process, we can extract $\beta^{-1} \ln 2$ of work, and $\beta^{-1} \ln 2$ of heat is absorbed by the
Thermodynamic and logical reversibilities revisited

Figure 2. (a) Information erasure with the two-box memory (figure 1(b)). The logical state of the memory is initially 0 or 1 with equal probability 1/2, which corresponds to one bit of information. To erase information, we remove the wall, and compress the box from right quasi-statically and isothermally. The logical state of the memory then becomes 0 with unit probability. While the information erasure is logically irreversible, this quasi-static erasure protocol achieves the Landauer bound (35), and is thermodynamically reversible. (b) Information erasure in the memory with the binary symmetric potential. The erasure in (a) with the two-box memory is an idealization of this erasure. (c) The time-reversal of the information erasure. The logical state is initially 0 with unit probability. It finally becomes 0 or 1 with equal probability 1/2, where the probability distribution is the same as that before the erasure. This process is the time-reversal of (a) in terms of the time evolution of the probability distribution.

memory from the bath. We next insert a wall at the center of the box, and the final logical state of the memory becomes 0 or 1 with equal probability 1/2. While the final logical state may be different from the pre-erasure logical state for the individual processes, the probability distribution of the final logical states is the same as that of the pre-erasure logical states. In fact, the protocol shown in figure 2(c) is the time-reversal of that in figure 2(a) in terms of the ensemble; the time evolutions of their probability distributions are the time-reversal of each other. We stress that the thermodynamic reversibility is defined in terms of the ensemble.

We discuss how to achieve the quasi-static limit in the information erasure, in particular for the removal process of the barrier. In the case of the two-box memory shown in figure 2(a), the velocity of the removal of the wall does not affect the probability distribution of the position of the particle, which is the same before and after the removal. Therefore, we can rapidly remove the wall even in the quasi-static limit. We note that, if
the particle is quantum, its wavefunction can be affected by the velocity of the removal of the wall [82].

On the other hand, in the case of the binary-potential memory shown in figure 2(b), the velocity of the removal of the barrier affects the probability distribution of the particle, because the barrier is not infinitely thin but has a finite width. In this case, one needs infinitely slow change in the height of the barrier to achieve the quasi-static limit. The higher the barrier is, the more time one needs to achieve the quasi-static limit, because the relaxation time over the two wells becomes exponentially larger as the barrier becomes higher. This makes it hard to achieve the quasi-static limit with the binary-potential memory in practice.

5. Thermodynamics of computation

In this section, we discuss the general formulation of thermodynamics of computation, and derive the general formulas that set the fundamental lower bounds of the work requirement and the heat emission during a computation. Moreover, we clarify the relationship between the thermodynamic and logical reversibilities in the general setup.

5.1. General Setup

We first consider the physical structure of the memory. In general, the logical states do not have one-to-one correspondence to the physical states of the memory; there may be a lot of microscopic physical states that correspond to a single logical state.

Let \( Y \) be the phase space of the memory, where each phase-space point \( y \in Y \) describes a microscopic physical state of the memory, and let \( M \) be the set of the possible input logical states. To relate the physical states to logical ones, we decompose \( Y \) into subspaces \( Y_m \) (\( m \in M \)), where \( Y_m \) and \( Y_n \) do not overlap with each other for \( m \neq n \), i.e., \( \bigcup_{m \in M} Y_m = Y \) and \( Y_m \cap Y_n = \emptyset \) (\( m \neq n \)) with \( \emptyset \) the empty set. If the phase-space point of the memory is in \( Y_m \) before the computation, we define that the input logical state is \( m \). We call \( Y_m \) an input logical subspace associated with \( m \).

Figure 3 shows simple examples of phase-space separations with binary potentials of the memory. Figure 3(a) shows a symmetric potential, which is the same as the memory illustrated in figure 1(a). In this case, 0 and 1 correspond to the left and right well, respectively, and the phase-space volume of \( Y_0 \) is the same as that of \( Y_1 \). Figure 3(b) shows an asymmetric potential, where 0 and 1 correspond to the left and right well, respectively. In this case, the phase-space volume of \( Y_0 \) is different from that of \( Y_1 \).

We consider probability distributions of the physical states in \( Y \). Let \( y \in Y \) be the initial phase-space point before the computation, \( P[y] \) be its probability, \( m \in M \) be the initial logical state such that \( y \in Y_m \), and \( P[m] \) be its probability that satisfies

\[
P[m] = \int_{y \in Y_m} dy P[y]. \tag{40}
\]

The probability distribution of \( y \) under the condition of \( m \) is written as \( P[y|m] \), which
Figure 3. (a) Separation of phase space $\mathcal{Y}$ into subspaces $\mathcal{Y}_0$ and $\mathcal{Y}_1$ with equal volumes, which respectively correspond to the left and right well of the binary symmetric potential. (b) Separation of phase space $\mathcal{Y}$ into subspaces $\mathcal{Y}_0$ and $\mathcal{Y}_1$ with different volumes, which respectively correspond to the left and right well of the asymmetric potential.

takes nonzero value only if $y \in \mathcal{Y}_m$. We then have

$$P[y] = \sum_{m \in \mathcal{M}} P[y|m]P[m]. \tag{41}$$

We also consider output logical states and phase-space points after the computation. Let $\mathcal{M}'$ be the set of the output logical states, and $\mathcal{Y}_{m'}$ be the logical subspace associated with $m' \in \mathcal{M}'$, where $\cup_{m' \in \mathcal{M}'} \mathcal{Y}_{m'} = \mathcal{Y}$ and $\mathcal{Y}_{m'} \cap \mathcal{Y}_{n'} = \phi$ ($m' \neq n'$). Let $y' \in \mathcal{Y}$ be the final phase-space point after the computation, $P'[y']$ be its probability, $m' \in \mathcal{M}'$ be the final logical state such that $y' \in \mathcal{Y}_{m'}$, and $P'[m']$ be its probability. The probability of $y'$ under the condition of $m'$ is written as $P'[y'|m']$, which satisfies

$$P'[y'] = \sum_{m' \in \mathcal{M}'} P'[y'|m']P'[m']. \tag{42}$$

5.2. Entropy balance

We next consider the changes in entropies during the computation. Before the computation, the Shannon entropy of the physical states is given by

$$S(\mathcal{Y}) = -\int_{y \in \mathcal{Y}} dy P[y] \ln P[y], \tag{43}$$

and the logical entropy of the input is given by

$$H(\mathcal{M}) = -\sum_{m \in \mathcal{M}} P[m] \ln P[m]. \tag{44}$$

The conditional Shannon entropy inside $\mathcal{Y}_m$ is given by

$$S(\mathcal{Y}|m) = -\int_{y \in \mathcal{Y}_m} dy P[y|m] \ln P[y|m], \tag{45}$$

whose ensemble average over $m \in \mathcal{M}$ is

$$S(\mathcal{Y}|\mathcal{M}) := \sum_m P[m]S(\mathcal{Y}|m). \tag{46}$$

doi:10.1088/1742-5468/2014/03/P03025
Due to a general formula in probability theory [80], these entropies satisfy
\[ S(Y) = H(M) + S(Y|M), \tag{47} \]
which implies that the total entropy is given by the sum of the logical entropy of \( M \) and the average of the conditional entropies of the \( Y_m \). Intuitively, the fluctuation over the whole phase space can be decomposed into that over the logical states and that over the internal physical states in the individual logical subspaces.

