Journal of Statistical Physics, Vol. 54, Nos. 5/6, 1989

Dissipation and Noise Immunity in Computation, Measurement, and Communication

Rolf Landauer[1]

Received March 30, 1988

We have known for almost three decades that the steps in a computer that require a minimal energy expenditure, that cannot be avoided by any means, are those that discard information. For more than half that period we have known that such steps are not essential; computation can be carried out through a sequence of logical 1:1 mappings. Computation, therefore, can be carried out with arbitrarily little dissipation per step, if done sufficiently slowly. Much more recently it has been emphasized that measurement and communication are similar to computation; it is only the information-discarding steps that have a lower bound on the dissipation. Such steps are not required in communication. In measurement, as shown by Bennett, they only become essential when we reset the meter for its next (or first) use. This paper is not a detailed exposition of all this, but only an annotated guide to the existing literature.

KEY WORDS: Dissipation; noise immunity; computation; measurement; communication.

1. INTRODUCTION

I will address energy dissipation requirements in three related but distinguishable areas: (1) measurement, (2) communication, and (3) computation. The distinction between communication and computation, for example, is made clear in Fig. 1. In a communications link, we want to get out what we have put in, changed one hopes, as little as possible in the transmission process. In contrast, the elementary logic step in a computer invokes a nonlinear interaction between two or more binary inputs.

I am dealing with three fields, all of which have an extensive literature. I cannot hope to discuss even one of these areas in a definitive manner in

[1] IBM Research Division, T. J. Watson Research Center, Yorktown Heights, New York 10598.

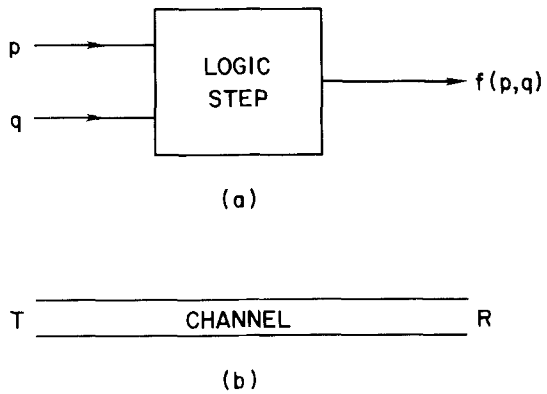

Fig. 1. Logic step in (a) utilizes a nonlinear interaction between p and q. The transmission channel in (b), it is hoped, reproduces its input at the output.

this paper, which is closer to an annotated bibliography than to a complete exposition. My intent, instead, is to emphasize the recent application of what has been learned about computation to the two other areas. These other areas are much older and can be dated to Maxwell's demon[1] in the case of the measurement problem and to Claude Shannon[2] for the communications channel. Thus, if we newcomers in the computational area tell the investigators in the two older areas that they had not gotten it quite right, we are likely to be making controversial assertions. I cannot expect to convince all readers immediately, but hope that our notions will be met with an open mind and a willingness to give them serious consideration.

The three areas listed have been burdened with excessively simple attempts to estimate minimal energy dissipation. These casual estimates have been widely accepted, without a demonstration that they were indeed minimal, or without adequate concern whether the arguments for minimal dissipation were applicable to *all possible methods* for carrying on the desired operation. For an entertaining and provocative assessment of this problem see ref. 3. I give a simple example. The quantum mechanical uncertainty principle, $\Delta E \, \Delta t \sim \hbar$, has frequently been invoked casually[4] to suggest that fast switching in computers requires an energy dissipation defined by the uncertainty principle. There are also much more detailed studies with a similar thrust.[5] But the uncertainty principle does not refer to energy dissipation, it refers to an energy spread. For this and other reasons[6] the uncertainty principle argument is invalid. Nevertheless, an incorrect argument can, conceivably, yield a correct result, and can only be given a definitive rebuttal through a counterexample. Such counterexamples, showing that conservative Hamiltonian systems can carry out computation, have been developed by Benioff, Feynman, Zurek, Deutsch, Peres, and others. For a more detailed discussion and critical evaluation of this field and an entry to the citation trail, see ref. 7. The most common error, however, characteristic to a varying extent of all three fields, is classical rather than quantum mechanical. Systems at a temperature T have thermal energies kT, per degree of freedom. It is then natural, but not necessarily correct, to assume that execution of a useful function requires energy comparable to or larger than that, and that this energy has to be dissipated.

