Information erasure

Barbara Piechocinska*
Los Alamos National Laboratory, T-6, Los Alamos, New Mexico 87544

(Received 27 September 1999; published 17 May 2000)

Landauer's principle states that in erasing one bit of information, on average, at least $k_B T \ln(2)$ energy is dissipated into the environment (where k_B is Boltzmann's constant and T is the temperature of the environment at which one erases). Here, Landauer's principle is microscopically derived without direct reference to the second law of thermodynamics. This is done for a classical system with continuous space and time, with discrete space and time, and for a quantum system. The assumption made in all three cases is that during erasure the bit is in contact with a thermal reservoir.

PACS number(s): 03.67.−a, 89.70.+c

I. INTRODUCTION

The purpose of this paper is to show that Landauer's principle [1] [which states that in erasing one bit (binary digit) of information one dissipates, on average, at least $k_B T \ln(2)$ of energy into the environment, where k_B is Boltzmann's constant and T is the temperature at which one erases] can be derived from microscopic considerations. The intention is to show this for classical and quantum systems without direct reference to the second law of thermodynamics. This, however, does not imply that we will be proving the second law of thermodynamics. Since the beginning of this century it has been known that it is possible to derive many of the inequalities in thermodynamics from the canonical distribution [2].

The introductory part of the paper will try to give the reader a clear picture of what we mean by erasure, what Landauer's principle is about, and what has been done in this field.

In Sec. II of this paper we will attempt to derive Landauer's principle from microscopic considerations for three separate cases: the classical continuous case, the classical discrete case, and the quantum case. In Sec. III we will discuss nondegenerate energy levels for the states of the bits and probability distributions of the states of bits in ensembles of bits.

A. Basic setup

The bit is a fundamental unit of information, the smallest item capable of indicating a choice. We will assume that all information is physically representable, and therefore all bits representing it are encoded in the states of physical systems [3]. The bits will have two distinguishable states that we will call ''zero'' and ''one.'' They will be in contact with an environment that we will be modeling as a heat reservoir at a fixed temperature *T*. There will also be an external parameter that will let us do work on the bit. This external parameter will be the means with which we will erase the bit. In all cases we will assume to have a large number of bits but they will be erased individually, one by one. We will view the large number of bits as an ensemble. To have a clearer picture of the entire process one could for instance think of the bit as being a spin-1/2 particle and of the external parameter as being a magnetic field that we can alter. This shows us that all we need is one heat reservoir and one external parameter for erasure. We do not need additional heat reservoirs or additional external parameters. Erasure is a reset operation. It can be defined either as ''restore to one'' or as ''restore to zero.'' In either case we are going from two possible states of the bit to one possible state.

B. Landauer's argument

The relationship between physical entropy and information may have been mentioned first by Szilard [4,5]. Based on the second law of thermodynamics Szilard introduced the idea that a measurement, information gain, should at some point be accompanied by an increase in entropy. This has been further discussed in terms of quantum mechanics by Zurek [6] and Lloyd [7].

In his original paper from 1961 [1], Landauer argues that since erasure is a logical function that does not have a single-valued inverse it must be associated with physical irreversibility and therefore require heat dissipation. He argues that a bit has one degree of freedom and the heat dissipation should be of order $k_B T$. More precisely, that since before erasure a bit can be in any of the two possible states and after erasure it can only be in one state this implies a change in information entropy of $-k \ln(2)$. Since entropy cannot decrease it must appear, Landauer argues, somewhere else as heat. Implicit in this argument is the crucial assumption that information entropy translates into physical entropy.

C. Later work on erasure

Bennett built on Landauer's principle in a paper on reversible computation [8] (see also [9]). He showed that every step in computation can be made reversible except for erasure (which also includes error correction). After these papers were published many other scientists wrote papers on Szilard's engines [10], different variations on Maxwell's demons [11,12], and on computations which all used Landauer's principle. In fact, their arguments often strongly depended on Landauer's principle. Later, papers were published which criticized Landauer's paper claiming that his proof is not rigorous enough. The reason for these objections was that Landauer's proof is only based on the second law of thermodynamics [10,13] and that it is not clear what

*Electronic address: bpiechocinska@hotmail.com the connection between information and thermodynamic entropy is [14]. In response to the criticism Shizume showed that Landauer's principle holds for a model in which a particle having Brownian motion in a time-dependent double potential well in which a white and Gaussian force is acting on the particle [15]. This was shown using the Fokker-Planck equation. His derivation is restricted to a specific model and only works in classical mechanics. Therefore, it is not as general as Landauer's principle.

Even though computation can be made reversible, and therefore not generate heat (at least in theory), it is still desirable to have a rigorous derivation of Landauer's principle for demonic reasons and because computers always need error-correction. Error-correction is not a reversible operation for a cyclic process with a finite amount of memory and just like erasure it requires heat generation.

II. MICROSCOPIC DERIVATION OF LANDAUER'S PRINCIPLE

In this section we will show the validity of Landauer's principle in three cases: for a classical system in continuous space and time, for a classical system in discrete space and time, and for a quantum system. In all three cases we will assume that the bits which are to be erased are in contact with a heat reservoir whose initial microstate is chosen from a canonical distribution. To make the microscopic derivation as general as possible we will not be using specific and detailed models of the bits and erasure. Instead we will be treating erasure as a thermodynamic process during which we can change an external parameter while the bit is in contact with a heat reservoir. To show Landauer's principle we will be using an ensemble of bits and averaging over the microscopic realizations of this process.

We will assume that the two states of the bit, the ''zero'' state and the ''one'' state, have equal energy. We make this assumption because it has been shown that for these systems all computational operations (except for erasure) can be made reversible and can, at least in theory, be performed using an arbitrarily small amount of work. In Sec. III B of this paper, we will generalize Landauer's principle to include the case where the two energy-states defining the ''zero'' and the ''one'' state are nondegenerate.

A. The continuous classical case

We will be making the following assumptions:

(i) Our system is classical.

