Optimal Work Extraction and the
Minimum Description Length Principle

Léo Touzo[1], Matteo Marsili[2], Neri Merhav[3], and Édgar Roldán[2]

[1]Department of Physics, École Normale Supérieure, 24 rue Lhomond, 75005 Paris, France

[2]The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy

[3]The Andrew and Erna Viterbi Faculty of Electrical Engineering, Technion - Israel Institute of Technology Technion City, Haifa 3200003, Israel

Abstract

We discuss work extraction from classical information engines (e.g., Szilárd) with N-particles, q partitions, and initial arbitrary non-equilibrium states. In particular, we focus on their *optimal* behaviour, which includes the measurement of a set of quantities $\boldsymbol{\Phi}$ with a feedback protocol that extracts the maximal average amount of work. We show that the optimal non-equilibrium state to which the engine should be driven before the measurement is given by the normalised maximum-likelihood probability distribution of a statistical model that admits $\boldsymbol{\Phi}$ as sufficient statistics. Furthermore, we show that the minimax universal code redundancy $\mathcal{R}^\star$ associated to this model, provides an upper bound to the work that the demon can extract on average from the cycle, in units of $k_\mathrm{B}T$. We also find that, in the limit of N large, the maximum average extracted work cannot exceed $H[\boldsymbol{\Phi}]/2$, i.e. one half times the Shannon entropy of the measurement. Our results establish a connection between optimal work extraction in stochastic thermodynamics and optimal universal data compression, providing design principles for optimal information engines. In particular, they suggest that: (i) optimal coding is thermodynamically efficient, and (ii) it is essential to drive the system into a critical state in order to achieve optimal performance.

In memory of Jorma J. Rissanen,
the father of the minimum description length principle,
and a great inspiration to many of us, for years to come.

1   Introduction and statement of the main results

Bits are emerging as a universal currency in a wide variety of sciences, ranging from information theory and coding theory, to statistical mechanics and thermodynamics [1, 2, 3, 4, 5]. In the spirit of Maxwell's demon, Szilárd's information engine [6] is considered by many as the founding stone of information thermodynamics. Szilárd's information engine executes a cyclic protocol by using one bit of information of a physical system, yielding to the extraction of $k_\mathrm{B}T \ln 2$ of energy from a thermal bath at temperature T, in apparent contradiction with the second law of thermodynamics [1, 2, 3].

1

Recent progress within the emerging field of stochastic thermodynamics has led to a deeper understanding of the relationship between information and entropy at mesoscopic scales, leading to e.g., the refinement of the second law of thermodynamics for systems with feedback control [7]. These works study fundamental thermodynamic bounds governing processes that convert information into work and/or heat, and vice versa, both in theory [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 5, 19], and experiment [20, 21, 22, 23, 24, 25, 26, 27, 28].

Since Maxwell's and Carnot's [29] pioneering work, optimizing the performance of heat and information engines has remained one of the key goals of thermodynamics. Substantial progress towards efficient rectification of fluctuations by mesoscopic machines has been achieved during the last decades within framework of stochastic thermodynamics [30, 31]. Important examples include finite-time protocols that enable optimization of different thermodynamic quantities, such as: (i) minimizing the average heat dissipation and work needed to drive a system out of equilibrium [32, 33, 34, 35, 36]; (ii) minimizing the uncertainty (e.g., Fano factor) of the work done on a generic non-equilibrium process [37]; (iii) "one shot" optimization leading to work extraction in every realization above a prescribed threshold value [38, 39]; (iv) enforcing a desired value for the probability to extract work by an arbitrarily large value above the free energy change [40, 41]; and (v) reversible feedback-control protocols which ensure an average work dissipation equal to the information gain during a non-equilibrium process [42, 43, 44]. For general non-equilibrium processes involving feedback control (e.g. information engines), it remains yet unclear whether there exist fundamental principles linking optimal work extraction and efficient data compression via e.g., optimal coding of the information used in feedback control. Until now, few insights have been developed yet at the interface between stochastic thermodynamics and coding, such as the relation between work and entropy production with sequence complexity [45, 46, 47] and channel capacity [48].

In this paper, we consider classical many-particle information engines that extract work from a cycle in which a certain microscopic quantity is measured. In particular, we first consider a variant of the N-particle Szilárd engine with q partitions in which the system is prepared in a non-equilibrium state before the measurement, see Fig. 1 for an illustration. An important question is the following: what is the maximal amount of work that the engine can extract on average from a measurement, and for which initial non-equilibrium distribution? We limit our discussion to work extraction protocols that, as in the original Szilárd engine, rely on reversible expansion of the partitions after the measurement. Our main result is that, for this class of systems, the average work that can be extracted cannot exceed the minimax universal code redundancy of the statistical model that admits the measured quantities as sufficient statistics, according to Minimum Description Length (MDL) principle [49], i.e.,

$$\langle W_{\mathrm{ext}} \rangle \leq k_{\mathrm{B}} T \, \mathcal{R}^{\star}, \tag{1}$$

with equality holding under optimal protocols. Here, W_{ext} denotes the work extracted in a cycle of the engine, which includes the work needed to prepare the non-equilibrium state and the one extracted from the feedback protocol after the measurement. The minimax universal code redundancy, $\mathcal{R}^{\star}$, is a measure of model complexity. Loosely speaking, it quantifies the amount of information (measured in nats) that the measurement provides about the parameters of the model.

Equation (1) relates a thermodynamic quantity, on the left hand side, to a central concept in source coding theory, on the right hand side. We explore this connection via a formal equivalence between a system of classical non-interacting particles and optimal universal coding of a sample of independent observations. In this equivalence, the maximization of the extracted work corresponds to the maximization of the likelihood. Indeed, we show that the non-equilibrium state to which

the engine should be driven before the measurement corresponds to the *normalised maximum likelihood* (NML) probability distribution, which is the coding distribution that provides optimal, universal data compression according to the MDL principle [49]. This code has been shown to have critical statistical–mechanical properties in a precise sense [50], showing once again the relevance of criticality in information processing systems [51]. For the N-particle Szilárd engine with q partitions we provide explicit expressions for $\mathcal{R}^*$ −Eq. (11)− and for the optimal non-equilibrium distribution −Eq. (10)− as well as numerical values in specific cases (Fig. 3).

Notably, we show that the connection between optimal work and optimal coding generalizes beyond the Szilárd engine to a broad class of models, provided that the equilibrium state of the engine is uniquely identified by the measurement. Thus, the measured quantities should be sufficient statistics of the model describing the equilibrium state. Our results suggest design principles of optimal information engines, such as: (i) how the equilibrium state should be defined, depending on the quantity that is measured, (ii) what is the non-equilibrium state to which the system should be driven; and (iii) the fundamental limits on the work that can be extracted. With respect to the latter, an interesting issue is how close our bound gets to the fundamental limit [7] $\langle W_{\text{ext}} \rangle \leq k_{\text{B}} T\, H[\boldsymbol{\Phi}]$, where $H[\boldsymbol{\Phi}]$ is the entropy of the measurement $\boldsymbol{\Phi}$. We find that, in the limit $N \to \infty$ the protocol studied here comes half-way to the upper bound, in the sense that $\langle W_{\text{ext}} \rangle / (k_{\text{B}} T\, H[\boldsymbol{\Phi}])$ converges to $1/2$ when $N \to \infty$. This is reminiscent of recent bounds derived on the sensory capacity [52], and on the relevance of sufficient statistics to achieve these bounds [53].
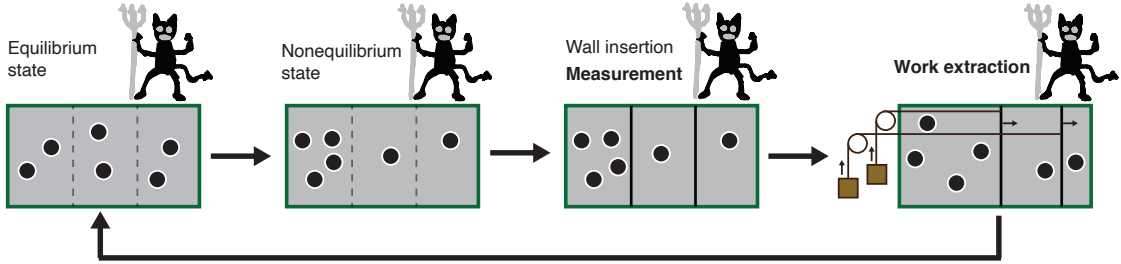

Figure 1: Illustration of a Szilárd engine with $N = 6$ particles and $q = 3$ partitions. The engine is composed by N non-interacting, distinguishable, classical point particles confined in a box that is in contact with a bath at temperature T. Initially, the system is in an equilibrium state with the N particles distributed homogeneously in the box. Next, the system is driven to a non-equilibrium state with inhomogeneous spatial distribution. An intelligent "demon" introduces walls in the box at prescribed locations, and measures the number of particles located in each of the resulting partitions. Using this measurement, the demon is able to lift weights (i.e. to extract work) from the partitions that expand reversibly to a new equilibrium state with equal pressure in each of the partitions. The cycle is then completed by removing the walls and letting the particles relax to equilibrium. Here, we investigate what is the maximum work that can be extracted on average over many repetitions of the cycle, and what is the non-equilibrium distribution at which the maximum average work is attained. For the Szilárd engine in the Figure, the maximum work extracted is given by $k_{\text{B}} T \mathcal{R}^*$, with $\mathcal{R}^* = \ln(3167/324) \simeq 2.28$ the minimax redundancy, following Eq. (11).

The rest of the paper is organised as follows: In Section 2, we revisit the classical Szilárd engine with N particles and q partitions [54], pose our main question, and draw a connection between optimal average work extraction and the MDL principle [49]. In Section 3, we map the Szilárd engine to a system of non-interacting particles in q energy levels and re-derive this connection in a broader framework. Finally, in Section 4, we discuss the implications and possible future extensions

of our work.

2 Minimum description length and work extraction in the N-particle Szilárd's engine

In this section, we revisit a variant of the classical Szilárd engine modelled by a classical ideal gas of N particles confined in the x-dimension of a one-dimensional box of length one and unit area in the perpendicular directions. This system has been extensively studied in [54], to which we refer for a detailed discussion. Here, we first review its main properties and introduce the notation used throughout the paper. Next, we discuss optimal work extraction and relate it to optimal coding of the information measured by the demon.

2.1 Definition of the model and notation

We consider an information engine operating cyclically and consisting of $N \geq 1$ point, non-interacting, particles confined in a box. The system is assumed to be initially in equilibrium with a heat bath at temperature T. The temperature of the heat bath is kept constant all along the cycle, hence we take $k_{\mathrm{B}}T = 1$ henceforth. In the initial state, the positions x_i ($i = 1, \ldots, N$) of the particles are dynamical variables spanning the whole box, with a uniform distribution. We will denote by $\mathbf{x} = (x_1, \ldots, x_N)$ the vector of the positions of all particles.

