Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Electronic Notes in Theoretical Computer Science 253 (2010) 3–16
Electronic Notes in Theoretical Computer Science
www.elsevier.com/locate/entcs

# From Such Simple a Beginning: The Momentous Consequences of Physics' Microscopic Reversibility for Communication and Computation—and Almost Anything Else

Tommaso Toffoli[1]

*Electrical and Computer Engineering*
*Boston University*
*Boston MA USA*

**Abstract**

Darwin concludes *The Origin of Species* with a splendid one-phrase poem,

> From so simple a beginning
> endless forms most beautiful and most wonderful
> have been, and are being, evolved.

Darwin's "simple beginning" may be identified, in today's terminology, with *dissipation*—evolution's basic fuel. All the rest is commentary—or, more precisely, corollary.

One can aptly apply Darwin's phrase to another kind of "simple beginning," from which as well "endless forms most beautiful and most wonderful have been, and are being, evolved." What I have in mind is a concept that is apparently the very antithesis of dissipation, namely, physics' fundamental assumption of *invertibility*—or "microscopic reversibility." To paraphrase Dobzhansky, no sensible step can be taken today in information, communication, and computer sciences, as well as in fundamental physics, except in the light of invertibility.

*Keywords:* invertibility, irreversibility, computation, dynamics, thermodynamics, entropy, second law of thermodynamics

## 1 Introduction

Darwin's "simple beginning" in this paper's abstract may be identified, in today's terminology, with *dissipation*—evolution's basic fuel. Evolution is, in essence, an advanced form of dissipative cascade—one which exhibits long-lived historical symmetry breakings, or, in Bennett's terms, displays great *logical depth*[4].

[1] tt@bu.edu


doi:10.1016/j.entcs.2010.02.002

Even though the laws of microscopic physics are presumed to be strictly invertible—that is, distinct microscopic trajectories forever remain distinct—dissipation means that, as time progresses, different *macroscopic* states tend to merge into a same macroscopic state. If you pour equal amounts of cream into two cups of tea, the coarse swirls that you initially get in the two cups may look quite different. As time progresses and the swirls churn about and break up into finer and finer swirls, the two macroscopic patterns will still look different if compared pixel by pixel, but overall they will tend to develop a similar fractal *texture*—the same statistics for the frequency of swirls of a given size, for the two-point density correlations of swirls, and so forth. The *dissipative cascade* by which coarse nonhomogeneities gradually reshuffle themselves into finer ones, yielding ever more complex structure and subtler correlations, is a prototype of Darwinian evolution—literally, a universe in a cup of tea. But, inexhorably, the bulk of the distribution of feature sizes shifts towards smaller and smaller swirls, sweeping most detail downwards to the mesoscopic level, until finally the only differences that are left between the states of the two cups are at the microscopic level. From the macroscopic viewpoint, the two different initial states have converged to a unique final state of *thermodynamic equilibrium*—in Kelvin's words, the "heat death of the universe."

It this scenario, the macroscopic progression is clearly *irreversible*—different states flow into one and the same state. Moreover, entropy—intuitively, the amount of disorder—grows monotonically until it attains the maximum allowed by the scenario's invariant constraints. [2]

Let us turn our attention now to a drastically stylized kind of universe, namely, John Conway's well-known game of LIFE[11], whose laws are, like those of our physics, homogeneous, isotropic, time-invariant, deterministic, and locally acting (no "action at a distance")—in sum, a discrete version of a field theory. (This kind of systems, called *cellular automata*, were independently proposed by John von Neumann[23] and Konrad Zuse[25] precisely to establish a common ground between automata theory and physics.) Let us initialize LIFE to a state of high but not maximal entropy—say, a region of mostly 1's ("cream") next to one of mostly 0's ("tea")—and set it going. What kind of behavior will we observe? (To help intuition, Fig. 1 provide a time-lapse-photography sequence of this evolution.)

Specifically, to compare this domesticated scenario with the genuine physical scenario discussed before, we may ask of each of them

(i) May complex structures and textures emerge at some point (possibly to be readsorbed)?

(ii) Is the macroscopic progression irreversible?

(iii) Is entropy monotonically increasing? and

(iv) Is the microscopic law invertible?

[2] The number of molecules of each kind, the total energy (if the system is isolated) or the temperature (if it is in contact with a thermostatic environment), the system's volume or pressure, etc.