(ii) The memory state is a symmetric double potential well where the states ''zero'' and ''one'' have the same energy before and after the erasure.

(iii) The input is randomly distributed (the number of ''zeros'' and ''ones'' is equal and there are no correlations between the bits).

(iv) During erasure the system is in contact with a thermal reservoir with initial states chosen from a canonical distribution.

(v) The interaction term in the Hamiltonian is negligibly small.

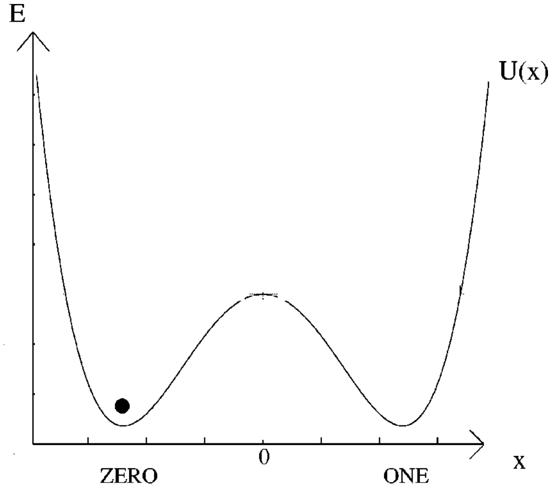

FIG. 1. Double potential well described by the function $U(x)$, where the state ''one'' is $x>0$ and the state ''zero'' is $x<0$. E (in J) is the energy and x (in m) is a distance.

We have an infinite ensemble of bits which we each model as a system with one continuous degree of freedom, x, subject to a symmetric double-well potential energy E, as shown in Fig. 1. The position of the particle in the double-well potential will determine the state of the bit. If the particle is found on the left-hand side of the potential $(x<0)$, then we will say that the bit is in the state ''zero.'' If it is found on the right-hand side of the well $(x>0)$, then we will say that the bit is in the state ''one.'' For it to be considered a useful bit one should also add that the energy barrier separating the two wells is much greater than k_BT. This way the bit is stable enough to store information for a longer period of time. If the energy barrier and k_BT were comparable in size then the particle in the potential well would have enough energy to jump between the two distinct states and the initial information would not be stored. This is not a desirable situation for computational purposes. This assumption, however, is not necessary for the purpose of showing Landauer's principle.

If we were given a bit like the one described above and wanted to erase it by putting it into the ''one'' state it should be clear that this could be done by coupling it to a heat reservoir and changing an external parameter.

To show that erasure implies heat dissipation we will use some of the results presented by Jarzynski in his paper on Clausius-Duhem processes [16]. Before the erasure we want half of the bits to be in the ''one'' state and the other half to be in the ''zero'' state. We assume that the ensemble of bits is in contact with a thermal reservoir where the temperature of the reservoir is low enough not to change the state of the bits $(k_BT \leqslant \Delta U)$. The system will instead reach a ''local'' thermal equilibrium in one of the half-wells. We therefore assume that the initial statistical state is described by the following distribution function for the bits before erasure:

$$\rho_{init}(x,p)=\frac{1}{Z}\exp\left\{-\beta\left[U(x)+\frac{p^2}{2m}\right]\right\} \tag{1}$$

and the statistical state after erasure can be described as

$$\rho_{final}(x,p)=\begin{cases}\dfrac{2}{Z}\exp\left(-\beta\left[U(x)+\dfrac{p^{2}}{2m}\right]\right) & \text{for } x>0\\ 0 & \text{for } x<0,\end{cases}\tag{2}$$

where x is the position, m the mass, p the momentum, $\beta=1/(k_BT)$, and $Z=\int\exp\{-[U(x)+p^2/2m]/k_BT\}dx\,dp$ is the partition function.

Since the total system (the bit and the heat reservoir) is a classical and isolated system it will evolve according to the following Hamiltonian:

$$H(x,p,\mathbf{x}_T,\mathbf{p}_T,t)=H(x,p)+H_T(\mathbf{x}_T,\mathbf{p}_T)+H_{int}(x,p,\mathbf{x}_T,\mathbf{p}_T),\tag{3}$$

where $H(x,p)$ is the Hamiltonian of the system, $H_T(\mathbf{x}_T,\mathbf{p}_T)$ is the Hamiltonian of the heat reservoir, and H_{int} is the Hamiltonian of interaction which we assume to be negligible in comparison with the other terms of the Hamiltonian. The $\mathbf{x}_T$ and $\mathbf{p}_T$ are the positions and momenta of the degrees of freedom that describe the heat reservoir. For easier notation let us use $\zeta=(x,p,\mathbf{x}_T,\mathbf{p}_T)$. Then the trajectory $\zeta(t)$, where t is time, will describe the evolution of all degrees of freedom for one realization of the erasure process. Let us assume that the erasure process takes a time τ and use the shorthand $\zeta^0=\zeta(0)$ and $\zeta^\tau=\zeta(\tau)$.

Following [16] let us now define a function, $\Gamma(\zeta^0,\zeta^\tau)$. Γ is defined for a given microscopic realization in the following way:

$$\Gamma(\zeta^0,\zeta^\tau)=-\ln[\rho_{final}(x^\tau,p^\tau)]+\ln[\rho_{init}(x^0,p^0)]+\beta\Delta E(\mathbf{x}_T^0,\mathbf{p}_T^0,\mathbf{x}_T^\tau,\mathbf{p}_T^\tau),\tag{4}$$

where $\Delta E=H_T(\mathbf{x}_T^\tau,\mathbf{p}_T^\tau)-H_T(\mathbf{x}_T^0,\mathbf{p}_T^0)$ is the change in the internal energy of the heat reservoir and β is as defined previously using the temperature of the heat reservoir. Γ is defined in this particular way because it has proven useful in the kind of calculations we want to perform. Γ is explicitly a function of the initial and final microstates of the system and reservoir during the course of one realization. However, since the final state ζ^τ can be viewed as a function of the initial state ζ^0 [because the evolution $\zeta(t)$ is deterministic], Γ can be viewed as a function of ζ^0 alone:

$$\Gamma(\zeta^0,\zeta^\tau)=\Gamma(\zeta^0,\zeta^\tau(\zeta^0)).\tag{5}$$

For the purpose of Landauer's principle, which is a statement about the average heat released into the environment, we will be interested in averaging over the statistical ensemble of realizations. We will also be interested in finding an inequality relating these averages. For these purposes we will now compute $\langle\exp(-\Gamma)\rangle$, where the angular brackets denote the average over the statistical ensemble of realizations. Since the evolution [governed by the Hamiltonian written in Eq. (3)] is deterministic $\langle\exp(-\Gamma)\rangle$ can be written as an integral over initial conditions ζ^0:

$$\langle\exp(-\Gamma)\rangle=\frac{1}{Z_T}\int\rho_{init}(x^0,p^0)\exp\left(-\frac{H_T(\mathbf{x}_T^0,\mathbf{p}_T^0)}{k_BT}\right)\times\exp(-\Gamma)\,d\zeta^0\tag{6}$$

$$=\frac{1}{Z_T}\int\rho_{init}(x^0,p^0)\frac{\rho_{final}(x^\tau,p^\tau)}{\rho_{init}(x^0,p^0)}\times\exp\left(-\frac{H_T(\mathbf{x}_T^0,\mathbf{p}_T^0)}{k_BT}\right)\times\exp\left(\frac{H_T(\mathbf{x}_T^0,\mathbf{p}_T^0)}{k_BT}-\frac{H_T(\mathbf{x}_T^\tau,\mathbf{p}_T^\tau)}{k_BT}\right)d\zeta^0\tag{7}$$

$$=\frac{1}{Z_T}\int\rho_{final}(x^\tau,p^\tau)\exp\left(-\frac{H_T(\mathbf{x}_T^\tau,\mathbf{p}_T^\tau)}{k_BT}\right)d\zeta^\tau$$

$$=\frac{Z_T}{Z_T}=1,\tag{8}$$

where $Z_T=\int\exp\{-[H_T(\mathbf{x}_T,\mathbf{p}_T)/k_BT]\}d\mathbf{x}_Td\mathbf{p}_T$. In the equations above we have changed the integration variables from $d\zeta^0$ to $d\zeta^\tau$. Since the evolution of our system is Hamiltonian the Jacobian associated with this change of variables is equal to 1. We thus have $\langle\exp[-\Gamma(\zeta^0,\zeta^\tau)]\rangle=1$, where the brackets indicated an average over an ensemble of realizations of the erasure process. Note that the equations above do not in any way imply that the final distribution is a canonical one. The function Γ is just a function of ζ^0 and ζ^τ and it happens to be chosen in such as way that the terms involving ζ^0 in the equations above cancel.

By the convexity of the exponential function $-\langle\Gamma\rangle\leqslant0$. Written explicitly the inequality becomes

$$\langle\ln[\rho_{final}(x^\tau,p^\tau)]\rangle-\langle\ln[\rho_{init}(x^0,p^0)]\rangle\leqslant\langle\beta\Delta E\rangle.\tag{9}$$

Written even more explicitly, using the distribution functions in Eqs. (1) and (2) and the fact that for any function $A(x,p)$,

$$\langle A(x^0,p^0)\rangle=\int\rho_{init}(x,p)A(x,p)dx\,dp\tag{10}$$

as well as

$$\langle A(x^\tau,p^\tau)\rangle=\int\rho_{final}(x,p)A(x,p)dx\,dp,\tag{11}$$

the left-hand side of the inequality above becomes a sum of two contributions.

For $x>0$

$$\int_0^\infty\frac{2}{Z}\exp(-\beta H)\ln[(2/Z)\exp(-\beta H)]dx\,dp-\int_0^\infty\frac{1}{Z}\exp(-\beta H)\ln[(1/Z)\exp(-\beta H)]dx\,dp\tag{12}$$

$$=\int_0^\infty2\alpha\ln(2\alpha)dx\,dp-\int_0^\infty\alpha\ln(\alpha)dx\,dp,\tag{13}$$

where $\alpha=(1/Z)\exp(-\beta H)$. For $x<0$

$$0\ln(0)-\int_{-\infty}^{0}(1/Z)\exp(-\beta H)\ln(1/Z)\exp(-\beta H)dx\,dp$$
$$=-\int_{-\infty}^{0}\alpha\ln(\alpha)\,dx\,dp. \tag{14}$$

The term 0 ln(0) is equal to 0 by l'Hôpital's rule. Since the initial distribution function is symmetric with respect to $x=0$, we have

$$\int_{0}^{\infty}\alpha\ln(\alpha)dx\,dp=\int_{-\infty}^{0}\alpha\ln(\alpha)dx\,dp. \tag{15}$$

Totally, for all x, the left-hand side of Eq. (9) becomes

$$\int_{0}^{\infty}2\alpha\ln(2\alpha)dx\,dp-2\int_{0}^{\infty}\alpha\ln(\alpha)dx\,dp$$
$$=\int_{0}^{\infty}2\alpha\ln\left(\frac{2\alpha}{\alpha}\right)dx\,dp=\int_{0}^{\infty}2\alpha\ln(2)dx\,dp. \tag{16}$$

Since the distribution function is normalized to unity Eq. (9) becomes

$$\ln(2)\leqslant\beta\langle\Delta E\rangle. \tag{17}$$

In defining Γ we defined ΔE as the change in the internal energy of the heat reservoir. We recognize that the interaction term in the Hamiltonian is necessary for the heat reservoir and the system to be able to exchange energy. The size of this term depends on the nature of the heat reservoir as well as the nature of the bit. We will now use the approximation that the interaction term in the Hamiltonian of system and heat reservoir is negligible. To determine how good an approximation this is one would need to specify the physical systems. If we write the equation of conservation of energy for the system and the heat reservoir it gives:

$$W=\Delta E+\Delta E_{system}, \tag{18}$$

where W is the work done on the system and heat reservoir, while ΔE_{system} is the change in the internal energy of the system. Due to the symmetry of ρ_{init} and ρ_{final} ΔE_{system} disappears when averaged over ($\langle\Delta E_{system}\rangle=0$). So, Eqs. (17) and (18) taken together give us

$$k_BT\ln(2)\leqslant\langle W\rangle. \tag{19}$$

This means that to erase one bit of information, on average, the work performed on the system has to be equal to or greater than k_BT ln(2),[1] or, equivalently, that the heat dissipation by the system into the heat reservoir has to be greater than or equal to k_BT ln(2).[2]