We also consider the entropies after the computation in the same manner:
\[ H'(M') = - \sum_{m' \in M'} P'[m'] \ln P'[m'], \tag{48} \]
\[ S'(Y) = - \int_{y' \in Y} P'[y'] \ln P'[y'], \tag{49} \]
\[ S'(Y|m') = - \int_{y' \in Y_{m'}} P'[y'|m'] \ln P'[y'|m'], \tag{50} \]
and
\[ S'(Y|M') := \sum_{m'} P'[m'] S'(Y|m'). \tag{51} \]
They also satisfy the decomposition formula:
\[ S'(Y) = H'(M') + S'(Y|M'). \tag{52} \]

Therefore, the total entropy change during the computation is decomposed as
\[ \Delta S = \Delta H + \Delta S_{\text{in}}, \tag{53} \]
where
\[ \Delta S := S'(Y) - S(Y), \tag{54} \]
\[ \Delta H := H'(M') - H(M), \tag{55} \]
\[ \Delta S_{\text{in}} := S'(Y|M') - S(Y|M). \tag{56} \]

Here, \( \Delta S \) is the change in the total entropy, \( \Delta H \) is the change in the logical entropy, and \( \Delta S_{\text{in}} \) is the average of the change in the physical entropy in the individual logical subspaces. We note that \( \Delta H \leq 0 \) holds for logically irreversible computations, while \( \Delta H = 0 \) holds for logically reversible computations.

As a special case, we consider the information erasure. We refer to one of the logical states in \( M' \) as the ‘standard state’, which we denote by \( 0 \in M' \). The information erasure is defined as the process in which the output logical state is in the standard state with unit probability (i.e., \( P'[0] = 1 \) and \( P'[m'] = 0 \) if \( m' \neq 0 \)) for any probability distribution \( P[m] \) of the input logical states. In this case, \( H'(M') = 0 \) holds by definition, and therefore \( \Delta H = -H(M') \). The change in the physical entropy is given by
\[ \Delta S_{\text{in}} = S'(Y'|0) - S(Y|M). \tag{57} \]
We note that \( \Delta S \neq -H(M') \) if \( \Delta S_{\text{in}} \neq 0 \).
To clarify the role of $\Delta S_{\text{in}}$, we consider a simple case with $M = M'$ and $\mathcal{Y}_m = \mathcal{Y}'_m$. We assume that the initial and final distributions inside the individual logical subspaces are the same (i.e., $P[y|m] = P'[y|m]$ for any $y$ and $m$), and therefore $S(\mathcal{Y}|m) = S'(\mathcal{Y}|m)$ holds for any $m$. On the other hand, the probability distribution over $M$ changes from $P[m]$ to $P'[m]$ during the computation. In this case, we have

$$
\Delta S_{\text{in}} = \sum_{m \in M} (P'[m] - P[m])S(\mathcal{Y}|m).
$$

(58)

If $S(\mathcal{Y}|m)$ does not depend on $m$, we have $\Delta S_{\text{in}} = 0$ for any $P[m]$ and $P'[m]$. For example, in the case of the symmetric memory in figure 3(a), $S(\mathcal{Y}|0) = S(\mathcal{Y}|1)$ holds if the conditional probability distributions in the individual wells are in thermal equilibrium. If $\Delta S_{\text{in}} = 0$, the internal fluctuations inside the individual logical subspaces do not contribute to the change in the total entropy (i.e., $\Delta S = \Delta H$).

In contrast, if $S(\mathcal{Y}|m)$ depends on $m$, $\Delta S_{\text{in}} \neq 0$ in general. For example, in the case of the asymmetric memory in figure 3(b), $S(\mathcal{Y}|0)$ and $S(\mathcal{Y}|1)$ are different to each other. If $\Delta S_{\text{in}} \neq 0$, the internal fluctuations inside the individual wells contribute to the change in the total entropy, and therefore $\Delta S \neq \Delta H$. The role of the asymmetry of the potential has been discussed in [31]–[35].

### 5.3. Generalized Landauer principle

We now consider the second law of thermodynamics for computation. We assume that the memory is attached to a single heat bath at inverse temperature $\beta$ during the computation. The total entropy production is given by equation (2), which leads to

$$
\Delta S_{\text{tot}} = \Delta H + \Delta S_{\text{in}} - \beta Q,
$$

(59)

where $Q$ is the average heat that is absorbed by the memory from the bath. Therefore, the second law (4) is given by

$$
\Delta H \geq \beta Q - \Delta S_{\text{in}},
$$

(60)

or equivalently,

$$
-\beta Q \geq -\Delta H - \Delta S_{\text{in}},
$$

(61)

which gives the fundamental lower bound of the heat emission into the bath during the computation. We refer to inequalities (60) and (61) as the generalized Landauer principle, and the right-hand side of (61) as the generalized Landauer bound. Several inequalities that are similar to or essentially equivalent to (61) have been obtained in [33]–[35].

In the special case of $\Delta S_{\text{in}} = 0$, inequality (61) reduces to

$$
-\beta Q \geq -\Delta H.
$$

(62)

In this case, if the computation is logically irreversible, the heat emission $-Q$ is non-negative because of $\Delta H \geq 0$.

In the case of the information erasure, the heat emission is bounded as

$$
-\beta Q \geq H(M) - \Delta S_{\text{in}},
$$

(63)

doi:10.1088/1742-5468/2014/03/P03025 18
where $\Delta S_{\text{in}}$ is the modification term to the original Landauer principle due to the change in the fluctuations inside the individual logical subspaces. If $\Delta S_{\text{in}} = 0$, we obtain

$$-\beta Q \geq H(M),$$

which is the conventional Landauer principle [7], [16]–[18].

In the conventional setup of the Landauer principle with a symmetric memory such as in figures 1(a) and 3(a), the decrease in the entropy of the logical states should be compensated for by the increase in the entropy of the bath, which is accompanied by at least $\beta^{-1}H(M)$ of heat emission into the bath. In contrast, in the case of the asymmetric memories such as figure 3(b), the decrease in the entropy of the logical states can be compensated for not only by the increase in the entropy of the bath, but also by that inside the individual logical subspaces. Here, the heat emission is determined only by the change in the entropy of the bath, but not by that inside the logical subspaces. Therefore, the lower bound of the heat emission can be different from $\beta^{-1}H(M)$ as shown in inequality (63); in particular, the heat emission can be smaller than $\beta^{-1}H(M)$ if $\Delta S_{\text{in}} > 0$.