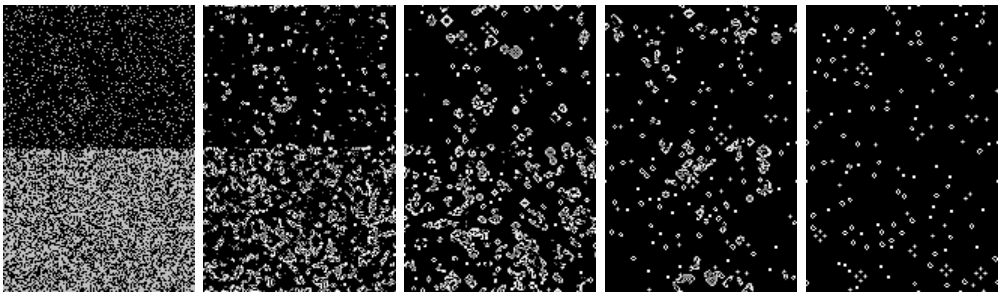

Fig. 1. LIFE started from a state of high, but not maximal, entropy; $t = 0,\ 2,\ 10,\ 258,\ 1054$.

The answers, in the two cases, are

| question | teacup | LIFE |
|---|---|---|
| 1 complex structures may emerge | YS | YS |
| 2 irreversible macroscopics | YS | YS |
| 3 monot. increasing entropy | YS | NO |
| 4 microscopic law invertible | YS | NO |

Note that, for questions 1 and 2, whether or not the microscopic law is invertible doesn't appear to make a critical difference; on the other hand, the answer to question 3 seems to go hand-in-hand with that to question 4. In fact, we shall see that the most symptomatic evidence for the second law of thermodynamics, namely, the very *irreversible* decay associated with the monotonical increase of coarse-grained entropy, is a characteristic indicator (in the sense of "if and only if") of microscopic *reversibility*. Paradoxically, it is the very "feature" of an invertible system to be unable to forget its initial microscopic state that inexhorably leads it—Alzheimer-wise—to progressively forget its initial macroscopic state! But I'm getting ahead of myself.

For the moment, let us just remark that the two kinds of system show remarkable differences in their life career. With an invertible dynamics ("teacup"), a maximally disordered initial state will forever remain so. With a noninvertible dynamics ("LIFE"), on the other hand, out of a maximally disordered initial state there will often spontaneously emerge texture (spatial statistical correlations) which will evolve (possibly growing macroscopic features of unbounded size) and ultimately converge to a "limit texture" through a transient that may range from short and banal to long and full of surprises—a veritable "history." In the invertible case, history can occur only if "fueled" by initial disequilibrium—and this fuel is obviously provided by the system's *state*. In the noninvertible case, fuel for history is also provided by the *dynamics* itself, which makes LIFE easier—if you pardon the pun. (Whether "life" achieved at this price can be as promising will have to be the matter for another paper.)

One could aptly apply Darwin's poetic phrase quoted above to another kind

of "simple beginning," from which as well "endless forms most beautiful and most wonderful have been, and are being, evolved." As you may guess, what I have in mind is a concept that is apparently the very antithesis of dissipation, namely, physics' fundamental assumption of *microscopic reversibility*. (But aren't opposites often found to be the two sides of the same coin?) As we shall see, microscopic reversibility—that is, the *strict invertibility* of a process—turns out to be an extraordinarily productive constraint, whose impact is particularly evident in information, communication, and computer sciences as well as in fundamental physics. To paraphrase Dobzhansky[9], no sensible step can be taken today in those disciplines except in the light of invertibility.

The backbone of chemistry is the conservation of atoms; of circuit theory, that of charge; and of hydrodynamics, that of water; in a similar manner, *information theory* and *analytical dynamics* are essentially suites of variations on the theme of "conservation of information." It's not that the composer is strictly forbidden to depart from that theme—that mechanics or information theory only deal with invertible systems—but the music tends to get keyless and boring unless the invertibility theme is recalled often, by imitation or by contrast.

## 2 Text compression, cryptography, and all that

For starters, let us consider a number of techniques that more or less knowingly we have all been taking advantage of.

