Connection between Information Theory and Statistical Mechanics

Vegard B. Sørdal

December 19, 2019

1 Information theory and entropy

1.1 Shannon entropy

Claude Shannon, while working at Bell Telephone Laboratories, developed in 1948 a mathematical measure of uncertainty, to quantify the loss of information in phone-line signals [1]. Supposedly while working on this measure he visited Von Neumann, and they had the following discussion:

> *My greatest concern was what to call it. I thought of calling it information, but the word was overly used, so I decided to call it uncertainty. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, "You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage".*

Shannon followed Von Neumann's advice, and called his measure the Shannon entropy. E.T. Jaynes has a clear derivation of Shannon entropy that we will follow from now on [2]. Assume we have a variable x that can take on discrete values $(x_1 \dots x_n)$. The process that determines what value x assumes can be represented by the corresponding probabilities $(p_1 \dots p_n)$, where p_i represents the probability that $x = x_i$. The goal is to derive a quantity $H(p_1 \dots p_n)$, which uniquely measures the amount of uncertainty represented by this probability distribution. Or in other words, a function that quantifies our lack of information about a system. It might seem difficult to create an unique and consistent measure of uncertainty. Remarkably, only by using three elemental conditions of consistency we can show that this quantity H is what we now call Shannon entropy. The three conditions are:

(1) H has to be a continuous function of the p_i's, or else an arbitrarily small change in their value would lead to a large change in the amount of uncertainty.

(2) If all p_i are equal, the quantity $h(n) = H(\frac{1}{n} \cdots \frac{1}{n})$ is a monotonic increasing function of n: If you don't know anything about the distribution, your uncertainty can only increase if the number of possible choices increases.

(3) The measure H has to be consistent, meaning that if there is more than one way of calculating its value they all have to give the same answer.

In the opening statement we said that x can assume any of the discrete values $(x_1 \dots x_n)$, thus we can not assign $p_i = 0$ for any x_i. Unless we *know* what value x is e.g., $p_k = 1$, we have to give a finite value for all p_i. But if we know that $p_k = 1$ then we have complete information about the distribution, and a function describing our lack of knowledge is nonsensical.

According to condition (3), we have a choice between giving the probabilities of the events $(x_1 \dots x_n)$ directly, or partitioning them in groups. We can group the first k of them, such that the group probability is $\omega_1 = (p_1 + \cdots + p_k)$, then group the next m so that the probability is $\omega_2 = (p_{k+1} + \cdots + p_{k+m})$, and so on. The amount of uncertainty of the composite events is then $H(\omega_1, \dots, \omega_N)$, where N is the total number of groups. The conditional probabilities of the events $(x_1 \dots x_k)$, given the composite event ω_1 is then $(p_1/\omega_1, \dots, p_k/\omega_1)$. Doing this for all the composite events, eventually brings us to the same state of knowledge as if all the p_i's had been given directly.

$$\begin{aligned} H(p_1 \dots p_n) &= H(\omega_1 \dots \omega_r) + \omega_1 H(p_1/\omega_1 \dots p_k/\omega_1) \\ &+ \omega_2 H(p_{k+1}/\omega_2 \dots p_{k+m}/\omega_2) + \cdots . \end{aligned} \tag{1}$$

That is, the uncertainty given by the p_i's, is the same as the uncertainty of composite events plus the conditional probability of each composite event. As an example, lets say we have $(p_1, p_2, p_3) = (1/2, 1/3, 1/6)$ and decide to form the two following groups; $\omega_1 = p_1 = 1/2$, and $\omega_2 = p_2 + p_3 = 1/2$. We then get

$$\begin{aligned} H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) &= H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} H\left(\frac{1}{2}\right) + \frac{1}{2} H\left(\frac{1/3}{1/2}, \frac{1/6}{1/2}\right) \\ &= H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} H\left(\frac{2}{3}, \frac{1}{3}\right) \end{aligned} \tag{2}$$