2. COMPUTATION

This subject is best summarized in a somewhat historical fashion. The 1950s was a period for speculation about limits, and these speculations were not much more than dimensional analysis. Brillouin's well-known book,[8] for example, despite its chapter on "The Problem of Computing," contains no references to the actual logic processes involved in a computer, e.g., to a logical "and" or a logical "or," and contains no references to a total working computer system, such as a Turing machine or a cellular automaton. A more careful approach commenced with ref. 9. There it was argued that the operations in a computer which have an unavoidable minimal energy dissipation were those which discarded information, i.e., those which did not represent a logical 1:1 mapping. Note that throughout the present discussion, unless otherwise specified, I will be considering classical systems with frictional forces proportional to velocity, as found in electricity and hydrodynamics. I will also assume that these systems exhibit thermal equilibrium noise. (Note, however, that some of the detailed considerations in the literature are stronger than that, and allow for a greater variety of noise sources.) Thus, if computation can be made equivalent to motion along a path, with each program and its initial data determining a unique and identifiable path, then the frictional forces and the resulting energy dissipation can be made arbitrarily small, by choosing a sufficiently low computational velocity. This is like the motion of an electron through a solid. Over a short period, in the presence of noise, the motion is diffusive. Over a long period, however, an arbitrarily small force is adequate to produce a predictable drift velocity. On the other hand, if we have paths that merge, representing information loss, then the phase space before the merging event is larger than after, and the system in equilibrium will prefer the unmerged state with its larger entropy. A minimal force and dissipation are required to be certain that the system will pass the place where the tracks merge. It was the brilliant perception of Bennett[10] that, in fact, general-purpose computation can be carried out by systems which are 1:1

at every step, and do not need merging of tracks. Wheeler and Zurek[11] have labeled this perception as "epoch-making."

The notion of reversible computation is still not known to, or accepted by, all those who comment on the subject.[12] A sufficient variety of investigators, however, with differing backgrounds and perspective have analyzed reversible computation to give the notion some credence. It is inadequate for the skeptic to demolish only one of the several and complementary viewpoints.

3. MODULATED POTENTIAL WELL

I allude here, briefly, to some aspects of one of the detailed embodiments for reversible computation. This utilizes particles in time-modulated potential wells shown in Fig. 2. The wells are assumed to be heavily damped; the velocity of the particles is proportional to the force, and inertial effects are unimportant. Different wells will be exposed to one of several different phases for the time modulation. Particles in a deeply bistable state locked into a "0" (left-hand well) or "1" (right-hand well) will be coupled via springs as shown in Fig. 3 to wells which are just undergoing the transition from the monostable state to the bistable state. The particles in wells going through the bifurcation transition will be pushed one way or the other according to the majority vote of the odd number of wells influencing them. As shown in refs. 13, all logic functions required in general-purpose computation can be executed via such "majority logic."

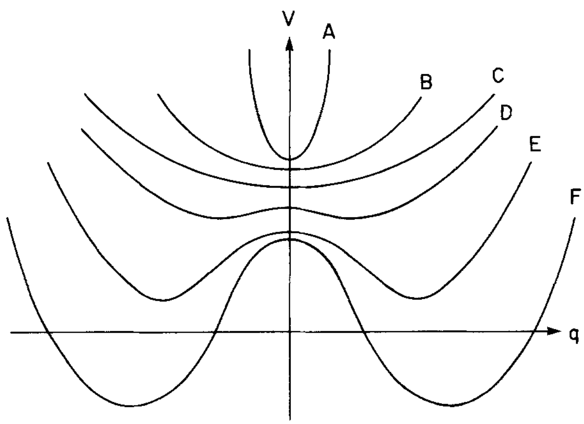

Fig. 2. Potential changing with time. Starts at A with a single minimum and ends up at F in a deeply bistable state, and then returns to A. Relative vertical displacement of curves is unimportant, and is selected for clarity.