During the operation of the engine, the box is divided into q partitions. The partitions divide the unit interval into q sub-intervals, $[y_{s-1}, y_{s-1} + \ell_s)$, where $y_0 = 0$, $y_s = y_{s-1} + \ell_s$ (for all $s = 1, \ldots, q$) and $y_q = 1$. The lengths of the partitions can take arbitrary positive values, consistent with the constraint $\sum_{s=1}^{q} \ell_s = 1$. We denote by $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_q)$ the set of lengths of each partition. We now introduce two key quantities that we will use in the following to describe the dynamics of the engine:

- The state of the system $\mathbf{s} = (s_1, \ldots, s_N)$ is a vector composed of a collection of N coarse-grained integer variables, with $s_i \in \{1, \ldots, q\}$ denoting the partition in which the i-th particle is located. Each particle, $i = 1, \ldots, N$, is assigned to a partition of the x axis, $s_i = s(x_i)$, where $s(x_i)$ is the smallest s for which $x_i > y_s$.

- The measured quantity $\mathbf{n} = (n_1, \ldots, n_q)$ is a vector composed of q integer variables $n_s \in \{0, \ldots, N\}$ denoting the number of particles located in the s-th partition. Each partition, $s = 1, \ldots, q$ contains $n_s = \sum_{i=1}^{N} \delta_{s_i, s}$ particles, with the total number of particles $\sum_{s=1}^{q} n_s = N$ conserved during the cycle. We henceforth refer to n_s ($s = 1, \ldots, q$) as occupation numbers.

We assume that $\boldsymbol{\ell}$ and N are known, while $\mathbf{n}$ is obtained through a measurement, as explained below[1]. Here, we focus on protocols that rely on the change of model parameters ($\boldsymbol{\ell}$ here). The cyclic operation of the engine is as follows (see Fig. 2):

---

[1]For simplicity, we use the notation $\mathbf{n} \equiv \mathbf{n}(\mathbf{s})$ (and thus $n_1 \equiv n_1(s_1, \ldots, s_N), \ldots, n_q \equiv n_q(s_1, \ldots, s_N)$) hence the dependence on $\mathbf{s}$ will be assumed and dropped henceforth. We will also use the notations for the sums $\sum_{\mathbf{s}} \equiv \sum_{s_1} \cdots \sum_{s_N}$ and $\sum_{\mathbf{n}} \equiv \sum_{n_1 \geq 0} \cdots \sum_{n_q \geq 0} \delta_{\sum_s n_s, N}$, where $\delta_{i,j}$ is the Kronecker delta.

A) Initially, particles are distributed uniformly in the box according to the equilibrium distribution. In terms of the variables $\mathbf{s}$, the equilibrium state corresponds to the distribution

$$P_{\text{eq}}(\mathbf{s}|\boldsymbol{\ell}) = \prod_{i=1}^{N} \ell_{s_i} = \prod_{s=1}^{q} \ell_s^{n_s}. \tag{2}$$

In other words, in the equilibrium state each s_i is drawn independently from the same distribution, with ℓ_s being the probability that $s_i = s$.

B) Next, the system is driven quasistatically into an arbitrary non-equilibrium distribution $P_{\text{neq}}(\mathbf{s})$.

C) In the next step, the demon inserts $q - 1$ rigid walls at positions $y_1, \ldots, y_{q-1}$, separating the box in q partitions. When the walls are inserted, the positions of the particles $\mathbf{s}$ remain constrained, i.e., frozen, for the rest of the cycle. The demon measures the number of particles in each partition $\mathbf{n}$. Note that the walls are introduced in the *same* positions in each cycle.

D) The demon applies a feedback protocol, based on the measurement $\mathbf{n}$, extracting work from the reversible expansion of the different partitions from their initial sizes $\boldsymbol{\ell}$ to new sizes $\boldsymbol{\ell}'$.

E) The cycle is completed by removing the walls from $\boldsymbol{\ell}'$ and resetting the system to the initial equilibrium state (A). Once the partitions are removed, the constraints are dropped and s_i $(i = 1, \ldots, q)$ become again dynamical variables.

We shall denote by $\langle f(\mathbf{s}) \rangle_{\text{neq}} = \sum_{\mathbf{s}} P_{\text{neq}}(\mathbf{s}) f(\mathbf{s})$, averages over a non-equilibrium distribution P_{neq}, and similarly $\langle f(\mathbf{s}) \rangle_{\text{eq}} = \sum_{\mathbf{s}} P_{\text{eq}}(\mathbf{s}) f(\mathbf{s})$. We denote by $W_{\text{ext}}(\mathbf{n})$ the stochastic work extracted in each operation of the cycle[2] and $\langle W_{\text{ext}} \rangle_{\text{neq}}$ its average over many cycles.

Previous work [54] discussed work extraction from equilibrium states of the form (2), for different partition sizes $\boldsymbol{\ell}$. This corresponds to the cycle A→C→D→E→A (see Fig. 2), which does not include the non-equilibrium driving A→B. Ref. [54] showed that: (i) $\langle W_{\text{ext}} \rangle_{\text{eq}}$ depends on the initial partition sizes $\boldsymbol{\ell}$ and (ii) the optimal choice of $\boldsymbol{\ell}$ allows to extract an amount of work $-$ from an initial equilibrium state $-$ that saturates at $\sim 0.8371(q-1)$ for N large. Apart from the case $N = 1, q = 2$, this protocol does not allow to saturate the second law inequality with feedback control [55] $\langle W_{\text{ext}} \rangle_{\text{eq}} \leq I(\mathbf{x}; \mathbf{n})$. It was shown that it is possible to saturate the inequality with different protocols if the particles interact weakly [11].

2.2 Work extraction

We now discuss the following optimization problem: what is the maximum work that can be extracted from an average over many cycles of an N-particle Szilárd engine, and for which *non-equilibrium* distribution the maximum average work extraction is achieved?

In our analysis, we will consider the work done by insertion (B→C) and removal (D→E) of the walls as negligible [56]. In addition the work $W_{\text{E}\to\text{A}}$ needed to restore the initial state is also zero, since once the walls are removed the partitions have no mechanical effect on the underlying gas. Thus the total work done has two contributions: $W_{\text{A}\to\text{B}}$ corresponding to the preparation of the non-equilibrium state, and $W_{\text{C}\to\text{D}}$ corresponding to the expansion of the walls $\boldsymbol{\ell} \to \boldsymbol{\ell}'$.

---

[2]We take here the thermodynamic convention by which $W > 0$ then work is done on the system, and $W < 0$ when it is extracted from the system. We thus define $W_{\text{ext}} = -W$.
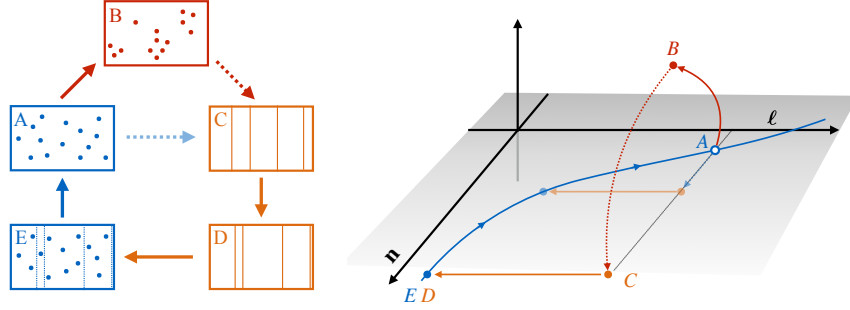
Figure 2: Schematic representation of the cycle performed by the information engine (left) and pictorial sketch of the corresponding transformations in parameter space (right). The initial equilibrium state (A) lies on the blue curve that designates the equation of state in the (hyper)plane of the parameters $\boldsymbol{\ell}$ (the partitions' lengths) and the (average of the) sufficient statistics $\mathbf{n}$ (the number of particle in each partition). In the preparation phase, the system is driven out of equilibrium (red arrow) to a state B (red dot/box). When partitions are introduced, and $\mathbf{n}$ is measured, the system is projected back on the plane to a different point C (dashed red arrow). The sufficient statistics are measured and the work extraction protocol is operated (orange arrow) driving the system to state D. Then the partitions are removed bringing the system back on the equilibrium curve (E). The cycle is completed by driving back the system to the equilibrium state A. In the setting of Ref. [54], the measurement is taken on the equilibrium state A (light blue arrow) and work is extracted (light orange arrow) from the resulting state.

The preparation of the non-equilibrium state requires an amount of work that is at least the free energy difference between the non-equilibrium distribution P_{neq} and the equilibrium state P_{eq}, given by the Kullback-Leibler (KL) divergence [57][3]:

$$\langle W_{\mathrm{A} \to \mathrm{B}} \rangle \geq \Delta F_{\mathrm{A} \to \mathrm{B}} = D\left[P_{\mathrm{neq}}(\mathbf{s}) || P_{\mathrm{eq}}(\mathbf{s}|\boldsymbol{\ell})\right] = \sum_{\mathbf{s}} P_{\mathrm{neq}}(\mathbf{s}) \ln \frac{P_{\mathrm{neq}}(\mathbf{s})}{P_{\mathrm{eq}}(\mathbf{s}|\boldsymbol{\ell})}. \tag{3}$$

To calculate the work done in the expansion C→D, we assume the law of ideal gases and unit volume, hence the initial pressure inside the box is $\mathsf{P}_{\mathrm{in}} = N$. The insertion of the walls by the demon puts the system in a state where the gas is in local equilibrium in each partition with a different pressure, $\mathsf{P}_s = n_s/\ell_s$ ($s = 1, 2, \ldots, q$). The forces exerted on the internal walls can be exploited to extract work, using the knowledge (i.e. measurement) of $\mathbf{n}$. These forces will vanish when the walls adjust such that the pressures in all partitions equalise and revert to the initial values, i.e., $\mathsf{P}'_s = n_s/\ell'_s = N$. This implies that the new partition sizes are given by $\hat{\ell}_s = n_s/N$. The work $-W_{\mathrm{C} \to \mathrm{D}}(\mathbf{n})$ extracted from the expansion of the gas can be written in two ways [54]:

$$-W_{\mathrm{C} \to \mathrm{D}}(\mathbf{n}) = \sum_{s=1}^{q} \int_{\ell_s}^{n_s/N} \mathrm{d}V \frac{n_s}{V} = \sum_{s=1}^{q} n_s \ln \frac{n_s}{N \ell_s} \tag{4}$$

$$= \ln \frac{P_{\mathrm{eq}}(\mathbf{s}|\hat{\boldsymbol{\ell}})}{P_{\mathrm{eq}}(\mathbf{s}|\boldsymbol{\ell})} = \max_{\boldsymbol{\ell}'} \ln \frac{P_{\mathrm{eq}}(\mathbf{s}|\boldsymbol{\ell}')}{P_{\mathrm{eq}}(\mathbf{s}|\boldsymbol{\ell})}. \tag{5}$$

_____

[3]This equation assumes that the particles in each partitions are in an equilibrium state, i.e., $P(\hat{x}|\mathbf{s}) = A(\mathbf{s}) \prod_{s=1}^{q} \delta_{s_i, s(x_i)}$, with $A(\hat{s})$ being a normalisation constant.

Eq. (5) implies that optimal work extraction is equivalent to adjusting the parameters $\hat{\boldsymbol{\ell}}$ so as to maximise the likelihood of $\mathbf{s}$ under model Eq. (2). The solution

$$\hat{\ell}_s(\mathbf{s}) \equiv \arg\max_{\boldsymbol{\ell}'} \ln P_{\mathrm{eq}}(\mathbf{s}|\boldsymbol{\ell}') = \frac{n_s}{N}, \tag{6}$$

coincides with the maximum likelihood estimators (MLE) of the parameters ℓ_s [4]. After this re-adjustment phase, the partitions are removed and the system's original equilibrium state is restored.

