Heat generation required by information erasure

Kousuke Shizume
University of Library and Information Science, 1-2 Kasuga, Tsukuba, Ibaraki 305, Japan
(Received 18 April 1995)

Landauer argued that the erasure of 1 bit of information stored in a memory device requires a minimal heat generation of $k_BT \ln 2$ [IBM J. Res. Dev. **5**, 183 (1961)], but recently several articles have been written to dispute the validity of his argument. In this paper, we deal with a basic model of the memory, that is, a system including a particle making the Brownian motion in a time-dependent potential well, and show that Landauer's claim holds rigorously if the random force acting on the particle is white and Gaussian. Our proof is based on the fact that the analogue of the second law of thermodynamics $dQ \leq k_BTdS$ holds rigorously by virtue of the Fokker-Planck equation, even if the potential is not static. Using the above result, we also discuss the counterargument of Goto *et al.* to Landauer's claim based on the quantum flux parametron.

PACS number(s): 05.40.+j, 05.20.−y, 89.70.+c

I. INTRODUCTION

In 1961, Landauer discussed the limitation of the efficiency of computers imposed by physical laws [1]. He argued that the erasure of 1 bit of information requires a minimal heat generation of $k_BT \ln 2$ based on the second law of thermodynamics. To be more precise, the argument is as follows. Consider a memory device that can hold one of the two values ONE and ZERO. Physically it is a system which has two stable states. When it is in one of the two states, it can be regarded as holding the value ONE, and when it is in another state, ZERO. The erasure of information stored in the memory means the operation RESTORE TO ONE (RTO), which sets the value to ONE, regardless of its initial value. Physically, RTO forces the system into the state corresponding to ONE regardless of its initial state. Moreover, this operation must not leave any trace of the initial value (in other words, any trace of the initial state) anywhere in the system. Landauer's claim means that RTO is inevitably accompanied by the heat generation of at least $k_BT \ln 2$.

For the sake of concreteness, let us introduce a basic model of a memory [1,2]. It is a binary device in which a particle makes the Brownian motion in a bistable potential well (see Fig. 1). When the particle is in the right-hand-side well, one may regard the device as taking the value ONE, and when it is in the left-hand-side well ZERO. In this model one can perform RTO by varying the shape of the potential well with time so that the particle ends in the right-hand-side well, regardless of its initial position. In this model Landauer's claim means that, in whatever way one varies the shape of the potential well with time, the work dissipated into the environment due to the friction cannot be less than $k_BT \ln 2$.

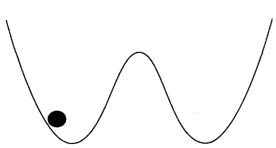

FIG. 1. The model of the memory. The particle is making a Brownian motion in the potential well.

Inspired by his study, a considerable number of studies have been made on the thermodynamics of information processing, which include Maxwell's demon problem [2,3], reversible computation [2,4], the proposal of the algorithmic entropy [5], etc. The thermal cost of more general information processes have also been investigated [6].

On the other hand, however, objections have been raised to his claim. Some authors indicate that his claim is based only on the second law of thermodynamics, and, although plausible, not very rigorous [7,8]. Other authors argued that the information has nothing to do with thermodynamical entropy at all [9]. One of the most interesting counterarguments was advanced by Goto *et al.* [9,10]. They argue that it is possible to erase information with infinitesimal heat generation by using the quantum flux parametron (QFP) [9–13] developed by themselves (see Sec. III). Their counterargument has special importance because one may be able to realize the basic model of the memory introduced above by using the QFP. Although some discussions have appeared thereafter [14,15], they are not quantitative or rigorous.

A major drawback in Landauer's discussion that makes these objections possible is that it is based only on the second law. Although the second law is very general, these objections seems to suggest that applying it to information processing requires more careful consideration. Since Landauer's claim has become a part of the foundations of the thermodynamics of information processing, more rigorous and quantitative discussion is desired.