Since H is continuous according to condition (1), it is sufficient to determine H for all rational values

$$p_i = n_i / \sum_i n_i, \qquad n_i = \text{integers}. \tag{3}$$

We can then regard each probability p_i's as a grouping of n_i equally likely events. We can group together any number of equally likely events, to create a composite event of arbitrary probability. Take as an example $N = 9$ equally likely events, and then form the following $n = 3$ groups; one group of $n_1 = 4$, one group of $n_2 = 3$, and one group of $n_3 = 2$. The composition law, Eq. (1) then becomes

$$h(9) = H\left(\frac{4}{9}, \frac{3}{9}, \frac{2}{9}\right) + \frac{4}{9} h(4) + \frac{3}{9} h(3) + \frac{2}{9} h(2), \tag{4}$$

where $h(n)$ is shorthand for

$$h(n) = H\left(\frac{1}{n}, \dots, \frac{1}{n}\right). \tag{5}$$

The general form of Eq. (1) with this notation becomes

$$h(\sum_i^n n_i) = H(p_i, \ldots, p_n) + \sum_i p_i h(n_i). \tag{6}$$

If we now choose all $n_i = m$, the equation further simplifies to

$$h(mn) = h(m) + h(n), \tag{7}$$

which can be shown [1] to have the unique solution

$$h(n) = K \log(n), \tag{8}$$

where K is an arbitrary constant. Combining this with Eq. (6) we get

$$\begin{aligned} H(p_1, \cdots, p_n) &= K \ln\Big(\sum n_i\Big) - K \sum_i p_i \ln(n_i) \\ &= K \ln\Big(\sum n_i\Big) - K \sum_i p_i \ln\left(p_i \sum_i^n n_i\right) \\ &= K \ln\Big(\sum n_i\Big) - K \sum_i p_i \ln p_i - K \sum_i p_i \ln\left(\sum_i n_i\right) \\ &= -K \sum_i p_i \ln p_i, \end{aligned} \tag{9}$$

which is the familiar form of the Shannon entropy, and this is only equation that satisfies the conditions we imposed. It then follows that for a given a probability distribution $(p_1, \cdots, p_n)$, the values of the p_i's that maximizes the Shannon entropy is the least biased and most "honest" description of a system, subject to the constraints imposed by our available information.

We can find the maximum of H, given that the probability is normalized, by using the method of Lagrange multipliers.

$$\begin{gathered} \nabla \left[H(p_1 \ldots p_n) - \lambda G(p_1 \ldots p_n)\right] = 0 \\ \Downarrow \\ \max \left\{H(p_1 \ldots p_n) \mid G(p_1 \ldots p_n) = 0\right\}, \end{gathered} \tag{10}$$

where $G(p_1 \ldots p_n) = \sum_i p_i - 1$. Performing the calculation of the gradient along one dimension p_k, we obtain

$$-\ln p_k - 1 - \lambda = 0 \tag{11}$$

$$p_k = e^{-(1+\lambda)} \tag{12}$$

which has to apply for all p_k. Putting this into the normalization constraint gives us

$$\sum_i^N e^{-(1+\lambda)} = N e^{-(1+\lambda)} = 1 \quad \rightarrow \quad \lambda = \ln(N) - 1, \tag{13}$$

with the final result

$$p_k = e^{-(1+\ln N - 1)} = e^{-\ln N} = \frac{1}{N}. \tag{14}$$

The implication is that $p_i = \frac{1}{N}$ is the *least biased* probability distribution for the points $(p_1 \dots p_n)$, and it indicates that we don't know anything about the distribution other than how many possible outcomes there are.