To illustrate the above situation, we consider the situation with $M = M' = \{0, 1\}$ and $Y_m = Y_m'$ for $m = 0, 1$. We assume that $S(Y|0) = S'(Y|0)$ holds, which implies that the probability distribution inside the standard state is the same before and after the erasure. We then have

$$\Delta S_{\text{in}} = S(Y|0) - (P[0]S(Y|0) + P[1]S(Y|1))$$

$$= P[1] (S(Y|0) - S(Y|1)).$$

(65)

Therefore, if $P[1] \neq 0$ and $S(Y|0) > S(Y|1)$, we have $\Delta S_{\text{in}} > 0$. It is natural to assume that $S(Y|0) \neq S(Y|1)$ in asymmetric memories.

5.4. Thermodynamic and logical reversibilities

We now summarize the relationship between the thermodynamic and logical reversibilities in the general setup. On the basis of the argument in section 5.2, the conventional Landauer principle discussed in section 4 needs to be modified in general, and the generalized Landauer principle is stated as follows:

The decrease in the entropy of the logical states during a logically irreversible computation can be compensated for not only by the increase in the entropy of the heat bath, but also by that of the physical states inside the individual logical subspaces, where only the former determines the amount of the heat emission.

Since the generalized Landauer principle (60) is equivalent to the second law of thermodynamics (3), the equality in (60) can be achieved in the quasi-static limit (i.e., in the case of the quasi-static computation) where the total entropy production is zero. In this case, the computational process is thermodynamically reversible. Therefore, we conclude that:

Any logically irreversible computation can be performed in a thermodynamically reversible manner in the quasi-static limit, and therefore, the thermodynamic and logical reversibilities are not equivalent with each other.
These are generalizations of the arguments in section 4. We again stress that our observations here do not contradict the conventional Landauer principle; the logical reversibility is related to the change in the entropy of the logical states, while the thermodynamic reversibility is related to the change in the entropy of the whole universe that consists of the logical and physical states of the memory and the physical states of the heat bath.

5.5. Work requirement for computation

We next consider the thermodynamic work needed for computation. Let $E[y]$ with $y \in \mathcal{Y}$ be the initial Hamiltonian of the memory. We define the conditional free energy in logical subspace $\mathcal{Y}_m$ as

$$F_m := -\beta^{-1} \int_{y \in \mathcal{Y}_m} dy \, e^{-\beta E[y]}.$$  \hspace{1cm} (66)

We assume that the memory is initially in the conditional canonical distribution in the individual logical subspaces, which is given by

$$P[y|m] = e^{\beta (F_m - E[y])}$$ \hspace{1cm} (67)

for $y \in \mathcal{Y}_m$, and otherwise $P[y|m] = 0$. The total probability distribution is then given by

$$P[y] = \sum_{m \in \mathcal{M}} P[m] \chi(y,m)e^{\beta(F_m - E[y])},$$ \hspace{1cm} (68)

where $\chi(y,m)$ is the characteristic function which takes one if $y \in \mathcal{Y}_m$ and zero otherwise. The conditional entropy is then given by

$$S(\mathcal{Y}|m) = \beta(E_m - F_m),$$ \hspace{1cm} (69)

where

$$E_m := \int_{y \in \mathcal{Y}_m} dy \, P[y|m]E[y].$$ \hspace{1cm} (70)

Therefore, we obtain

$$S(\mathcal{Y}|\mathcal{M}) = \beta(E - F),$$ \hspace{1cm} (71)

where

$$E := \sum_{m \in \mathcal{M}} P[m]E_m = \int_{y \in \mathcal{Y}} dy \, P[y]E[y]$$ \hspace{1cm} (72)

is the average energy, and

$$F := \sum_{m \in \mathcal{M}} P[m]F_m$$ \hspace{1cm} (73)

is the average free energy.
We also consider the final Hamiltonian $E'[y']$ after the computation. Correspondingly, every argument about the final state is parallel to that about the initial one in the previous paragraph. To show that a quantity is about the final states after the computation, we use the notation of a prime.

We then have

$$\Delta S_{\text{in}} = \beta (\Delta E - \Delta F),$$

(74)

where $\Delta E := E' - E$ is the change in the average energy, and $\Delta F := F' - F$ is the change in the average free energy. We note that, if the final distribution is different from the conditional canonical distribution, we can show an inequality

$$\Delta S_{\text{in}} \leq \beta (\Delta E - \Delta F),$$

(75)

where the equality in (75) is achieved if and only if the output state is in the conditional canonical distribution. Therefore, the total entropy production (59) satisfies

$$\Delta S_{\text{tot}} \leq \beta (W - \Delta F) + \Delta H,$$

(76)

where $W$ is the work performed on the memory, and we have used the first law of thermodynamics

$$\Delta E = Q + W.$$

(77)

Therefore, by applying the second law (3), we obtain the generalized Landauer principle in terms of the work:

$$\beta W \geq -\Delta H + \beta \Delta F,$$

(78)

which gives the minimal work needed for the computation. In the case of information erasure, inequality (78) reduces to

$$\beta W_{\text{eras}} \geq H(\mathcal{M}) + \beta \Delta F,$$

(79)

where $\Delta F$ is the modification term to the conventional Landauer principle (36). As shown in equation (74), there are two contributions to $\Delta F$: the changes in the entropy and the energy inside the individual logical subspaces, where the former is $\Delta S_{\text{in}}$ and the latter is $\Delta E$. We can also regard $W - \Delta F$ as the energy cost needed for the information erasure, whose lower bound is given by the conventional Landauer bound $\beta^{-1} H(\mathcal{M})$. Several inequalities that are similar to or essentially equivalent to the generalized Landauer principle (79) have been obtained in [33]–[35].

In general, we call the memory as symmetric if $F_m$ does not depend on $m$. The memory shown in figure 3(a) is symmetric in this sense because of $F_0 = F_1$. In contrast, the memory shown in figure 3(b) is asymmetric because of $F_0 \neq F_1$. When $\mathcal{M} = \mathcal{M}', \mathcal{Y}_m = \mathcal{Y}'_{m'}$ with $m = m'$, $F_m = F_{m'}$, with $m = m'$, and $F_m$ does not depend on $m$, then we have $\Delta F = 0$ for any $P[m]$ and $P'[m]$.