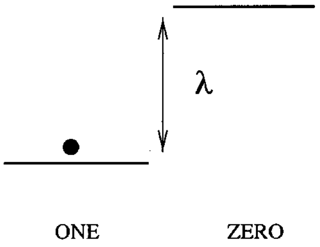

FIG. 2. Two-state system where the left state describes the ''zero'' state and the right state describes the ''one'' state. The difference in energy between the states is λ, which we treat as an external parameter.

B. The discrete classical case

We will now consider a case in which the evolution of the system (or bit) is modeled as a Markov process. In some sense the stochastic dynamics which we use here is less fundamental than the Hamiltonian evolution considered in the preceding section, or the quantum evolution studied in the next section. Nevertheless, Markov evolution is very often used to model a system in contact with a heat reservoir, and it is instructive to see that using this approach, we obtain exactly the same result (Landauer's principle) as when using more fundamental equations of motion. Both the discrete classical case and the quantum case are closely related to the information theoretic treatment of classical Markovian and quantum versions of Maxwell's demon presented by Lloyd [17].

Let us model our bit as a system with two classical states, analogous to quantum energy levels. One of the levels corresponds to the state ''one'' and the other is the state ''zero'' (see Fig. 2). We will use a discrete rather than a continuous time parameter: at each time step the system can either stay in its current state, or ''jump'' to the other state. In other words both our time and space will be discrete. To calculate the heat dissipation into the environment during erasure we will use the method of Crook described in [18].

We assume that in the beginning the energy difference between these levels, λ, is zero, and that half of the bits are in the state ''one'' while the other half are in the state ''zero.'' Then, during the erasure procedure, we change the value of λ in discrete steps, thus performing work on the bit. During the time of erasure we couple the two-state system to a heat reservoir with a certain temperature, T. We could for instance imagine that we separate the energy levels in such a way as to make one of the energy levels have a much higher energy than k_BT. This would guarantee that the transition from the lower energy level to the higher energy level is

[1] Note that this is an approximation because we neglected the interaction term in the Hamiltonian.

[2] This, however, is a precise statement. We do not need to neglect the interaction term in the Hamiltonian for this to be true. It follows from the microscopic definition of ''heat dissipated,'' which we have used.

highly improbable. If we wait for a sufficiently long period of time the probability of finding the system (bit) in the higher energy state will be extremely small. This is one way of seeing how erasure could work in this particular case. Note, however, that the discrete two-state system described in this section is not restricted to this particular scheme of erasure. After erasure the energy difference between the two levels is again set to zero. We will assume that during erasure every step in the process is independent of the other steps (the Markov approximation), and therefore we can write the probability of going from state i_0 to state i_N with all the intermediate states and λ's as:

$$P(i_0 \xrightarrow{\lambda_1} i_1 \xrightarrow{\lambda_2} \cdots \xrightarrow{\lambda_N} i_N) = P(i_0 \xrightarrow{\lambda_1} i_1) P(i_1 \xrightarrow{\lambda_2} i_2) \cdots P(i_{N-1} \xrightarrow{\lambda_N} i_N) \quad (20)$$

using the notation of [18]. Every step in the erasure process can be divided into two parts. The first part consists of changing λ from λ_t to λ_{t+1}. The subscript t indicates the discrete time step. In changing λ, we perform external work on the bit, for instance, in the first time step $Work = E(i_0,\lambda_1) - E(i_0,\lambda_0)$, where $E()$ is the energy of the system. During the second part of the step in the erasure process the bit evolves from one state, i_t, to the next state, i_{t+1}. The corresponding heat absorbed by the system can be written as $\text{Heat} = E(i_1,\lambda_1) - E(i_0,\lambda_1)$ for the first time step. Looking at the entire process we can calculate the total work performed on the system, W, and the total heat absorbed by the system, Q,

$$W = \sum_{t=0}^{N-1} E(i_t,\lambda_{t+1}) - E(i_t,\lambda_t), \quad (21)$$

$$Q = \sum_{t=1}^{N} E(i_t,\lambda_t) - E(i_{t-1},\lambda_t). \quad (22)$$

We assume that the transition probabilities obey detailed balance. Detailed balance in general can be written as ([19,20])

$$\frac{P(i_0 \xrightarrow{\lambda_1} i_1)}{P(i_0 \xleftarrow{\lambda_1} i_1)} = \exp\{-\beta[E(i_1,\lambda_1) - E(i_0,\lambda_1)]\}, \quad (23)$$

where $P(i_t \xrightarrow{\lambda} i_{t+1})$ is the probability of transition from the state i_t to the state i_{t+1} with the external parameter λ. Using this definition for our particular case we can write

$$\frac{P(i_0 \xrightarrow{\lambda_1} i_1) P(i_1 \xrightarrow{\lambda_2} i_2) \cdots P(i_{N-1} \xrightarrow{\lambda_N} i_N)}{P(i_0 \xleftarrow{\lambda_1} i_1) P(i_1 \xleftarrow{\lambda_2} i_2) \cdots P(i_{N-1} \xleftarrow{\lambda_N} i_N)} \quad (24)$$

$$= \frac{\exp[-\beta E(i_1,\lambda_1)]\exp[-\beta E(i_2,\lambda_2)] \cdots \exp[-\beta E(i_N,\lambda_N)]}{\exp[-\beta E(i_0,\lambda_1)]\exp[-\beta E(i_1,\lambda_2)] \cdots \exp[-\beta E(i_{N-1},\lambda_N)]} \quad (25)$$

$$= \exp\left(-\beta \sum_{t=1}^{N-1} [E(i_t,\lambda_t) - E(i_{t-1},\lambda_t)]\right) = \exp(-\beta Q). \quad (26)$$