When we send an attachment with an email message, it is usually compressed in order to save bandwidth. Most ordinary texts can easily be squeezed down to between one-third and two-thirds of their original size. What is essential is that this transformation be *reversible*, so that, by using an inverse transformation, the recipient may able to reconstruct the original *verbatim*. We speak in this case of *lossless* compression—since, intuitively, no "information" is lost. But in this scenario it would be hard to establish an independent meaning for that word in isolation; all we can say is that the entire phrase "no information is lost" just means the same as "the transformation is reversible."

In cryptography, the purpose is to transmit a message in an encoded form, so that its contents will be hidden to everyone—including the material bearer of the message—but recoverable in its entirety by a selected recipient. So, even though the encoding must be reversible—as in the case of lossless compression—the means to carry out the reverse process must be denied to anyone but the recipient. In our age, cryptography is used on an industrial scale, as when you pay your phone bill through the Internet. So encoding and decoding algorithms are standard and essentially *public-domain*. All that remains confidential is a *key*, that is, a small *ad hoc* piece of passive data—a *parameter*—that is fed to the algorithm to customize it for an individual client or transaction.

To insure reversibility, the decoding key should be matched to the encoding key, and thus, at least in principle, reconstructible from it. Yet, like the access code for

the office copier, it may be hard to keep an encoding key confidential for long. From what we just said, this would seem to imply that once the encoding key is known the decoding key as well would soon cease to be a secret. The problem is gotten around *in practice* by making the relation between encoding key and decoding key, if not irreversible (which would be a contradiction in terms), at least *extremely hard to reverse*[16].

There is a parallel here between the *reversibility in principle* postulated for microscopic physics and the *irreversibility de facto* of macroscopic phenomena, which we are continually reminded of by the second law of thermodynamics. We'll have to say more about this in §4.

## 3 Invertibility and information conservation

From a category-theory viewpoint (see [12]), invertibility is in essence nothing but "informationlossless" (ach!)—that is, intuitively, the capability for a dynamical process to transform the state of a system without losing information that would allow one to reconstruct the previous state. This connection between a *transformation*, which operates on the individual microscopic states of a system, and *information*, which is a quantitative attribute of a "state of knowledge," [3] is conceptually subtle, and will be elucidated below.

Briefly, a *dynamical system* consists of set $X$ of *states* and a rule, called transition function or *dynamics*, that to each element of $X$ associates one element of that same set called its *successor* (this is indicated, in Fig. 2, by a directed arc). A dynamics is *invertible* if each state has exactly one predecessor, [4] as in Fig. 2b.

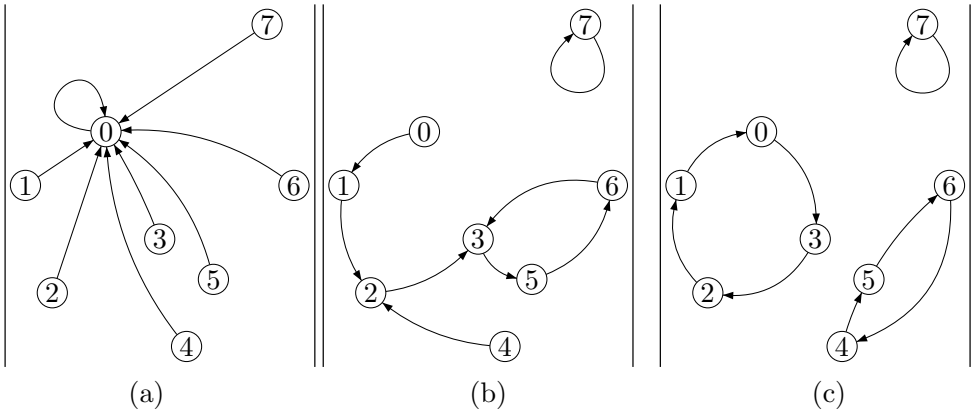


Fig. 2. Three dynamics on the same set of eight states. Dynamics (b)—a typical case—is *noninvertible*, as there are states, such as 0 and 4, that have no predecessors, and others, such as 2 and 3, that have more than one predecessor. (One can say that states 1 and 4, for example, "merge" into state 2.) Dynamics (a) is an extreme case of noninvertibility, as *all* states merge into a single one. At the opposite extreme, dynamics (c) is *invertible*, as there are *no* merges—each state has one and only one predecessor.

[3] Namely, one specifying how likely it is, for any of the conceivable states, that the system be in that state.