In this paper we present a sufficient condition for Landauer's claim. Our discussion is based on the Fokker-Planck equation, so that it is less general than Landauer's discussion but more definite and instructive. We show that Landauer's claim holds rigorously in the basic model of the memory introduced above if the random force acting on the particle is a white and Gaussian noise force.

More specifically, we show rigorously that, no matter how one changes the potential well with time to perform RTO, the average of the total heat generated during the operation will be greater than $k_BT\ln 2$ if the random force acting on the particle is white and Gaussian. In our proof, which is given in Sec. II, we *do not presume* the second law of thermodynamics as Landauer did. Because of our assumption about the nature of the random force, the time dependence of the distribution function of the particle is described by the Fokker-Planck equation (FPE) [16] as is well known. We show that the analogue of the second law of thermodynamics $dQ \leq k_BTdS$ *is derived from the FPE, even if the potential is time dependent.* The minimal heat generation caused by RTO can be obtained as a direct result of the analogue of the second law.

Based on this result, Goto *et al.*'s counterargument is discussed in Sec. III and it is shown that, if the behavior of the Josephson junction used in the QFP can be described by the FPE, it is impossible to erase information with infinitesimal heat generation even by using the QFP. We also comment on some other discussions relating to Landauer's claim. We summarize our main results in Sec. IV. We use the units $k_B = 1$ below.

II. THE MINIMUM HEAT GENERATION

We study the model of the memory involving a particle making the Brownian motion in a time dependent potential well. The motion of the particle is described by the Langevin equation [16]

$$m\frac{d^2x}{dt^2} + m\gamma\frac{dx}{dt} = -\frac{\partial V(x,t)}{\partial x} + F_R(t), \tag{2.1}$$

where m is the mass of the particle, γ is the friction constant, $V(x,t)$ is the potential that is assumed to be time dependent, and $F_R(t)$ is the thermal random force. We assume the random force $F_R(t)$ to be a white and Gaussian noise force satisfying

$$\langle F_R(t_1)F_R(t_2)\rangle = 2m\gamma T\delta(t_1 - t_2), \tag{2.2}$$

where T is the temperature of the environment of the memory. Because of this assumption, the motion of the distribution function $f(x,u,t)$ of the particle in position and velocity space is described by the FPE [16]

$$\frac{\partial}{\partial t}f(x,u,t) = \left[-\frac{\partial}{\partial x}u + \frac{\partial}{\partial u}\left(\gamma u + \frac{1}{m}\frac{\partial V(x,t)}{\partial x}\right) + \frac{\gamma T}{m}\frac{\partial^2}{\partial u^2}\right]f(x,u,t). \tag{2.3}$$

If the potential V is static, the well-known H theorem is derived from (2.3), and the analog of the second law of thermodynamics,

$$dQ \leq TdS, \tag{2.4}$$

is derived based on the H theorem. [The exact meaning of Eq. (2.4) is given by Eq. (2.6) below.] Moreover, several important thermodynamic concepts are incorporated into the stochastic process framework [17,18]. In our case, however, the H theorem *does not hold* because V is time dependent. (It is easy to show it. One can change the distribution f cyclically by changing the potential V cyclically. In this case any functional of f such as H should also change cyclically, so that it should increase at some period.) Nevertheless Eq. (2.4) is valid even in this case. In fact, letting $\dot{Q}$ be the ensemble average of the energy given to the particle by the environment per unit time (the dot over Q means "per unit time") and S be the Shannon–von Neumann entropy

$$S \equiv -\int_{-\infty}^{\infty} dxdu f\ln f \tag{2.5}$$

of the distribution function at time t, the inequality

$$\dot{Q} \leq T\frac{dS}{dt} \tag{2.6}$$

holds at any time.