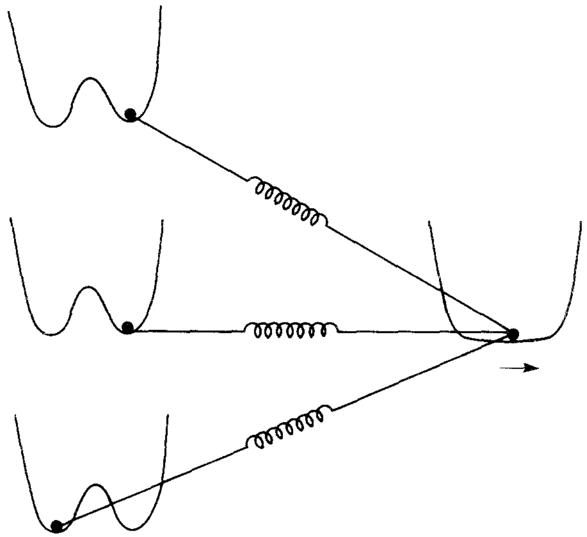

Fig. 3. Three wells in deeply bstable state, on the left, coupled to the one about to undergo transtion to bistability. The bottom well on the left is a "lost" vote.

The scheme we are describing here was originally proposed in refs. 13 as a method of using parametrically excited nonlinear circuits to do computation. It was adapted in ref. 14 to the mechanical potential wells invoked here, and extended to a proposal involving Josephson junctions, and permitting reversible computation, by Likharev.[15] These schemes were reviewed in ref. 16. If the wells are modulated slowly enough, then the particles in the wells will remain close to the Boltzmann distribution. The well forces in the deeply bistable state are kept large compared to the forces exerted by the coupling devices. These spring forces, in turn, are selected so as to bias the Boltzmann distribution enough to give a high probability for the particle being influenced ending up in the desired state. Thus, for any given choice of potentials and coupling devices, there will be a small residual error probability for ending up on the undesired side of the barrier. But by suitable design choices, this error can be made as small as desired. Note that I have referred to "springs," and these are shown in the figures. The word "spring" is intended as an abbreviation for a device that couples the relative displacement from the center of one well to the displacement from the center of the other well. The interaction of the information-bearing particle with its time-dependent potential is not a source of dissipation. Dissipation occurs as a result of the motion of the information-bearing particle against viscous forces. This source of dissipation can be made as small as desired by a sufficiently slow motion. An additional loss occurs if we need a many-to-one mapping, and this cannot be minimized by slow execution. Such information loss can occur, for example, if three stages, as shown in Fig. 3, influence one subsequent one. After the left-hand wells in Fig. 3 are restored to their monostable state, information about the existence and identity of a dissident well is no longer available. Likharev[15] recognized that such steps could easily be avoided. They will, in any case, not be needed in the subsequent examples.

Consider Fig. 4, in which information from the left well on the left side of the diagram is transferred to the well on the right side of the diagram. One can consider this as an example of a simple measurement: Which left-hand well is occupied? The information transfer process from one well to the next does not require a minimal dissipation. Does that mean that the total measurement cycle is without any minimal dissipation? The word *measurement* is, of course, ambiguous. There are many sophisticated theoretical discussions of measurement[17] which do not define the object of their discussion and do not say how to differentiate between a measurement and a dead horse. A typical ingredient, however, of a measurement is the resetting of the *meter* to a standardized state after it is decoupled from the system being measured. This is *destruction of information*, and does require minimal dissipation. This dissipation is enough to save the second law, in the operation of Maxwell's demon.[18] We do not need to look for additional dissipation in the information transfer step, from the system to the meter.[18] Many of the analyses of Maxwell's demon, however, did not understand the need for dissipation in the resetting step, and looked for it in the information transfer step. They found it, of course, but did not ask themselves whether their procedures were really minimally dissipative. This erroneous view was so deeply imbedded that the generation of this view, perceived as a correct view, even became a subject in the history of science.[19] For a typical uncritical scientific (rather than historical) discussion see ref. 20. For a more detailed discussion of measurement, see refs. 18 and 21.

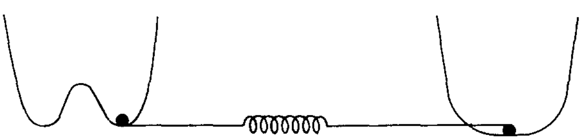

Fig. 4. Particle in deeply bistable potential well on the left, coupled to a particle on the right. The particle on the right is in a well about to undergo transition to bistability. The spring is symbolic; it is the relative displacements from the center of the respective potentials which are coupled.

