The Szilard engine revisited: Entropy, macroscopic randomness, and symmetry breaking phase transitions

Juan M. R. Parrondo[a)]
Departamento de Física Atómica, Molecular y Nuclear, Universidad Complutense de Madrid, 28040-Madrid, Spain

(Received 2 January 2001; accepted 29 May 2001; published 31 August 2001)

The role of symmetry breaking phase transitions in the Szilard engine is analyzed. It is shown that symmetry breaking is the only necessary ingredient for the engine to work. To support this idea, we show that the Ising model behaves exactly as the Szilard engine. We design a purely macroscopic Maxwell demon from an Ising model, demonstrating that a demon can operate with information about the macrostate of the system. We finally discuss some aspects of the definition of entropy and how thermodynamics should be modified to account for the variations of entropy in second-order phase transitions. © *2001 American Institute of Physics.* [DOI: 10.1063/1.1388006]

The Maxwell demon and the Szilard engine are *gedanken* experiments that are crucial to the search for a microscopic explanation of the second law of thermodynamics and to the elucidation of how entropy and information are related. Here we show that one of the key ingredients of the Szilard engine is a symmetry breaking phase transition. Following this idea, we design a purely macroscopic Maxwell demon from an Ising model, demonstrating that a demon can operate with information about the macrostate of the system, without violation of the Kelvin–Planck statement of the second law.

I. INTRODUCTION

The Szilard engine is one of the most relevant sequels of the well-known Maxwell demon.[1,2] Maxwell devised his demon to show the probabilistic nature of the second law of thermodynamics: a being capable of measuring the position and velocity of the molecules of a gas could in principle violate the second law. Operating a door in an adiabatic wall between two gases at different temperatures, the demon could induce a flow of energy from the cold to the hot gas. The conclusion is that information about the microscopic details of a system allows one to beat the second law.

The Szilard engine[1,2] exhibits the relevant features of the Maxwell demon, i.e., the trade-off between entropy and information, but its setup is simpler to analyze. The reason is that the information needed to operate the engine is very precise. The engine consists of a box with a single-particle gas, i.e., a particle that thermalizes in any collision with the walls. A piston can be introduced (or removed) either at the middle of the box or at two opposite walls (see Fig. 1).

The engine operates as follows. We insert the piston in the middle of the box and *measure* in which side the particle gets trapped. Then we let the gas expand reversibly and remove the piston. In the expansion the gas performs work:

$$W=\int_{V/2}^{V} P\,dV = kT\ln 2. \tag{1}$$

This work can be used, for instance, to lift a weight and store $kT\ln 2$ as potential energy. The energy is taken from the thermal bath, since the internal energy of the gas is constant. Therefore, the Szilard engine extracts energy from a single thermal bath and performs work, in contradiction with the second law of thermodynamics.

Notice that, for the engine to work properly, it is absolutely necessary to know in which side the particle gets trapped. In this way, we can exert a pressure on the piston equal and opposite to the pressure of the gas and let it expand quasistatically. On the other hand, if the direction of the pressure were not correct, the gas would expand irreversibly and Eq. (1) would not hold. As in the original Maxwell demon, the Szilard engine can beat the second law of thermodynamics only if some information about the state of the system is available.

The literature on the Szilard engine, as well as on the Maxwell demon, has focused mainly on the heat dissipation accompanying the measurement, i.e., the acquisition of information, and/or accompanying the erasure of this information.[1–5] As an exception, Magnasco presented in Ref. 6 an interesting analysis of the topology of the phase space of the engine.

Nevertheless, none of these papers has analyzed one of the obscure points of the Szilard engine, namely, that it consists of microscopic and macroscopic degrees of freedom interacting with each other. This mixture of micro (the particle) and macro (the piston) makes the Szilard engine a rather difficult and unclear problem for many physicists, even for those working on statistical mechanics.

In this paper I address this question, by giving a novel interpretation to one of the steps of the Szilard engine. The insertion of the piston in the middle of the box can be interpreted as a spontaneous symmetry breaking. The Hamiltonian of the particle is symmetric under the permutation of the two sides of the box. However, the particle gets trapped in only one of the sides. This is equivalent to what happens

a)Electronic mail: parr@seneca.fis.ucm.es

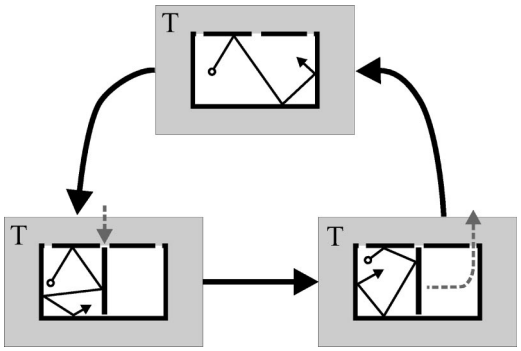

FIG. 1. The Szilard engine.

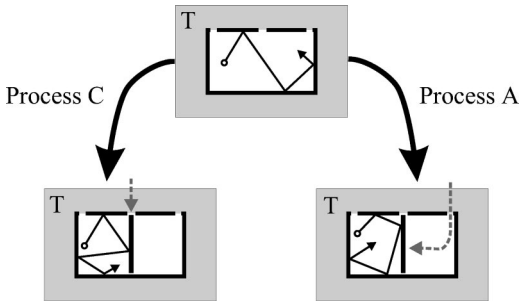

FIG. 2. Processes A and C in the Szilard engine.

in an Ising model when it is driven from a paramagnetic to a ferromagnetic phase in the absence of external magnetic field.

We will see in the following that all the astounding facts of the Szilard engine are reproduced in the Ising model and in any system exhibiting second-order phase transitions.

The benefit of this new interpretation is twofold. On the one hand, it helps us to understand better the Szilard engine and the relationship between entropy and information, since we will reach the same conclusions without the use of single-particle gases interacting with pistons. We will show, for instance, that a Szilard engine can be operated with information about the macrostate of the system: The necessary ingredient is information, but it is not relevant if this information is microscopic or macroscopic. On the other hand, our interpretation reveals that the consequences of the relationship between entropy and information and the intriguing aspects of the Szilard engine are not restricted to academic and artificial constructions, such as the Maxwell demon and the Szilard engine itself, but they are present in any spontaneous symmetry breaking, that is to say, almost everywhere in nature.