A cycle of the information engine includes the preparation of the non-equilibrium state and the work extraction phase. The total amount of work that can be extracted from a cycle is at most

$$\langle W_{\mathrm{ext}}\rangle_{\mathrm{neq}} \quad \leq \quad -\Delta F_{\mathrm{A}\to\mathrm{B}} - \langle W_{\mathrm{C}\to\mathrm{D}}\rangle_{\mathrm{neq}} \tag{7}$$

$$= \quad -\sum_{\mathbf{s}} P_{\mathrm{neq}}(\mathbf{s}) \ln \frac{P_{\mathrm{neq}}(\mathbf{s})}{P_{\mathrm{eq}}(\mathbf{s}|\boldsymbol{\ell})} + \sum_{\mathbf{s}} P_{\mathrm{neq}}(\mathbf{s}) \ln \frac{P_{\mathrm{eq}}(\mathbf{s}|\hat{\boldsymbol{\ell}})}{P_{\mathrm{eq}}(\mathbf{s}|\boldsymbol{\ell})} \tag{8}$$

$$= \quad -\sum_{\mathbf{s}} P_{\mathrm{neq}}(\mathbf{s}) \ln \frac{P_{\mathrm{neq}}(\mathbf{s})}{P_{\mathrm{eq}}(\mathbf{s}|\hat{\boldsymbol{\ell}})}. \tag{9}$$

We remark here that $P_{\mathrm{eq}}(\mathbf{s}|\hat{\boldsymbol{\ell}})$ is not normalized because $\hat{\boldsymbol{\ell}}$ depends on $\mathbf{s}$, through $\mathbf{n}$. The corresponding normalized distribution reads

$$P^*(\mathbf{s}) \equiv \frac{P_{\mathrm{eq}}(\mathbf{s}|\hat{\boldsymbol{\ell}}(\mathbf{s}))}{\sum_{\mathbf{s}'} P_{\mathrm{eq}}(\mathbf{s}'|\hat{\boldsymbol{\ell}}(\mathbf{s}'))} = e^{-\mathcal{R}^\star} \prod_{s=1}^{q} \left(\frac{n_s}{N}\right)^{n_s}. \tag{10}$$

The parameter $\mathcal{R}^\star$ in the normalization constant $e^{-\mathcal{R}^\star}$ is given by

$$\mathcal{R}^\star = \ln\left[N! \sum_{\mathbf{n}} \prod_{s=1}^{q} \frac{(n_s/N)^{n_s}}{n_s!} \right], \tag{11}$$

where the sum in (11) is restricted to all vectors of positive integers $\mathbf{n}$ such that $\sum_{s=1}^{q} n_s = N$. On substituing Eqs. (10) and (11) in Eq. (9), we obtain

$$\langle W_{\mathrm{ext}}\rangle_{\mathrm{neq}} \leq \mathcal{R}^\star - D[P_{\mathrm{neq}}(\mathbf{s})||P^\star(\mathbf{s})] \leq \mathcal{R}^\star, \tag{12}$$

where the first inequality saturates when the process A→B is quasistatic. In this case, the maximum average work, $\langle W_{\mathrm{ext}}\rangle_{\mathrm{neq}} = \mathcal{R}^\star$, is attained when $D[P_{\mathrm{neq}}(\mathbf{s})||P^\star(\mathbf{s})] = 0$, i.e. $P_{\mathrm{neq}}(\mathbf{s}) = P^\star(\mathbf{s})$, and thus when the non-equilibrium distribution is given by the distribution (10). Note that neither P^* nor $\mathcal{R}^\star$ depend on the initial state $\boldsymbol{\ell}$ of the engine.

We observe that the information engine does not measure the whole sample $\mathbf{s}$, but only the counts $\mathbf{n}$, which are the *sufficient statistics* of the equilibrium model (2). The constant $\mathcal{R}^*$ can be estimated asymptotically when $N \gg q$ (see e.g. [58]), and

$$\mathcal{R}^\star \simeq \frac{q-1}{2} \ln N + \frac{q}{2} \ln \pi - \ln \Gamma(q/2). \tag{13}$$

Therefore, the work extracted in the cycle grows with N as $[(q-1)/2]\ln N$. This contrasts with the case where work is extracted from the equilibrium state, where the work depends on $\boldsymbol{\ell}$ and the maximal work is bounded above by a constant, when $N \to \infty$ [54].

---

[4] As done for $\mathbf{n}$, we will drop in the following for simplicity the dependency of $\hat{\boldsymbol{\ell}}(\mathbf{s})$ on $\mathbf{s}$ and simply write $\hat{\boldsymbol{\ell}}$.

For thermodynamic processes with feedback control, the second law of thermodynamics [7] sets a limit to the amount of work that can be extracted, which cannot exceed the entropy of the measured quantity. In the present case, this bound reads $\langle W_{\text{ext}} \rangle \leq k_{\text{B}} T \, H[\boldsymbol{n}]$, where

$$H[\boldsymbol{n}] = -\sum_{\boldsymbol{n}} P^*(\boldsymbol{n}) \ln P^*(\boldsymbol{n}), \tag{14}$$

is the entropy of the measurement $\boldsymbol{n}$, and

$$P^*(\boldsymbol{n}) = e^{-\mathcal{R}^\star} N! \prod_{s=1}^{q} \frac{(n_s/N)^{n_s}}{n_s!}. \tag{15}$$

It is interesting to see how close the protocol discussed above comes to this limit. In the limit $N \to \infty$, we find (see Appendix A)

$$H[\boldsymbol{n}] \simeq (q-1) \ln N + \frac{q}{2} \ln \pi - \ln \Gamma \left(\frac{q}{2} \right) + q[\psi(q/2) - \psi(1/2)], \qquad N \gg q, \tag{16}$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function. Taking Eq. (13) into account, we find that

$$\lim_{N \to \infty} \frac{\mathcal{R}^*}{H[\boldsymbol{n}]} = \frac{1}{2}. \tag{17}$$

This result is reminiscent of those derived in Refs. [52, 53] on sensory capacity of bipartite systems. We observe that Eq. (17) has a simple interpretation in information theoretic terms, in the present setting: When $\boldsymbol{n}$ is measured, its entropy is reduced from $H[\boldsymbol{n}] \simeq (q-1) \ln N$ nats to zero. After the walls are removed, the uncertainty on the occupation numbers $\boldsymbol{n}$ increases. Indeed in state E, n_s are known to a relative precision $\delta n_s / n_s \sim 1/\sqrt{N}$, which requires $\frac{q-1}{2} \ln N$ nats[5]. Hence half of the information harvested in the measurement is lost when the walls are released (D→E).

We exemplify our result in Fig. 3 by providing the values of the optimal non-equilibrium distribution P^* −Eq. (10)− its corresponding average extracted work $k_{\text{B}} T \mathcal{R}^*$ −Eq. (11)− and the entropy $H[\boldsymbol{n}]$, in Szilárd engines with $q = 2$ partitions of equal size $\ell_1 = \ell_2 = 1/2$ for three different number of particles, $N = 1, 2, 3$. The $N = 1$ particle Szilárd engine is the only example for which $P^\star = P_{\text{eq}}$ and thus $\mathcal{R}^* = H[\boldsymbol{n}] = \ln 2$, as expected. The maximum extractable work is $k_{\text{B}} T \ln(5/2)$ for $N = 2$ and $k_{\text{B}} T \ln(26/9)$ for $N = 3$. This is 32% and 81% larger than the work that can be extracted from an initial equilibrium state for $N = 2$ and $N = 3$, respectively [54], and it is 87% and 81% of the entropy of the measurement $H[\boldsymbol{n}]$, respectively. Notably, for both $N = 2$ and $N = 3$ Szilárd engines, we find that the optimal $P^\star(\boldsymbol{s})$ is such that states where particles are distributed asymmetrically, i.e. with particles clustered in one of the boxes, are considerably more likely than under the equilibrium distribution[6]. This result suggests a well-defined procedure to achieve optimal work in Szilárd engines by means of attractive interaction potentials between the particles.

---

[5]Remember that one of the occupation numbers is fixed by the constraint $\sum_s n_s = N$

[6]We observe that, when all the particles are in the same partition, $P_{\text{eq}}(\boldsymbol{n}|\hat{\boldsymbol{\ell}})$ attains its maximal value, i.e. $\max_{\boldsymbol{n}} P_{\text{eq}}(\boldsymbol{n}|\hat{\boldsymbol{\ell}}) = 1$. This implies that the maximal work that can be extracted equals the min-entropy, i.e. $\mathcal{R}^* = -\ln \max_{\boldsymbol{n}} P^*(\boldsymbol{n})$. The min-entropy, which is the Rényi entropy of infinite order, appears in optimization problems of "one-shot" thermodynamics [39]. Fig. 3 illustrates this relation for $N = 1, 2, 3$ and $q = 2$.

N	q	(n_1, n_2)	$P^*(n_1, n_2)$	$\mathcal{R}^*$	$H[\mathbf{n}] - \mathcal{R}^*$	$\mathcal{R}^*/H[\mathbf{n}]$
1	2	[● |]	1/2	$\ln 2 \simeq 0.69$	0	1
		[| ●]	1/2			
2	2	[●● |]	2/5	$\ln \dfrac{5}{2} \simeq 0.92$	$\dfrac{1}{5}\ln 2 \simeq 0.14$	$\simeq 0.87$
		[● | ●]	1/5			
		[| ●●]	2/5			
3	2	[●●● |]	9/26	$\ln \dfrac{26}{9} \simeq 1.06$	$\dfrac{8}{26}\ln \dfrac{9}{4} \simeq 0.25$	$\simeq 0.81$
		[●● | ●]	4/26			
		[● | ●●]	4/26			
		[| ●●●]	9/26			

Figure 3: Example values for the optimal non-equilibrium distribution ($P^\star$, Eq. (10)) and its corresponding maximum average work in $k_{\mathrm{B}}T$ units ($\mathcal{R}^\star$, Eq. (11)) for $N = 1, 2, 3$ particle Szilárd engines with $q = 2$ partitions of equal size. The fourth column reports the probability $P^*(n_1, n_2)$ of finding n_1 and n_2 particles on the left and on the right partitions, respectively.

2.3 Relation to the Minimum Description Length (MDL) principle

Equations (10) and (11) are intimately related to the solution of an apparently very different problem, the problem of optimal universal coding, which sets the stage to the MDL principle. We limit our discussion to a brief account and refer to Ref. [49] for a more detailed discussion.

Let $\mathbf{s} = (s_1, \ldots, s_N)$ be a message generated by an information source that we wish to compress as much as possible. This entails to translating $\mathbf{s}$ into a sequence of nats[7]. A code is a rule that assigns to each $\mathbf{s}$ a sequence of nats by a one–to–one mapping. The *description length*, $E(\mathbf{s})$, is the number of nats of the compressed sequence. The key to achieve efficient codes is to assign short codewords to frequent patterns. This has to be done satisfying Kraft-McMillan inequality

$$\sum_{\mathbf{s}} e^{-E(\mathbf{s})} \leq 1, \tag{18}$$

which imposes a fundamental limit to uniquely decodable codes [59].