[4] Here we restrict our argument to *finite* systems. To properly deal with infinite systems one would have to introduce technicalities from topology and measure theory, arriving, in that more general context, at substantially the same conclusions.

In Fig. 2, suppose we express our knowledge about the state of the system at a certain moment by a *probability distribution* $P = \langle p_1, p_2, \ldots p_8 \rangle$; such a state of knowledge is often called a "macroscopic state," to distinguish it from the specification of a *single element* of the system, called "microscopic state."

A dynamics over a set $X$ automatically induces a dynamics over the (much larger) set of all distributions over $X$. If all we knew about the initial state was, for instance, that "it had an even label"—0, 2, 4, or 6—our lack of information would be measured by the *entropy* $X$ of the distribution $P = \langle \frac{1}{4}, 0, \frac{1}{4}, 0, \frac{1}{4}, 0, \frac{1}{4}, 0 \rangle = \frac{1}{4}\langle 1, 0, 1, 0, 1, 0, 1, 0 \rangle$, namely,

$$X = \sum_{i=0,\ldots,7} p_i \log_2 \frac{1}{p_i} = 4 \cdot \tfrac{1}{4} \log_2 4 = 2\,\text{bit}.$$

Table (1) shows how the distribution $P$ (we have left out the normalization factor 1/4) evolves with dynamics (a), (b), and (c) as we "turn the crank":

| $t$ | $P_a$ | | $P_b$ | | $P_c$ | |
|---|---|---|---|---|---|---|
| | 0123 | 4567 | 0123 | 4567 | 0123 | 4567 |
| 0 | 1.1. | 1.1. | 1.1. | 1.1. | 1.1. | 1.1. |
| 1 | 4... | .... | .112 | .... | .1.1 | 11.. |
| 2 | 4... | .... | ..11 | .2.. | 1.1. | .11. |
| 3 | 4... | .... | ...1 | .12. | .1.1 | 1.1. |
| 4 | 4... | .... | ...2 | .11. | 1.1. | 11.. |
| 5 | 4... | .... | ...1 | .21. | .1.1 | .11. |
| 6 | 4... | .... | ...1 | .12. | 1.1. | 1.1. |
| 7 | 4... | .... | ...2 | .11. | .1.1 | 11.. |
| 8 | 4... | .... | ...1 | .21. | 1.1. | .11. |
| 9 | 4... | .... | ...1 | .12. | .1.1 | 1.1. |
| 10 | 4... | .... | ...2 | .11. | 1.1. | 11.. |
| 11 | 4... | .... | ...1 | .21. | .1.1 | .11. |
| | ⋮ | | ⋮ | | ⋮ | |

(1)

Note that, as a (microscopic) state is turned into a new state by the dynamics, the weight associated with it in the distribution $P$ "flows" with it along the arcs of Fig. 2; whenever two arcs merge, the contents of the corresponding two "bins" of the distribution are poured together into a single bin—so that the distribution itself gets coarser and its entropy accordingly *decreases*. In any dynamical system, therefore, the microscopic, or "fine-grained," entropy is always a *monotonically decreasing* function of time. In fact, even if were not told anything about the system's initial state, we may be able to pin down with increasing accuracy its state at a later time—by just *waiting*. That is, if we precisely know the *dynamics* and how much time has elapsed from system initialization, *that alone* may be enough to give us some knowledge of the current state!

An extreme example of irreversibility is that of dynamics (a), where *all* states flow into a single state in one step. In this case, even if we had total lack of knowledge about the initial state (distribution $\frac{1}{8}\langle 1, 1, 1, 1, 1, 1, 1, 1 \rangle$, of maximal entropy), already after one step we'd be sure the system would have settled into that distinguished state (distribution $\frac{1}{8}\langle 8, 0, 0, 0, 0, 0, 0, 0 \rangle$, of *zero* entropy).

