# Reversibility and Stability of Information Processing Systems

W. H. Zurek
*Theoretical Astrophysics, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, and Institute for Theoretical Physics, University of California, Santa Barbara, California 93106*
(Received 26 April 1984)

Classical and quantum models of dynamically reversible computers are considered. Instabilities in the evolution of the classical "billard ball computer" are analyzed and shown to result in a one-bit increase of entropy per step of computation. "Quantum spin computers," on the other hand, are not only microscopically, but also operationally reversible. Readoff of the output of quantum computation is shown not to interfere with this reversibility. Dissipation, while avoidable in principle, can be used in practice along with redundancy to prevent errors.

PACS numbers: 89.70.+c, 05.20.−y, 06.50.−x, 89.80.+h

Szilard[1] used the second law to establish that a bit of information can be acquired at the price of no less than a bit of entropy increase. Von Neumann[2] and Brillouin[3] have extrapolated Szilard's conclusion to conjecture that a similar price must be paid for a single step of information processing. Indeed, Landauer[4] has demonstrated that computer programs which contain fundamentally, *logically irreversible* steps lead unavoidably to dissipation of

$$\Delta F \sim kT \ln 2 \tag{1}$$

of free energy per step of typical computation.[5] The above estimate, while orders of magnitude below the dissipation levels of present-day computers, raises the following question: Is there a fundamental price which must be paid for processing of information? Bennett[6,7] has been able to settle this issue by devising models of physical systems which compute, and yet dissipate arbitrarily small amounts of free energy. Programs for such computers avoid dangers of logical irreversibility by recording all the steps taken along the path of the computation. Indeed, every program can be recast in such a logically reversible manner.[6,8] However, Bennett's models are thermodynamic in nature; random Brownian motion pushes the computation forward, and reversibility is achieved only in the limit of infinitesimally slow computation. One is then led to inquire: Is it possible to compute arbitrarily fast with arbitrarily small dissipation? Below we shall analyze classical and quantum dynamically reversible computer models, which are needed to answer this question.

Fredkin and Toffoli[8,9] proposed the *billiard ball computer* (bbc). It uses elastic collisions between hard-core disklike particles, "billiards," moving on a two-dimensional "table" to perform a computation. The dynamics of the bbc is no doubt, microscopically reversible. So is, however, dynamics of the Boltzmann gas. Is it possible that the computation of the bbc is also effectively irreversible?

Trajectories of the individual particles in a run of a bbc must follow the "grid" of the cubic lattice. This ideal case could be achieved only at the expense of an infinite initial free energy, as the initial conditions for each billiard would have to be set with infinite precision.[10] Let us therefore suppose that initial momenta and positions are set with arbitrarily small errors $\Delta p_x = \Delta p_y = \Delta p$ and $\Delta x = \Delta y = \Delta q$, which, for simplicity, are assumed to be identical for all particles. For a bbc with $N$ particles in two dimensions the entropy of this initial configuration is given by the logarithm of the volume subtended by the $4N$-dimensional error box in the phase space:

$$V^{(0)} = \prod_{j=1}^{N} \Delta p_j^2 \Delta q_j^2 = (\Delta p \, \Delta q)^{2N}, \tag{2}$$

$$S = k \ln( V^{(0)}/h^{2N}) = 2Nk \ln(\Delta q \, \Delta p/h). \tag{3}$$

Error boxes of individual particles evolve as a consequence of collisions: A small error in angle increases with each particle-particle collision as

$$\Delta\phi' \simeq \Delta\phi(1 + l/r), \tag{4}$$

where $l$ is the free path and $r$ is the hard-core radius.[4] The uncertainty of the momenta of the billiards grows by the same factor. The volume of the phase space subtended by *individual* error boxes after $n$ collisions will then be given by

$$V^{(n)} = V^{(0)} \prod_{i=1}^{n} (1 + l_i/r). \tag{5}$$

The above estimate deliberately neglects *correlations* between billiards which are introduced in the course of the collisions.[11] In a typical bbc $l$ is at least of the order of $r$. The increase of entropy in a single collision, which can be regarded as a "single step of computation," is then

$$\Delta S \cong k \ln(1 + l/r) \cong k \ln 2. \tag{6}$$

This limit is equivalent to Eq. (1) for $T = 0$.[12] It is worth stressing that above it was arrived at through reasoning very different from that of Szilard.

Errors introduced by the coupling to additional degrees of freedom or by an imperfect realization of interactions have been disregarded so far. We were solely concerned with "software errors" in preparation of the initial state—the "program" for the bbc. There appears to be, however, no reason why, at least within the context of classical mechanics, such "hardware errors" cannot be made arbitrarily small. Moreover, as long as they are small, they enter only as an additional source of the already discussed software errors.