As a simple example of information erasure with an asymmetric memory, we consider a model of the memory shown in figure 4(a) that is in contact with a heat bath at inverse temperature $\beta$. If the barrier is much higher than the thermal fluctuation, this model is
idealized by a memory with two boxes with different volumes as shown in figure 4(b). Let \( t : 1 - t \) \((0 < t < 1)\) be the ratio of the volumes of the boxes, which characterizes the ratio of the phase-space volumes inside the individual logical states. If the memory is symmetric, \( t = 1/2 \). We assume that \( \mathcal{M} = \mathcal{M}' = \{0, 1\} \) and \( \mathcal{Y}_m = \mathcal{Y}_m' \) with \( m = 0, 1 \), and write \( p := P[0] \). We note that \( t \) is in general different from \( p \), because the initial state is not necessarily in global thermal equilibrium over the whole phase space. The protocol of the information erasure is as follows (see also figure 4(c)). We first move the wall quasi-statically to the position where the ratio of the two volumes is given by \( p : 1 - p \). During this process, we perform the work of \( \beta^{-1}[p \ln(t/p) + (1 - p) \ln(1 - t/1 - p)] \) on average. We then remove the wall without any work. We next compress the box from the right quasi-statically, and the ratio of the two volumes is given by \( t : 1 - t \) in the final stage, where the logical state is ‘0’ with unit probability. During this compression, we perform the work of \(-\beta^{-1} \ln t \). The total work performed on the memory is then given by

\[
\beta W_{\text{eras}} = -p \ln p - (1 - p) \ln(1 - p) + \left(1 - p \ln \frac{1 - t}{t}\right). \tag{80}
\]

We note that the initial logical entropy is given by \( H(\mathcal{M}) = -p \ln p - (1 - p) \ln(1 - p) \). The free-energy difference between two logical states are given by \( F_0 - F_1 = \beta^{-1} \ln[(1-t)/t] \), and therefore \( \Delta F = \beta^{-1}(1 - p) \ln((1-t)/t) \). Therefore, equation (80) achieves the generalized Landauer bound in (79), which implies that this protocol of the information erasure is thermodynamically reversible.
Figure 5. (a) A memory with a single-well potential, whose logical state is only the standard state \( m' = 0 \). (b) A memory with a single box, which is an idealization of the single-well memory in (a). (c) Information erasure from the symmetric two-box memory (figure 1(b)) to the single-box memory (b). The initial logical state of the memory is \( m = 0 \) or \( 1 \) with equal probability \( 1/2 \), which corresponds to one bit of information. We then remove the wall so that the final logical state becomes \( m' = 0 \), which can be regarded as the information erasure that achieves the generalized Landauer bound in (79). (d) Information erasure where the initial logical state is \( 0 \) (1) with probability \( p \) \((1 - p)\). We first move the wall quasi-statically and isothermally so that the ratio of the volumes of the boxes becomes \( p : 1 - p \). We then remove the wall, and the final logical state becomes \( m' = 0 \) with unit probability. This erasure also achieves the generalized Landauer bound in (79).

We next consider a simple example of the information erasure with \( \mathcal{M} \neq \mathcal{M}' \). We assume that \( \mathcal{M} = \{0, 1\} \) and that the memory is initially binary symmetric as shown in figure 1(a), which is idealized by the two-box memory with equal volumes as shown in figure 1(b). We then assume that \( \mathcal{M}' = \{0\} \) in the final stage, i.e., the output logical state is only the standard state; the corresponding potential model is shown in figure 5(a). This is idealized by a memory with a single box as shown in figure 5(b). For simplicity, we set \( P[0] = 1/2 \). In this case, the information erasure is just the removal of the wall as shown in figure 5(c), and therefore \( W_{\text{eras}} = 0 \). This erasure protocol achieves the generalized Landauer bound in (79), since \( H(\mathcal{M}) = \ln 2, F_0 - F_0' = F_1 - F_1' = \beta^{-1} \ln 2 \), and \( \Delta F = -\beta^{-1} \ln 2 \). Therefore, this erasure protocol is thermodynamically reversible. In fact, if we insert a wall at the center of the single box after the erasure, the probability distribution of the logical states returns to the initial one, where a particle is in one of the two boxes with equal probability \( 1/2 \).
If \( p := P[0] \neq 1/2 \) in the above model, we quasi-statically move the wall to the position where the ratio of the volumes \( p : 1 - p \), and then remove the wall (see all figure 5(d)). In this case, the work performed on the memory is given by \( \beta W_{\text{er}} = - \ln 2 - p \ln p - (1 - p) \ln(1 - p) \). Since \( H(M) = - p \ln p - (1 - p) \ln(1 - p) \) and \( \beta \Delta F = - \ln 2 \), this erasure protocol also achieves the generalized Landauer bound in (79).

6. Two memories

In this section, we consider the cases that the memory consists of two sub-memories. In particular, we discuss measurement and feedback control between the two sub-memories, which constitute the typical setup of Maxwell’s demon [1].

6.1. General argument

We assume that any logical state in \( M \) is a pair of two logical states of sub-memories that we refer to as S and D; any input logical state \( m \in M \) is written as \( m = (s, d) \) where \( s \) and \( d \) are the logical states of memories S and D, respectively. Let \( M^S (M^D) \) be the set of logical states of S (D) with \( M = M^S \times M^D \), where \( s \in M^S \) and \( d \in M^D \).

We also consider the physical states of the memories. Let \( Y^S (Y^D) \) be the set of physical states of S (D) with \( Y = Y^S \times Y^D \). Let \( y := (u, v) \) for any initial physical state \( y \in Y \) with \( u \in Y^S \) and \( v \in Y^D \). Let \( Y_m := Y^S_s \times Y^D_d \).

The conditional probability of \((u, v)\) under the condition of \((s, d)\) is given by \( P[u, v|s, d] \) that takes nonzero value only if \( u \in Y^S_s \) and \( v \in Y^D_d \). By applying the argument in section 5.1 to the present situation, we have

\[
P[u, v] = \sum_{(s, d) \in M} P[u, v|s, d]P[s, d], \tag{81}
\]

whose Shannon entropy is given by equation (47).