At the other extreme—dynamics (c)—we have the special case of a system that happens to be invertible: since in this case states never merge, states and associated bins are merely *permuted* by the dynamics. Since entropy is a symmetric function of all its arguments (that is, it does not depend on the order of the bins), and *all* that is changed by the dynamics in the present case is the bins' order, the entropy remains strictly *constant.*[5] So, if the dynamics is invertible, then *information is conserved.*

## 4 The second law of thermodynamics

To complete the argument of the previous section—that the concepts of invertibility and conservation of information are equivalent—we need to prove the converse of the above, namely, that if a dynamics conserves the "amount of information" of all possible "states of knowledge" (in other words, conserves entropy for *all* possible distributions), then the dynamics itself is invertible. It would be hard to test all possible distributions, since, these, unlike the states, are infinite in number. However, it turns out that to prove our statement it is sufficient to verify its validity on only a *finite* subset, namely, the "basis" consisting of those distributions where all the mass is concentrated, in equal parts, on just two bins. In fact, if any two arcs merge in the graph, this will be revealed by a drop of entropy in one of those basis distributions. ■

What we have obtained so far is something one might call the "*weak* second law of thermodynamics" ("weak" because so tautological), which states that

W2LT: *The* [fine-grained] *entropy of an invertible system is constant.*

We can compare this with the (traditional) "second law of thermodynamics," which, in a form equivalent to Clausius' original formulation,[6] states that

T2LT: *The* [thermodynamic] *entropy of an isolated system grows monotonically.*

Here, as per Clausius' original definition (1854), *entropy* is a quantity characterized (up to an additive constant) by the relation

$$\Delta S = Q\left(\frac{1}{T_1} - \frac{1}{T_0}\right)$$

where $Q$ is the amount of energy that passes in the form of *heat* from a body at temperature $T_1$ to one at $T_0$. Note that the temperature of a body is well-defined only if the latter is as close as desired to *thermal equilibrium.* By "state" here was intended a *thermodynamic state*, characterized by the values of its macroscopically measurable properties such as volume, pressure, temperature, chemical composition, etc.; a few decades later, a thermodynamic state would have been interpreted as an *equivalence class* of *microscopic* states, namely, all those microstates sharing those particular values and thus consistent with that macroscopic description.

[5] A constant function is only a special case of a monotonically decreasing function.

[6] "Heat generally cannot flow spontaneously from a material at lower temperature to a material at higher temperature."

Boltzmann's intuition was to interpret the entropy $S$ of a thermodynamic state directly in terms of the *number* $\Omega$ of microstates consistent with the given macrostate:

$$S = k \ln \Omega.$$

This led to a generalization of the concept of entropy, since now one could come up with satisfactory characterizations of systems that are *not* at equilibrium—and thus may not have a temperature—but still determine their entropy in a consistent way by just *counting* the number of microstates that went along with that characterization.

This new freedom opened the floodgates to a whole slew of new "entropies," each tied to what information one had and was willing or capable to use in order to determine the makeup (and thus the microstate count) of a macrostate as it evolved in time. Specifically, one had "coarse-grained" entropies, that depended on what kinds of measurements one felt capable (or incapable) of performing. It is not so much that entropy became a *subjective* quantity, as that it became clear that it is a property of a *description* of a system rather than an intrinsic aspect of the system itself. Some of these entropies have the property that they monotonically increase only *in the mean*—the instantaneous value may fluctuate about this mean and so occasionally decrease, in contradiction with T2LT.

Even worse, one had Poincaré's paradoxical "recurrence theorem," whereby the microstate of a large but finite system driven by an invertible dynamics would eventually come back, possibly after an astronomical long time, to the original state. In fact, as clear from Fig. 2c, in these systems *every* state lies on a cyclic trajectory. Consequently, *any macrostate*, *qua* a class of microstates, would eventually [7] return to the exact initial macrostate. If we treat entropy as an (even approximately) measurable property of a physical state, we'd expect it to grow rapidly (with respect to this astronomical scale) towards its maximum value, then coast very close to it for almost the entire cycle, and then, just as rapidly as it went up, come down to the initial value at the end of the cycle. then have to admit that at a moment very distant in time its entropy would have come be back to the initial value.

It is clear that many of these "statistical-mechanical" variants of entropy lead to violations of the traditional second law. Either this law is to remain valued today merely for its antiquarian interest, or something else has to give.

We have proposed an interpretation that views the state of a system as the "most honest state of knowledge" [8] that one can give for the current time [9] based *exclusively* on one's knowledge (from one's own measurements or from somebody else's reports) of the initial state, one knowledge of the microscopic laws, and a statement of the kind and amount of memory and computational resources that one is able or willing to devote to integrating the equations of motion from that initial state and for those laws.

[7] Though in a time equal to the least common multiple of the lengths of all the orbits involved, and thus exponential in the size of the macrostate.