1.2 Thermodynamic and logical reversibility

Consider a general system where the total phase space is Γ, and the phase space coordinates are described by the vectors $\gamma \in \Gamma$. The system is surrounded by a heat bath at inverse temperature $\beta = 1/k_B T$. A transformation that maps some initial phase space distribution $\{\,\gamma_i\,\} \equiv \Gamma_i \subset \Gamma$ to some final distribution $\{\,\gamma_f\,\} \equiv \Gamma_f \subset \Gamma$, is then due to some physical process. The Shannon entropy of the initial and final state is given by

$$S_i = -\int_{\gamma\in\Gamma_i} d\gamma\; p(\gamma)\ln p(\gamma) \quad \text{and} \quad S_f = -\int_{\gamma\in\Gamma_f} d\gamma\; p(\gamma)\ln p(\gamma), \tag{15}$$

where $p(\gamma)$ is the probability of the state represented by the phase space point γ. If Q is the average heat absorbed by the system under the transformation, the total entropy production (i.e., system + environment) is then given by

$$\Delta S_{tot} = \underbrace{(S_f - S_i)}_{\Delta S} - \beta Q. \tag{16}$$

According to the second law of thermodynamic, total entropy change is bounded below at zero

$$\Delta S_{tot} \geq 0 \quad \rightarrow \quad \Delta S \geq \beta Q. \tag{17}$$

A physical process which achieves equality in this bound, is considered a thermodynamically reversible process. Notice that the flow of entropy between system and bath is possible for reversible processes, if the amount of heat absorbed by the system is equal to its entropy change. This is because the absorption of heat by the system results in a decrease in the environment entropy according to $\Delta S_{env} = -Q\beta$.

If the phase space points of our system is distributed according to the canonical distribution, the probabilities $p(\gamma)$ is given by

$$p(\gamma) = e^{\beta\left(F - E(\gamma)\right)}, \tag{18}$$

where $E(\gamma)$ is the energy associated with the phase space point γ, and $F = -\ln Z$ is the free energy associated with the distribution of phase space points $\{\,\gamma\,\}$. With this probability the entropy of the initial and final state becomes

$$S_{(i/f)} = \beta\left(U_{(i/f)} - F_{(i/f)}\right), \tag{19}$$

where $U_{(i/f)} = \langle E_{(i/f)} \rangle_C$ is the canonical ensemble average of the energy. Using the first law of thermodynamics, $\Delta U = \Delta W + \Delta Q$, where W is the average work performed on the system, we find that the second law of thermodynamic in this form becomes

$$\Delta S_{tot} = \beta\left(W - \Delta F\right) \geq 0 \quad \rightarrow \quad W \geq \Delta F, \tag{20}$$

where $\Delta F = F_f - F_i$. We see that if the input work we perform on the system is equal to its change in its free energy, the process is reversible.

Phase space trajectories can not cross each other, because if they could the phase space point at the intersection does not have a deterministic Hamiltonian evolution. The point could evolve according to either trajectory, so we would lose information about its past. This concept is closely related to logical reversibility. Consider a set of logical input states I, and logical output states O. Lets for simplicity consider one single bit of information, that can be in one of two logical states $\{\,0, 1\,\}$. A logical process, or a computation C, can then be described as a transformation between the input state and the output state $C : I \to O$. An example of an irreversible process is then the ERASE operation, which is defined by

$$\text{ERASE}: \quad 0 \to 0, \quad 1 \to 0. \tag{21}$$

No matter which state you were in (0 or 1), you end up in the same state (0), and lose any information about the past. An example of a reversible process is the NOT operation, which is defined as

$$\text{NOT}: \quad 0 \to 1, \quad 1 \to 0. \tag{22}$$

In this case, given the output, you always know the input. A logically reversible process can be defined as one that, for any output logical state, a unique input logical state exists [3]. Meaning that for every logical state in O, there exists a reversal of C, which is defined as $C^{-1} : O \to I$.

Now let the input and output states be two probability distributions instead of a single bit. We defined them as $P_O(n_o)$ for $n_o \in O$ and $P_I(n_i)$ for $n_i \in I$, with normalized probabilities. After the operation C, the distribution on O is given by

$$P_0(n_o) = \sum_{n_i:\ C(n_i)=n_o} P_I(n_i), \tag{23}$$

where the sum is taken over all n_i which satisfies $C(n_i) = n_o$. If the process is reversible, then there is one unique n_i for each n_o, giving us