The mutual information plays a crucial role in the presence of two memories; it characterizes the correlation between them [79, 80]. The mutual information between the physical states and that between the logical states are respectively given by

\[
I(Y^S : Y^D) = S(Y^S) + S(Y^D) - S(Y), \tag{82}
\]

\[
I(M^S : M^D) = H(M^S) + H(M^D) - H(M). \tag{83}
\]

The mutual information between the internal states is

\[
I(Y^S : Y^D|s, d) = S(Y^S|s) + S(Y^D|d) - S(Y|s, d), \tag{84}
\]

whose ensemble average over \((s, d)\) is given by

\[
I(Y^S : Y^D|M) := \sum_{(s, d) \in M} I(Y^S : Y^D|s, d)P[s, d]
= S(Y^S|M^S) + S(Y^D|M^D) - S(Y|M). \tag{85}
\]

doi:10.1088/1742-5468/2014/03/P03025
Therefore, we obtain

\[ I(\mathcal{Y}^S : \mathcal{Y}^D) = I(\mathcal{M}^S : \mathcal{M}^D) + I(\mathcal{Y}^S : \mathcal{Y}^D | \mathcal{M}), \tag{86} \]

which implies that the total correlation between the two memories can be decomposed into the correlation between their logical states and that between their internal physical states in the individual logical subspaces. We note that \( I(\mathcal{Y}^S : \mathcal{Y}^D | \mathcal{M}) = 0 \) if \( P[u, v|s, d] = P[u|s]P[v|s, d] \). We also note that

\[
\begin{align*}
S(\mathcal{Y}^S) &= H(\mathcal{M}^S) + S(\mathcal{Y}^S | \mathcal{M}^S), \tag{87} \\
S(\mathcal{Y}^D) &= H(\mathcal{M}^D) + S(\mathcal{Y}^D | \mathcal{M}^D). \tag{88}
\end{align*}
\]

The sum of equations (86), (87), and (88) leads to equation (47).

We also consider the probability distributions, the Shannon entropy, and the mutual information after the computation in the parallel manner to those before the computation; for example, we write \( \mathcal{M}': = \mathcal{M}_S \times \mathcal{M}_D, \) \( m' := (s', d') \) with \( s' \in \mathcal{M}_S^S \) and \( d' \in \mathcal{M}_D^D \). To show that a quantity is about the states after computation, we use the notation of a prime.

The change in the total Shannon entropy is then given by

\[ \Delta S = \Delta H^S + \Delta H^D - \Delta I + \Delta S^S_{in} + \Delta S^D_{in} - \Delta I_{in}, \tag{89} \]

where

\[
\begin{align*}
\Delta S &:= S'(\mathcal{Y}') - S(\mathcal{Y}), \\
\Delta H^S &:= H'(\mathcal{M}^S) - H(\mathcal{M}^S), \quad \Delta H^D := H'(\mathcal{M}^D) - H(\mathcal{M}^D), \\
\Delta I &:= I'(\mathcal{M}^S : \mathcal{M}^D) - I(\mathcal{M}^S : \mathcal{M}^D), \\
\Delta S^S_{in} &:= S'(\mathcal{Y}^S | \mathcal{M}^S) - S(\mathcal{Y}^S | \mathcal{M}^S), \quad \Delta S^D_{in} := S'(\mathcal{Y}^D | \mathcal{M}^D) - S(\mathcal{Y}^D | \mathcal{M}^D), \\
\Delta I_{in} &:= I'(\mathcal{Y}^S : \mathcal{Y}^D | \mathcal{M}') - I(\mathcal{Y}^S : \mathcal{Y}^D | \mathcal{M}). \tag{90-94}
\end{align*}
\]

Here, \( \Delta H^S \) and \( \Delta H^D \) describe the changes in the logical entropies, \( \Delta I \) describes the change in the logical correlation, \( \Delta S^S_{in} \) and \( \Delta S^D_{in} \) describe the change in the internal physical entropies in the individual logical subspaces, and \( \Delta I_{in} \) describes the change in the internal correlation between the individual logical subspaces. Therefore, the total entropy production is given by

\[ \Delta S_{tot} = \Delta H^S + \Delta H^D - \Delta I + \Delta S^S_{in} + \Delta S^D_{in} - \Delta I_{in} - \beta Q. \tag{95} \]

We next assume that the energies of two memories are additive before the computation, \( E[y] = E^S[u] + E^D[v] \), where \( E^S[u] \) (\( E^D[v] \)) is the Hamiltonian of S (D). This assumption implies that the interaction Hamiltonian between the memories is negligible before the computation. The corresponding free energies are given by

\[
\begin{align*}
F^S_s &:= -\beta^{-1} \ln \int_{u \in \mathcal{Y}_S} e^{-\beta E^S[u]}, \quad F^D_d := -\beta^{-1} \ln \int_{v \in \mathcal{Y}_D} e^{-\beta E^D[v]}. \tag{96}
\end{align*}
\]

In this case, the conditional distribution is given by

\[ P[u, v|s, d] = P[u|s]P[v|d]. \tag{97} \]

doi:10.1088/1742-5468/2014/03/P03025
where, if \( u \in Y_s \) and \( v \in Y_d \),
\[
P[u|s] = e^{\beta(F_s^u - E_s[u])}, \quad P[v|d] = e^{\beta(F_d^v - E_d[v])}.
\]
(98)

We also consider the energies in the memories after the computation, where the energies are also assumed to be additive. The argument about energy and the free energy after the computation is parallel to that before the computation. To show that a quantity is about the states after the computation, we use the notation of a prime. We note that the energies are not necessarily additive during the computation, since the interaction Hamiltonian between the memories may become nonzero during the computation.

If the initial probability distribution is the conditional probability distribution, the total entropy production satisfies equation (76). Therefore, we obtain the lower bound of the work that is needed for the computation with two memories:
\[
W \geq -\Delta H + \Delta F;
\]
(99)
where
\[
\Delta H = \Delta H^S + \Delta H^D - \Delta I,
\]
(100)
\[
\Delta F := \Delta F^S + \Delta F^D,
\]
(101)
\[
\Delta F^S := \sum_{s \in \mathcal{M}^S} P[s] F_s^S, \quad \Delta F^D := \sum_{d \in \mathcal{M}^D} P[d] F_d^D.
\]
(102)

In the special case of \( \Delta H^S = \Delta H^D = 0 \) and \( \Delta F = 0 \), we have
\[
-W \leq -\Delta I,
\]
(103)
which is the work extraction by using the correlation, which can be demonstrated by Maxwell’s demon [1].

6.2. Measurement and feedback

In this section, we consider measurement and feedback as a special case of the argument in section 6.1 with the assumption of the initial and final conditional canonical distributions.

We first consider a measurement process, where memory D performs a measurement on memory S. In other words, the information in memory S is copied to memory D. The logical state of D is initially the standard state \( 0^D \) with unit probability, and the logical entropy of S is initially given by \( H^S \). We assume that the initial logical entropy in D and the initial mutual information between the logical states are both zero. We also assume that the two memories are in the conditional canonical distributions before and after the measurement.