The proof of Eq. (2.6) is straightforward. By virtue of the energy conservation law, $\dot{Q}$ is given by

$$\dot{Q} = \frac{d\langle E\rangle}{dt} - \dot{W}, \tag{2.7}$$

where

$$E \equiv \frac{mu^2}{2} + V(x,t), \tag{2.8}$$

$\langle X\rangle$ denotes the ensemble average of any function X,

$$\langle X\rangle \equiv \int_{-\infty}^{\infty} dxdu f(x,u,t)X(x,u,t), \tag{2.9}$$

and $\dot{W}$ is the average work done by the potential V on the particle per unit time. Since the work given to the particle at a position x by V per unit time is $\partial V(x,t)/\partial t$ [19], $\dot{W}$ is given by

$$\dot{W} = \left\langle\frac{\partial V(x,t)}{\partial t}\right\rangle = \int_{-\infty}^{\infty} dxdu f(x,u,t)\frac{\partial V(x,t)}{\partial t}. \tag{2.10}$$

By inserting Eqs. (2.8)–(2.10) into (2.7) and using the FPE, we obtain

$$\dot{Q} = \int_{-\infty}^{\infty} dxdu \frac{\partial f(x,u,t)}{\partial t}V(x,u,t) = \gamma(T - \langle mu^2\rangle). \tag{2.11}$$

The time derivative of S is given by differentiating Eq. (2.5) with respect to t and using the FPE as follows:

$$\frac{dS}{dt} = \gamma\left[\frac{T}{m}\left\langle\left(\frac{\partial \ln f}{\partial u}\right)^2\right\rangle - 1\right]. \tag{2.12}$$

Therefore, by means of (2.11) and (2.12),

$$\dot{Q} - T\frac{dS}{dt} = -\frac{\gamma}{m}\left\langle\left(T\frac{\partial \ln f}{\partial u} + mu\right)^2\right\rangle \leq 0. \tag{2.13}$$

Thus we obtain Eq. (2.6).

Equation (2.6) allows us to use reasoning formally identical to that used in thermodynamics. Especially, letting $\Delta Q_{out}(t_i, t_f)$ be the average energy dissipated into the environment between any times t_i and t_f, the lower bound of $\Delta Q_{out}(t_i, t_f)$ is determined by the value of Shannon–von Neumann entropy of the distribution function at those times as follows:

$$\Delta Q_{out}(t_i, t_f) = \int_{t_i}^{t_f} (-\dot{Q})dt \geq T[S(t_i) - S(t_f)]. \tag{2.14}$$

Now we can calculate the lower bound of the heat generation caused by the erasure of 1 bit of information by using (2.14). The following argument may be regarded as a refinement of Landauer's original discussion [1]. Consider an ensemble consisting of $N(\gg 1)$ memories. Let us assume that at time t_i every memory in the ensemble stores 1 bit of information. This means that the potential well of each memory forms the double well form (Fig. 1), and the distribution of each particle is localized in either the right-hand-side well (when the stored value is ONE) or the left-hand-side well (when it is ZERO). Let $f_1(x,u)$ and $f_0(x,u)$ be the distribution functions of the particle when the memory takes the values ONE and ZERO, respectively, and also p_1N and p_0N be the number of memories whose values are ONE and ZERO, respectively. Then the number of particles whose positions and velocities are within $x \sim x + dx$ and $u \sim u + du$, respectively, is given by

$$Np_0 f_0(x,u)dxdu + Np_1 f_1(x,u)dxdu$$
$$= N[p_0 f_0(x,u) + p_1 f_1(x,u)]dxdu. \tag{2.15}$$

Therefore the Shannon–von Neumann entropy S_{init} of the ensemble per memory at time t_i is given by

$$S_{init} = -\int_{-\infty}^{\infty} dudx(p_0 f_0 + p_1 f_1)\ln(p_0 f_0 + p_1 f_1). \tag{2.16}$$