[8] In the tradition of de Finetti[8] and Jaynes[13].

[9] Or *any* specified time, whether in the future or in the past.

It turns out [10] that, with this approach, one can state a "strong second law of thermodynamics," namely,

S2LT: *The* [honest] *entropy of an invertible system grows monotonically,*

which is identical in formulation to the traditional law, [11] but whose content is *logically equivalent* to the *invertibility of the underlying dynamics*—just as in the weak law. In other words, if a system is *microscopically reversible*, then its macroscopic behavior will strictly obey the second law of thermodynamics (incidentally, displaying macroscopic *irreversibility* and all that); conversely, if it obeys this law, then it *must* be microscopically reversible. *As simple as that!*

## 5 Invertible computing

In all endeavors in which we try to get the most from nature—think of agriculture, athletics, or steam power—we can't go very far if we don't ask questions about "the nature" of nature—what are the fundamental resources and constraints—and have the humility to accept the answers.

Computation, numeric or symbolic, whether done in one's mind, with pencil and paper, or on an electronic computer, is a way to concretely run abstract logic arguments on a concrete physical substrate. Today, such a substrate is invariably macroscopic; that is, both the logical variables—the *bits*—and the logic operators—the *gates*—are realized by lumps of matter consisting, even in state-of-the-art implementations, of millions of atoms each[24,6]. [12]

To reduce the intrinsic noisiness, due to both fabrication tolerances (coarseness of edges, inhomogeneity of composition) and operating conditions (thermal noise, signal crosstalk) of such a substrate and approximate the ideal precision of logic, a large amount of redundancy is used. For example, in a nonvolatile magnetic medium such as a hard disk, to guarantee—against thermal noise—a mean relaxation time of at least ten years for the states of the individual bistable elements ("bits"), these have to be realized as metastable states with a stored energy of at least 65 times the value $kT$ characteristic of thermal agitation impacts—where $T$ is the temperature of the environment and $k$ Boltzmann's constant. When the state of a bit of magnetic memory is changed, the energy of the original metastable state is typically dumped and a new metastable state is built out of high-grade energy freshly drawn from the power supply.

In principle, most of this energy could be *recycled*, much as when one brings a car from a state of motion to one of rest not by converting the coherent kinetic energy of the car into incoherent thermal agitation by means of the cheapest dissipation device, namely the friction brake, and dumping it as heat, but by converting it into electric energy by means of a braking dynamo and storing it into a battery

[10] This argument is briefly sketched in [20]; a more scholarly version is in preparation.

[11] But note that growth is now monotonical vis-à-vis the *absolute* time distance between initial and final states, since we are allowed to extrapolate the system's evolution either forward or *backwards* in time.

[12] This for just the active volume, without counting ancillary (e.g., wires) and support (mechanical substrate, power distribution, etc.) structures.

as high-grade chemical energy. Of course, dynamo and battery represent a much more expensive infrastructure than a simple disk brake (and the surrounding air to cool it); the alternative is between a large one-time investment in permanent infrastructure and a small daily operating charge—till the end of time—for "fuel delivery" and "garbage collection."

Assume an interest rate of zero, so that, by amortizing the investment over a large enough number of years, one could set the yearly premium as close to zero as desired; then the resources to be invested in the recycling infrastructure are no longer a concern. In this scenario, could one dream of achieving 100% energy recycling efficiency when *writing* a bit to memory—i.e., overwriting the old value with an arbitrary new one? Suppose that the new value is 0; we are thus envisaging a 1-bit register whose old value may have been 0 or 1 but whose new value must definitely be 0, with a transition graph ("clear bit" operation) as in (2)a.

$$\begin{array}{cc|cc} \multicolumn{2}{c|}{\text{a}} & \multicolumn{2}{c}{\text{b}} \\ \text{old} & \text{new} & \text{old} & \text{new} \\ 0 & 0 & 0, r & 0, r \\ 1 & 1 & 1, r & 0, s \end{array} \qquad (2)$$

This is clearly a *noninvertible* operation, and as such cannot represent the microscopic evolution of a physical system—or at least of a *complete* one: something must be missing. One way out of this difficulty is to imagine that the system actually consists of *two* components, only one of which (the "logic" component $X$) is shown in (2)a. In reality, another component, $A$ (for "ancilla"), is ganged with $X$ such that whenever there is a state merge in $X$ there is a concomitant state split in $A$, as shown in (2)b, so that overall the transition is still *invertible*. Note that if we only look at the logic component $X$ of this ganged pair we still observe the desired noninvertible behavior. On the other hand, if we only look at the ancilla $A$, we see a *nondeterministic* transition—state $r$ may remain unchanged or go to a different state $s$. Typically, the ancilla will be a thermal reservoir, and the clearing of a logic bit, entailing a loss of entropy of 1 bit in the $X$ subsystem, is *inescapably* accompanied by a further randomization—a gain of at least one bit—of the thermal reservoir; this is *Landauer's principle*[15].