Let us now also take into consideration the initial and final probabilities and write

$$\frac{P_0(i_0) P(i_0 \rightarrow i_N)}{P_N(i_N) P(i_0 \leftarrow i_N)} = \frac{P_0(i_0)}{P_N(i_N)} \exp(-\beta Q) = \exp\{\ln[P_0(i_0)] - \ln[P_N(i_N)] - \beta Q\}. \quad (27)$$

Just like in the classical continuous case, we are interested in the average over all realizations. If the probabilities are normalized we can see that

$$\langle \exp\{-\ln[P_0(i_0)] + \ln[P_N(i_N)] + \beta Q\}\rangle \quad (28)$$

$$= \sum_{i_0,\ldots i_N} P_0(i_0) P(i_0 \rightarrow i_N) \exp\{-\ln[P_0(i_0)] + \ln[P_N(i_N)] + \beta Q\} \quad (29)$$

$$= \sum_{i_0,\ldots i_N} P_0(i_0) P(i_0 \rightarrow i_N) \frac{P_N(i_N) P(i_0 \leftarrow i_N)}{P_0(i_0) P(i_0 \rightarrow i_N)} \quad (30)$$

$$= \sum_{i_0,\ldots i_N} P_N(i_N) P(i_0 \leftarrow i_N) = 1, \quad (31)$$

where $\langle\,\rangle$ denote the average value. To find the appropriate inequality we use the convexity of the exponential function and we write

$$-\langle \ln[P_0(i_0)]\rangle + \langle \ln[P_N(i_N)]\rangle + \langle \beta Q\rangle \leqslant 0. \quad (32)$$

In the case of erasure, if we erase by restoring to ''one'' then the initial probability will be $P_0(0) = P_0(1) = 1/2$ and the final probabilities will be $P_N(1) = 1$ and $P_N(0) = 0$. Putting these values into Eq. (32) and using the fact that for an arbitrary function $A(i)$, $\langle A(i_0)\rangle = \Sigma_i P_0(i) A(i)$ and $\langle A(i_N)\rangle$

$=\Sigma_i P_N(i) A(i)$, and keeping in mind that the probabilities are normalized gives us the following inequality:

$$-\ln(1/2)+\ln(1)+\langle\beta Q\rangle\leq 0 \tag{33}$$

or just

$$\ln(2)\leq-\beta\langle Q\rangle. \tag{34}$$

In Eq. (21) we defined Q as heat absorbed by the system. This means that $-Q$ is the heat dissipated into the heat reservoir by the system. We know that the total average work done on the system can be written as

$$\langle W\rangle=\langle\Delta E\rangle-\langle Q\rangle, \tag{35}$$

where ΔE is the change in the energy of the system. Since we start with $\lambda_0=0$ and end with $\lambda_N=0$ the change in the energy will be zero. This leaves us with $\langle W\rangle=-\langle Q\rangle$. Combining this with the inequality (34) give us

$$k_B T\ln(2)\leq\langle W\rangle, \tag{36}$$

which tells us that the average work we have to do on the system to erase one bit has to be greater than or equal to $k_B T\ln(2)$. Or, again, equivalently, we could say that the heat dissipated into the heat reservoir has to be equal to or greater than $k_B T\ln(2)$.

C. The quantum case

First let us give a concrete example of what systems could be involved in erasure of a quantum system. The bit could be a two-level atom that initially has two degenerate energy states. The heat reservoir could be described as a photon reservoir with harmonic oscillators. We would couple it to the atom in the beginning of the erasure procedure and decouple it at the end. The external parameter could be a magnetic field that we can alter as we please. The field would be switched on in the beginning and, during erasure, it would split up the two initially degenerate energy states into two different energy states. If the energy difference between the two states became large enough so that the photon reservoir would not be able to excite the atom into the higher energy state, after a while the atom would find itself in the lower energy state. Then the field would be switched off and the energy levels of the two-state atom would become degenerate again. At this point the erasure would be complete. Another way to imagine ''quantum erasure'' could be to use a spin-1/2 particle as a quantum bit.

We will assume that once the erasure itself is complete, the reservoir becomes weakly coupled to some unspecified environment, causing it (the reservoir) to decohere. We will furthermore assume that this coupling is such that the energy eigenstates of the reservoir form the so-called preferred basis states [21], so that, once decoherence has set in, we can view the reservoir to be in a definite energy eigenstate. Effectively, the role of the environment in this situation is that of an ''outside observer,'' who measures the final energy of the reservoir. Such a measurement is necessary if the heat absorbed by the reservoir is to be a well-defined quantity.

In our derivation we will be using a two-state quantum system, not necessarily in a pure state. Therefore we will use a density matrix, $\hat{\rho}$, to describe its statistical state. In the case where we have a 2×2 density matrix we can always write it as

$$\hat{\rho}_{2\times 2}=a\hat{\rho}_a+b\hat{\rho}_b\,. \tag{37}$$

In other words we could say that the density matrix is diagonal in some basis (that does not have to be the energy basis). We can interpret the statistical state described by the density matrix of Eq. (37) as follows: the system is either in state $|a\rangle$ or in state $|b\rangle$, with probability a and b, respectively. The statistical state of any two-state system can be described by a density matrix with the properties outlined above.

We will also be using a heat reservoir. As usual we assume it to be initially in thermal equilibrium. This allows us to write its density matrix as

$$\hat{\rho}_{\hat{H}}=\frac{\exp(-\beta\hat{H})}{tr[\exp(-\beta\hat{H})]}, \tag{38}$$

where $\hat{H}$ is the Hamiltonian of the heat reservoir. We can interpret the $\hat{\rho}_{\hat{H}}$ by imagining the reservoir to be in a definite energy eigenstate. The probability of finding the heat reservoir in its energy eigenstate $|E_n\rangle$ with the eigenvalue E_n is

$$\text{Prob}(|E_n\rangle)=P_n=\frac{\exp(-\beta E_n)}{\sum_m \exp(-\beta E_m)}=\frac{\exp(-\beta E_n)}{Z}. \tag{39}$$

We also have an external parameter, $\lambda(t)$. This parameter serves the purpose of splitting up the two degenerate energy-eigenstates of the two-state system.