Let us consider messages $\mathbf{s}$ that are generated by drawing at random and independently each character s_i from a probability distribution, $f(s_i|\boldsymbol{\theta})$, that depends on a vector of $M \geq 1$ parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)$. If the parameters $\boldsymbol{\theta}$ are known, the optimal description length that can be

---

[7]For later convenience, we measure information in nats and take logarithms in the natural basis.

achieved (ignoring integer length constraints) is [59]

$$E(\mathbf{s}|\boldsymbol{\theta}) = -\ln P(\mathbf{s}|\boldsymbol{\theta}) \equiv -\sum_{i=1}^{N} \ln f(s_i|\boldsymbol{\theta}). \tag{19}$$

MDL deals with optimal coding in the case where $\boldsymbol{\theta}$ are unknown. Then, for a given message $\mathbf{s}$, the optimal compression is achieved minimising the description length over $\boldsymbol{\theta}$, i.e. $E(\mathbf{s}|\hat{\boldsymbol{\theta}}) = \min_{\boldsymbol{\theta}}\{-\ln P(\mathbf{s}|\boldsymbol{\theta})\}$. A code with description length $E(\mathbf{s})$ entails to an excess code length for this sample, of $\mathcal{R}(E, \mathbf{s}) = E(\mathbf{s}) - E(\mathbf{s}|\hat{\boldsymbol{\theta}})$ nats, which is called the *redundancy*[8]. The MDL identifies optimal codes with those achieving the minimal redundancy, for the worst possible sample [49]

$$\mathcal{R}^{\star} = \min_{E} \max_{\mathbf{s}} \mathcal{R}(E, \mathbf{s}), \tag{20}$$

where the minimum is taken over all $E(\mathbf{s})$ satisfying the Kraft-McMillan inequality (18). A code that solves this problem is associated with the length function, $E^*(\mathbf{s}) = -\ln P^*(\mathbf{s})$, where

$$P^*(\mathbf{s}) = e^{-\mathcal{R}^{\star}} P(\mathbf{s}|\hat{\theta}(\mathbf{s})), \tag{21}$$

which is called the normalised maximum likelihood (NML) probability distribution and

$$\mathcal{R}^{\star} = \ln \sum_{\mathbf{s}} P(\mathbf{s}|\hat{\theta}(\mathbf{s})) \tag{22}$$

$$\simeq \frac{M}{2} \ln \frac{N}{2\pi} + \int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})}, \tag{23}$$

is the *minimax universal code redundancy* [60], that we shall call minimax redundancy for short, in what follows. Equation (23) provides the large N leading behaviour [60], where M is the number of parameters (i.e., the dimension of the vector $\boldsymbol{\theta}$) and $J(\boldsymbol{\theta})$ is the Fisher information matrix. The latter is a $M \times M$ matrix with elements

$$J_{a,b}(\boldsymbol{\theta}) = \sum_{\boldsymbol{s}} P(\boldsymbol{s}|\boldsymbol{\theta}) \left[\frac{\partial}{\partial \theta_a} \ln P(\mathbf{s}|\boldsymbol{\theta})\right] \left[\frac{\partial}{\partial \theta_b} \ln P(\mathbf{s}|\boldsymbol{\theta})\right]. \tag{24}$$

Notice that, a sample $\mathbf{s}$ of length N allows to estimate each parameter to a precision of $\delta\theta_a \sim 1/\sqrt{N}$, and hence it provides $\frac{1}{2} \ln N$ nats of information for each parameter, in agreement with the first term of Eq. (23). In a loose sense, $\mathcal{R}^{\star}$ quantifies the amount of information that the sequence $\mathbf{s}$ provides on the parameters $\boldsymbol{\theta}$ of the model from which it has been drawn.[9]

The NML provides a *universal code* because it achieves a compression per data point which is as good as the compression that would be achieved asymptotically as $N \to \infty$ with the optimal choice of $\boldsymbol{\theta}$. Indeed, as shown by Rissanen [61], $\mathcal{R}^{\star}$ grows only as $\ln N$ (Eq. 23), so $\mathcal{R}^{\star}/N \to 0$ as $N \to \infty$ [10]. Note that the minimax redundancy, $\mathcal{R}^{\star}$ in Eq. (22) is independent of the sample $\mathbf{s}$. Therefore it provides a measure of the complexity of the model f, which can be used in model selection.

---

[8]$\mathcal{R}$ is also called *regret* in the literature of predictive learning theory, in the context of learning with expert advice.

[9]This statement can be made more precise within a Bayesian framework, where Eq. (23) is the asymptotic value of the mutual information $I(\mathbf{s}; \theta)$ between a sample and the parameters, under Jeffrey's prior.

[10]Moreover, the convergence of $\mathcal{R}^{\star}/N$ to zero is at the fastest possible rate when $\boldsymbol{\theta}$ is unknown (in a very strong sense) [61].

Eq. (2) is a parametric model that depends on the parameters $\theta_s = \ell_s$, $s = 1, \ldots, q-1$ with $M = q-1$ ($\ell_q = 1 - \ell_1 - \ldots - \ell_{q-1}$ fixed by the normalisation constraint). Its minimax redundancy (22) coincides with the maximal average work extracted in a cycle of the N-particle Szilárd engine (11), and the NML (21) coincides with the optimal non-equilibrium distribution $P^\star$ in Eq. (10). This is the main result of this paper; it connects information thermodynamics with the MDL principle. We provide in Table 1 a "dictionary" of analogies between different quantities in information thermodynamics and coding that stem from our theory.

In the rest of the paper, we shall see that the subtle connection between thermodynamics and coding unveiled here is a very general result that applies to a broad class of parametric models. As we will see, this is most easily seen if the original Szilárd engine is described in terms of a model of a system of non-interacting classical particles in q energy levels.

information engine	coding (MDL)
microscopic state $\mathbf{s}$	message $\mathbf{s}$
particle coordinate s_i	character s_i
measurement $\mathbf{n}$ ($\mathbf{\Phi}$)	sufficient statistics $\mathbf{n}$ ($\mathbf{\Phi}$)
protocol parameters $\boldsymbol{\ell}$	model's parameters $\boldsymbol{\theta}$
equilibrium state $P_{\text{eq}}(\mathbf{s}|\boldsymbol{\ell})$	parametric model $P(\mathbf{s}|\boldsymbol{\theta}) = e^{-E(\mathbf{s}|\boldsymbol{\theta})}$
non-equilibrium state $P_{\text{neq}}(\mathbf{s})$	(exponentiated) coding function $e^{-E(\mathbf{s})}$
optimal non-equilibrium state $P^*(\mathbf{s})$	optimal coding distribution $P^*(\mathbf{s})$ (NML)
maximal work extracted $\langle W_{\text{ext}} \rangle$	minimax universal code redundancy $\mathcal{R}^*$
energy levels $\epsilon_s = -\ln \ell_s$	description length ϵ_s for character s

Table 1: Correspondence between different concepts in information engines and in MDL coding.

3 Physical implementation and further extensions

3.1 A physical realization of the N-particle Szilárd engine

We now discuss, in the light of stochastic thermodynamics, a physical analog of a Szilárd engine which enlightens extensions of our theory to more general scenarios. The model consists of N classical non-interacting particles in q energy levels $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_q)$ immersed in a thermal bath with $k_{\mathrm{B}}T = 1$. We let s_i denote the energy level occupied by the ith particle and n_s be occupation number of the sth energy level. Each level can be occupied by a maximum of N particles, with the constraint $\sum_{s=1}^{q} n_s = N$, i.e. conservation of the total number of particles. The equilibrium probability of a configuration $\mathbf{s} = (s_1, \ldots, s_N)$ is

$$P_{\mathrm{eq}}(\mathbf{s}|\boldsymbol{\epsilon}) = \frac{e^{-\mathcal{H}(\mathbf{s})}}{Z(\boldsymbol{\epsilon})} = e^{-[\mathcal{H}(\mathbf{s}) - \mathcal{F}_{\mathrm{eq}}(\boldsymbol{\epsilon})]}, \tag{25}$$

where $\mathcal{H}(\mathbf{s}) = \sum_{i=1}^{N} \epsilon_{s_i} = \sum_{s=1}^{q} n_s \epsilon_s$ is the total energy (Hamiltonian) of the particles, $Z(\boldsymbol{\epsilon}) = \left(\sum_{s=1}^{q} e^{-\epsilon_s} \right)^N$ the partition function, and $\mathcal{F}_{\mathrm{eq}}(\boldsymbol{\epsilon}) = -\ln Z(\boldsymbol{\epsilon})$ the equilibrium free energy. Setting

$$\ell_s = \frac{e^{-\epsilon_s}}{\sum_{s'=1}^{q} e^{-\epsilon_{s'}}} \tag{26}$$

for all $s = 1, \ldots, q$, Eq. (25) becomes identical to Eq. (2), where ℓ_s is the probability that a particle is found in partition s of the Szilárd engine. Hence the system of particles described by Eq. (25) is formally equivalent to a Szilárd engine with a box of size $L = \sum_{s=1}^{q} e^{-\epsilon_s}$ in the x-direction and partitions of size $e^{-\epsilon_s}$. Note that $L = 1$ can be enforced by an appropriate shift of all energy levels and that would imply $\mathcal{F}_{\mathrm{eq}}(\boldsymbol{\epsilon}) = 0$.
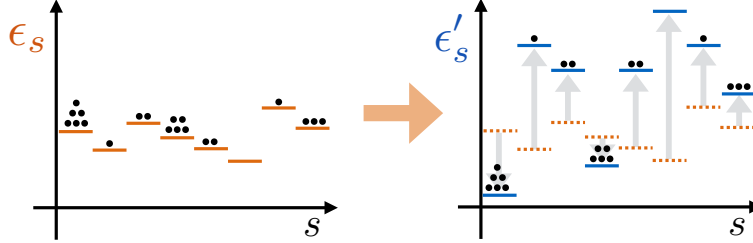

Figure 4: Re-adjustment of energy levels during the work extraction transformation C→D. The occupation of energy levels, $\mathbf{n}$, stays the same during this transformation.

Within this analogy, the non-equilibrium driving in the Szilárd engine A→B requires an amount of work of at least, see Eq. (3),

$$\langle W_{\mathrm{A} \to \mathrm{B}} \rangle \geq \Delta F_{\mathrm{A} \to \mathrm{B}} = \langle \ln P_{\mathrm{neq}}(\mathbf{s}) \rangle_{\mathrm{neq}} + \langle \mathcal{H}(\mathbf{s}) \rangle_{\mathrm{neq}} - \mathcal{F}_{\mathrm{eq}}(\boldsymbol{\epsilon}). \tag{27}$$

The work extraction phase C→D corresponds to adjusting the energy levels $\boldsymbol{\epsilon} \to \boldsymbol{\epsilon}'$ without changing the population $\mathbf{n}$ of the levels (see Fig. 4). Since particles do not jump between different levels,

no heat is released in this transformation[11], so the work equals the change in internal energy $W_{\mathrm{C\to D}}(\mathbf{s}) = \mathcal{H}'(\mathbf{s}) - \mathcal{H}(\mathbf{n})$, where $\mathcal{H}'(\mathbf{s}) = \sum_{i=1}^{N} \epsilon'_{s_i} = \sum_{s=1}^{q} n_s \epsilon'_s$.