The origin of the increase of entropy in the dynamically reversible bbc is the same as in the Boltzmann gas. Thus, even though the volume of the system in the phase space is constant, in accord with Liouville's theorem, the volume traced out by the error boxes of individual particles increases exponentially. It is important to stress that the estimate for the entropy produced per collision, Eq. (6), is derived in the limit of a small initial error. Therefore, the dissipation rate of approximately one bit per step of computation cannot be made smaller even for an infinitesimally small initial error. Instability of the trajectory poses a very serious threat to the operational reversibility of the bbc. It seems unlikely that it could be eliminated by natural modifications of its design.[13]

One may object to these conclusions by noting that the simultaneous reversal of all the velocities would force the bbc to return from the large effective volume $V^{(n)}$ into a much smaller, initial $V^{(0)}$. The problem with this objection is of the operational nature: It cannot be implemented without the additional increase of entropy. To see this, consider a reversal which is to be accomplished by means of flat "mirrors," positioned with the accuracy $\Delta x_j^{(R)}$ and aimed perpendicular to the *actual* velocities of the particles with the accuracy $\Delta\phi_j^{(R)}$. The accuracy of the reversal can then be characterized by the volume of the error box,

$$V_R = \prod_{j=1}^{N} (\Delta p_j^{(R)} \Delta q_j^{(R)})^2,$$

where $\Delta p_j^{(R)} = p\,\Delta\phi_j^{(R)}$. If the system is to return into the volume $V^{(0)}$ after $n$ "reverse" collisions, one must at least require that

$$V_R < V^{(0)}(V^{(0)}/V^{(n)}). \tag{7}$$

However, after $n$ collisions positions and velocities of the individual particles in the system are known only with the accuracy characterized by $V^{(n)}$. Therefore, in order to position mirrors with an accuracy called for by the restrictions on $V_R$, one must *measure* momenta and positions of particles. Such measurement, by Szilard's argument, requires an expenditure of entropy of no less than

$$\Delta S = k \ln(V^{(n)}/V_R) = 2k \ln(V^{(n)}/V^{(0)}). \tag{8}$$

Thus, even though the system can be forced to return into the initial error box, the reversal can be accomplished only at the expense of entropy production which is precisely such as to make dissipation of less than a bit per collision, Eq. (6), out of the question. It is perhaps interesting to note that a similar argument could have been used by Boltzmann in his famous defense of the $H$ theorem against some of the time-reversal arguments of Loschmidt.

Quantum computers were suggested and analyzed by Benioff[14] and Feynman.[15] Typically, they consist of $N$ interacting two-state (spin-$\frac{1}{2}$-like) systems. We shall call them *quantum spin computers* (qsc's). The initial "input" state of a qsc is a quantum binary string, e.g.,

$$|\Phi_{\text{in}}\rangle = |0\rangle_1 |1\rangle_2 \cdots |0\rangle_N.$$

States $|0\rangle$ and $|1\rangle$ span Hilbert spaces of individual binary elements. Computation is accomplished through the unitary evolution:

$$|\Phi_{\text{out}}\rangle = \exp\left(-i\int_0^T \hat{H}\,dt\right)|\Phi_{\text{in}}\rangle = \hat{U}(T)|\Phi_{\text{in}}\rangle.$$

Above we have set $\hbar = 1$. The "output" state of a qsc also has a binary string structure.[16] In the course of the computation, intermediate states are, in general, superpositions of binary strings. In most of the designs they recover the product structure in the $\{|0\rangle, |1\rangle\}$ basis with a certain periodicity $\tau$.

As in the case of a bbc, one can inquire about the effect of small software errors in $|\Phi_{\text{in}}\rangle$, as well as about the role of hardware errors in the Hamiltonian. An additional question, which does not arise for classical computers, concerns the readoff of $|\Phi_{\text{out}}\rangle$. Will the measurement of the state of the quantum computer lead to obstacles in making computation reversible? Fortunately, the answer to this question is straightforward. Both $|\Phi_{\text{in}}\rangle$ and $|\Phi_{\text{out}}\rangle$ have simple product structures in a properly designed qsc.[17] Therefore, the measurement of the string of $|0\rangle$'s and $|1\rangle$'s will not result in the "reduction of the wave packet."[18] Consequently, readoff does not interfere with the reversal.