The overlapping of $f_1(x,u)$ and $f_0(x,u)$ should be negligible, because if not it is impossible to decide the value of the memory by the measurement of the position of the particle. Thus

$$S_{init} \approx -\int_{-\infty}^{\infty} dudx[p_0 f_0 \ln(p_0 f_0) + p_1 f_1 \ln(p_1 f_1)]$$
$$= p_0 S[f_0] + p_1 S[f_1] + S[p_0, p_1], \tag{2.17}$$

where

$$S[f_k] \equiv -\int_{-\infty}^{\infty} dxdu f_k \ln f_k \quad (k = 0 \text{ or } 1), \tag{2.18}$$

and

$$S[p_0, p_1] \equiv -\sum_{i=0,1} p_i \ln p_i. \tag{2.19}$$

$S[f_k]$ is the entropy due to the distribution of the particle in the phase space when the memory stores a definite value k, and $S[p_0, p_1]$ is that due to the distribution of the values of the memories. If RTO is performed, and then the values of all memories become ONE at time t_f, the entropy S_{final} of the ensemble per memory at time t_f is given by

$$S_{final} = S[f_1]. \tag{2.20}$$

By inserting Eqs. (2.17) and (2.20) into (2.14), we obtain the lower bound of the heat generated by RTO as follows:

$$\Delta Q^{\text{RTO}}_{Lower\ bound} = T(S[p_0, p_1] + p_0 S[f_0] + p_1 S[f_1] - S[f_1]). \tag{2.21}$$

Then if

$$S[f_0] = S[f_1] \tag{2.22}$$

holds, Eq. (2.21) is reduced to

$$\Delta Q^{\text{RTO}}_{Lower\ bound} = TS[p_0, p_1]. \tag{2.23}$$

The right-hand side of Eq. (2.23) is exactly the product of the temperature and the amounts of erased information, and gives $T\ln 2$ if $p_0 = p_1 = 1/2$. Thus we have derived Landauer's claim rigorously in our model.

In the above discussion, we assumed Eq. (2.22) to derive Eq. (2.23). But, in fact, the assumption can be dispensed with, if we take into account the thermal cost required to *set* a definite value for the memory after the erasure. (Landauer seems to have noted this point in Ref. [1]. However, because he gave no detailed explanation, we discuss it here for completeness.) The entropy of the ensemble per memory after setting a definite value for each memory is given by

$$S_{set} = p_0 S[f_0] + p_1 S[f_1], \tag{2.24}$$

where it is assumed that the ratio between the memories whose values are ZERO and those whose values are ONE is the same as that before the erasure. [$S[p_0, p_1]$ does not appear in the right-hand side of Eq. (2.24) because each memory has a definite value.] Therefore the lower bound of the work dissipated into the environment during the erasure-setting process is given by

$$T(S_{init} - S_{set}) = TS[p_0, p_1] \tag{2.25}$$

without the assumption (2.22). This result can be generalized to the case where each memory can take M (≥ 2) values in a straightforward manner.

III. DISCUSSION

In this section we first discuss Goto *et al.*'s counterargument based on the QFP, and then comment on some interesting discussions relating to Landauer's claim.

The QFP is a Josephson logic device developed by Goto *et al.*, which uses magnetic flux to hold and transfer information. The output flux Φ from the QFP can be controlled by the "input flux" Φ_s and the "activation flux" Φ_a, because it is governed by the Langevin-type equation of motion

$$2C\frac{d^2\Phi}{dt^2} + \frac{2}{R}\frac{d\Phi}{dt} + \frac{\partial V}{\partial \Phi} = -I_R \ , \tag{3.1}$$

where

$$V = -2E_J \cos\left(\frac{2\pi\Phi}{\Phi_0}\right)\cos\left(\frac{2\pi\Phi_a}{\Phi_0}\right) + \frac{(\Phi-\Phi_s)^2}{2L_L} \tag{3.2}$$

($\Phi_0 = h/2e$), I_R is the resistive thermal noise current, and C, R, L_L, and E_J are constants. See Refs. [9–13] for details. The essential point is that one can make V either a single well or double well in form by controlling the activation flux Φ_a. When $\Phi_a = 0$, V forms a single well, and when $\Phi_a = \Phi_0/2$ it forms a double well. Therefore one may be able to realize the basic model of the memory introduced in Sec. I by using the QFP.