Let us imagine the erasure procedure as follows.

(1) At time $t=0$ the bit begins in some statistical state described by the 2×2 density matrix

$$\hat{\rho}_{init}=\begin{pmatrix}1/2 & 0\\ 0 & 1/2\end{pmatrix} \tag{40}$$

in the energy eigenstate basis. (Therefore it can be viewed as starting in either the ''zero'' or the ''one'' state, with equal probability.) The reservoir begins in a definite eigenstate of $\hat{H}$, of energy E_n, with a thermal probability distribution [Eq. (39)]. The initial value of λ is zero.

(2) At time $t=0^+$ we couple the bit to the reservoir.

(3) Between times $t=0^+$ and $t=\tau$ we change the value of the external parameter in some way which we believe will cause the bit to get erased. At the end, $t=\tau$, we make sure that the value of λ is once again zero.

(4) At time $t=\tau^+$ we decouple the bit from the reservoir.

Assuming that the erasure was successful, the bit will now (with excellent probability) be in the pure state corresponding to ''one.'' The reservoir will be in some statistical state, typically described by a density matrix which is not diagonal in the energy eigenbasis. This is where we invoke our assumption about the reservoir being weakly coupled to an external environment which causes it to decohere: the effect of the decoherence is to cause the off-diagonal elements of the density matrix to vanish, without changing the diagonal ones. Thus, at the end, the reservoir will again be in one of the energy eigenstates, with a probability determined by the diagonal density matrix elements.

We can therefore say that the bit begins in either the ''zero'' or the ''one'' state, and ends in the ''one'' state, whereas the reservoir begins in some state $|n\rangle$ and ends in a state $|m\rangle$. Then we can define

$$Q=E_n-E_m \tag{41}$$

as being the heat lost by the heat reservoir. Furthermore, let $|i\rangle$ and $|f\rangle$ denote the initial and final states of the bits, and let $P_{init}(i)=P_i$ and $P_{final}(f)=P_f$ denote the probability distributions of these bits. Then, by assumption, we have

$$P_{init}(i)=1/2 \quad \text{for} \quad i=0,1, \tag{42}$$

$$P_{final}(f)=\begin{cases} 0 & \text{for} \quad f=0, \\ 1 & \text{for} \quad f=1. \end{cases} \tag{43}$$

Finally, let us define an observable Γ, as

$$\Gamma=\ln(P_i)-\ln(P_f)-\beta(E_n-E_m). \tag{44}$$

We can now calculate $\langle\exp(-\Gamma)\rangle$ where the angled brackets denote the average of the function written between them,

$$\langle\exp(-\Gamma)\rangle=\sum_{n,m,i,f} P_iP_n|U_{f,m,i,n}|^2 \times\exp[-\ln(P_i)+\ln(P_f)+\beta(E_n-E_m)] \tag{45}$$

$$=\sum_{n,m,i,f} P_i|U_{f,m,i,n}|^2\frac{P_f}{P_i}\frac{\exp(-\beta E_n)}{Z} \times\exp(-\beta E_m)\exp(\beta E_n) \tag{46}$$

$$=\frac{1}{Z}\sum_{f,m} P_f\exp(-\beta E_m)\sum_{i,n} |U_{f,m,i,n}|^2, \tag{47}$$

where $U_{f,m,i,n}$ corresponds to $\langle f,m|U(\tau)|i,n\rangle$ and it is the time evolution operator. At the same time it is a unitary matrix and therefore has the property that the sum of the absolute value squared of the elements in a column or a row is equal to 1. $|U_{f,m,i,n}|^2$ is the probability for finding the bit and reservoir in final states $|f\rangle$ and $|m\rangle$ given initial states $|i\rangle$ and $|n\rangle$. We then see that

$$\langle\exp(-\Gamma)\rangle=1. \tag{48}$$

Just like in the classical continuous case, Eqs. (45)–(47) do not in any way imply that the final distribution of reservoir states is canonical. The convexity of the exponential function gives us $-\langle\Gamma\rangle\leqslant 0$, which written more explicitly using Eq. (44) is

$$-\langle\ln(P_i)\rangle+\langle\ln(P_f)\rangle+\langle\beta Q\rangle\leqslant 0. \tag{49}$$

Putting in the assumed values of $P_{init}(i)$ and $P_{final}(f)$, we get

$$k_BT\ln(2)\leqslant-\langle Q\rangle. \tag{50}$$

From Eq. (41) we can deduce that $-Q$ is the heat dissipated into the heat reservoir.

Looking at both the system and heat reservoir we can define work the way it is defined in the classical case, namely,

$$W=\Delta E_{heat}+\Delta E_{system}. \tag{51}$$

Just like in the classical continuous case the above equation is valid under the assumption that the interaction energy between the heat reservoir and the two-state system is negligible. Again, the validity of this approximation depends on the physical systems used. As an example of a physical system we could look at nuclear magnetic resonance experiments for quantum computation where trichloroethylene was dissolved in chloroform. We will see that the interaction constant between the qubits and the chlorine is smaller than 1 Hz. This makes the interaction term negligible. Since the two energy-eigenstates of the two-state system are degenerate both before and after erasure we can say that the total change in its energy $\Delta E_{system}=0$. The change in the internal energy of the heat reservoir can be defined as $\Delta E_{heat}=E_m-E_n=-Q$. Putting these values into Eq. (51) gives us $W=-Q$. Putting them into Eq. (50) we see that

$$k_BT\ln(2)\leqslant\langle W\rangle. \tag{52}$$

This means that we can equally well[3] say that the work we have to do on the system in order for it to erase has to be at least $k_BT\ln(2)$.