Both preparation of $|\Phi_{\rm in}\rangle$ and readoff of $|\Phi_{\rm out}\rangle$ may introduce "software errors." Consider, for example,

$$|\Phi_{\rm in}\rangle = |0\rangle_1 |1\rangle_2 \cdots \times [(1-|\alpha|^2)^{1/2}|1\rangle_k + \alpha|0\rangle_k] \cdots |0\rangle_N. \qquad (9)$$

A small software error $(\sim\alpha)$ in the encoding of the $k$th bit of the program causes only an equally small error in the output:

$$|\Phi'_{\rm out}\rangle = (1-|\alpha|^2)^{1/2}|\Phi_{\rm out}\rangle + \alpha|\Psi\rangle, \qquad (10)$$

where $|\Psi\rangle$ is an output of the computation which starts with the state differing from $|\Phi_{\rm in}\rangle$ in the $k$th bit. This follows from the linearity of $\hat{U}(T)$ and can be generalized to the case when there are errors in more than one bit:

$$|\Phi'_{\rm in}\rangle = \left(1-\sum_{i=1}^{N}|\alpha_i|^2\right)^{1/2}|\Phi_{\rm in}\rangle + \sum_{i=1}^{N}\alpha_i|\Phi_i\rangle. \qquad (11)$$

Above, $|\Phi_i\rangle$ spans the Hilbert space of all possible states. Again, by linearity,

$$|\Phi'_{\rm out}\rangle = (1-\sum_{i=1}^{N}|\alpha_i|^2)^{1/2}|\Phi_{\rm out}\rangle + \sum_{i=1}^{N}\alpha_i\hat{U}(T)|\Phi_i\rangle. \qquad (12)$$

Software errors are *conserved* in a qsc. Therefore, quantum computers appear to be reversible in a more fundamental sense than the classical bbc.

Let us now consider hardware errors of a qsc. Suppose that the actual Hamiltonian $\hat{H}'$ differs from the ideal $\hat{H}$ by a (small) error $\hat{h}$:

$$\hat{H}' = \hat{H} + \hat{h}. \qquad (13)$$

We can estimate the influence of the hardware error $\hat{h}$ by calculating

$$|\langle\Phi_{\rm out}|\Phi'_{\rm out}\rangle|^2 = |\langle\Phi_{\rm in}|e^{i\hat{H}t}e^{-i(\hat{H}+\hat{h})t}|\Phi_{\rm in}\rangle|^2 = 1-\lambda^2t^2+\ldots, \qquad (14)$$

where

$$\lambda^2 = \langle\Phi_{\rm in}|\hat{h}^2|\Phi_{\rm in}\rangle - |\langle\Phi_{\rm in}|\hat{h}|\Phi_{\rm in}\rangle|^2.$$

Above, we have assumed that $\lambda t \ll 1$ and that higher-order terms with powers of $|\langle\Phi_{\rm in}| \times [\hat{H},\hat{h}]|\Phi_{\rm in}\rangle|$ are still smaller. (Indeed, this is the sense in which $\hat{h}$ ought to be small.)

My calculation demonstrates that, as in the "watchdog effect,"[19] the probability of the computational error induced by a small hardware error $\hat{h}$ increases only quadratically with time. This suggests a way to stabilize quantum computation with measurements of the intermediate states of the quantum computer at these instants when its state has a product structure. For, the probability to find the state perturbed from what it ought to be is diminished by $n$ measurements as $[1-(\lambda t/n)^2]^n > 1-(\lambda t)^2$ as long as $\lambda t < 1$.[20]

The computation of a qsc is stable and does not increase entropy in a manner which appears so inevitable in classical many-body systems. This qualitative distinction between a bbc and a qsc can be traced back to the structure of their respective phase spaces. The Hilbert space of a qsc is discrete. Each orthogonal state corresponds to a distinct possible input. There is literally no room for the errors. This rules out instabilities encountered in a classical bbc, where the ideal input states are only a subset (of measure zero for $\hbar \to 0$) of all the possible states. One can, of course, design a quantum computer which in the classical limit would "recover" all the instabilities of a bbc. Needless to say, this would not be an optimal design of a reversible quantum computer.

However, there is a useful way to employ additional room in the phase space in both quantum and classical computers. Consider an ensemble of $N$ identical qsc's which perform the same computation. The *redundancy* of such an ensemble can be used to reset states of each qsc to the "average" state every few steps of computation. This will reduce the random errors by $N^{1/2}$. An equivalent strategy can be also employed in the bbc. There, the state can be periodically measured to verify how far it has strayed from the ideal trajectory. This is possible because of the redundancy: Trajectories which correspond to different "programs" are far apart in the phase space and can be distinguished. Errors can be reduced by restarting the computer on the appropriate trajectory. However, measurements and resetting require dissipation. The minimal cost of such an operation is expected to follow from Eq. (6). While this strategy does not accomplish reversibility, the way in which it combines redundancy with dissipation is reminiscent of the strategies employed in the "real world" computers.