The paper is organized as follows. In Sec. II, the energetics of two processes in the Szilard engine are analyzed. Section III is a brief review of the concept of spontaneous symmetry breaking and the Ising model. In Sec. IV, two processes in the Ising model which are equivalent to the processes studied in Sec. II are introduced. Section V discusses the implications of the above-mentioned results on the definition of entropy and on the general validity of the Second Law. Finally, in Sec. VI, some conclusions and a list of open problems are presented.

II. TWO PROCESSES IN THE SZILARD ENGINE

Consider the Szilard gas and the processes A and C described in Fig. 2. In C, the piston is inserted in the middle of the box and the particle gets trapped in one of the sides. In A, the piston is introduced in the rightmost wall and moved slowly to the middle of the box. Then, C is the first step of the Szilard cycle and A is the inversion of the rest of the cycle (cf. Figs. 1 and 2).

Let us investigate the energetics of these two processes, i.e., the energy transfer between the particle and its surroundings. The particle exchanges energy with two external systems: the thermal bath, and some *external agent* that handles the piston, exerting pressure when it is needed. As in thermodynamics, I call *heat*, Q, the energy transferred from the thermal bath to the particle in a given process and *work*, W, the energy transferred from the system to the external agent. Finally, if $\mathcal{U}$ is the internal energy of the particle, the first law of thermodynamics

$$\Delta\mathcal{U}=Q-W \tag{2}$$

holds for any process.

In our particular case, process C does not require any work, or at least the work can be arbitrarily small. On the other hand, process A involves a compression of the single-particle gas to half of its volume and in this compression, if carried out quasistatically, a work $kT\ln 2$ is done by the external agent. Therefore, as defined previously, work in each process is given by

$$W_A=-kT\ln 2, \quad W_C=0. \tag{3}$$

The internal energy of the particle remains constant since the two processes are isothermal. Thus, the heat in each process is

$$Q_A=-kT\ln 2, \quad Q_C=0, \tag{4}$$

i.e., along A, energy is transferred from the system to the thermal bath.

The difference in the energetics of A and C is the key point of the Szilard engine. The engine is nothing but the cycle CA^{-1}, where A^{-1} is the inverse of process A. The energetics of A^{-1} is $W_{A^{-1}}=-W_A$ and $Q_{A^{-1}}=-Q_A$, if and only if A^{-1} is the *true* inversion of A, i.e., if the external agent exerts a pressure equal to the pressure of the gas and thus the expansion is done adiabatically. In this case, we have $W_{CA^{-1}}=kT\ln 2$. However, notice that, after process C, the system can end with the particle on any of the two sides of the box, whereas after A the particle is certainly on the left-hand side. Therefore, the cycle CA^{-1} cannot be implemented reversibly in the cases where the particle is on the right-hand side after C. In these cases, if the external agent insists in conducting process A^{-1} and consequently exerts a pressure to the right, then the piston will not move. Therefore, the Szilard engine consists of C followed by A^{-1} if the particle gets trapped on the left-hand side and followed by the mirror image of A^{-1} if the particle gets trapped on the right-hand side. A measurement is then necessary between C and A^{-1}.

Notice that so far the discussion has been restricted to energy. The consequences of the above-mentioned results on the definition of entropy will be explored in Sec. V.

III. SYMMETRY BREAKING TRANSITIONS

I have split the Szilard cycle into processes A and C, and showed that the paradoxical nature of the engine lies in the energetics of these two processes.

As mentioned in Sec. I, process C can be seen as a spontaneous symmetry breaking and process A as a forced or nonspontaneous symmetry breaking. In fact, symmetry breaking is the only necessary ingredient to reproduce all the relevant features of the Szilard engine.

Let us recall first what a spontaneous symmetry breaking is. If $\mathcal{H}(x)$ is the Hamiltonian of a system, x being a point in the phase space Γ, statistical mechanics prescribes that the probability density for the equilibrium state of the system at temperature T is given by the Gibbs distribution:

$$\rho_T(x) = \frac{e^{-\beta\mathcal{H}(x)}}{Z}, \tag{5}$$

where $\beta = 1/kT$, k is the Boltzmann constant, and $Z = \int_\Gamma e^{-\beta\mathcal{H}}$ is the partition function. From Eq. (5) we see that $\rho_T(x)$ has the same symmetries as $\mathcal{H}(x)$. Nevertheless, in some cases, a macroscopic system is not described by the Gibbs distribution. The phase space splits into n pieces, $\Gamma_1, \Gamma_2, \ldots, \Gamma_n \subset \Gamma$ and the macroscopic system is confined within one of them.[7] The distribution that describes the system is (see Appendix A for a discussion of the meaning of these distributions)

$$\rho_i(x) = \frac{e^{-\beta\mathcal{H}(x)}}{Z_i}\mathcal{X}_{\Gamma_i}(x), \tag{6}$$

where $\mathcal{X}_A(x)$ is the indicator function of the set $A \subset \Gamma$, i.e., $\mathcal{X}_A(x) = 1$ if $x \in A$ and $\mathcal{X}_A(x) = 0$ if $x \notin A$, and Z_i is the partition function restricted to Γ_i. The distributions $\rho_i(x)$, called *macroscopic phases*, have fewer symmetries than the Hamiltonian. The partition of the phase space, called *coexistence of macroscopic phases*, occurs for some values of the temperature and the parameters of the Hamiltonian. A *spontaneous symmetry breaking transition* occurs when the system is driven to a region of coexistence of phases along a process which does not favor any of the macroscopic phases. The phase is then chosen by thermal fluctuations. The selected phase can be interpreted as an amplification of microscopic fluctuations. One could say that it is a transfer of randomness from the microscopic to the macroscopic world resulting in an emergence of *macroscopic randomness*. If the system is driven to the region of coexistence of phases along a process which favors one of the phases, we say that the system undergoes a *nonspontaneous* or *forced symmetry breaking transition*. In this case, the chosen macroscopic phase depends on the past of the system.

The reader could immediately recognize process C as a spontaneous symmetry breaking transition and process A as a forced symmetry breaking.

The globally coupled Ising model is one of the simplest systems exhibiting coexistence of macroscopic phases.[7] Its Hamiltonian is

$$\mathcal{H}(\{s_i\};J,B) = -\frac{J}{N}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} s_i s_j - B\sum_{i=1}^{N} s_i, \tag{7}$$

where the spins take values $s_i = \pm 1$, with $i = 1,2,\ldots,N$. It depends on two parameters: the coupling J between spins and the external field B. It is called *globally coupled* because every spin interacts with all the others.