When the partitions are removed D→E the occupation numbers $\mathbf{n}$ equilibrate to the new energy levels without any work expenditure[12] ($W_{\mathrm{D\to E}} = 0$). The cycle is completed when the system is driven back to the initial equilibrium E→A, which requires a work[13] $W_{\mathrm{E\to A}} = \mathcal{F}_{\mathrm{eq}}(\boldsymbol{\epsilon}) - \mathcal{F}_{\mathrm{eq}}(\boldsymbol{\epsilon}')$.

Taken all together, the work extracted in a cycle satisfies the inequality

$$
\begin{aligned}
W_{\mathrm{ext}}(\mathbf{n}, \boldsymbol{\epsilon}') \quad \leq \quad & -\Delta F_{\mathrm{A\to B}} - W_{\mathrm{C\to D}}(\mathbf{s}) - W_{\mathrm{E\to A}} && (28) \\
= \quad & -\langle \ln P_{\mathrm{neq}}(\mathbf{s})\rangle_{\mathrm{neq}} - \langle \mathcal{H}(\mathbf{s})\rangle_{\mathrm{neq}} + \mathcal{H}(\mathbf{s}) - \mathcal{H}'(\mathbf{s}) + \mathcal{F}_{\mathrm{eq}}(\boldsymbol{\epsilon}') && (29) \\
= \quad & -\langle \ln P_{\mathrm{neq}}(\mathbf{s})\rangle_{\mathrm{neq}} + \mathcal{H}(\mathbf{s}) - \langle \mathcal{H}(\mathbf{s})\rangle_{\mathrm{neq}} + \ln P_{\mathrm{eq}}(\mathbf{s}|\boldsymbol{\epsilon}') && (30)
\end{aligned}
$$

where we have used the definition (25) in the last step. The only term that depends on the new energy levels $\boldsymbol{\epsilon}'$ in Eq. (30) is the last one. Therefore the lower bound to the work extracted is maximal for

$$
\hat{\boldsymbol{\epsilon}}(\mathbf{s}) = \arg\max_{\boldsymbol{\epsilon}'} \left[\ln P_{\mathrm{eq}}(\mathbf{s}|\boldsymbol{\epsilon}')\right]. \tag{31}
$$

Note that work optimization corresponds to maximizing the likelihood, i.e. setting $\epsilon'_s = \hat{\epsilon}_s(\mathbf{s}) = -\ln(n_s/N)$ which are the MLE of the distribution (25) (see Eq. 5). Also note that the optimal energy levels $\hat{\boldsymbol{\epsilon}}$ depend on $\mathbf{s}$ through the measurement $\mathbf{n}$.

Taking the average of Eq. (30) with $\boldsymbol{\epsilon}' = \hat{\boldsymbol{\epsilon}}$ over P_{neq}, we find the following upper bound for the work extracted averaged over many cycles (see Eq. 9),

$$
\begin{aligned}
\langle W_{\mathrm{ext}}\rangle \quad = \quad & -\left\langle \ln \frac{P_{\mathrm{neq}}(\mathbf{s})}{P_{\mathrm{eq}}(\mathbf{s}|\hat{\boldsymbol{\epsilon}})}\right\rangle_{\mathrm{neq}} \\
= \quad & \mathcal{R}^\star - D[P_{\mathrm{neq}}(\mathbf{s})||P^\star(\mathbf{s})] \leq \mathcal{R}^\star, && (32)
\end{aligned}
$$

where again

$$
P^*(\mathbf{s}) = e^{-\mathcal{R}^\star} P_{\mathrm{eq}}(\mathbf{s}|\hat{\boldsymbol{\epsilon}}(\mathbf{s})), \tag{33}
$$

is the NML and

$$
\mathcal{R}^\star = \ln\left[\sum_{\mathbf{s}} P_{\mathrm{eq}}(\mathbf{s}|\hat{\boldsymbol{\epsilon}}(\mathbf{s}))\right]. \tag{34}
$$

The first inequality in (32) is saturated in the quasistatic limit whereas the second inequality in (32) is saturated when the non-equilibrium state is drawn from the NML distribution (33).

3.2   Optimal work extraction measuring sufficient statistics: theory

We now ask the question: how do these results generalise when the measured quantities $\boldsymbol{\Phi}$ are different from $\mathbf{n}$? In order to address this question, we consider a model where the energy levels

$$
\epsilon_s(\boldsymbol{\theta}) = \epsilon_s^0 - \boldsymbol{\theta} \cdot \boldsymbol{\phi}(s) = \epsilon_s^0 - \sum_{m=1}^{M} \theta_m \phi_m(s), \tag{35}
$$

---

[11]Note that if the energy levels $\{\epsilon'_s\}$ are such that $\sum_s e^{-\epsilon'_s} \neq 1$, this process also involves an expansion of the Szilárd box as a whole followed by a compression (or *vice versa*). This global expansion/compression does not occur in the optimal protocol.

[12]When partitions are removed ($D \to E$), an amount $Q(\mathbf{s}) = \sum_s [\langle n_s\rangle_{\mathrm{eq}'} - n_s]\epsilon'_s$ of heat is released to the thermal bath, where the average $\langle \ldots\rangle_{\mathrm{eq}'}$ is taken on the equilibrium distribution with energies $\boldsymbol{\epsilon}'$.

[13]We shall see that $W_{\mathrm{E\to A}} = 0$ when the engine operates under the optimal protocol.

depend on the parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)$ through the M functions $\phi(s) = (\phi_1(s), \ldots, \phi_M(s))$. Here the dot denotes the scalar product and ϵ_s^0 specifies the energy of level s when it is unoccupied ($n_s = 0$). For a microscopic state $\boldsymbol{s}$, the Hamiltonian now takes the value

$$\mathcal{H}(\mathbf{s}|\boldsymbol{\theta}) = \mathcal{H}_0(\mathbf{s}) - \boldsymbol{\theta} \cdot \boldsymbol{\Phi}(\boldsymbol{s}), \tag{36}$$

with

$$\mathcal{H}^0(\mathbf{s}) = \sum_{i=1}^{N} \epsilon_{s_i}^0 = \sum_{s=1}^{q} n_s \epsilon_s^0, \tag{37}$$

and the vector

$$\boldsymbol{\Phi}(\boldsymbol{s}) = \left(\sum_{s=1}^{q} \phi_1(s) n_s, \ldots, \sum_{s=1}^{q} \phi_M(s) n_s \right). \tag{38}$$

The equilibrium distribution becomes

$$P_{\text{eq}}(\mathbf{s}|\boldsymbol{\theta}) = e^{-\mathcal{H}^0(\mathbf{s}) + \boldsymbol{\theta} \cdot \boldsymbol{\Phi}(\mathbf{s}) + \mathcal{F}_{\text{eq}}(\boldsymbol{\theta})}, \tag{39}$$

with the corresponding equilibrium free energy given by

$$\mathcal{F}_{\text{eq}}(\boldsymbol{\theta}) = -\ln \left[\sum_{\mathbf{s}} e^{-\mathcal{H}^0(\mathbf{s}) + \boldsymbol{\theta} \cdot \boldsymbol{\Phi}(\mathbf{s})} \right]. \tag{40}$$

From the viewpoint of mathematical statistics, the parametric probability distribution (39) belongs to an *exponential family*. Exponential families have special properties in that they exhibit explicitly the sufficient statistics, i.e., those functions of the data whose knowledge is sufficient to estimate the parameters from a sample. In the present context, notice that $P_{\text{eq}}(\boldsymbol{s}|\boldsymbol{\theta}) = \prod_{i=1}^{N} f(s_i|\boldsymbol{\theta})$ is the probability of a sequence $\boldsymbol{s}$ of N independent draws s_i from the distribution $f(s|\boldsymbol{\theta}) = e^{-\epsilon_s^0 + \boldsymbol{\theta}\phi(s) - \varphi(\boldsymbol{\theta})}$, with $\varphi(\boldsymbol{\theta}) = \mathcal{F}_{\text{eq}}(\boldsymbol{\theta})/N$. For this distribution, $\boldsymbol{\Phi}(\boldsymbol{s})$ is a sufficient statistics, in the sense that it contains all necessary information about $\boldsymbol{\theta}$ [59], i.e. the mutual information between $\boldsymbol{s}$ and the parameters $\boldsymbol{\theta}$ equals that between $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$:

$$I(\boldsymbol{s}; \boldsymbol{\theta}) = I(\boldsymbol{\theta}; \boldsymbol{\Phi}). \tag{41}$$

Exponential families are unique, in that the Pitman-Koopman-Darmois theorem [62] ensures that the set of sufficient statistics of a parametric model is independent of N only if the model has the exponential form of Eq. (39). As we shall see, this is a crucial step in the work extraction cycle from measurements of sufficient statistics.

An information engine based on the model (39) performs a cycle analogous to that shown in Fig. 2, with the sole exceptions that here the parameters of the model are $\boldsymbol{\theta}$ instead of $\boldsymbol{\ell}$ and the values of the statistics $\boldsymbol{\Phi}(\mathbf{s})$ are measured instead of $\mathbf{n}(\mathbf{s})$. Furthermore, work is extracted by adjusting the conjugate parameters $\boldsymbol{\theta} \to \boldsymbol{\theta}'$. The work extracted in a cycle can be calculated in exactly the same way as in the discussion leading to Eq. (30). Note that, by Eqs. (36, 39), W_{ext} depend on $\boldsymbol{s}$ only through the measured quantity Φ. The work extracted is maximal for

$$\hat{\boldsymbol{\theta}}(\mathbf{s}) = \arg\max_{\boldsymbol{\theta}'} \left[\ln P_{\text{eq}}(\mathbf{s}|\boldsymbol{\theta}') \right] = \arg\max_{\boldsymbol{\theta}'} \left[\boldsymbol{\theta}\boldsymbol{\Phi}(\boldsymbol{s}) - \mathcal{F}_{\text{eq}}(\boldsymbol{\theta}) \right] \tag{42}$$

which is again the MLE. The crucial point is that the measured quantity $\boldsymbol{\Phi}$ should provide sufficient information to compute the values $\hat{\boldsymbol{\theta}}$ of the optimal parameters in the work extraction phase.

Therefore $\boldsymbol{\Phi}$ must be a sufficient statistics of model Eq. (39) and, by the Pitman-Koopman-Darmois theorem [62], $P_{\text{eq}}(\mathbf{s}|\boldsymbol{\theta})$ needs to have the exponential form of Eq. (39), for this to be possible for all N. Taking the expected value of the optimal extracted work, following the same steps of the previous section, we arrive at

$$\langle W_{\text{ext}} \rangle \leq \mathcal{R}^\star - D[P_{\text{neq}}(\mathbf{s})||P^\star(\mathbf{s})], \tag{43}$$

where P^* is the NML distribution (21) and $\mathcal{R}^*$ is the minimax redundancy of Eq. (22). This shows that the maximal work that can be extracted from a cycle, on average, equals the minimax redundancy $\mathcal{R}^*$, and that the non-equilibrium state that achieves this bound is the NML, i.e $P_{\text{neq}} = P^*$. The only change, with respect to the previous section, is that the re-adjustment of energy levels $\boldsymbol{\epsilon}(\boldsymbol{\theta})$ is constrained by the functional form of Eq. (35), to those that can be achieved by fine-tuning the parameters $\boldsymbol{\theta}$, conjugate to the measured quantity.