Now let us, in Fig. 4, after the information transfer, restore the left-hand well to the monostable state of Fig. 2. This is not erasure; the information originally in the left-hand well still exists in the right-hand well. If we had restored the left-hand well to a monostable state without coupling to the right-hand well, then the information in the left-hand well would indeed have been destroyed and this would be associated with a minimal dissipation of $kT \log_e 2$. This dissipation, of $kT \log_e 2$, is associated with the fact that both the left-hand well and the right-hand well at the bottom of Fig. 2 are mapped into the same central final well. The phase space associated with the twofold choice of wells has to appear elsewhere, i.e., as a heating up of the irrelevant degrees of freedom. (The exact result $kT \log_e 2$ assumes equal likelihood of the "0" and "1" states.) But if we restore the left-hand well in the presence of the biasing force exerted by the right-hand well, then no irreversible event (beyond that required by the lateral motion of the particle, which can be minimized to any desired extent) takes place. After all, we can go back again to the bistable state for the well on the left of Fig. 4, and thus recover our earlier state.

After restoring the well on the left of Fig. 4 to its monostable state, and with our bit left in the well on the right-hand side of Fig. 4, we have moved the bit from one well to the next. One can think of these two wells as part of a long chain, and continue to move the bit to the right. Clearly, this is communication, and in contrast to the prevailing notions in that field,[22] done without having to spend $kT \log_e 2$ per transmitted bit.

The prevailing notions, requiring a dissipation of $kT \log_e 2$ per transmitted bit, are based on ref. 2, which states (and I am here paraphrasing a few sentences from ref. 23): "An important *special case* occurs when the noise is *added* to the signal and is independent of it (in the probability sense)." In that case, Shannon finds

$$C = W \log_2[N^{-1}(P+N)] \tag{1}$$

C is the channel capacity, W the bandwith, P the average received power, and N the average noise power. Thermal noise, for a classical transmission line with additive equilibrium noise, is given by $N = kTW$. Equation (1) yields a maximum for C/P at small P given by $C = P/(kT \log_e 2)$. Thus, $kT \log_e 2$ energy per bit is required, though it is not clear that this has to be dissipated. Equation (1) has an obvious plausibility. N determines the distance between distinguishable signals. The larger the N, the fewer the number of possible distinguishable messages. But communication does not need wave propagation. Mail is communication. Communication, furthermore, need not use degrees of freedom in which the effects of noise are added linearly to that of the signal in the transmission process and in the detection process used at the receiving end. There is an obvious advantage to the use of states of local stability to denote information, in contrast to the use of states in which noise causes unhindered diffusion from one state to another. This intuitively apparent distinction was discussed analytically in ref. 9. Furthermore, even if channel capacity expressions describe the energy required in the message, does the energy have to be dissipated?

There are a number of other examples which demonstrate that communication can be done with arbitrarily small dissipation per bit[6,21,23]; I will not repeat them here.

4. MEASUREMENT AND COMMUNICATION

In the preceding section, via modulated potential devices, I have argued Bennett's point that measurement requires unavoidable energy dissipation only when information is destroyed, i.e., in the meter-resetting step. I also argued that information transmission requires no minimal energy dissipation. This conclusion was implicit in earlier discussions of the computational process, e.g., in ref. 14, but has now been made totally explicit. All of our three areas of concern, computation, measurement, and communication, can be lumped and characterized by the assertion that unavoidable minimal energy dissipation is required only when information is discarded.

5. SUPPLEMENTARY REMARKS

I allude here briefly to two further notions not discussed in detail in the preceding account. We have learned that supposed computer limitations related to kT, and related to the uncertainty principle,[6,21] can be circumvented. There are other limitations, however, related to such cosmological questions as: How many degrees of freedom in the universe can we really couple together into an effective computing structure? If that is limited, then continuum mathematics with its need for unlimited sequences of operations is not executable, and is therefore not a satisfactory basis for physical laws. This theme was introduced in ref. 24 and revisited on a number of later occasions, e.g., ref. 7.

A second theme not taken up in this discussion relates to the ultimate sources of friction, irreversibility, and noise. Why do we have friction for the tire on the road, but not for the electron in the benzene molecule? I believe that this can, in turn, be tied to the invalidity of continuum mathematics to which I have just alluded. This admittedly very speculative and exploratory suggestion has been presented in refs. 21 and 25.