The system exhibits coexistence of two macroscopic phases when $B = 0$ and $J/kT > 1$. One of the phases is restricted to Γ_+, the set of configurations $\{s_i\}$ with positive global magnetization $M \equiv \Sigma_i s_i > 0$, and the other is restricted to Γ_-, the set of configurations with negative magnetization. Each phase breaks the symmetry $\{s_i\} \to \{-s_i\}$ that the Hamiltonian possesses for $B = 0$.

When temperature is lowered, keeping $B = 0$, from an initial value above the critical temperature $T_c \equiv J/k$, a second-order phase transition occurs at $T = T_c$. Below T_c the system is in one of the two macroscopic phases. None of the phases is favored along the process, since $B = 0$. Therefore, the system chooses the macroscopic phase at random or, more precisely, undergoes a spontaneous symmetry breaking.

The globally coupled Ising model also exhibits first-order phase transitions when the field crosses $B = 0$ below T_c. The external field breaks the symmetry $\{s_i\} \to \{-s_i\}$ of the Hamiltonian and, if for instance the coexistence region is reached decreasing a positive field, the macroscopic phase is the one with positive magnetization. This is a forced or nonspontaneous symmetry breaking.

To reproduce in the Ising model the two processes A and C discussed in Sec. II for the Szilard engine, we need to induce a spontaneous symmetry breaking at constant temperature (remember that processes A and C in the Szilard engine are isothermal). This can be achieved if we tune the coupling J at constant temperature T. The spontaneous symmetry breaking occurs then for a critical coupling $J_c \equiv 1/kT$, and for $B = 0$ and $J > J_c$ the system exhibits coexistence of phases. Notice that the Ising model is commonly used as a model for ferromagnetic materials, where the coupling cannot be tuned and the symmetry breaking is achieved by decreasing the temperature. On the other hand, here we need isothermal symmetry breaking transitions and then we are forced to modify the coupling J at constant temperature. This makes the system less realistic. However, we are not interested at this point in providing a physically realizable model of the Szilard engine, but only in showing the role of symmetry breaking transitions in the problem.

IV. TWO PROCESSES IN THE ISING MODEL

Consider the following two processes on the plane (J,B) (see Fig. 3).

Process A: Starting at (0,0), the field is increased up to $B_f > 0$, then the coupling is increased up to $J_f > J_c$, then the field is decreased down to zero.

Process C: starting at (0,0), the coupling is increased up to $J_f > J_c$, keeping $B = 0$.

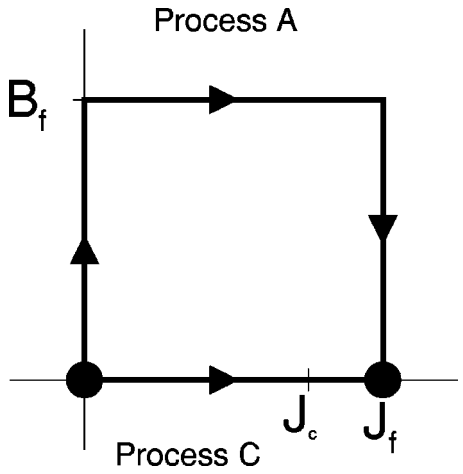

FIG. 3. Processes A and C in the Ising model. The two closed circles are the initial and final states of both processes.

The two processes are quasistatic in the following sense: They are slow enough for the system to relax within *each possible macroscopic phase*, but fast enough for the system to remain in one of the two phases (see Appendix A for a detailed explanation).

Applying to process A the formalism described in Appendix B, one finds the following energetics, up to order kT:

$$W_A = -\mathcal{F}(T,J_f,0) + \mathcal{F}(T,0,0) - kT\ln 2, \tag{8}$$

where $\mathcal{F}(T,J,B) = -kT\ln Z(\beta,J,B)$ and $Z(\beta;J,B) = \Sigma e^{-\beta\mathcal{H}}$ is the partition function of the system. $Z(\beta;J,B)$ and $\mathcal{F}(T,J,B)$ must be considered here as mere mathematical definitions and we should refrain from attributing any physical meaning to them at this stage of the discussion. For process C one has

$$W_C = -\mathcal{F}(T,J_f,0) + \mathcal{F}(T,0,0). \tag{9}$$

Therefore, $W_A - W_C = -kT\ln 2$, i.e., the external agent has to do more work to complete process A than to complete C, exactly as in the Szilard engine.

The whole discussion on the Szilard engine in Secs. I and II can be applied to the Ising model. For instance, one can design a cyclic engine as CA^{-1}.

Let us first analyze the inverse processes A^{-1} and C^{-1} in detail. The inversion of C does not present any difficulty. The energetics of C^{-1} is simply $W_{C^{-1}} = -W_C$ and $Q_{C^{-1}} = -Q_C$.

On the other hand, if we try to invert A, *the sign of the field must be the same as the sign of the initial magnetization of the system*. If we start to increase a positive field on a system with negative magnetization, the system becomes metastable, it runs along one of the branches of a hysteresis cycle and eventually relaxes irreversibly to the stable state for some value of the field B (see Fig. 4).

The most general case is when we have an ensemble of systems. If initially a fraction α of them have negative magnetization, the energetics of A^{-1} is given by

$$W_{A^{-1}} = -W_A - \alpha\frac{A_{\rm hys}}{2}, \tag{10}$$

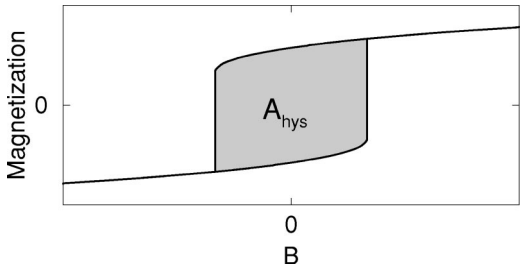

FIG. 4. Hysteresis cycle in the Ising model.

where $A_{\rm hys}$ is the area of the hysteresis cycle at $J = J_f$, as shown in Fig. 4.

The hysteresis phenomenon is not present in the Szilard engine. However, it has similar consequences to exerting the pressure in the wrong direction along the expansion, since in both cases the system evolves irreversibly doing less work.

Consider now the equivalent to the Szilard engine, i.e., the cycle CA^{-1} on an ensemble of Ising models. Its energetics (per system) is immediately obtained from Eqs. (8)–(10):

$$W_{CA^{-1}} = W_C + W_{A^{-1}} = kT\ln 2 - \alpha\frac{A_{\rm hys}}{2}, \tag{11}$$

where α is the fraction of systems with magnetization of the same sign as the field in A^{-1}. There are two consequences of this expression.