In brief, a 1-bit increase in the entropy of a thermal reservoir at temperature $T$ implies the thermalization of a quantity $kT \ln 2$ of energy. Thus, the higher the temperature $T$ of the environment, the higher the energetic cost of getting rid of the previous value when writing a new value to a memory location; intuitively, the cost of getting rid of one unit of unwanted information ("garbage") is proportional to how much garbage is already "out there;" as a matter of fact, if we don't take active measures like cooling, it is the "out there" that will occasionally dump garbage onto our own backyard!

How serious is the practical impact of Landauer's principle? How large is the fee that it extracts from us, in comparison with other charges that today we already routinely pay to nature to have our computations done?

In general, the serviceability of any kind of computer is maintained by a "pump" (power supply and heat sink) that continually injects fresh predictability into the system, typically in the form of high-grade energy, and drains away accumulated unpredictability, in the form of heat. Today, the unpredictability generated within the system comes from two sources. By far the larger fraction is due to imperfection of the macroscopic mechanisms: ripples of escaped mechanical energy must be damped by frictional methods (e.g., cable terminators) so that they won't show up in unwanted places and interfere with critical operations. The need for damping can be reduced by more precise design and fabrication of a computer's primitive logic elements—transmission lines and gates. In fact, frictional losses, which used to be billions of times larger that those ascribable to "logic" losses (those that, as we've just seen, must accompany any noninvertible logic step), have been reduced so much in the last two decades that now they are only a few orders of magnitude larger than the latter. At this rate, in a decade or two, logic losses will emerge as a *substantial fraction* of the total computational budget.

Moreover, frictional sources of dissipation can in principle be lowered below any set threshold by achieving sufficient control of the physical processes that underlie mechanical computation; in this domain at least, Moore's law has no end in sight. Logic sources of dissipation, on the other hand, are intrinsic to the computational procedure one intends to carry out, and can only be reduced by jointly

- Introducing logic primitives that, unlike the ubiquitous NAND gate, are *invertible* (cf. [3,18,10,2]);
- Rewriting the algorithms so as to make use of those invertible primitives and avoid as far as possible the need for noninvertible logic operations; [13] and
- Contriving realistic physical effects that implement those invertible primitives in a physically invertible (or at least thermodynamically reversible) way. [14]

The above three tasks characterize the field of *invertible computing*.

Invertible computing is of course a prerequisite for *quantum computing* (see [2]), which must run on that more specialized form of invertibility that is a fundamental feature of quantum mechanics, namely, *unitarity*.

## 6 Green is not a pic-nic

The tension that we've seen above between invertibility and computation reveals itself again when one tries to make discrete models of local interactions in distributed (i.e., spatially extended) systems. Here, instead of differential equations one makes recourse to *recurrence relations* (of which cellular automata are an example). It turns out that most of these recurrence schemes cannot be plausible candidates, even

[13] Whether general-purpose computation can be carried out *at all* by an invertible mechanism or within an invertible medium had been seriously questioned (cf., e.g., [7,1]), until it was finally shown that such a feat is in principle possible, though generally at the cost of more complex computational infrastructure[3,17,18,10,5].

[14] Fine-grained, massively parallel computational schemes, such as invertible cellular and lattice-gas automata, have turned out to be very productive conceptual tools not only for physical models of physics but also for imagining physical models of invertible computation[19,14,21].

as toy models, for ultimate, microscopic physical interactions, because they need to be "powered" to work.[15] The issue arises of whether, given a simple recurrence scheme that happens to describe a desired behavior but requires power to operate, one can devise an equivalent alternative scheme that can operate autonomously, "without batteries," and thus be a more plausible model of a primitive physical effect. Recent research[22] shows that this desideratum is in principle achievable (except for a few, well-characterized exceptional cases), though typically requiring deployment of *much more extensive and complex mechanisms*.