$$P_0(n_o) = P_I(C^{-1}(n_o)). \tag{24}$$

The input and output logical entropies are given by

$$H_I = -\sum_{n_i \in I} P_I(n_i) \ln P_I(n_i), \tag{25}$$

and

$$H_O = -\sum_{n_o \in O} P_O(n_o) \ln P_O(n_o). \tag{26}$$

For reversible operations, defined by Eq. (24), we see that the logical entropy does not change

$$\begin{aligned} H_O &= -\sum_{n_o \in C(O)} P_I(C^{-1}(n_o)) \ln P_I(C^{-1}(n_o)) \\ &= -\sum_{n_i \in I} P_I(n_i) \ln P_I(n_i) = H_I. \end{aligned} \tag{27}$$

In the general case, including non-reversible operation, the entropy difference becomes

$$H_O - H_I = \sum_{n_o} P_O(n_0) \sum_{n_i:\ C(n_i)=n_o} \frac{P_I(n_i)}{P_O(n_o)} \ln \frac{P_I(n_i)}{P_O(n_o)} \tag{28}$$

2 Erasing information: Landauer's principle

As discussed in the introduction, Landauer's solution to the apparent violation of the second law of thermodynamics by the Szilard engine was the fact that one has to erase the information obtained by the measurement [4, 5, 6]. All physical systems designed to perform logical operations have specific physical states (microstates) which correspond to the logical states. A one-bit memory can be modeled as a single-particle-box with a barrier in the center, as shown in Fig. 1(a). The two logical states are a particle found on the left side of the barrier (0) or a particle found on the right side of the barrier (1). In this model the logical states $\{\,0,1\,\}$ correspond to the physical states

$$0 \equiv \{\, x \in [-L/2, 0], |p| = \sqrt{2mE}\,\}\,, \tag{29}$$

and

$$1 \equiv \{\, x \in [0, L/2], |p| = \sqrt{2mE}\,\}\,. \tag{30}$$

Landauer argued that logically irreversible processes, which reduce the logical state space, must therefore also compress the physical state space. This compression of phase space results in an increase in entropy, in the form of heat dissipation [7, 8]. An example of a logical irreversible process is the ERASE operation discussed earlier ($0 \to 0, 1 \to 0$). The physical implementation of this protocol on the SPB memory is shown in Fig. 1(b). The memory is initially in either of the two logical states $\{\,0,1\,\}$. We then remove the barrier from the center of the box, and insert it in the far right-hand side of the box. While the barrier back towards the center, the collisions between the particle and the barrier exerts an effective pressure on the barrier. Therefore an amount of work is required to push the barrier, which is transferred to the heat bath via the thermal contact between the particle and environment. When the barrier reaches the center of the box, the particle is always found in a physical state corresponding to the logical state 0.

Before the erasure, the probability of 0 and 1 are equally 1/2, giving a logical entropy $H_i = \ln 2$. After the erasure has been performed, the probability of 0 is 1, so the logical entropy is $H_f = 0$. The difference in logical entropy is therefore $\Delta H = H_f - H_i = -\ln 2$. Since the logical entropy has to be treated on the same level as physical entropy, we have $\Delta S = \Delta H$, and from the second law of thermodynamics (Eq. (17)) we obtain

$$-\ln 2 \geq \beta Q \tag{31}$$

where Q is the heat dissipated into the environment. Since the internal energy does not change during the isothermal erasure we have, according to the first law of thermodynamics, $W = -Q$. Therefore the work needed to erase one bit of information is given by