Also in this case, this result is independent of the equilibrium parameters $\boldsymbol{\theta}$ in which the engine is initialised and the optimal work extracted grows asymptotically as $\frac{M}{2} \ln N$ with the number of particles, for $N \gg 1$ (see Eq. (23)).

Let us now see how close the maximal extracted work approaches the fundamental limit imposed by the second law of thermodynamics, $\langle W_{\text{ext}} \rangle \leq k_B T H[\boldsymbol{\Phi}]$ [7], where $H[\boldsymbol{\Phi}]$ is the entropy of the measurement. In the limit of large $N \gg M$, to leading order, we find (see Appendix A)

$$H[\boldsymbol{\Phi}] \simeq M \ln N + \ln \int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})} + \frac{\int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})} \ln \sqrt{\det J(\boldsymbol{\theta})}}{\int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})}}. \tag{44}$$

This confirms that, as in Eq. (17), $\mathcal{R}^*/H[\boldsymbol{\Phi}] \to 1/2$ as $N \to \infty$. The intuition at the basis of this result is the same as in Section 2.2: The total reduction in the uncertainty on $\boldsymbol{\Phi}$ brough about by the measurement is $H[\boldsymbol{\Phi}] \simeq M \ln N$. A part of this is lost when the walls are removed, and the system relaxes to state E. In state E, the relative fluctuations of each component of $\boldsymbol{\Phi}$ are of order $1/\sqrt{N}$, hence the loss amounts to $\frac{1}{2} \ln N$ nats per parameter. This result tallies with the findings of Ref. [53] in bipartite systems, that the measurement of sufficient statistics implies losses in thermodynamic efficiency.

3.3 Optimal work extraction measuring sufficient statistics: examples

Let us consider the case $q = 4$ and let us label the state of the i-th particle with a spin representation $s_i = \vec{\sigma}_i = (\sigma_i^x, \sigma_i^y)$, with $\sigma_i^x = \pm 1$ and $\sigma_i^y = \pm 1$ $(i = 1, \ldots, N)$. We let $\boldsymbol{\sigma} = (\vec{\sigma}_1, \ldots, \vec{\sigma}_N)$ denote the state of the N particle system in this representation. For the model

$$P(\boldsymbol{\sigma}|J) = \frac{e^{J\Phi(\boldsymbol{\sigma})}}{(4\cosh J)^N}, \qquad \Phi(\boldsymbol{\sigma}) = \sum_{i=1}^{N} \sigma_i^x \sigma_i^y, \tag{45}$$

the box sizes and the energy levels for a given value of the coupling constant J are given by

$$\ell_{\boldsymbol{\sigma}}(J) = \frac{e^{J\sigma^x\sigma^y}}{4\cosh J}, \qquad \epsilon_{\boldsymbol{\sigma}}(J) = -J\sigma^x\sigma^y,$$

respectively. As shown in Fig. 5(a), $\ell_{\boldsymbol{\sigma}}(J)$ and $\epsilon_{\boldsymbol{\sigma}}(J)$ cannot be adjusted independently In particular, under this model, the partition between the box $[-, +]$ and box $[+, -]$ cannot move. Optimal
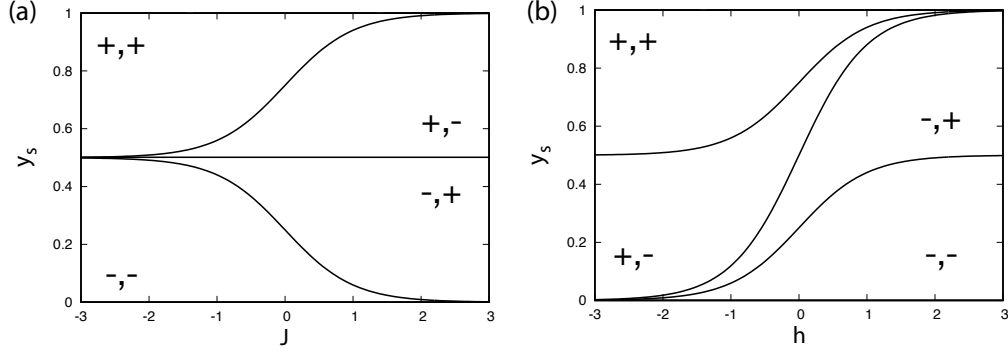
15

Figure 5: Two examples of a Szilárd box with $q = 4$ partitions, labeled $[-,-]$, $[-,+]$, $[+,-]$ and $[+,+]$ from bottom to top. (a) The model in Eq. (45), corresponding to partitions that can move as the coupling constant J varies. (b) The model in Eq. (48), corresponding to partitions that can move as the intensity of the magnetic field h varies. Different lines correspond to the coordinates y_s ($s = 1, 2, 3, 4$) of the different walls for different values of the parameters.

work extraction demands that the parameter J be adjusted to the MLE value

$$\hat{J}(\boldsymbol{\sigma}) = \arg\max_{J} \left[J\boldsymbol{\Phi}(\boldsymbol{\sigma}) - \ln(4\cosh J) \right] = \arg\tanh\left[\frac{\Phi(\boldsymbol{\sigma})}{N} \right]. \tag{46}$$

The maximal work that can be extracted from a cycle, in this case equals the one that can be extracted from a Szilárd box with $q = 2$ partitions and N particles (see Fig. 3 and Eq. (13) for $N \gg 1$).

Analogous considerations can be applied to the model

$$P(\boldsymbol{\sigma}|h) = \frac{e^{h\mathbf{M}^z(\boldsymbol{\sigma})}}{(4\cosh h)^N}, \qquad \mathbf{M}^x(\boldsymbol{\sigma}) = \sum_{i=1}^{N} \sigma_i^x, \tag{47}$$

where again the MLE value is given by $\hat{h}(\boldsymbol{\sigma}) = \arg\tanh[\mathbf{M}^x(\boldsymbol{\sigma})/N]$. The partitions are shown in Fig. 5(b) for this case, as a function of h. Notice that the two cases differ only by a permutation of the partitions, hence the value of $\mathcal{R}^*$ is the same as above.

In a system where both $\mathbf{M}^x$ and $\boldsymbol{\Phi}$ can be measured, the optimal non-equilibrium state to which the system should be driven would be given by the NML of the model

$$P(\boldsymbol{\sigma}|h) = \frac{e^{h\mathbf{M}^z(\boldsymbol{\sigma})+J\Phi(\boldsymbol{\sigma})}}{(16\cosh h\cosh J)^N}, \qquad \mathbf{M}^x(\boldsymbol{\sigma}) = \sum_{i=1}^{N} \sigma_i^x. \tag{48}$$

The energy levels $\epsilon_{\boldsymbol{\sigma}}(h, J)$ depend on both h and J, which provides more flexibility in adjusting partitions. Hence more work can be extracted. Detailed calculation shows that, in this case, twice as much work can be extracted, compared to the previous cases.

16

4 Discussion

The result of this paper lends itself to a number of considerations:

4.1 Efficient communication as ideal information engines

The process of measuring a quantity Φ can be seen as a communication problem between two parties, Alice and Bob. In order to focus on thermodynamic efficiency, we model the channel over which Alice and Bob communicate, as an information engine. The messages that Alice and Bob exchange are a sequence $\mathbf{s} = (s_1, \ldots, s_N)$ of symbols from a finite alphabet, where s_i are drawn i.i.d. from a certain distribution. The exchange is realized by an information engine that operates at a constant temperature. Let $\Phi(\mathbf{s})$ be the value of the quantity that is measured. Alice and Bob both have a device, that is initially in state $P_0(\mathbf{s})$. They can choose to code their messages with a distribution $P_{\mathrm{neq}}(\mathbf{s})$. This requires Alice to spend $\Delta F = D[P_{\mathrm{neq}}(\mathbf{s})||P_0(\mathbf{s})]$ units of $k_B T$ to generate messages, but it allows Bob to harvest work $-W(\Phi)$ from the measurement of $\Phi(\mathbf{s})$. If Alice and Bob exchange messages both ways it makes sense for them to agree on a design that minimises the average energetic cost $\Delta F + \langle W \rangle$ of the communication, over many cycles. Our result implies that (i) P_0 should be of the form Eq. (39); (ii) $P_{\mathrm{neq}} = P^*$ should be the corresponding NML, and (iii) some work can be extracted for each message, that on average cannot exceed the minimax redundancy $\mathcal{R}^\star$ times $k_B T$. The channel should work as follows: A value of $\mathbf{\Phi}$ is drawn at random from $P^*(\mathbf{\Phi})$. Knowing $\mathbf{\Phi}$, Alice can prepare the engine with partitions of sizes dictated by $\hat{\boldsymbol{\theta}}(\mathbf{\Phi})$ and distribute the N particles in such a way that $\mathbf{\Phi}(\boldsymbol{s})$ takes the desired value $\mathbf{\Phi}$. Then she can operate a quasistatic transformation to bring back the partitions to the sizes that correspond to the equilibrium values $\boldsymbol{\theta}$, spending ΔF units of $k_B T$. Then Alice can pass the information engine to Bob, who would measure $\mathbf{\Phi}$ and extract work from the isothermal expansion of the partitions.

Interestingly, the costs and benefits across the communication channel are shared unequally in this optimal protocol. This argument neglects the cost needed to erase the memory in which $\mathbf{\Phi}$ is stored, which is at least equal to the Shannon entropy $H[\mathbf{\Phi}]$ of $\mathbf{\Phi}$. This makes sense in a communication setting, where $\mathbf{\Phi}$ is valuable information that does not need to be erased. In addition, we show that this communication protocol provides an energy budget that can cover half of the memory cost, in the limit of long messages.

4.2 The importance of being critical

The NML codes have been recently shown to be "critical" in a very precise sense. First, Ref. [50] has shown that, a typical sample drawn from P^* exhibits *statistical criticality* [63], i.e. the distribution of n_s is very broad. Second, studying large deviations of the in-sample description length per character, $\hat{H}[s] = -\sum_{s=1}^{q} \hat{\ell}_s \ln \hat{\ell}_s$, for samples generated from P^*, Ref. [50] has shown that P^* sits exactly at a critical point. This is because there are no codes that can achieve a better compression than P^*. It is interesting that extracting the maximal amount of work from the cycle of an information engine, requires driving the system to a critical state. This provides hints for possible experimental protocols towards more efficient information engines.

4.3 Which observables should be measured?

The above discussion suggests that most efficient communication is achieved for sufficient statistics Φ such that $\mathcal{R}^\star$ is largest. This result provides further insights with respect to recent work [53]

which pointed to the relevance of the existence of sufficient statistics in information-thermodynamic efficiency. We have shown that, for fixed number of particles N and parameters M, the maximal work extracted by a broad class of information engines depends on the value $c = \int d\theta \sqrt{\det \mathbf{J}(\theta)}$ of the constant term in Eq. (23), which is reminiscent of the so-called "thermodynamic length" introduced in Refs. [35, 64, 65]. Observables for which the corresponding value of c is larger should allow for more energy efficient measurement devices. For example Ref. [58] has studied the term c for spin models with interactions of arbitrary order and found that pairwise models achieve large values of c.