Goto *et al.* [9,10] argued that the heat generated by QFP per clock period is given by

$$P \equiv \int \frac{\dot{\Phi}^2}{R} dt = H_0 f_c, \tag{3.3}$$

where H_0 is a constant and f_c is the clock cycle, and that by making f_c small one can make P as small as one wishes in contradiction to Landauer's claim. However, since Eq. (3.1) is mathematically the same as Eq. (2.1), we can apply our result obtained in Sec. II. Therefore we can conclude that, if the noise current can be regarded as white and Gaussian, it is impossible even for the QFP to erase 1 bit of information with less heat generation than $k_B T \ln 2$.

We should keep in mind, however, that we need the assumption about the nature of the noise current to derive the above conclusion. The assumption that the noise current is white and Gaussian, or, in other words, the motion of the Josephson junction used in the QFP can be described by the FPE, is known to be valid in many cases [16,20,21]. Nevertheless, it may be possible to construct a QFP with a Josephson junction whose current noise is not white and Gaussian. In this case, our discussion is not enough to contradict Goto *et al.*'s counterargument and we must investigate whether the discussion given in the previous section can be generalized to the case where the random force is not white and Gaussian. It needs further investigation.

There is one more point to be discussed concerning their discussion. As stated above, they take the total heat P generated by the resistors and discuss whether it is possible or not to make it lower than $k_B T \ln 2$. In the model discussed in Sec. II the quantity corresponding to P is

$$P' \equiv \int_{t_i}^{t_f} \gamma m u^2 dt. \tag{3.4}$$

However, the resistors not only generate but also absorb heat from the environment. Then "heat generation required by information erasure" means the difference between the heat generated and that absorbed by the memory during the erasure process. In other words, it means the work done by the time-dependent potential and dissipated into the environment. It is for this reason that we should use, and have used, $\Delta Q_{out}(t_i, t_f)$ rather than P' as the definition of the "heat generation required by information erasure." It may be worth pointing out that the relationship between the quantities P' and $\Delta Q_{out}(t_i, t_f)$ is obtained, by using Eq. (2.11), as

$$P' = \Delta Q_{out}(t_i, t_f) + \gamma T (t_f - t_i). \tag{3.5}$$

As a result, the inequality

$$P' > \Delta Q_{out}(t_i, t_f) \tag{3.6}$$

always holds as expected.

Igeta [22] argued, against Landauer's claim, that "the physical entropy does not change because there is no thermodynamic difference between *Zero* and *One*. Each of them is definite and has no statistical factors. Also, the physical entropy of the states *Zero* and *One* can be the same by physical symmetry." His last statement corresponds to Eq. (2.22). However, our result supports Landauer's claim and contradicts Igeta's argument. The point is that the ensemble before the erasure contains both of memories whose values are ONE and those whose values are ZERO. As a result, S_{init} given by Eq. (2.17) contains the term (2.19) which does not appear in S_{final}.

Fahn [23] argued, in connection with the analysis of Maxwell's demon (especially Szilard's engine), that "there is an entropy symmetry between the measurement and erasure steps, whereby the two steps additively share a constant entropy change, but the proportion that occurs during each of the two steps is arbitrary." His "measurement" process corresponds to our "setting process" in Sec. II, so that our result on the thermal cost of the erasure-setting process agrees with his argument.

Finally we comment on Schneider's note [24] on Landauer's claim. According to Schneider, the information erasure consists of two steps: priming and setting. The device is in one of the two stable states at the beginning, so that one must add some energy to the memory device to alter that state at first. He called this step the "priming step" and argued that information is lost at this step. On the next "setting step," the device is guided by preset inputs to fall into the standard state and the energy added at the priming step is dissipated. He argued that Landauer lumped these two steps into a single step.