III. DISCUSSION

A. Probability distributions

In all three cases we have assumed that the initial probability distributions have been 1/2. This is equivalent to saying that in the string of bits about to be erased half of the bits are in the ''zero'' state and half in the ''one'' state. This distribution happens to correspond to thermalized bits. However, in general, the initial string of bits can have any probability distribution. The amount of heat dissipated into the environment will then depend on that distribution. As an example we can take the case where all bits are already in one state only. The equations in the derivations of Landau-

[3] Again, just like in the classical continuous case, this is not an exact statement but an approximation because the interaction term in the Hamiltonian is neglected.

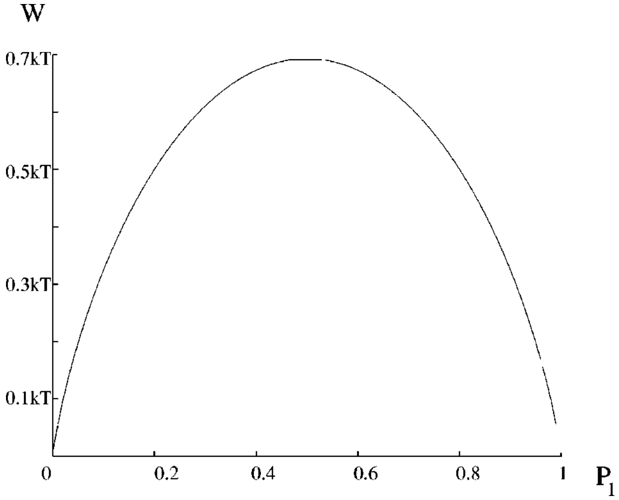

FIG. 3. Graph showing the average minimal amount of work, $P_{final}\ln P_{final}-P_{initial}\ln P_{initial}=\langle W\rangle$, required to erase a bit given the initial probability, $P_{initial}(1)$, of finding the bit in the state ‘‘one.’’ W (in J) is the work and P_1 is the probability to find the bit initially in state ‘‘one.’’

er’s principle tell us that this kind of ‘‘erasure’’ can be done without any dissipation of heat. All other initial distributions will require some heat dissipation and the one which will require the largest dissipation will be the one where half of the states are in one state. More specifically, for both the classical cases considered above (Secs. II A and II B), we found that

$$\langle W\rangle \geqslant T\Delta S, \tag{53}$$

where ΔS is equal to minus k_B times the change in the information entropy between the initial and final statistical states of the bit. In Fig. 3 we plot this lower bound on $\langle W\rangle$, as a function of the probability to find the bit initially in state ‘‘one.’’[4] We see that the distribution that requires the greatest amount of work (or heat dissipation) is the case $P_1=1/2$, in which the initial states are distributed equally between ‘‘zero’’ and ‘‘one.’’

In the quantum case we could have assumed that the initial density matrix for the bit, written in its energy eigenbasis, is some arbitrary matrix

$$\hat{\rho}=\begin{pmatrix} c & d \\ d* & e \end{pmatrix}. \tag{54}$$

This would give us a different initial distribution from the case we considered $(c=e=1/2,\ d=0)$. To see what the minimal heat dissipation would be for a particular density matrix we would have to diagonalize it first,

$$\hat{\rho}=\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}. \tag{55}$$

The diagonal density matrix is expressed using basis states $|a\rangle$ and $|b\rangle$. We now interpret a and b as initial probabilities:

$$P_{init}(i')=\begin{cases} a & \text{for } i'=a, \\ b & \text{for } i'=b, \end{cases} \tag{56}$$

where the prime indicates that we are using a basis different from the energy eigenstate basis. For the final state of the bit, we still have

$$P_{final}(f)=\begin{cases} 0 & \text{for } f=0, \\ 1 & \text{for } f=1 \end{cases} \tag{57}$$

in the energy basis. Desipite the difference in the basis set used to describe the initial state of the bit, the calculation in Eqs. (45)–(47) goes through as before but with $i\rightarrow i'$ and we end with

$$\beta\langle W\rangle \geqslant -\sum_{i'} P_{init}(i')\ln P_{init}(i')=-a\ln a-b\ln b, \tag{58}$$

assuming perfect erasure. This result tells us that $\langle W\rangle$ is bounded from below by minus the change in the von Neumann entropy of the bit. As in the classical case, the greatest value of this lower bound occurs when $a=b=1/2$, in which case $\langle W\rangle \geqslant k_B T\ln 2$.

Note, however, that if we are using an algorithm for erasure where we assume to receive a string with a random distribution and the string we actually receive has a different distribution, for instance, all ‘‘ones,’’ this does not mean that we will automatically be erasing without heat dissipation. To do this we will need to change the algorithm used for erasure.

B. Nondegenerate energy levels

All along we have assumed that the energy levels of the two states of the bits are equal before and after erasure. But what happens if we omit this assumption? Based on Landauer’s argument one would have to say that his principle should apply in this case as well. To see in what form it still applies let us imagine the following system. We have an infinite ensemble of bits with degenerate energy values, like in the classical case with discrete space and time. We assume them to be populated so that half of them are in the ‘‘zero’’ state and half in the ‘‘one’’ state. We then use some external parameter to lower the energy of the ‘‘zero’’ state by ΔE. This is our point of departure (a in Fig. 4).

To do the erasure, which we will assume to be defined as restore to ‘‘zero,’’ we start by raising state ‘‘zero’’ in all of the bits to the energy of state ‘‘one’’ (b in Fig. 4). To do this we will on average have to do $1/2\Delta E$ work. Then we go through with the erasure procedure which we have shown to require $k_BT\ln(2)$ of work (c in Fig. 4). Assuming that the erasure was perfect all the bits will be in the state ‘‘zero’’ (d in Fig. 4). To go back to the original state we lower the energy of the state ‘‘zero’’ by ΔE (e in Fig. 4). This will on average return the energy ΔE. Summing up the energy put in and gotten out of the system, on average, we have

[4] There is no particular reason for us having chosen the state ‘‘one’’ here. We could just as easily have chosen to write the state ‘‘zero.’’