REFERENCES

1. J. C. Maxwell, *Theory of Heat* (Longmans, Green, London, 1871), p. 308.
2. C. E. Shannon, *Bell Syst. Techn. J.* **27**:397, 623 (1984).
3. M. D. Sturge, in *No Way*, P. J. Davis and D. Park, eds. (Freeman, New York, 1987), p. 111.
4. P. W. Smith, *Phil. Trans. R. Soc. Lond. A* **313**:349 (1984); J. D. Meindl, *IEEE Trans. Elect. Devices* **ED-31**:1555 (1984); J. D. Meindl, in *Cutting Edge Technologies* (National Academy Press, Washington, D.C., 1984), p. 5.
5. L. B. Levitin, *Int. J. Theor. Phys.* **21**:299 (1982); H. J. Bremermann, *Int. J. Theor. Phys.* **21**:203 (1982).
6. R. Landauer, *Int. J. Theor. Phys.* **21**:283 (1982).
7. R. Landauer, *Found. Phys.* **16**:551 (1986).
8. L. Brillouin, *Science and Information Theory* (Academic Press, New York, 1956).
9. R. Landauer, *IBM J. Res. Dev.* **5**:183 (1961).
10. C. H. Bennett, *IBM J. Res. Dev.* **17**:525 (1973).
11. J. A. Wheeler and W. H. Zurek, eds., *Quantum Theory and Measurement* (Princeton University Press, Princeton, New Jersey, 1983), p. 782.
12. K. Obermayer, G. Mahler, and H. Haken, *Phys. Rev. Lett.* **58**:1792 (1987); J. Maddox, *Nature* **327**:97 (1987); S. Hawking, *New Sci.* **115**:46 (July 9, 1987); O. Costa de Beauregard, *Found. Phys.*, to be published; J. Rothstein, in *Performance Limits in Communication*, J. K. Skwirzynski, ed. (Kluwer Academic, Dordrecht, in press); W. Porod, R. O. Grondin, D. K. Ferry, and G. Porod, *Phys. Rev. Lett.* **52**:232 (1984); W. Porod, R. O. Grondin, D. K. Ferry, and G. Porod, *Phys. Rev. Lett.* **53**:1206 (1984); C. Mead and L. Conway, *Introduction to VLSI Systems* (Addison-Wesley, Reading, Massachusetts, 1980), Chapter 9; H. M. Hastings and S. Waner, *Biosystems* **17**:241 (1985); E. M. Drogin, *Defense Electron.* **1986**(March):31.
13. J. von Neumann, *Non-linear capacitance or inductance switching, amplifying and memory organs*, U.S. Patent 2,815,488; E. Goto, *J. Elec. Commun. Eng. Jpn.* **38**:770 (1955).
14. R. W. Keyes and R. Landauer, *IBM J. Res. Dev.* **14**:152 (1970).
15. K. K. Likharev, *Int. J. Theor. Phys.* **21**:311 (1982); K. K. Likharev, S. V. Rylov, and V. K. Semenov, *IEEE Trans. Magn.* **21**:947 (1985).
16. R. Landauer, in *Der Informationsbegriff in Technik und Wissenschaft*, O. G. Folberth, C. Hackl, eds. (R. Oldenbourg, Munich, 1986), p. 139.
17. D. M. Greenberger, ed., *New Techniques and Ideas in Quantum Measurement Theory* (New York Academy of Science, 1986); M. Roth and A. Inomata, eds., *Fundamental Questions in Quantum Mechanics* (Gordon and Breach, New York, 1986).
18. C. H. Bennett, *Int. J. Theor. Phys.* **21**:905 (1982); C. H. Bennett, *Sci. Am.* **255**:108 (1987); C. H. Bennett, *IBM J. Res. Dev.* **32**:16 (1988).
19. E. E. Daub, *Hist. Phil. Sci.* **1**:213 (1970).
20. O. Costa de Beauregard, *Time, The Physical Magnitude* (Reidel, Dordrecht, 1987), p. 153.
21. R. Landauer, in *Signal Processing*, S. Haykin, ed. (Prentice-Hall, Englewood Cliffs, New Jersey, in press).
22. H. Marko, *Kybernetik* **2**:274 (1965); F. T. S. Yu, in *Optics and Information Theory* (Wiley, New York, 1976); J. R. Pierce, E. C. Posner, and R. Rodemich, *IEEE Trans. Inf. Theory* **27**:61 (1981).
23. R. Landauer, *Appl. Phys. Lett.* **51**:2056 (1987).
24. R. Landauer, *IEEE Spectrum* **4**:105 (1967).
25. R. Landauer, *Z. Phys. B* **68**:217 (1987).