The present considerations support the conclusions of Bennett[6,7] and Landauer[21]: Reversible computation is compatible with the laws of physics. The quantum spin computers of Benioff and Feynman prove that it can be arbitrarily fast and dissipationless, even though practical realizations of qsc's are likely to be forbiddingly difficult. The billiard ball computer of Fredkin and Toffoli—although plagued by instabilities—provides additional argument in favor of the fundamentally reversible nature of the computation. For, if the computation were, for some reason, fundamentally irreversible, it should be impossible to map it onto a system with microscopically reversible dynamics.[4] Ultimately, Szilard's limit, Eq. (1), does not apply because, in contrast to a measuring apparatus, the *computer cannot be used to gain new information.* However, *dissipation can be used to correct errors* in a strategy which also involves redundancy. In this practical context the analogy with the noisy channel, and, therefore, the limit set by Eqs. (1) and (6), can be meaningful.

This paper was stimulated by discussions with the participants of the workshop on the physics of the information-processing structures held at the Aspen Center for Physics. The research was supported in part by the National Science Foundation under Grant No. PHY77-27084, supplemented by funds from the National Aeronautics and Space Administration.

[1]L. Szilard, Z. Phys. **53**, 840 (1929).

[2]J. Von Neumann, in *Theory of Self-Reproducing Automata,* edited by A. W. Burks (Univ. of Illinois Press, Urbana, 1966), p. 66.

[3]L. Brillouin, *Science and Information Theory* (Academic, London, 1956), pp. 261–264.

[4]R. Landauer, IBM J. Res. Dev. **14**, 152 (1961).

[5]J. D. Bekenstein, Phys. Rev. Lett. **46**, 623 (1981), and, even more recently, W. Porod, R. O. Grondin, D. K. Ferry, and G. Porod, Phys. Rev. Lett. **52**, 232 (1984), have also argued in favor of a similar estimate. Their arguments are based on the analogy between the computation and a noisy communication channel. For a critical assessment of this analogy see, e.g., R. Landauer, Int. J. Theor. Phys. **21**, 283 (1982).

[6]C. H. Bennett, IBM J. Res. Dev. **17**, 525 (1973).

[7]C. H. Bennett, Int. J. Theor. Phys. **21**, 905 (1982).

[8]E. Fredkin and T. Toffoli, Int. J. Theor. Phys. **21**, 219 (1982).

[9]T. Toffoli, Int. J. Theory. Phys. **21**, 165 (1982).

[10]Even if practical difficulties could not prevent one from making $\Delta q$ and $\Delta p$ small, quantum indeterminacy would impose a limit $\Delta q \, \Delta p \geqslant \hbar/2$. Therefore, some error in setting up an initial "program" of the bbc is inevitable.

[11]N. S. Krylov, *Works on Foundations of Statistical Physics* (Princeton Univ. Press, Princeton, 1979), pp. 196–210.

[12]If the bbc were allowed to evolve for a sufficiently long time in isolation, individual particles would assume Maxwell-Boltzmann velocity distribution with $kT = p^2/2m$, where $p$ is the initial momentum and $m$ is the mass of each billiard. Therefore, $k \ln 2$ increase of entropy can be said to correspond to $\sim kT \ln 2$ increase of free energy per step of computation. This assertion should be regarded only as a suggestive, but somewhat misleading analogy. The billiard ball computer is very far from equilibrium.

[13]"Unnatural" modifications including "square balls" and holonomic constraints, are possible (Ref. 7). This proves an important point: Classical physics can be, at least in principle, used as a basis for the design of operationally reversible, dynamical computers. Their physical realizations are, however, bound to introduce additional degrees of freedom, e.g., rotation of the squares and/or vibrations around the "physical" constraint in the above example. This leads one to inquire whether all the "natural" classical systems which are complex enough to be capable of computation are also inherently unstable in a bbc-like manner.

[14]P. A. Benioff, Int. J. Theor. Phys. **21**, 177 (1982), and Phys. Rev. Lett. **48**, 1581 (1982).

[15]R. P. Feynman, private communication.

[16]In some qsc's (in particular, in Feynman's computer) the design allows parts of the computer which are not essential for the computation to be in some general superposition state both at the beginning and at the end of the computation.

[17]Unless, for a certain group of input bits, one adopts a superposition $\alpha_j|0\rangle_j + \beta_j|1\rangle_j$, to perform a "Monte Carlo" calculation in which a random choice is made during the readoff, after the computation is already done.

[18]Interaction Hamiltonians which can accomplish such a nondemolition readoff of a spin can be found, e.g., in W. H. Zurek, Phys. Rev. D **26**, 1862 (1981).

[19]B. Misra and E. C. G. Sudarshan, J. Math. Phys. **18**, 756 (1977); K. Kraus, Found. Phys. **11**, 547 (1981), and references therein.

[20]This strategy may involve an additional expense of free energy because of the stringent timing requirements (readoffs at just the right instants). It will also be difficult to implement in Feynman's computer because, in the course of the computation, it is always in a superposition state (Ref. 16). Both of these difficulties can be relaxed by letting the computer produce intermediate outputs which can be stabilized with far less concern for timing.

[21]Landauer, Ref. 5.