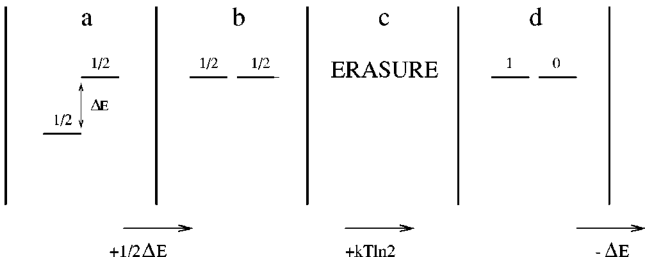

FIG. 4. Schematic picture of erasure for an ensemble of bits with different energy values for the ''zero'' state and the ''one'' state. The numbers above the lines showing the energy levels show the probability of finding the bit in that state. In going from a to b we need to add $1/2\Delta E$ amount of work on average. In c we are erasing the bit which requires $k_B T \ln(2)$ amount of work. In going from d to e we can recover ΔE of work. The average total amount of work we can recover in this kind of erasure is $\Delta E - k_B T \ln(2)$.

$$W_{out} = -1/2\Delta E - k_B T \ln(2) + \Delta E = 1/2\Delta E - k_B T \ln(2). \tag{59}$$

This means that in a system with a ground state and an excited state, where erasure is equivalent to restore to ''zero,'' with ''zero'' being the ground state, we can on average only hope to recover as much as $1/2\Delta E - k_B T \ln(2)$ of useful energy, ΔE being the difference in energies between the two states. The term ''useful energy'' is here used to denote energy that is not necessarily heat. We could have decided to define erasure as restore to ''one''. That, however, would not have given us any extra energy. We would have had to add as much as $1/2\Delta E + k_B T \ln(2)$ to the system in that procedure.

One could argue that it is not always physically possible to change the difference in energy between the levels so that one will have degeneracy. To show that even in this case one cannot get any more ''usable'' energy out of the system than $1/2\Delta E - k_B T \ln(2)$ we can use any of the above discussed models (the continuous classical case, the discrete classical case, or the quantum case) and perform the necessary calculations. The calculations will give us the same result as the thought experiment above.

In summary, we can say that for a system with equal energy levels the work required for erasure is equal to the heat dissipated into the environment, $k_B T \ln(2)$. For a system with different energy levels as the ''zero'' and ''one'' states we do not have that equality. At best we can get out half of the energy difference between the states minus the heat dissipated into the environment which will still be $k_B T \ln(2)$.

C. Suggestions for future research

The microscopic derivation of Landauer's principle could perhaps be made more general if we could drop the assumption that the states of the thermal reservoir, with which the bits are in contact, are chosen from a canonical distribution. Perhaps it is enough to assume that the states of the thermal reservoir are chosen form a microcanonical distribution. The microcanonical ensemble would still provide us with a well defined temperature.

Landauer's principle gives us a fundamental lower bound on the amount of heat dissipated into the environment in the process of erasure. It would be interesting to see if this lower bound can actually be reached physically. One can certainly imagine a process where erasure is done on an infinite time scale where one reaches the lower bound. But is it physically realizable? If so, would it be practical for computational purposes?

ACKNOWLEDGMENTS

This paper was written at Los Alamos National Laboratory. I would like to to thank Wojciech H. Zurek, N. Balazs, Christof Zalka, and Tanmoy Bhattacharya for new ideas and valuable and very interesting discussions about erasure. I would especially like to thank Chris Jarzynski for his guidance.

[1] R. Landauer, IBM J. Res. Dev. **3**, 183 (1961).

[2] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Charles Scribner's Sons, New York, 1902), Chap. XIII.

[3] R. Landauer, in *Statistical Physics, Invited Papers from STATPHYS 20, 20th IUPAP International Conference on Statistical Physics UNESCO and Sorbonne Paris, 1998*, edited by A. Gervois, D. Iagolnitzer, M. Moreau, and Y. Pomeau (North-Holland, Elsevier, 1998).

[4] L. Szilard, Z. Phys. **53**, 840 (1929) [Translation in Wheeler and Zurek (Ref. [5])].

[5] J. A. Wheeler and W. H. Zurek, *Quantum Theory and Measurement* (Princeton University Press, Princeton, NJ, 1983).

[6] W. H. Zurek, in *Maxwell's Demon, Entropy, Information, Computing* (Princeton University Press, Princeton, NJ, 1990).

[7] S. Lloyd, Phys. Rev. A **56**, 3374 (1997).

[8] C. H. Bennett, IBM J. Res. Dev. **17**, 525 (1987).

[9] C. H. Bennett and R. Landauer, Sci. Am. **253**, 38 (1985).

[10] J. Berger, Int. J. Theor. Phys. **29**, 9 (1990).

[11] W. H. Zurek, e-print quant-ph/9807007.

[12] H. S. Leff and A. F. Rex, *Maxwell's Demon, Entropy, Information, Computing* (Ref. [6]).

[13] D. Wolpert, Phys. Today **45**, 98 (1992).

[14] E. Goto, N. Yoshida, K. F. Loe, and W. Hioe, in *Proceedings of the 3rd International Symposium on the Foundations of Quantum Mechanics, Tokyo*, edited by H. Ezawa, Y. Murayama, and S. Nomura (Physical Society Japan, Tokyo, 1990), p. 412.

[15] K. Shizume, Phys. Rev. E **52**, 3495 (1995).

[16] C. Jarzynski, e-print cond-mat/9802249.

[17] S. Lloyd, Phys. Rev. A **39**, 5378 (1989).

[18] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).

[19] S. R. de Groot and P. Mazur, *Nonequilibrium Thermodynamics* (North-Holland, Amsterdam, 1962).

[20] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987), p. 165.

[21] W. H. Zurek, Prog. Theor. Phys. **89**, 281 (1993).