First, if instead of an ensemble we take a single system and measure its magnetization after C to decide the sign of the field, then $\alpha = 0$ and $W_{CA^{-1}} = kT\ln 2 > 0$, i.e., the system is extracting energy from the thermal bath and converting it into work. We recover the same result as in the Szilard engine but now with a genuine macroscopic system. Thus, we have a Maxwell demon with the important novelty that he needs to measure a *macroscopic quantity*.

Second, for an ensemble, $\alpha = 1/2$, and we still can beat the second law unless

$$A_{\rm hys} \geqslant 4kT\ln 2. \tag{12}$$

This inequality is a by-product of this theory and clarifying its origin is one of the open problems of the present work.

V. ENTROPY AND MACROSCOPIC UNCERTAINTY

The above discussion has focused on energy. In this Section the consequences of the previous results on the definition of entropy will be explored.

The change of entropy in the thermal bath along a process is given by $\Delta S_{\rm bath} = -Q/T$, whereas the entropy of the external agent is constant because its interaction with the system is purely mechanical. Then the change of the total entropy is

$$\Delta S_{\rm total} = -\frac{Q}{T} + \Delta S_{\rm syst}. \tag{13}$$

The second law of thermodynamics tells us that, if a process is reversible, $\Delta S_{\rm total} = 0$, and, if it is irreversible, $\Delta S_{\rm total} > 0$. In particular, for a cyclic process, $\Delta S_{\rm syst} = 0$ hence $Q \leqslant 0$.

This is the Kelvin–Planck statement of the second law: *it is not possible to extract energy from a single thermal bath in a cyclic process*.

However, Eq. (13) and the second law lead to contradictions when applied to processes A and C. Let us recall first some properties of the heat transferred from the thermal bath to the system in each process, as calculated in Sec. IV and in Appendix B:

$$Q_C = -Q_{C^{-1}}, \quad Q_A = -Q_{A^{-1}}, \qquad Q_C - Q_A = kT\ln 2. \tag{14}$$

In any cycle $\Delta S_{\text{syst}} = 0$. Then, using (13) and (14), $\Delta S_{\text{total}}^{CC^{-1}} = \Delta S_{\text{total}}^{AA^{-1}} = 0$. Therefore, AA^{-1} and CC^{-1} are reversible and so are their components, A, A^{-1}, C, and C^{-1}. On the other hand $\Delta S_{\text{total}}^{AC^{-1}} = k\ln 2$, hence AC^{-1} is irreversible. Notice that no measurement is necessary in any of the previous cycles.

Moreover, if A and C are reversible, then $\Delta S_{\text{total}}^{A} = \Delta S_{\text{total}}^{C} = 0$, and from (13) and (14), we obtain $\Delta S_{\text{syst}}^{C} = \Delta S_{\text{syst}}^{A} + k\ln 2$. On the other hand, whenever the system ends with positive magnetization after C, the initial and final states of both processes A and C are the same from a physical point of view.

These contradictions are usually explained with the following definition for the thermodynamic entropy of the system:

$$S_{\text{syst}}^{(\text{ens})} = -k\langle \ln \rho_{\text{ens}} \rangle, \tag{15}$$

where ρ_{ens} is the probability distribution describing an ensemble of systems. After process C, $\rho_{\text{ens}} = (\rho_+ + \rho_-)/2$ where ρ_+ and ρ_- are the probability distribution of the two macroscopic phases (see Sec. III). On the other hand, after A, $\rho_{\text{ens}} = \rho_+$. Then, $S_{\text{syst}}^{(\text{ens})}$ is $k\ln 2$ bigger after C than after A.

This picture is, however, rather unsatisfactory if we deal with single systems instead with ensembles, since ρ_{ens} becomes a subjective quantity. For instance, the physical state of an Ising model after process A is the same as after C if the final magnetization is positive. The only difference between these two situations is that we ignore the magnetization after C. Then $S_{\text{syst}}^{(\text{ens})}$, as defined in Eq. (15), is a subjective quantity for single systems. Mathematically, this can be expressed as

$$S_{\text{syst}}^{(\text{ens})} = -k\langle \ln \rho_{\text{single}} \rangle + kH. \tag{16}$$

Here, ρ_{single} is the invariant measure that gives the temporal average of any magnitude and it is a fully objective distribution for a single system (see Appendix A). H is the ignorance or uncertainty that we have about the macrostate of the system. It is evaluated (in a unit called *nats*) using Shannon formula:[8] $H = -\Sigma_i\, p_i \ln p_i$, where p_i is the probability of having an instance i (in the Szilard and Ising case, after C, $H = \ln 2$).

Moreover, in this interpretation not only entropy is subjective but also the concept of reversibility. Consider C^{-1} on a single system: It is reversible if we do not know the initial macroscopic magnetization and it is irreversible if we do know it. This was already pointed out by Bennett in Ref. 4.

A few words are in order about the objectivity of the invariant measure ρ_{single}. For an ergodic system, this invariant measure is the fraction of time the system spends in a given region of the phase space. Therefore, it is an objective distribution for a single system and does not depend on the information at our disposal. In particular, it does not change under a measurement. However, the invariant measure has some well-known limitations: it does not describe the instantaneous state of the system and it can be considered as a description of the system only for periods of time long enough to ensure the validity of the ergodic property. In our present discussion, we are dealing with equilibrium systems or systems undergoing quasistatic processes. In both cases, the ergodic theorem holds and we can consider ρ_{single} as a fully objective description of a single system at any stage of these processes (see Appendix A for further details).

Here a simpler interpretation of the above-mentioned results is proposed, using the invariant measure ρ_{single}. In this new interpretation, entropy is an objective magnitude for single systems, but we are forced to admit that it decreases along certain processes, in contradiction with *some* formulations of the second law. However, the main limitation imposed by the second law, namely, the Kelvin–Planck statement, remains valid, since these processes cannot be used to construct cycles. Ishioka and Fuchikami, in Ref. 9, have reached similar conclusions. The assumptions for this interpretation are the following.

(1) The thermodynamic entropy of a system is given by

$$S_{\text{syst}} \equiv -k\langle \ln \rho_{\text{single}} \rangle. \tag{17}$$

(2) If an external agent induces, in a quasistatic and isothermal way, a spontaneous symmetry breaking with n phases, the total entropy (the sum of the entropies of the system, thermal bath, and external agent) decreases by $k\ln n$. These processes will be called *anti-irreversible* (in Ref. 9 the term *partitioning processes* is used instead) and they correspond to the *creation of macroscopic randomness*.