$$W \geq k_B T \ln 2. \tag{32}$$

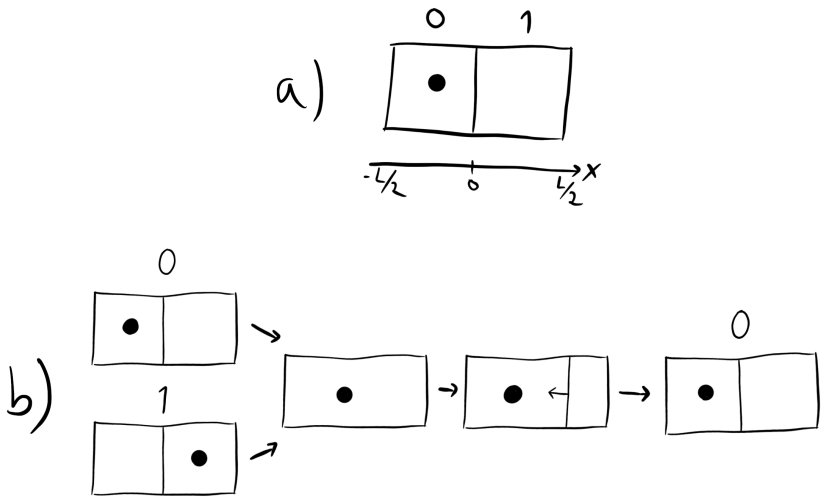

Figure 1: Illustration of a binary memory, modeled as a SPB of width L, and two logical states; left side of the barrier (0) and right side of the barrier (1). In b) we show a physical implementation of the ERASE operation.

This equation is known as Landauer's principle. Equality is achieved if the erasure is performed adiabatically, in such a way that the memory is always in equilibrium with the environment while we push the barrier towards the center. A quasi-static isothermal compression requires an amount of work given by

$$W = \int_{V/2}^{V} \frac{k_B T}{V'} dV' = k_B T \ln 2, \tag{33}$$

and is , therefore, an example of a physical erasure protocol that reaches equality in the Landauer bound. The Landauer principle has in recent years been experimentally verified in a number of different systems [9, 10, 11].

In general, a logical state does not have a one-to-one mapping to a unique physical state. Rather, a logical state is a subset of the full phase space, $\Gamma_{0/1} \subset \Gamma$, and corresponds to many different microstates. By definition $\Gamma = \Gamma_0 \cup \Gamma_1$, and $\Gamma_0 \cap \Gamma_1 \equiv \emptyset$. If this was not the case, the two logical states would have indeterminate members which could not be definitely associated with either state. In the previous case the logical state 0 is associated with the subspace $\Gamma_0 : \{ x \in [-L/2, 0] \}$, while the logical state 1 is given by $\Gamma_1 : \{ x \in [0, L/2] \}$. Ignoring the irrelevant y-coordinate and momentum $\vec{p} = p_x \vec{x} + p_y \vec{y}$, we denote the probability distribution of the total phase space by $P(x)$. The probability distribution of the logical states, P_L, is then given by

$$P_L(i) = \int_{x \in \Gamma_i} P(x)\ dx, \qquad i = 0, 1. \tag{34}$$

The conditional probability of the microstate x given the logical state i is therefore

$$P(x|i) = P(x)/P_L(i). \tag{35}$$

The total entropy, S, is given by the integral over the total phase space

$$S = \int_\Gamma P(x) \ln P(x) \; dx, \tag{36}$$

while the logical entropy is given by

$$H = -\sum_i P_L(i) \ln P_L(i). \tag{37}$$

Following the discussion in Section 1.1, we can group the microstates into composite events, i.e. the logical states $\Gamma_{0/1}$ in this case. The total entropy can then be written as the entropy of the logical states, plus the conditional entropy $S(\Gamma_i|i)$ weighted by the logical state probabilities

$$S = -\sum_i P_L(i) \ln P_L(i) - \sum_i P_L(i) S(\Gamma_i|i), \tag{38}$$

where

$$S(\Gamma_i|i) = \int_{x\in\Gamma_i} P(x|i) \ln P(x|i) \; dx. \tag{39}$$

We see that the total entropy can be decomposed into two terms, where one is the logical entropy H, and the other is the average conditional entropy $S_{in} = \sum_i P_L(i) \; S(\Gamma_i|i)$, which we identify as the internal physical entropy in the logical subspaces

$$S = H + S_{in}. \tag{40}$$

Using this decomposed version of the total entropy, we can calculate contribution of each term for an ERASURE operation. For this operation the initial logical probability distribution is $P_L(0) = P_L(1) = 1/2$, while the final one is $P'_L(0) = 1$ and $P'_L(1) = 0$, which gives us a change in logical entropy $\Delta H = -\ln 2$. The change in internal entropy is