4.4 Experimental applications

Our results shed light on potential future experimental applications in stochastic thermodynamics. Until now, experimental realizations of one-particle Szilárd engines have been achieved using e.g. colloidal particles trapped with optical tweezers [20, 21, 22] and feedback traps [24], DNA hairpins [28], and single-electron transistors [23]. Most of these experiments rely on the accurate control of a single degree of freedom, e.g. the position of a colloid along the x axis, the energy level occupied by one extra electron in a single-electron transistor, etc. This leads to optimal work extractions of the order of $k_{\mathrm{B}}T$ by means of current experimental techniques.

We have shown that Szilárd engines extracting work above the "$k_{\mathrm{B}}T$ limit" can be realized physically, e.g. with N-particle systems in q energy levels. Our formalism provides exact formulas for the non-equilibrium state at which an N-particle information engine should be driven and how much work can be extracted from the measurement of an arbitrary coarse-grained physical quantity. The potential theoretical extensions to bosonic systems and experimental realizations of these results with e.g. polaritons or photons trapped in optical cavity networks [66, 67], are two interesting avenues of future research.

5 Acknowledgments

A The entropy of the measurement

Let us compute the entropy of measurement in the case of a generic sufficient statistics $\boldsymbol{\Phi}$ and then specialise to the case of the standard Szilárd engine, where $\boldsymbol{\Phi} = \boldsymbol{n}$. For finite N, $\boldsymbol{\Phi}$ takes values in a discrete set. The entropy of $\boldsymbol{\Phi}$ is defined as:

$$H[\boldsymbol{\Phi}] = -\sum_{\boldsymbol{\Phi}} P^*(\boldsymbol{\Phi}) \ln P^*(\boldsymbol{\Phi}) \tag{49}$$

where

$$P^*(\boldsymbol{\Phi}) = \sum_{\boldsymbol{s}} P^*(\boldsymbol{s})\delta\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}(\boldsymbol{s})\right) = e^{-\mathcal{R}} \sum_{\boldsymbol{s}} P(\boldsymbol{s}|\hat{\boldsymbol{\theta}}(\boldsymbol{s}))\delta\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}(\boldsymbol{s})\right). \tag{50}$$

For large $N \gg M$ we can use the relation

$$P(s|\hat{\boldsymbol{\theta}}(s)) \simeq \left(\frac{N}{2\pi}\right)^{M/2} \int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})} P(s|\boldsymbol{\theta}) \tag{51}$$

which is easily verified by performing the integral on the right hand side by the saddle point method (see e.g. [50]). Here and below the $\simeq$ sign stands for an asymptotic equality, that holds in the limit $N \to \infty$, in the sense that terms that are neglected on the right hand side vanish as $N \to \infty$. Using also

$$\mathcal{R} \simeq \frac{M}{2} \ln N + c, \qquad c \equiv \ln \int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})}, \tag{52}$$

we rewrite Eq. (50) as

$$P^*(\boldsymbol{\Phi}) \simeq e^{-c} \int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})} P(\boldsymbol{\Phi}|\boldsymbol{\theta}) \tag{53}$$

with

$$P(\boldsymbol{\Phi}|\boldsymbol{\theta}) = \sum_s P(s|\boldsymbol{\theta}) \delta\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}(s)\right) = e^{\boldsymbol{\theta}\boldsymbol{\Phi} + \mathcal{F}_{\mathrm{eq}}(\boldsymbol{\theta}) + \mathcal{S}(\boldsymbol{\Phi})}, \tag{54}$$

where

$$\mathcal{S}(\boldsymbol{\Phi}) = \ln \sum_s \delta\left(\boldsymbol{\Phi} - \boldsymbol{\Phi}(s)\right) \tag{55}$$

is the entropy (i.e. the log of the number) of configurations with $\boldsymbol{\Phi}(s) = \boldsymbol{\Phi}$. All quantities in the exponential of Eq. (54) are extensive. Let us introduce to the intensive variable

$$\bar{\boldsymbol{\phi}} = \frac{\boldsymbol{\Phi}}{N}. \tag{56}$$

For N large, $\bar{\boldsymbol{\phi}}$ takes values on a dense set, so we change sums over $\boldsymbol{\Phi}$ into integrals over $\bar{\boldsymbol{\phi}}$

$$\sum_{\boldsymbol{\Phi}} (\ldots) \simeq N^M \int d\bar{\boldsymbol{\phi}} (\ldots), \tag{57}$$

where the factor N^M comes from the Jacobian of the transformation $\boldsymbol{\Phi} \to \bar{\boldsymbol{\phi}}$. Correspondingly, distributions change into probability density functions, as

$$P(\boldsymbol{\Phi}|\boldsymbol{\theta}) \simeq N^{-M} P(\bar{\boldsymbol{\phi}}|\boldsymbol{\theta}), \quad P^*(\boldsymbol{\Phi}) \simeq N^{-M} p^*(\bar{\boldsymbol{\phi}}). \tag{58}$$

We also introduce the intensive functions

$$\varphi(\boldsymbol{\theta}) = \frac{1}{N} \mathcal{F}_{\mathrm{eq}}(\boldsymbol{\theta}), \qquad \sigma(\bar{\boldsymbol{\phi}}) = \frac{1}{N} \mathcal{S}(N\bar{\boldsymbol{\phi}}). \tag{59}$$

With this change of variables, the entropy now reads:

$$H[\boldsymbol{\Phi}] \simeq M \ln N - \int d\bar{\boldsymbol{\phi}} \, p^*(\bar{\boldsymbol{\phi}}) \ln p^*(\bar{\boldsymbol{\phi}}) \tag{60}$$

Now, $p(\bar{\boldsymbol{\phi}}|\boldsymbol{\theta})$ is a sharply peaked function on the value $\hat{\boldsymbol{\phi}}(\boldsymbol{\theta})$ that dominates the saddle point, i.e. such that $\sigma'(\bar{\boldsymbol{\phi}}) = \boldsymbol{\theta}$. Notice that $-\sigma(\bar{\boldsymbol{\phi}})$ is the Legendre transform of $\varphi(\boldsymbol{\theta})$ and vice-versa. In the limit $N \to \infty$ we can write

$$p(\bar{\boldsymbol{\phi}}|\boldsymbol{\theta}) \simeq \delta\left(\bar{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}(\boldsymbol{\theta})\right) \tag{61}$$

19

Therefore

$$p^*(\bar{\boldsymbol{\phi}}) \simeq e^{-c} \int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})} \delta\left(\bar{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}(\boldsymbol{\theta})\right) \tag{62}$$

$$= e^{-c} \int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})} \frac{\delta\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\bar{\boldsymbol{\phi}})\right)}{\det J(\boldsymbol{\theta})} \tag{63}$$

$$= e^{-c} \frac{1}{\sqrt{\det J(\hat{\boldsymbol{\theta}}(\bar{\boldsymbol{\phi}}))}} \tag{64}$$

$$= e^{-c} \sqrt{\det K(\bar{\boldsymbol{\phi}})} \tag{65}$$

where, in Eq. (62), we changed variables $\bar{\boldsymbol{\phi}} \to \boldsymbol{\theta}$ in the delta function, using the fact that the Jacobian is given by

$$J(\boldsymbol{\theta}) = \frac{d\hat{\boldsymbol{\phi}}}{d\boldsymbol{\theta}}, \tag{66}$$

which is the Fisher Information matrix. The matrix K is instead the Jacobian of the inverse transformation $\boldsymbol{\theta} \to \bar{\boldsymbol{\phi}}$

$$K(\bar{\boldsymbol{\phi}}) = \frac{d\hat{\boldsymbol{\theta}}(\bar{\boldsymbol{\phi}})}{d\bar{\boldsymbol{\phi}}} = J^{-1}\left(\hat{\boldsymbol{\theta}}(\bar{\boldsymbol{\phi}})\right). \tag{67}$$

Note that Eq. (65) is correctly normalised, because, changing variables $\boldsymbol{\theta} \to \bar{\boldsymbol{\phi}}$ in Eq. 52), one finds

$$c = \ln \int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})} \tag{68}$$

$$= \ln \int d\bar{\boldsymbol{\phi}} \det K(\bar{\boldsymbol{\phi}}) \sqrt{\det J\left(\hat{\boldsymbol{\theta}}(\bar{\boldsymbol{\phi}})\right)} \tag{69}$$

$$= \ln \int d\bar{\boldsymbol{\phi}} \sqrt{\det K(\bar{\boldsymbol{\phi}})}. \tag{70}$$

Therefore in the limit $N \to \infty$

$$H[\boldsymbol{\Phi}] \simeq M \ln N + c - e^{-c} \int d\bar{\boldsymbol{\phi}} \sqrt{\det K(\bar{\boldsymbol{\phi}})} \ln \sqrt{\det K(\bar{\boldsymbol{\phi}})} \tag{71}$$

$$= M \ln N + c + e^{-c} \int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})} \ln \sqrt{\det J(\boldsymbol{\theta})} \tag{72}$$

$$= M \ln N + \ln \int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})} + \frac{\int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})} \ln \sqrt{\det J(\boldsymbol{\theta})}}{\int d\boldsymbol{\theta} \sqrt{\det J(\boldsymbol{\theta})}} \tag{73}$$

which, together with Eq. (52), implies that

$$\lim_{N \to \infty} \frac{\mathcal{R}}{H[\boldsymbol{\Phi}]} = \frac{1}{2}. \tag{74}$$

In the case of the Szilárd engine with $\boldsymbol{\Phi} = \boldsymbol{n}$, we have (see e.g. [58])

$$\det J(\boldsymbol{\ell}) = \prod_{s=1}^{q} \ell_s^{-1} \delta\left(\sum_{s=1}^{q} \ell_s - 1\right). \tag{75}$$

20

We can use the identity

$$\int_0^1 d\boldsymbol{\ell} \prod_{s=1}^q \ell_s^{\alpha-1} \delta\left(\sum_{s=1}^q \ell_s - 1\right) = \frac{\Gamma(\alpha)^q}{\Gamma(q\alpha)} \equiv e^{\Upsilon(\alpha)}, \tag{76}$$

which also defines the function $\Upsilon(\alpha)$. Upon inspection, we find that $c = \Upsilon(1/2)$ and the constant term in Eq. (73) is given by

$$-\frac{d\Upsilon}{d\alpha}\bigg|_{\alpha=1/2} = q\left[\psi(q/2) - \psi(1/2)\right],$$

as in Eq. (16).

References

[1] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM Journal of Research and Development*, vol. 5, pp. 183–191, July 1961.

[2] C. H. Bennett, "The thermodynamics of computation—a review," *Int. J. Theor. Phys.*, vol. 21, pp. 905–940, Dec 1982.

[3] H. Leff and A. F. Rex, *Maxwell's Demon 2 Entropy, Classical and Quantum Information, Computing.* CRC Press, 2002.

[4] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, "Thermodynamics of information," *Nature physics*, vol. 11, no. 2, p. 131, 2015.

[5] D. H. Wolpert, C. Kempes, P. F. Stadler, and J. A. Grochow, eds., *The energetics of computing in life and machines.* SFI Press, 2019.

[6] L. Szilard, "Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen," *Zeitschrift für Physik*, vol. 53, no. 11-12, pp. 840–856, 1929.