(3) Along the inverse of an anti-irreversible process, the total entropy increases by $k\ln n$. These processes will be called *quasi-irreversible* or simply irreversible.

Process C is anti-irreversible and C^{-1} is quasi-irreversible. The reason for the names is the following: C^{-1} cannot be truly reversed because, after $C^{-1}C$, the initial magnetization could be opposite to the final one due to the emergence of macroscopic randomness along C. Processes A and A^{-1} are reversible in the standard sense, i.e., total entropy does not change. The reader can check that every combination of processes A, C, and their inversions are explained with the three rules previously mentioned. Moreover, entropy and revers-

ibility become fully objective concepts. Notice also that the proposed modification only affects entropy by a quantity of order k, which vanishes in the thermodynamic limit. Nevertheless, the order of magnitude of the energy involved in the Maxwell demon or in any of its variants is kT, and k for the entropy (see Sec. VI for a comment on this point).

The measurement process can also be explained with this new thermodynamics. Consider, as a model of a system and a measurement device, the Hamiltonian:

$$\mathcal{H}(\{s_i^{(1)}\},\{s_i^{(2)}\};J_1,J_2,J_{12}) = -\frac{J_1}{N}\sum_{j>i}^{N} s_i^{(1)}s_j^{(1)} - \frac{J_2}{N}\sum_{j>i}^{N} s_i^{(2)}s_j^{(2)} - \frac{J_{12}}{N}\sum_{i,j=1}^{N} s_i^{(1)}s_j^{(2)},$$

which corresponds to two coupled Ising models, 1 (system) and 2 (measurement device or ''pointer''). The following table shows the behavior of the total entropy S_{total}, as defined by (13) and (17), and the macroscopic uncertainty H (in nats), along two isothermal and quasistatic processes. In the table, S_{total}^0 is the total entropy in the initial state:

	Step	$S_{\text{total}}-S_{\text{total}}^0$	H	Step	$S_{\text{total}}-S_{\text{total}}^0$	H
1)	$J_1:0\to J_f$	$-k\ln 2$	$\ln 2$	$J_1:0\to J_f$	$-k\ln 2$	$\ln 2$
2)	$J_{12}:0\to J_f$	$-k\ln 2$	$\ln 2$	$J_{12}:0\to J_f$	$-k\ln 2$	$\ln 2$
3)	$J_2:0\to J_f$	$-k\ln 2$	$\ln 2$	$J_2:0\to J_f$	$-k\ln 2$	$\ln 2$
4)	$J_1:J_f\to 0$	$-k\ln 2$	$\ln 2$	$J_{12}:J_f\to 0$	$-k\ln 2$	$\ln 2$
5)	$J_{12}:J_f\to 0$	$-k\ln 2$	$\ln 2$	$J_1:J_f\to 0$	0	$\ln 2$
6)	$J_2:J_f\to 0$	0	0	$J_2:J_f\to 0$	$k\ln 2$	0

Both processes involve a spontaneous symmetry breaking (step 1), copying the outcome (steps 2 and 3), and erase the copy and the original (steps 4–6).

The first process (left-hand column) can be interpreted as a reversible measurement. Measurement can be defined in a rather general way as any procedure which allows one to drive a system from the region of coexistence of phases to a region of noncoexistence along a reversible process, i.e., avoiding the critical point as well as the possibility of hysteresis. This is done in step 4 of the first column, where J_1 is decreased down to zero along a reversible process. As a result, the total entropy is lowered by $k\ln 2$ in the first five steps. Notice also that, to drive the whole system $1+2$ to its initial state, we have to *reset* the *measurement device* 2, by crossing again a critical point, i.e., along a quasi-irreversible process (step 6). We thus recover Bennett's interpretation of the Szilard engine.[4]

I have included the other process (the right-hand column in the table) to show how subtle the measurement and the erasure processes can be. If subsystem 1 is uncoupled before driven to its initial state, then it crosses a critical point and the entropy increases. Step 5 in the right-hand column is quasi-irreversible, because initially the magnetizations of 1 and 2 have the same sign, and, if step 5 were reversed, the final magnetizations would be uncorrelated. A similar effect of the correlation between the particle and the measurement device in the Szilard engine was pointed out by Fahn in Ref. 5.

VI. CONCLUSIONS AND OPEN PROBLEMS

It has been shown here that spontaneous symmetry breaking is the key ingredient in the Szilard engine, and a similar engine can be devised with any system exhibiting second-order phase transitions. As an example the Ising model has been used. The example has revealed that in the trade-off between entropy and information, the latter does not need to be information about the microscopic state of a system.

I have proposed a modification of thermodynamics to achieve a fully objective definition of entropy. Two objections can be raised against this thermodynamics. The first one is that energy is an extensive property, i.e., of order NkT, and terms of order $kT\ln 2$ are negligible and even much smaller than the energy fluctuations. This objection applies to any Maxwell demon but it is not sufficient to exorcize it. The reason is that the demon can repeat the cycle as many times as he wants, converting a macroscopic amount of heat into work.

The second objection is that the increase of entropy can be derived from nonequilibrium theories, such as the Fokker–Planck formalism. If $\mathbf{q}$ are the (overdamped) degrees of freedom of a system, the probability distribution obeys the Fokker–Planck equation (FPE):

$$\partial_t \rho(\mathbf{q},t) = -\boldsymbol{\nabla}\cdot\mathbf{J}(\mathbf{q},t), \tag{18}$$

where the current is $\mathbf{J}(\mathbf{q},t)=[-\boldsymbol{\nabla}\mu(\mathbf{q},t)]\rho(\mathbf{q},t)$ and the chemical potential is defined as $\mu(\mathbf{q},t)\equiv V(\mathbf{q},t)+kT\ln\rho(\mathbf{q},t)$. From these equations one can derive the following identity:[6,10]

$$\begin{aligned} -k\partial_t\int d\mathbf{q}\,\rho(\mathbf{q},t)\ln\rho(\mathbf{q},t) &= \frac{1}{T}\int d\mathbf{q}\,V(\mathbf{q},t)\partial_t\rho(\mathbf{q},t) + \frac{1}{T}\int d\mathbf{q}\,\frac{|\mathbf{J}(\mathbf{q},t)|^2}{\rho(\mathbf{q},t)} \\ &= \frac{\dot{Q}}{T} + \frac{1}{T}\int d\mathbf{q}\,\frac{|\mathbf{J}(\mathbf{q},t)|^2}{\rho(\mathbf{q},t)}. \end{aligned} \tag{19}$$

If the left-hand side of Eq. (19) is interpreted as $\dot{S}_{\text{syst}}$, the change of the entropy of the system per unit of time, then the total change of entropy, $\dot{S}_{\text{total}}=-\dot{Q}/T+\dot{S}_{\text{syst}}$, is always positive. A similar result can be obtained for underdamped degrees of freedom.[11] How then have we obtained $\dot{S}_{\text{total}}<0$ for some processes involving phase transitions? The answer is that the distribution that appears in the FPE (18) is ρ_{ens} and not ρ_{single}. Then, the FPE is not appropriate to describe single macroscopic systems in the region of coexistence of phases.