$$\begin{aligned}
\Delta S_{in} &= -\sum_i P'_L(i) S'(\Gamma_i|i) + \sum_i P_L(i) S(\Gamma_i|i) \qquad (41)\\
&= -S'(\Gamma_0|0) + \frac{1}{2} S(\Gamma_0|0) + \frac{1}{2} S(\Gamma_1|1) \\
&= -\int_{x\in\Gamma_0} \frac{P'(x)}{P'_L(0)} \ln \frac{P'(x)}{P'_L(0)} + \frac{1}{2} \int_{x\in\Gamma_0} \frac{P(x)}{P_L(0)} \ln \frac{P(x)}{P_L(0)} + \frac{1}{2} \int_{x\in\Gamma_1} \frac{P(x)}{P_L(1)} \ln \frac{P(x)}{P_L(1)} \\
&= -\int_{x\in\Gamma_0} dx \; P'(x) \ln P'(x) + \int_{x\in\Gamma} dx \; P(x) \ln P(x) + \int_{x\in\Gamma} dx \; P(x) \ln(2).
\end{aligned}$$

If we assume the initial and final phase space probabilities are equilibrium distributions, with $P(x) = \frac{1}{L}$ and $P'(x) = \frac{1}{L/2}$ we obtain

$$\begin{aligned}
\Delta S_{in} &= -\int_{-L/2}^{0} dx \; \frac{2}{L} \ln \frac{2}{L} + \int_{-L/2}^{L/2} dx \; \frac{1}{L} \ln \frac{1}{L} + \ln 2 \qquad (42)\\
&= -\ln \frac{2}{L} + \ln \frac{1}{L} + \ln 2 = 0
\end{aligned}$$

Therefore the total change in entropy when adiabatically erasing one bit of information is

$$\Delta S = \Delta H + \Delta S_{in} = -\ln 2, \tag{43}$$

and the generalized Landauer principle can be expressed as

$$\Delta H + \Delta S_{in} \geq \beta Q. \tag{44}$$

3 Obtaining information: Measurement

A measurement is to make a copy of the state of a system onto a memory. For the measurement of the state of a Szilard engine, we need a binary memory. We consider the total phase space (system + memory) to be $\Gamma = \Gamma_S \cup \Gamma_M$, where Γ_S and Γ_M is the phase space of the system and memory, respectively. Let $s \in S = \{0, 1\}$ and $m \in M = \{0, 1\}$ be the logical states of the system and memory, respectively. Their physical states is denoted by $x_s \in \Gamma_S$ and $x_m \in \Gamma_M$. The conditional probability of finding the total system in the physical state (x_s, x_m) given the logical states (s, m) is then $P(x_s, x_m|s, m)$, and the probability of the physical state is given by

$$P(x_s, x_m) = \sum_{s,m} P(x_s, x_m|s, m)P(s, m) \tag{45}$$

To characterize the correlation between the memory and the system, we introduce the mutual information. The mutual information quantifies how much information we obtain about one subsystem when observing another subsystem; If the mutual information is zero, the state of the memory and system is independent of each other. The mutual information between the physical states are given by

$$I_{in}(\Gamma_S; \Gamma_M) = S_{in}(\Gamma_S) + S_{in}(\Gamma_M) - S_{in}(\Gamma), \tag{46}$$

while for the logical states we have

$$I_H(S; M) = H(S) + H(M) - H(S \otimes M), \tag{47}$$

where $S \otimes M$ is the total logical state, i.e. $00, 01, 10, 11$. The mutual information between the internal states, given the logical states s and m, are given by

$$I_{in}(\Gamma_S; \Gamma_M|s, m) = S_{in}(\Gamma_S|s) + S_{in}(\Gamma_M|m) - S_{in}(\Gamma|s, m). \tag{48}$$

Taking the average over s and m, we obtain

$$I_{in}(\Gamma_S; \Gamma_M|L) = S_{in}(\Gamma_S|S) + S_{in}(\Gamma_M|M) - S_{in}(\Gamma|S \otimes M). \tag{49}$$