The two memories interact with each other, and make a correlation between the logical states. We assume that the logical state of S does not change in time during the measurement. When both of the two memories are one bit, a typical example of such measurement is given by CNOT (18), where the measurement is error-free (i.e., the copy of information is perfect).
We denote by $I$ the correlation between the logical states after the measurement, and denote by $H^D$ the logical entropy of D after the measurement. We note that the mutual information satisfies the following inequalities [79, 80]:

$$0 \leq I \leq H^S, \quad 0 \leq I \leq H^D.$$  

(104)

Applying equation (76) to the measurement, the total entropy production is given by

$$\Delta S_{\text{tot}} = \beta (W_{\text{meas}} - \Delta F^D_{\text{meas}}) + H^D - I,$$

(105)

where $W_{\text{meas}}$ is the work performed on the memories, and $\Delta F^D_{\text{meas}}$ is the change in the average free energy of D. Therefore, we obtain

$$W_{\text{meas}} \geq I - H^D + \Delta F^D_{\text{meas}},$$

(106)

which gives the minimal work needed for the measurement. Several inequalities that are essentially equivalent to (106) have been obtained in [35, 65, 71].

If $I = H^D$ holds and memory D is symmetric (i.e., $\Delta F^D_{\text{meas}} = 0$), inequality (106) reduces to

$$W_{\text{meas}} \geq 0,$$

(107)

which is the bound discussed by Bennett [10]. Since the energies of the memories do not change in time during the measurement if the memory is symmetric, inequality (107) is equivalent to

$$-Q_{\text{meas}} \geq 0,$$

(108)

where $Q_{\text{meas}}$ is the heat absorbed by the memories during the measurement. Inequalities (107) and (108) imply that the lower bounds of the work requirement and the heat emission during the measurement are both zero.

We consider a simple example where the logical states of S and D before and after the measurement are all one bit, which is a conventional setup of the measurement [10]. We assume that S is symmetric (i.e., $F_0^S = F_1^S$) before and after the measurement; in this case, memory S can be modeled by two boxes (see also figure 6(a)). Before the measurement, the logical state $(s, d)$ is $(0, 0)$ or $(1, 0)$ with equal probability 1/2. The measurement protocol is given by CNOT where S is the control bit and D is the target bit, and then the final logical state is $(0, 0)$ or $(1, 1)$ with equal probability 1/2. In this case, $I = H^D = \ln 2$ and $\Delta F^D_{\text{meas}} = 0$. Therefore, $W_{\text{meas}} = 0$ and $-Q_{\text{meas}} = 0$ are achieved in the quasi-static limit, where the measurement is thermodynamically reversible as well as logically reversible.

We summarize the thermodynamic and logical reversibilities for the conventional setup of measurement in table 2, which is contrastive to table 1 for the information erasure.

We next consider a feedback control process followed by the measurement, where memory D performs feedback control on memory S. This situation can be regarded as a typical setup of Maxwell’s demon (see also section 6.3), where D is the demon and S is the engine to be controlled. We assume that the two memories are in the conditional canonical distributions before and after the feedback control. The logical state of S changes depending on the logical state of D after the measurement. We assume that memory D
Figure 6. (a) Model of measurement with two symmetric memories. The two memories are initially not correlated. The logical state of D quasi-statically changes only if the logical state of S is initially in 1, and the memories finally have one bit of correlation. This measurement achieves the equality in (107), and is thermodynamically reversible. (b) Model of feedback control, where the two symmetric memories initially share one bit of correlation. The logical state of S quasi-statically changes only if the logical state of D is initially in 1. Due to the feedback control, the initial correlation vanishes and the logical entropy of S is decreased by one bit. This feedback protocol achieves the equality in (110), and is thermodynamically reversible. (c) Model of feedback control, where the two symmetric memories initially share one bit of correlation. The left (right) box of S is expanded quasi-statically and isothermally if the initial logical state of D is 0 (1), so that the final logical subspace of S is the whole phase space. Due to the feedback control, the initial correlation vanishes and \( \beta^{-1} \ln 2 \) of work is extracted from S, which is equivalent to the case of the Szilard engine. This feedback protocol also achieves the equality in (110), and is thermodynamically reversible.

Table 2. Summary for the conventional setup of measurement.

|                      | Quasi-static | Finite velocity |
|----------------------|--------------|-----------------|
| Thermodynamically    | Reversible   | Irreversible    |
| Logically            | Reversible   | Reversible      |
| Heat emission        | =0           | >0              |

does not evolve in time during the feedback control. For simplicity, we assume that the final correlation between the logical states between the memories is zero after the feedback control, such that the change in the logical mutual information is given by \(-I\). We also assume that the output logical state of S is the standard state 0\(^S\) with unit probability after feedback control such that the change in the logical entropy of S is given by \(-H^S\). When both of the two memories are one bit, a typical example of such feedback control is also given by CNOT (19).
Applying equation (76) to the feedback control, the total entropy production is given by

$$\Delta S_{\text{tot}} = \beta (W_{\text{fb}} - \Delta F^S_{\text{fb}}) - H^S + I,$$

(109)

where $W_{\text{fb}}$ is the work performed on the memories, and $\Delta F^S_{\text{fb}}$ is the change in the average free energy of S. Therefore, we obtain

$$W_{\text{fb}} \geq -I + H^S + \Delta F^S_{\text{fb}},$$

(110)

which gives the minimal work needed for the feedback control. Several inequalities that are essentially equivalent to (110) have been obtained in [51, 54, 61, 65, 71]; in these previous works, $H^S + \Delta F^S_{\text{fb}}$ is denoted by, for example, just $\Delta F^S_{\text{fb}}$.

We consider a simple situation where the logical states of S and D before and after feedback are all one bit. We assume that S is symmetric (i.e., $F^S_0 = F^S_1$) before and after the feedback (see also figure 6(b)). Before feedback, the logical state $(s,d)$ is $(0,0)$ or $(1,1)$ with equal probability 1/2. The feedback protocol is given by CNOT that is logically reversible, where D is the control bit and S is the target bit, and then the final logical state is $(0,0)$ or $(0,1)$ with equal probability 1/2. In this case, $I = H^S = \ln 2$ and $\Delta F^S_{\text{fb}} = 0$ hold. Therefore, inequality (110) reduces to $W_{\text{fb}} \geq 0$. In the quasi-static limit, $W_{\text{fb}} = 0$ is achieved, which implies that we can reduce the entropy of S (i.e., $H^S = \ln 2$) without performing any positive amount of work on S. We note that, in the conventional thermodynamics, we need a positive amount of work to isothermally decrease the entropy. In contrast, in the present situation, the mutual information plays the role of the resource to decrease the entropy of S.

We next consider another situation that there is a single logical state $0^S$ in S after the feedback, which corresponds to the whole phase space of S (see also figure 6(c)). In this case, $I = H^S = \ln 2$ and $\Delta F^S_{\text{fb}} = -\beta^{-1} \ln 2$ hold. Therefore, inequality (110) reduces to $W_{\text{fb}} \geq -\beta^{-1} \ln 2$. In the quasi-static limit, $W_{\text{fb}} = -\beta^{-1} \ln 2$ is achieved, where we extract $\beta^{-1} \ln 2$ of work by the feedback control; the mutual information is the resource of the work extraction. This situation is equivalent to the case of the conventional Szilard engine [3].