One of the open problems of the present work is to characterize ρ_{single} and derive an evolution equation similar to the FPE. Other open problems are: (a) analyzing the role of hysteresis and the origin of inequality (12); (b) extending the above-given discussion to the breaking of a continuous symmetry, where an infinite number of macroscopic phases coexist; (c) including the external agent in the Hamiltonian as a set of macroscopic degrees of freedom; and (d) exploring the

implications of the decrease of entropy along anti-irreversible processes, especially in cosmology.

ACKNOWLEDGMENTS

This work has been supported by the Dirección General de Enseñanza Superior (DGES, Spain), Grant No. PB-97-0076 and by a Grant from the *New Del Amo Program*.

APPENDIX A: PROBABILITY DISTRIBUTIONS AND MACROSCOPIC STATES

In this appendix, the meaning of the probability distributions ρ_{ens} and ρ_{single} used in the text is explained in some detail.

Suppose a classical system with Hamiltonian $\mathcal{H}(x;\alpha)$ and a process where the parameter α is changed from α_0 to $\alpha_0+\Delta\alpha$, at constant velocity, during a time interval $[0,t_f]$. Then $\alpha(t)=\alpha_0+t\Delta\alpha/t_f$ is the value of the parameter at time $t\in[0,t_f]$. If $x(t)$ is the trajectory of the microstate of the system, then the energy transfer between the system and the external agent which modifies α, i.e., the work done by the system along the process, is

$$\delta W=-\int_0^{t_f}dt\,\dot{\alpha}(t)\frac{\partial\mathcal{H}(x(t);\alpha)}{\partial\alpha}\bigg|_{\alpha=\alpha(t)}$$
$$=-\frac{\Delta\alpha}{t_f}\int_0^{t_f}dt\,\frac{\partial\mathcal{H}(x(t);\alpha)}{\partial\alpha}\bigg|_{\alpha=\alpha_0}+o(\Delta\alpha^2). \quad \text{(A1)}$$

The first term in δW is $\Delta\alpha$ times a time average of $\partial\mathcal{H}/\partial\alpha$. If t_f is large enough to apply the ergodic theorem to this time average,[12] we obtain

$$\delta W=-\left\langle\frac{\partial\mathcal{H}(x;\alpha)}{\partial\alpha}\bigg|_{\alpha=\alpha_0}\right\rangle\Delta\alpha+o(\Delta\alpha^2), \quad \text{(A2)}$$

where the average is taken over $\rho_{\text{single}}(x)$, the invariant measure on the subregion of Γ where the system is ergodic.[12] In the text, we distinguish this distribution from the distribution $\rho_{\text{ens}}(x)$ followed by an ensamble of systems *a la* Gibbs.[7,13] Both coincide except in the region of coexistence of phases, i.e., when the system is no longer ergodic in the whole phase space Γ, but only in a certain region Γ_i. This distinction is crucial for the arguments presented in the paper.

It is remarkable that the ergodic theorem, up to the best of the author's knowledge, has never been applied to processes. Since Boltzmann, the ergodic theorem has been invoked to prove irreversibility, either as the relaxation of a system to equilibrium or as the increase of entropy (see Refs. 12, 13, and references therein). However, the second law, at least in the Kelvin–Planck statement, is about processes and their energetics and, as we have seen here, the ergodic theorem arises in a very natural way when dealing with slow processes.

APPENDIX B: ENERGETICS OF PROCESSES *A* AND *C*

Consider a system whose Hamiltonian $\mathcal{H}(x;\mathbf{R})$ depends on a set of external parameters collected in a vector $\mathbf{R}$. We are interested in the energetics of a process along which the system is in contact with a thermal bath at temperature T and the parameters are changed by an external agent as $\mathbf{R}(t)$ with $t\in[0,\mathcal{T}]$.

The expressions for work and heat per unit of time along this process are[11,14]

$$\dot{Q}=\int_\Gamma dx\,\mathcal{H}(x;\mathbf{R}(t))\frac{\partial\rho(x;t)}{\partial t},$$
$$\dot{W}=-\int_\Gamma dx\,\rho(x;t)\frac{\partial\mathcal{H}(x;\mathbf{R}(t))}{\partial t}, \quad \text{(B1)}$$

where Γ is the phase space of the system and $\rho(x;t)$ the probability density for the state x. Heat and work, as given by Eq. (B1), obey the first law of thermodynamics: $\dot{\mathcal{U}}=\dot{Q}-\dot{W}$.

If the process is quasistatic, $\mathcal{T}\to\infty$, the probability density at time t depends only on the value of the external parameters at t, i.e., $\rho(x;t)=\rho(x;\mathbf{R}(t))$. In this case, the heat and the work in the whole process are given by

$$Q=\int_A\delta Q(\mathbf{R}),\quad W=\int_A\delta W(\mathbf{R}), \quad \text{(B2)}$$

where A is the path that $\mathbf{R}(t)$ describes along the process and the infinitesimal work and heat are given by

$$\delta Q(\mathbf{R})=\int_\Gamma dx\,\mathcal{H}(x;\mathbf{R})\frac{\partial\rho(x;\mathbf{R})}{\partial\mathbf{R}}\cdot d\mathbf{R},$$
$$\delta W(\mathbf{R})=-\int_\Gamma dx\,\rho(x;\mathbf{R})\frac{\partial\mathcal{H}(x;\mathbf{R})}{\partial\mathbf{R}}\cdot d\mathbf{R}. \quad \text{(B3)}$$

Notice that the expression for the work coincides with Eq. (A2) if the probability distribution is ρ_{single}. This is the distribution that we will take in the following.