In a similar way that we decomposed the total entropy into the logical entropy and average conditional entropy in Section 2, we can decompose the total mutual information into the correlation between the logical states and the average conditional mutual information between the physical states:

$$I_{in}(\Gamma_S; \Gamma_M) = I_H(S; M) + I_{in}(\Gamma_S; \Gamma_M|S \otimes M). \tag{50}$$

Taking the mutual information into account, the total change in entropy ΔS_{tot} after some arbitrary thermodynamic interaction between the system and the measurement apparatus is given by

$$\Delta S_{tot} = \underbrace{\Delta H^S + \Delta H^M - \Delta I_H}_{\text{logical entropy } \Delta H} + \underbrace{\Delta S_{in}^S + \Delta S_{in}^M - \Delta I_{in}}_{\text{internal entropy } \Delta S_{in}} - \underbrace{\beta Q}_{\text{heat}}, \tag{51}$$

where the superscript S and M indicates the system and entropy, respectively. Going back to the erasure process and Eq. (44), we see that if the internal

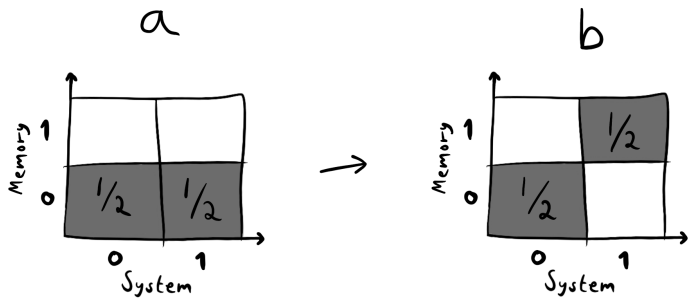

Figure 2: Illustration of the combined phase space of a Szilard engine and a single-particle-box memory. (a) shows the initial state where the memory is in the standard state 0 and the system is in either the 0 or the 1 state. The transition from (a) to (b) is an example of an error-free measurement, where the both the system and memory is either in the logical states 00 or 11.

entropy does not change during the erasure (i.e., the initial and final phase space distribution are equilibrium distributions), we obtain

$$\Delta H^S + \Delta H^M - \Delta I_H \geq \beta Q. \tag{52}$$

After the full cycle of measurement, expansion, and deletion of memory, the logical states of the system and the memory is the same as the initial ones. Therefore $\Delta H^S = \Delta H^M = 0$, and since the internal energy does not change we also have $Q = -\Delta W$. Using this we obtain yet another version of Landauer's principle

$$W \geq \Delta I_H / \beta. \tag{53}$$

The work required to delete the information in a memory, is given by the mutual information between the system and the memory. In the case of a perfect measurement we have $\Delta I_H = \ln 2$, which means that the minimum work we have to pay to erase the memory is the same as the work we obtain from the Szilard engine.

3.1 Measurement errors

Measurement errors reduce the mutual information between the system and memory, and therefore the work required to delete the memory. However, as we argue in paper 2, it is not possible to saturate the bound in Eq. (53) when measurement errors are present. This is due to an irreversible entropy production not accounted for, which we will describe briefly in the following. If the

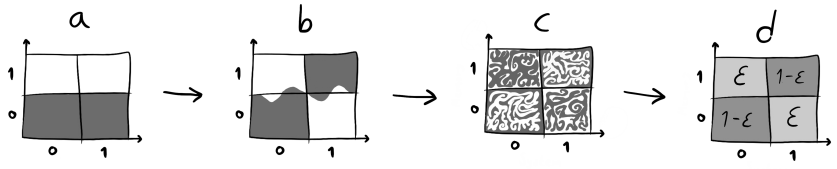

Figure 3: Illustration of the total phase space of a Szilard engine and a single-particle-box memory. (a) shows the initial state where the memory is in the standard state 0, while the system is in either 0 or 1 with probability 1/2. A measurement error occurs in (b1)/(b2), and once the barrier is inserted so the phase space can not flow between the quadrants, the phase space of the incorrectly mapped states evolve chaotically according to Hamiltonian dynamics as shown in (c). Coarse graining of the phase space after the time evolution results in the final phase space distribution shown in (d).