### 6.3. On Maxwell’s demon

We consider Maxwell’s demon as a special example of our general argument. A typical situation of Maxwell’s demon consists of measurement and feedback control. In the setup of section 6.2, memory D plays the role of the demon and memory S is the system to be measured and controlled. By summing up inequalities (106) and (110), we obtain

$$W_{\text{meas}} + W_{\text{fb}} \geq H^S - H^D + \Delta F^S_{\text{meas}} + \Delta F^D_{\text{fb}},$$

(111)

where the contribution of the mutual information vanishes in the right-hand side. If we perform the information erasure from D after the feedback control, the total work is given by

$$W_{\text{meas}} + W_{\text{fb}} + W_{\text{eras}} \geq H^S + \Delta F^S_{\text{meas}},$$

(112)

where we have used the generalized Landauer principle (79). Here, $H^S + \Delta F^S_{\text{meas}}$ in the right-hand side can be regarded as an effective free energy of S, which vanishes in, for
example, the case of the Szilard engine discussed in section 6.2. If $H^S + \Delta F^S_{\text{meas}} = 0$ holds, inequality (112) reduces to

$$W_{\text{meas}} + W_{\text{fb}} + W_{\text{eras}} \geq 0,$$

which implies that we cannot extract any work from the entire process.

We stress that, before the information erasure, the total entropy productions (105) and (109) are always non-negative (i.e., $\Delta S_{\text{tot}} \geq 0$) for the individual processes of measurement and feedback. This confirms that measurement and feedback control are individually consistent with the second law of thermodynamics, without considering the information erasure.

As shown in section 6.2, both of measurement and feedback are logically reversible in typical situations, and can be thermodynamically reversible in the quasi-static limit. On the other hand, the information erasure is logically irreversible, while it can be thermodynamically reversible in the quasi-static limit. Therefore, the entire process of Maxwell’s demon including the erasure is logically irreversible, but can be thermodynamically reversible in the quasi-static limit, where $\Delta S_{\text{tot}} = 0$ holds for the individual processes of the measurement, feedback control, and information erasure.

7. Conclusions

In this paper, we have discussed the relationship between computation and the second law of thermodynamics. In particular, we have clarified the fundamental relationship between the thermodynamic and logical reversibilities.

In section 2, we discussed the concept of thermodynamic reversibility. A physical process is reversible if and only if its time-reversal is not prohibited by the second law of thermodynamics. To be more precise, a physical process is thermodynamically reversible if the total entropy production in the whole universe (2) is zero during the process (i.e., $\Delta S_{\text{tot}} = 0$). The total entropy production consists of the increases in the entropy of the system and that of the heat bath.

In section 3, we discussed the concept of logical reversibility. A computational process is logically reversible if and only if it is an injection. The Shannon entropy of the logical states decreases if a computation is irreversible, while it does not change if a computation is reversible. A typical example of irreversible computation is the information erasure.

In section 4, we discussed the conventional setup of the information erasure with a binary symmetric memory (figure 1). On the basis of the second law of thermodynamics, we have confirmed the conventional Landauer principle, which states that at least $\beta^{-1} \ln 2$ of heat should be emitted into the heat bath during the information erasure of one bit of information. We have clarified that the information erasure can be thermodynamically reversible in the quasi-static limit, where the heat emission equals $\beta^{-1} \ln 2$. The crucial observation here is that the thermodynamic reversibility is related to the entropy production in the whole universe, while the logical reversibility is related only to the change in the entropy of the logical states. Such a thermodynamically reversible erasure is illustrated in figure 2.
In section 5, we discussed the general theory of the thermodynamics of computation. In particular, we derived a generalized Landauer principle (61). In the case of an asymmetric memory such as in figures 3(b) and figure 4(a), the lower bound of the heat emission should be modified from the original Landauer bound; the second term in the right-hand side of (61) originates from the asymmetry of the memory. In fact, the amount of the heat emission is determined by the increase in the entropy of the heat bath, while the change in the entropy of the logical states of the memory can be compensated for not only by the change in the entropy of the heat bath, but also by that of the internal physical states of the memory if it is asymmetric. We have clarified the general relationship between the thermodynamic and logical reversibilities, which is summarized in section 5.4; any logically irreversible computation can be thermodynamically reversible in the quasi-static limit.

In section 6, we discussed the case that there are two memories, and derived a generalized Landauer principle for two memories, which includes the mutual information. In particular, we discussed measurement and feedback control, which play crucial roles in the case of Maxwell’s demon.

Our results quantitatively extend the conventional Landauer principle to general situations including asymmetric memories, and clarify the subtle and fundamental relationship between the thermodynamic and logical reversibilities. The derived inequalities are based on the second law of thermodynamics, where the information contents are included as well as thermodynamic quantities. The thermodynamics of computation could serve as the foundation to analyze computation at the level of thermal fluctuations with, for example, artificial nanodevices and biological nanomachines.

Acknowledgments

The author is grateful to Masahito Ueda, Jordan M Horowitz, and Juan M R Parrondo for valuable discussions, and to Naoto Shiraishi for careful reading of the paper. This work was supported by JSPS KAKENHI Grant Nos 25800217 and 22340114, and by Platform for Dynamic Approaches to Living System from MEXT, Japan.

References

[1] Leff H S and Rex A F (ed), 2003 Maxwell’s Demon 2: Entropy, Classical and Quantum Information, Computing (Princeton, NJ: Princeton University Press)
[2] Maxwell J C, 1871 Theory of Heat (London: Appleton)
[3] Szilard L, 1929 Z. Phys. 53 840
[4] Brillouin L, 1951 J. Appl. Phys. 22 334
[5] Brillouin L, 1956 Science and Information Theory (New York: Academic)
[6] Toyabe S, Sagawa T, Ueda M, Muneyuki E and Sano M, 2010 Nature Phys. 6 988
[7] Landauer R, 1961 IBM J. Res. Dev. 5 183
[8] Landauer R, 1996 Science 272 1914
[9] Bennett C H, 1973 IBM J. Res. Dev. 17 525
[10] Bennett C H, 1982 Int. J. Theor. Phys. 21 905
[11] Bennett C H, 2003 Stud. Hist. Phil. Mod. Phys. 34 501
[12] Zurek W H, 1989 Nature 341 119
[13] Zurek W H, 1989 Phys. Rev. A 40 4731

doi:10.1088/1742-5468/2014/03/P03025
Thermodynamic and logical reversibilities revisited