The most familiar implementation of the above-mentioned expressions is obtained when the state of the system is the Gibbs distribution, $\rho_T(x;\mathbf{R})\equiv e^{-\beta\mathcal{H}(x;\mathbf{R})}/Z(\beta,\mathbf{R})$. For this particular case, Eq. (B3) reduces to

$$\delta Q(\mathbf{R})=T\frac{\partial S(T,\mathbf{R})}{\partial\mathbf{R}}\cdot d\mathbf{R},\quad \delta W(\mathbf{R})=-\frac{\partial\mathcal{F}(T,\mathbf{R})}{\partial\mathbf{R}}\cdot d\mathbf{R}, \quad \text{(B4)}$$

where

$$S(T,\mathbf{R})=-k\int_\Gamma dx\,\rho_T(x;\mathbf{R})\ln[\rho_T(x;\mathbf{R})]$$
$$\mathcal{F}(T,\mathbf{R})=-kT\ln Z(\beta,\mathbf{R}) \quad \text{(B5)}$$

are, respectively, the free energy and the entropy of the system.

For isothermal processes $\delta W(\mathbf{R})$ is an exact differential and therefore the integral in (B2) reduces to

$$W=-\mathcal{F}(T,\mathbf{R}(\mathcal{T}))+\mathcal{F}(T,\mathbf{R}(0)), \quad \text{(B6)}$$

i.e., the difference between the initial and the final free energy.

Although processes A and C considered in the text are isothermal and quasistatic, the state $\rho(x;\mathbf{R})$ is not equal to $\rho_T(x;\mathbf{R})$ in the region of coexistence of macroscopic phases,

due to (6). Consequently their energetics, up to terms of order kT, differ from the one prescribed by standard equilibrium thermodynamics.

Let us consider the energetics of the process A introduced in Sec. IV. We will split the process in three steps: along step 1 the field is increased from zero to B_f keeping $J=0$; along step 2, the coupling is increased from zero to J_f keeping $B=B_f$; along step 3, the field is decreased from B_f to zero keeping $J=J_f$. We will apply (B2) and (B3) with the following choice for the state $\rho(\{s_i\};J,B)$ along the three steps:

$$\rho(\{s_i\};J,B)=\begin{cases}\rho_T(\{s_i\};J,B) & \text{if } J=0 \text{ or } B=B_f \\ & \text{(steps 1 and 2)} \\ \rho_+(\{s_i\};J,B) & \text{if } J=J_f \\ & \text{(step 3)}\end{cases}. \tag{B7}$$

This choice implies that the system is in the phase of positive magnetization during the third step. The energetics, up to order kT, does not depend on where precisely the system changes form ρ_T to ρ_+. The above-given prescription has been chosen for simplicity. The replacement of ρ_T by ρ_+ is only significant at the end of step 3, i.e., when $J=J_f$ and $B\simeq 0$ and the system is close to the region of coexistence of macroscopic phases. In the rest of the plane (J,B) the energetics is the same whether one uses ρ_T or ρ_+.

Along steps 1 and 2, Eq. (B6) can be applied, since the state is ρ_T:

$$W^{(12)}=-\mathcal{F}(T,J_f,B_f)+\mathcal{F}(T,0,0). \tag{B8}$$

To evaluate the work performed along the third step it is convenient to define the partition function restricted to configurations with positive magnetization:

$$Z_+(\beta,J,B)=\sum_{\{s_i\}}\Theta(\Sigma_j s_j)e^{-\beta\mathcal{H}(\{s_i\};J,B)}, \tag{B9}$$

and the corresponding free energy:

$$\mathcal{F}_+(T,J,B)=-kT\ln Z_+(\beta,J,B). \tag{B10}$$

The work performed along the third step can be evaluated in a similar way as for the Gibbs state:

$$W^{(3)}=-\mathcal{F}_+(T,J_f,0)+\mathcal{F}_+(T,J_f,B_f). \tag{B11}$$

The total amount of work is

$$\begin{aligned}W_A&=W^{(12)}+W^{(3)}\\&=-\mathcal{F}_+(T,J_f,0)+\mathcal{F}(T,0,0)+o(e^{-CN}),\end{aligned}$$

where we have used the fact that $\mathcal{F}_+(T,J_f,B_f)-\mathcal{F}(T,J_f,B_f)$ is of order e^{-CN}, C being a positive number, for sufficiently large B_f (see Appendix C). On the other hand, for $B=0$ the restricted partition function verifies[15]

$$Z_+(\beta,J,0)=\frac{Z(\beta,J,0)}{2} \tag{B12}$$

and then

$$\mathcal{F}_+(T,J_f,0)=\mathcal{F}(T,J_f,0)+kT\ln 2. \tag{B13}$$

Therefore, the work along process A becomes:

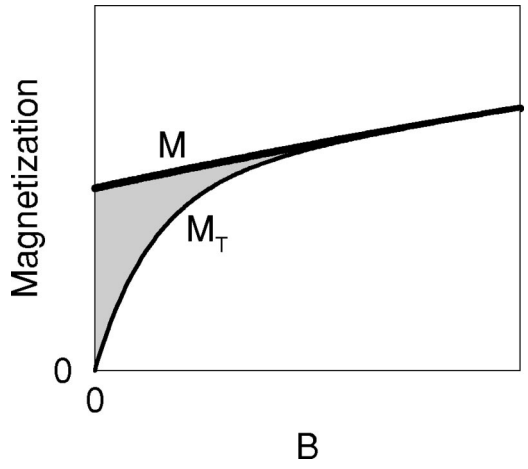

FIG. 5. Magnetization M as a function of the field B. The narrower line is the result of averaging over the state ρ_T. The gray area is the difference between the work along step 3 using ρ_T and ρ_+.

$$W_A=-\mathcal{F}(T,J_f,0)+\mathcal{F}(T,0,0)-kT\ln 2, \tag{B14}$$

which is Eq. (8) in Sec. IV.

Notice that the work in Eq. (B14) is the one given by equilibrium thermodynamics, Eq. (B6), *minus* an extra term $kT\ln 2$. This term is the novelty of this calculation and the key point of the analysis along the paper. It will be instructive to see in more detail how it arises.