system is a standard Szilard engine, and the memory is a single-particle-box as before, there are four distinct logical states $(00, 01, 10, 11)$. In Fig. 2 we show a schematic of the full phase space of the (system + memory). Here we reduce the dimension of the phase space to the only relevant degree of freedom (the x-coordinate). Therefore the horizontal axis represents the x-coordinate of the particle in the system, while the vertical axis represents the x-coordinate of the particle in the memory. The total phase space is divided into four quadrants, each of which represents one of four logical states, associated with which side of the box the particle is in the system and memory. The initial state of the system + memory is shown in Fig. 2(a), where the memory is in a standard state 0, while the system is either in the state 0 or 1 with probability 1/2. If an error-free measurement is performed on the system and copied into the memory, the full phase space evolve into what is shown in Fig. 2(b). The internal entropy and the logical state of the system and memory is identical; both the memory and the system is either in state 0 or 1 with the same phase space distribution.

Consider now the schematic in Fig. 3, showing the phase space evolution of this model when measurement errors are present. The initial state shown in Fig. 3(a), is the same standard state as in the error-free measurement. If the system is now put into contact with the measurement apparatus and copied into the memory, some of the system states are incorrectly mapped to the memory. This incorrect mappings come from the cases where the actual position of the particle in the system does not agree with what was recorded in the memory, and is shown in Fig. 3(b), i.e., the phase space points in 01 are wrongly mapped and should fill in the empty space in 00. When the barrier is inserted the phase space points can no longer cross the boundaries between the four quadrants. However, the phase space continues to evolve according to deterministic Hamiltonian dynamics, resulting in a complicated structure of the phase space as shown in Fig. 3(c). Nevertheless, since the time evolution obeys Liouville's theorem, the entropy of Fig. 3(c) is still the same as in Fig. 3(b). To reach the final state with uniform phase space distributions, shown in Fig. 3(d), we have to coarse-grain the phase space. We therefore lose information about the

complicated phase space structure. It is this coarse-graining that introduces an irreversible measurement entropy given by

$$S_\varepsilon = -\varepsilon \ln \varepsilon - (1-\varepsilon)\ln(1-\varepsilon), \tag{54}$$

where ε is the probability of measurement error.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[2] E. T. Jaynes, *Probability theory: The logic of science*. Cambridge university press, 2003.

[3] T. Sagawa, "Thermodynamic and logical reversibilities revisited," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2014, p. P03025, mar 2014.

[4] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM journal of research and development*, vol. 5, no. 3, pp. 183–191, 1961.

[5] R. Landauer, "Minimal energy requirements in communication," *Science*, vol. 272, no. 5270, pp. 1914–1918, 1996.

[6] R. Landauer, "The physical nature of information," *Physics letters A*, vol. 217, no. 4-5, pp. 188–193, 1996.

[7] C. H. Bennett, "Logical reversibility of computation," *IBM journal of Research and Development*, vol. 17, no. 6, pp. 525–532, 1973.

[8] C. H. Bennett, "Notes on landauer's principle, reversible computation, and maxwell's demon," *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, vol. 34, no. 3, pp. 501–510, 2003.

[9] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, "Experimental verification of landauerâs principle linking information and thermodynamics," *Nature*, vol. 483, no. 7388, p. 187, 2012.

[10] Y. Jun, M. Gavrilov, and J. Bechhoefer, "High-precision test of landauerâs principle in a feedback trap," *Physical review letters*, vol. 113, no. 19, p. 190601, 2014.

[11] A. O. Orlov, C. S. Lent, C. C. Thorpe, G. P. Boechler, and G. L. Snider, "Experimental test of landauer's principle at the sub-kbt level," *Japanese Journal of Applied Physics*, vol. 51, no. 6S, p. 06FE10, 2012.