[14] Plenio M B and Vitelli V, 2001. Contemporary Phys. 42 25
[15] Maruyama K, Nori F and Vedral V, 2009 Rev. Mod. Phys. 81 1
[16] Shizume K, 1995 Phys. Rev. E 52 3495
[17] Pechcinska B, 2000 Phys. Rev. A 61 062314
[18] Esposito M and Van den Broeck C, 2011 Europhys. Lett. 95 40004
[19] Dillenschneider R and Lutz E, 2009 Phys. Rev. Lett. 102 210601
[20] Hilt S, Shabir S, Anders J and Lutz E, 2011 Phys. Rev. E 83 030102
[21] Lambson B, Carlton D and Bokor J, 2011 Phys. Rev. Lett. 107 010604
[22] Aurell E, Gawędzki K, Mejia-Monasterio C, Mohayaee R and Muratore-Ginanneschi P, 2012 J. Stat. Phys. 147 487
[23] Diana G, Bagci G B and Esposito M, 2013 Phys. Rev. E 87 012111
[24] Bérut A, Arakelyan A, Petrosyan A, Ciliberto S, Dillenschneider R and Lutz E, 2012 Nature 483 187
[25] Bérut A, Petrosyan A and Ciliberto S, 2013 Europhys Lett. 103 60002
[26] Matsueda H, Goto E and Loe K-F, 1997 RIMS Kôkyûroku 1013 187
[27] Allahverdyan A E and Nieuwenhuizen T M, 2001 Phys. Rev. E 64 0561171
[28] Horhammer C and Buttner H, 2008 J. Stat. Phys. 133 1161
[29] Vaccaro J A and Barnett S M, 2011 Proc. R. Soc. A 467 1770
[30] Maroney O J E, 2005 Stud. Hist. Phil. Mod. Phys. 36 355
[31] Norton J D, 2005 Stud. Hist. Phil. Mod. Phys. 36 375
[32] Barkeshli M M, 2005 arXiv:cond-mat/0504323
[33] Maroney O J E, 2009 Phys. Rev. E 79 031105
[34] Turgut S, 2009 Phys. Rev. E 79 041102
[35] Sagawa T and Ueda M, 2009 Phys. Rev. Lett. 102 250602
[36] Sagawa T and Ueda M, 2011 Phys. Rev. Lett. 106 189901 (erratum)
[37] Jarzynski C, 1997 Phys. Rev. Lett. 78 2690
[38] Crooks G E, 1998 J. Stat. Phys. 90 1481
[39] Crooks G E, 1999 Phys. Rev. E 60 2721
[40] Jarzynski C, 2000 J. Stat. Phys. 98 77
[41] Seifert U, 2005 Phys. Rev. Lett. 95 040602
[42] Kawai R, Parrondo J M R and Van den Broeck C, 2007 Phys. Rev. Lett. 98 080602
[43] Crooks G E, 2011 J. Stat. Mech. P07008
[44] Jarzynski C, 2013 Prog. Math. Phys. 63 145
[45] Kurchan J, 2000 arXiv:cond-mat/0007360
[46] Tasaki H, 2000 arXiv:cond-mat/0009244
[47] Campisi M, Hänggi P and Talkner P, 2011 Rev. Mod. Phys. 83 771
[48] Sagawa T, 2012 arXiv:1202.0983

As a chapter of: Nakahara M and Tanaka S (ed), 2012 Lectures on Quantum Computing, Thermodynamics and Statistical Physics (Singapore: World Scientific)

[49] Nielsen M A, Caves C M, Schumacher B and Barnum H, 1998 Proc. R. Soc. A 454 277
[50] Touchette H and Lloyd S, 2000 Phys. Rev. Lett. 84 1156
[51] Touchette H and Lloyd S, 2004 Physica A 331 140
[52] Sagawa T and Ueda M, 2008 Phys. Rev. Lett. 100 080403
[53] Cao F J and Feito M, 2009 Phys. Rev. E 79 041118
[54] Jacobs K, 2009 Phys. Rev. A 80 012322
[55] Sagawa T and Ueda M, 2010 Phys. Rev. Lett. 104 090602
[56] Fujitani Y and Suzuki H, 2010 J. Phys. Soc. Japan 79 104003
[57] Horowitz J M and Vaikuntanathan S, 2010 Phys. Rev. E 82 061120
[58] Horowitz J M and Parrondo J M R, 2011 Europhys. Lett. 95 10005
[59] Morikuni Y and Tasaki H, 2011 J. Stat. Phys. 143 1
[60] Abreu D and Seifert U, 2011 Europhys. Lett. 94 10001
[61] Horowitz J M and Parrondo J M R, 2011 New J. Phys. 13 123019
[62] Sagawa T and Ueda M, 2012 Phys. Rev. E 85 021104
[63] Abreu D and Seifert U, 2012 Phys. Rev. Lett. 108 030601
[64] Stiller S, Sivak D A, Dell A J and Crooks G E, 2012 Phys. Rev. Lett. 109 120604
[65] del Rio L, Aberry J, Remmer R, Dahlsten O and Vedral V, 2011 Nature 474 61
[66] Sagawa T and Ueda M, 2012 Phys. Rev. Lett. 109 180602
[67] Granger L and Kantz H, 2013 Europhys. Lett. 101 50004
[68] Ito S and Sagawa T, 2013 Phys. Rev. Lett. 111 180603
[69] Deffner S and Jarzynski C, 2013 Phys. Rev. X 3 041003

doi:10.1088/1742-5468/2014/03/P03025
Thermodynamic and logical reversibilities revisited

[69] Tasaki H, 2013 arXiv:1308.3776
[70] Horowitz J M, Sagawa T and Parrondo J M R, 2013 Phys. Rev. Lett. 111 010602
[71] Sagawa T and Ueda M, 2013 New J. Phys. 15 125012
[72] Mandal D and Jarzynski C, 2012 Proc. Nat. Acad. Sci. USA 109 11641
[73] Barato A C and Seifert U, 2013 Europhys. Lett. 101 60001
[74] Strasberg P, Schaller G, Brandes T and Esposito M, 2013 Phys. Rev. Lett. 110 040601
[75] Mandal D, Quan H T and Jarzynski C, 2013 Phys. Rev. Lett. 111 030602
[76] Barato A C and Seifert U, 2013 arXiv:1308.4598
[77] Callen H B, 1985 Thermodynamics and an Introduction to Thermostatistics 2nd edn (New York: Wiley)
[78] Kondepudi D and Prigogine I, 1998 From Heat Engines to Dissipative Structures (New York: Wiley)
[79] Shannon C, 1948 Bell Syst. Technical J. 27 379
  Shannon C, 1948 Bell Syst. Technical J. 27 623
[80] Cover T M and Thomas J A, 1991 Elements of Information Theory (New York: Wiley)
[81] Moore C and Mertens S, 2011 The Nature of Computation (Oxford: Oxford University Press)
[82] Kim S W, Sagawa T, Liberato S D and Ueda M, 2011 Phys. Rev. Lett. 106 070401

doi:10.1088/1742-5468/2014/03/P03025