Along the third step, J is kept constant. Therefore, the work is given by

$$\begin{aligned}W^{(3)}&=-\int_{B_f}^{0}\left\langle\frac{\partial\mathcal{H}(\{s_i\};J,B)}{\partial B}\right\rangle dB\\&=-\int_{B_f}^{0}dB\,M(T,J,B)dB,\end{aligned} \tag{B15}$$

where $\langle\cdot\rangle$ is the average over $\rho(\{s_i\};J,B)$ and $M(T,J,B)$ is the magnetization of the system:

$$M(T,J,B)\equiv\left\langle\sum_{i=1}^{N}s_i\right\rangle. \tag{B16}$$

Observe that the magnetization is different if one uses ρ_T instead of ρ_+ In Fig. 5, these two values of the magnetization are plotted against the field B. The magnetization for ρ_T vanishes for $B=0$, since $\rho_T(\{s_i\},J,0)$ is a symmetric state. On the other hand, $\rho_+(\{s_i\},J,0)$ is nonsymmetric and the corresponding magnetization at zero field is positive. As a consequence, the work calculated using ρ_T differs from the one calculated with ρ_+. The former is bigger than the latter and the difference is equal to the gray area in Fig. 5. This area, as has been shown previously, is $kT\ln 2$.

Let us turn to the energetics of process C. In this case, the system crosses the critical point and chooses between one of the two possible macroscopic phases, ρ_+ or ρ_-. Assuming that the system chooses ρ_+, the state of the system along the process is given by

$$\rho(\{s_i\};J,0)=\begin{cases}\rho_T(\{s_i\};J,B) & \text{if } J<J_c \\ \rho_+(\{s_i\};J,B) & \text{if } J\geqslant J_c\end{cases}. \tag{B17}$$

As in the calculation for process A, the precise location of the replacement of ρ_T by ρ_+ does not affect the final conclusions. This replacement can be done even smoothly with the same results. In fact, the energetics is the same as if calculated using ρ_T, for symmetry reasons.

Along the process, the field is constant, $B=0$. Therefore,

$$W_C = -\int_0^{J_f} \left\langle \frac{\partial \mathcal{H}(\{s_i\};J,0)}{\partial J} \right\rangle dJ. \tag{B18}$$

The partial derivative is clearly symmetric under the transformations $s_i \to -s_i$. Consequently, the average is the same for ρ_T and ρ_+ and the work W_C is equal to the one given by the Gibbs state [cf. Eq. (B6)], i.e.,

$$W_C = -\mathcal{F}(T,J_f,0) + \mathcal{F}(T,0,0), \tag{B19}$$

which is Eq. (9) in Sec. IV.

APPENDIX C: BOUNDS FOR THE DIFFERENCE BETWEEN FREE ENERGIES

To complete the derivation of Eq. (B14), we have to prove that, for strong enough fields,

$$\ln \frac{Z(\beta,J,B)}{Z_+(\beta,J,B)} = \ln\left(1 + \frac{Z_-(\beta,J,B)}{Z_+(\beta,J,B)}\right) = o(e^{-CN}), \tag{C1}$$

where C is a positive constant.

We present here a rough lower bound for the quotient of partition functions, yet sufficient for our purposes. The partition functions can be written in the form:

$$Z_\pm(\beta,J,B) = \sum_{S=-N}^{N} \binom{N}{(S+N)/2} \exp\left[\beta\left(\frac{J(S^2-N)}{2N} + BS\right)\right] \times \Theta(\pm S), \tag{C2}$$

where $S=\Sigma_i s_i$. If $B>J/2$, the argument of the exponential is negative for all $S\leqslant 0$, then:

$$Z_-(\beta,J,B) < 2^{N-1}. \tag{C3}$$

If $C \equiv \beta(B+J/2) - \ln 2 > 0$, it is sufficient to bound Z_+ by the term with $S=N$:

$$Z_+(\beta,J,B) > \exp[\beta N(B+J/2-J/(2N))]. \tag{C4}$$

We finally have

$$\frac{Z_-(\beta,J,B)}{Z_+(\beta,J,B)} < e^{\beta J/2}\exp[-\beta N(B+J/2-\ln 2/\beta)] < e^{\beta J/2}e^{-CN} \tag{C5}$$

with $C>0$. The above boundary is very rough, since there are other terms in Z_+ exponentially larger than the one with $S=N$.

Summarizing, if B_f and J_f in the cycle of section IV verify $B_f > J_f/2 > J_c/2$ and $C \equiv \beta(B_f+J_c/2) - \ln 2 = \beta B_f + 1/2 - \ln 2 > 0$, then

$$\mathcal{F}_+(T,J_f,B_f) - \mathcal{F}(T,J_f,B_f) = kT \ln \frac{Z(\beta,J_f,B_f)}{Z_+(\beta,J_f,B_f)} = o(kTe^{-CN}). \tag{C6}$$

[1] H. S. Leff and A. F. Rex, *Maxwell's Demon: Entropy, Information, Computing* (Adam Hilger, Bristol, 1990).
[2] L. Szilard, Z. Phys. **53**, 840–856 (1929). English translation reprinted in Ref. 1, p. 124.
[3] R. Landauer, IBM J. Res. Dev. **5**, 183–191 (1961). Reprinted in Ref. 1, p. 188.
[4] C. H. Bennett, Int. J. Theor. Phys. **21**, 905–940 (1982). Reprinted in Ref. 1, p. 213.
[5] P. N. Fahn, Found. Phys. **26**, 71–93 (1996).
[6] M. O. Magnasco, Europhys. Lett. **33**, 583–588 (1996).
[7] K. Huang, *Statistical Mechanics* (Wiley, New York, 1987).
[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[9] S. Ishioka and N. Fuchikami, in *Unsolved Problems of Noise and Fluctuations*, Adelaide, Australia, 1999, AIP Conf. Proc. **511**, edited by D. Abbott and L. B. Kish (AIP, New York, 2000), pp. 329–340.
[10] J. M. R. Parrondo, B. Jiménez, and R. Brito, in *Stochastic Processes in Physics, Chemistry, and Biology*, edited by J. A. Freund and T. Pöschel (Springer, Berlin, 2000), pp. 38–49.
[11] K. Shizume, Phys. Rev. E **52**, 3495–3499 (1995).
[12] I. E. Farquhar, *Ergodic Theory in Statistical Mechanics* (Wiley, New York, 1964).
[13] J. Lebowitz, Phys. Today **46** (9), 32–38 (1993).
[14] K. Denbigh, Chem. Br. **17**, 168–185 (1981). Reprinted in Ref. 1, p. 109.
[15] N. Fuchikami, H. Iwata, and S. Ishioka, J. Phys. Soc. Jpn. **68**, 3751–3754 (1999).