[7] T. Sagawa, "Thermodynamics of Information Processing in Small Systems," *Prog. Theor. Phys.*, vol. 127, pp. 1–56, 01 2012.

[8] T. Sagawa and M. Ueda, "Minimal energy cost for thermodynamic information processing: measurement and information erasure," *Phys. Rev. Lett.*, vol. 102, no. 25, p. 250602, 2009.

[9] T. Sagawa and M. Ueda, "Nonequilibrium thermodynamics of feedback control," *Phys. Rev. E*, vol. 85, no. 2, p. 021104, 2012.

[10] D. Mandal and C. Jarzynski, "Work and information processing in a solvable model of Maxwell's demon," *PNAS*, vol. 109, no. 29, pp. 11641–11645, 2012.

[11] J. M. Horowitz and J. M. R. Parrondo, "Designing optimal discrete-feedback thermodynamic engines," *New Journal of Physics*, vol. 13, p. 123019, dec 2011.

[12] S. W. Kim, T. Sagawa, S. De Liberato, and M. Ueda, "Quantum szilard engine," *Phys. Rev. Lett.*, vol. 106, no. 7, p. 070401, 2011.

[13] K.-H. Kim and S. W. Kim, "Information from time-forward and time-backward processes in szilard engines," *Phys. Rev. E*, vol. 84, no. 1, p. 012101, 2011.

[14] Z. Lu, D. Mandal, and C. Jarzynski, "Engineering Maxwell's demon," *Physics Today*, vol. 67, no. 8, pp. 60–61, 2014.

[15] A. Barato and U. Seifert, "Unifying three perspectives on information processing in stochastic thermodynamics," *Physical review letters*, vol. 112, no. 9, p. 090601, 2014.

[16] E. Lathouwers and J. Bechhoefer, "When memory pays: Discord in hidden markov models," *Phys. Rev. E*, vol. 95, no. 6, p. 062144, 2017.

[17] J. Bengtsson, M. N. Tengstrand, A. Wacker, P. Samuelsson, M. Ueda, H. Linke, and S. Reimann, "Quantum szilard engine with attractively interacting bosons," *Phys. Rev. Lett.*, vol. 120, no. 10, p. 100601, 2018.

[18] R. Chetrite, M. Rosinberg, T. Sagawa, and G. Tarjus, "Information thermodynamics for interacting stochastic systems without bipartite structure," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2019, no. 11, p. 114002, 2019.

[19] S. Still, "Thermodynamic cost and benefit of memory," *Physical Review Letters*, vol. 124, no. 5, p. 050601, 2020.

[20] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, "Experimental demonstration of information-to-energy conversion and validation of the generalized jarzynski equality," *Nature physics*, vol. 6, no. 12, pp. 988–992, 2010.

[21] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, "Experimental verification of landauer?s principle linking information and thermodynamics," *Nature*, vol. 483, no. 7388, pp. 187–189, 2012.

[22] É. Roldán, I. A. Martinez, J. M. R. Parrondo, and D. Petrov, "Universal features in the energetics of symmetry breaking," *Nature Physics*, vol. 10, no. 6, pp. 457–461, 2014.

[23] J. V. Koski, V. F. Maisi, T. Sagawa, and J. P. Pekola, "Experimental observation of the role of mutual information in the nonequilibrium dynamics of a maxwell demon," *Phys. Rev. Lett.*, vol. 113, p. 030601, Jul 2014.

[24] Y. Jun, M. Gavrilov, and J. Bechhoefer, "High-precision test of landauer's principle in a feedback trap," *Phys. Rev. Lett.*, vol. 113, p. 190601, Nov 2014.

[25] M. c. v. Gavrilov and J. Bechhoefer, "Erasure without work in an asymmetric double-well potential," *Phys. Rev. Lett.*, vol. 117, p. 200601, Nov 2016.

[26] M. Gavrilov, R. Chétrite, and J. Bechhoefer, "Direct measurement of weakly nonequilibrium system entropy is consistent with gibbs–shannon form," *PNAS*, vol. 114, no. 42, pp. 11097–11102, 2017.

[27] G. Paneru, D. Y. Lee, T. Tlusty, and H. K. Pak, "Lossless brownian information engine," *Phys. Rev. Lett.*, vol. 120, no. 2, p. 020601, 2018.

[28] M. Ribezzi-Crivellari and F. Ritort, "Large work extraction and the landauer limit in a continuous maxwell demon," *Nature Physics*, vol. 15, no. 7, pp. 660–664, 2019.

[29] S. Carnot, *Réflexions sur la puissance motrice du feu.* No. 26, Vrin, 1978.

[30] K. Sekimoto, *Stochastic energetics*, vol. 799. Springer, 2010.

[31] U. Seifert, "Stochastic thermodynamics, fluctuation theorems and molecular machines," *Reports on progress in physics*, vol. 75, no. 12, p. 126001, 2012.

[32] K. Sekimoto and S.-i. Sasa, "Complementarity relation for irreversible process derived from stochastic energetics," *Journal of the Physical Society of Japan*, vol. 66, no. 11, pp. 3326–3328, 1997.

[33] T. Schmiedl and U. Seifert, "Optimal finite-time processes in stochastic thermodynamics," *Physical review letters*, vol. 98, no. 10, p. 108301, 2007.

[34] H. Then and A. Engel, "Computing the optimal protocol for finite-time processes in stochastic thermodynamics," *Physical Review E*, vol. 77, no. 4, p. 041105, 2008.

[35] D. A. Sivak and G. E. Crooks, "Thermodynamic metrics and optimal paths," *Physical review letters*, vol. 108, no. 19, p. 190602, 2012.

[36] M. V. Bonança and S. Deffner, "Optimal driving of isothermal processes close to equilibrium," *The Journal of chemical physics*, vol. 140, no. 24, p. 244119, 2014.

[37] A. P. Solon and J. M. Horowitz, "Phase transition in protocols minimizing work fluctuations," *Physical review letters*, vol. 120, no. 18, p. 180605, 2018.

[38] J. Åberg, "Truly work-like work extraction via a single-shot analysis," *Nature communications*, vol. 4, no. 1, pp. 1–5, 2013.

[39] N. Y. Halpern, A. J. Garner, O. C. Dahlsten, and V. Vedral, "Introducing one-shot work into fluctuation relations," *New Journal of Physics*, vol. 17, no. 9, p. 095003, 2015.

[40] V. Cavina, A. Mari, and V. Giovannetti, "Optimal processes for probabilistic work extraction beyond the second law," *Scientific reports*, vol. 6, no. 1, pp. 1–13, 2016.

[41] O. Maillet, P. A. Erdman, V. Cavina, B. Bhandari, E. T. Mannila, J. T. Peltonen, A. Mari, F. Taddei, C. Jarzynski, V. Giovannetti, *et al.*, "Optimal probabilistic work extraction beyond the free energy difference with a single-electron device," *Physical review letters*, vol. 122, no. 15, p. 150604, 2019.

[42] J. M. Horowitz and J. M. Parrondo, "Thermodynamic reversibility in feedback processes," *EPL (Europhysics Letters)*, vol. 95, no. 1, p. 10005, 2011.

[43] J. M. Horowitz and J. M. Parrondo, "Designing optimal discrete-feedback thermodynamic engines," *New Journal of Physics*, vol. 13, no. 12, p. 123019, 2011.

[44] L. Granger, L. Dinis, J. M. Horowitz, and J. M. Parrondo, "Reversible feedback confinement," *EPL (Europhysics Letters)*, vol. 115, no. 5, p. 50007, 2016.

[45] É. Roldán and J. M. Parrondo, "Estimating dissipation from single stationary trajectories," *Physical review letters*, vol. 105, no. 15, p. 150607, 2010.

[46] É. Roldán and J. M. Parrondo, "Entropy production and kullback-leibler divergence between stationary trajectories of discrete systems," *Physical Review E*, vol. 85, no. 3, p. 031129, 2012.

[47] N. Merhav, "Sequence complexity and work extraction," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, no. 6, p. P06037, 2015.

[48] A. Frishman and P. Ronceray, "Learning force fields from stochastic trajectories," *Physical Review X*, vol. 10, no. 2, p. 021009, 2020.

[49] P. D. Grünwald and A. Grunwald, *The minimum description length principle*. MIT press, 2007.

[50] R. J. Cubero, M. Marsili, and Y. Roudi, "Minimum description length codes are critical," *Entropy*, vol. 20, no. 10, p. 755, 2018.

[51] R. J. Cubero, J. Jo, M. Marsili, Y. Roudi, and J. Song, "Statistical criticality arises in most informative representations," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2019, no. 6, p. 063402, 2019.

[52] D. Hartich, A. C. Barato, and U. Seifert, "Sensory capacity: An information theoretical measure of the performance of a sensor," *Physical Review E*, vol. 93, no. 2, p. 022116, 2016.

[53] T. Matsumoto and T. Sagawa, "Role of sufficient statistics in stochastic thermodynamics and its implication to sensory adaptation," *Physical Review E*, vol. 97, no. 4, p. 042103, 2018.

[54] J. Song, S. Still, R. D. H. Rojas, I. P. Castillo, and M. Marsili, "Optimal work extraction and mutual information in a generalized szilárd engine," *arXiv preprint arXiv:1910.04191*, 2019.

[55] T. Sagawa and M. Ueda, "Generalized jarzynski equality under nonequilibrium feedback control," *Phys. Rev. Lett.*, vol. 104, p. 090602, Mar 2010.

[56] J. M. Parrondo, "The szilard engine revisited: Entropy, macroscopic randomness, and symmetry breaking phase transitions," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 11, no. 3, pp. 725–733, 2001.

[57] M. Esposito and C. Van den Broeck, "Second law and landauer principle far from equilibrium," *EPL (Europhysics Letters)*, vol. 95, no. 4, p. 40004, 2011.

[58] A. Beretta, C. Battistin, C. De Mulatier, I. Mastromatteo, and M. Marsili, "The stochastic complexity of spin models: Are pairwise models really simple?," *Entropy*, vol. 20, no. 10, p. 739, 2018.

[59] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[60] J. J. Rissanen, "Fisher information and stochastic complexity," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 40–47, 1996.

[61] J. Rissanen, "Universal coding, information, prediction, and estimation," *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 629–636, 1984.

[62] B. O. Koopman, "On distributions admitting a sufficient statistic," *Transactions of the American Mathematical Society*, vol. 39, no. 3, pp. 399–409, 1936.

[63] T. Mora and W. Bialek, "Are biological systems poised at criticality?," *Journal of Statistical Physics*, vol. 144, no. 2, pp. 268–302, 2011.

[64] G. E. Crooks, "Measuring thermodynamic length," *Physical Review Letters*, vol. 99, no. 10, p. 100602, 2007.

[65] E. H. Feng and G. E. Crooks, "Far-from-equilibrium measurements of thermodynamic length," *Physical review E*, vol. 79, no. 1, p. 012104, 2009.

[66] A. Reiserer and G. Rempe, "Cavity-based quantum networks with single atoms and optical photons," *Reviews of Modern Physics*, vol. 87, no. 4, p. 1379, 2015.

[67] M. J. Hartmann, F. G. Brandao, and M. B. Plenio, "Strongly interacting polaritons in coupled arrays of cavities," *Nature Physics*, vol. 2, no. 12, pp. 849–855, 2006.