However, the priming step is not essential for the information erasure. The information is assumed to be already lost before the erasure, that is, the value of the memory is already uncertain before the erasure. (If not so, one can set the value to ONE expending no energy [1].) In a typical erasure process energy is added to the particle and then dissipated almost simultaneously when the distribution of the particle is compressed.

IV. SUMMARY

We have investigated the lower bound of the heat generation required by information erasure in the case of a basic model of the memory, that is, a Brownian particle in a moving potential well. We have shown that, if the random force acting on the particle can be regarded as white and Gaussian, Landauer's claim that the erasure of 1 bit of information is accompanied by the heat generation of at least $k_B T \ln 2$ holds rigorously. Next we have discussed Goto *et al.*'s counterargument and concluded that if the resistive thermal current noise involved in the QFP is white and Gaussian, or, in other words, if the behavior of the Josephson junction can be described by the Fokker-Planck equation, it is impossible to erase information with infinitesimal heat generation even by using the QFP.

ACKNOWLEDGMENTS

The author would like to thank T. Arimitsu, H. Hayakawa, H. Hasegawa, T. Iwai, K. Sasaki, S. Takagi, and Y. Takahashi for helpful discussions.

[1] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961); reprinted in Ref. [3].
[2] C.H. Bennet, Int. J. Theor. Phys. **21**, 905 (1982); reprinted in Ref. [3].
[3] H.S. Leff and A.F. Rex, *Maxwell's Demon: Entropy, Information, Computing* (Princeton University Press, Princeton, 1990).
[4] E. Fredkin and T. Toffoli, Int. J. Theor. Phys. **21** , 219 (1982).
[5] W.H. Zurek, Phys. Rev. A **40**, 4731(1989).
[6] R. Landauer and J.W.F. Woo, J. Appl. Phys. **42**, 2301 (1971).
[7] J. Berger, Int. J. Theor. Phys. **29**, 985 (1990).
[8] D. Wolpert, Phys. Today **45** (3), 98 (1992).
[9] 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 (Phys. Soc. Jpn., Tokyo, 1990), p. 412.
[10] E. Goto, W. Hioe, and M. Hosoya, Physica C **185-189**, 385 (1991).
[11] E. Goto and K.F. Loe, *DC Flux Parametron* (World Scientific, Singapore, 1986).
[12] Y. Harada, H. Nanane, N. Miyamato, U. Kawabe, and E. Goto, IEEE Trans. Magn. **MAG-23**, 3801 (1987).
[13] W. Hioe and E. Goto, *Quantum Flux Parametron* (World Scientific, Singapore, 1991).
[14] R. Landauer, Physica C **208**, 205 (1993).
[15] E. Goto and N. Yoshida (unpublished).
[16] H. Risken, *The Fokker-Planck Equation*, 2nd. ed. (Springer-Verlag, New York, 1989).
[17] H. Hasegawa, Prog. Theor. Phys. **57**, 1523 (1977).
[18] H. Hasegawa, T. Nakagomi, M. Mabuchi, and K. Kondo, J. Stat. Phys. **23**, 281 (1980).
[19] L.D. Landau and E.M. Lifshitz, *Mechanics* (Pergamon, Oxford, 1969).
[20] A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect* (Wiley, New York, 1982), and references therein.
[21] E. Ben-Jacobi, E. Mottola, and G. Schoen, Phys. Rev. Lett. **51**, 2064 (1983).
[22] K. Igeta, in *Proceedings of the Workshop on Physics and Computation PhysComp'92* (IEEE Computer Society Press, Los Alamitos, CA, 1992), p. 184.
[23] P.N. Fahn, in *Proceedings of the Workshop on Physics and Computation PhysComp'94* (IEEE Computer Society Press, Los Alamitos, CA, 1994), p. 217.
[24] T.D. Schneider, Nanotechnology **5**, 1 (1994).