These results formalize a reasonable suspicion about ordinary *recycling*. Much energy can be recycled, for example, by collecting used glass bottles as close as possible to the source—say, by putting a recycling bin in a cafeteria. A little more energy could be recovered (or, more precisely, less entropy generated) if one provided separate bins for clear bottles, brown bottles, and green bottles—but clearly at the cost of more infrastructure (floor space, more bins, trucking trips, etc.). A point will be reached where the sheer cost of providing and maintaining the extra infrastructure will offset the raw energy-savings benefits.

The design-and-outlay difficulties mentioned above, which beset schemes for invertible computation within an extended fine-grained medium that attempts to model microscopic physics, turn out to be much less serious if instead of "generic invertibility" one could restrict oneself to a more specialized form of invertibility, as found in *second-order* systems[21] (an example is the symplectic structure of classical mechanics). For that matter, why is physics symplectic (or, in the quantum formulation, unitary)? Did God himself realize that generic invertibility was too demanding a specification to design a universe with, and backed up a little bit? Or—a question we are investigating at this moment—do we get the second-order "flavor" of invertibility (with inertia and all that) *for free* from generic invertibility as an *emergent feature*, when we look at the world at a slightly coarser resolution than that of the most fundamental physics?

## 7 Conclusions

When you first hear of Landauer's principle (§5), your first reaction is likely to be, "What a bummer! If it were not for physics' built-in invertibility, it would be possible to get rid of garbage for free!" Or, "In a universe governed by Conway's rule of LIFE, which is *noninvertible*, it is possible to have such a thing as a 'glider gun,' apparently a *perpetual motion machine of the second order*—a source of unlimited free energy. Why wasn't our world make like that?"

Well, if you had to design a universe in which something as interesting as yourself would hopefully eventually spontaneously emerge,

[15] To give an analogy, one can model an oscillator by means of an LC circuit. But concrete, macroscopic inductors and capacitors are lossy and thus cannot sustain constant-amplitude oscillations. However, by putting an LC circuit in the feedback loop of that miraculous, universal device that is the operational amplifier (OA) one can synthesize an ideal oscillator from nonideal L's and C's. The catch is that, while an LC circuit by itself doesn't need a power supply, the latter is required for the operation of the operation of an OA-assisted circuit.

Q: *Would you make its microscopic laws invertible or not?*

I don't know the answer to this question, though I've been thinking about it for many years, and still I lean toward 'yes' on Mondays, Wednesdays, and Fridays, 'no' on Tuesdays, Thursdays, and Saturdays, and "I don't want to think about it" on Sundays. [16]

I had a similar difficulty with another question related to computation in non-invertible media, and in desperation I had resolved that when I die the first thing I would ask of God would be "Please tell me the answer to that!" But then one night I saw the light all by myself, and with the help of a couple of friends we worked out a simple, final, complete solution[22]. [17] This goes to prove that questions of this kind are after all scientific, serious, and answerable.

When he and I were conjecturing about question Q above, Norman Margolus seriously quipped that invertibility is "The way to make your quarter play longest." If I wanted to be lazy and use some sort of anthropic principle, my own answer to that question would be, "Well, I'm here; so invertibility is certainly *one* way to get something interesting emerge." Winston Churchill would probably have answered, "Invertibility is the worst form of government for a universe, except for all those others that God must have tried before."

It is clear that the issue of invertibility has much to do with how to preserve information (and, as Edward Fredkin remarked, once you accept special relativity there is no intrinsic difference between *memory* and *communication*), pamper process, and get rid of no longer needed information. In a few years, preoccupations of this kind have moved from the stratospheric level of pure mathematics to the specialized but still rather stylized ballet of pure Alice, honest Bob, and wicked Eve in a cryptographic soap opera, down to the ever more pressing technological business of more efficient computing; and are finally becoming terribly relevant to how best to manage to live in a world with ten billion people. For each of these, by symmetry, their own affairs represent precious information to be protected and indispensable processing to be fostered, while everyone else's are most often indifferent but bothersome noise when not competitive threat. With "incompressibility" of information and finiteness of resources, how can we invent a way for each of us to *live like a king*, with our own castle and park, without denying almost everyone else—like in the feudal periods that characterized most of "civilized" history—similar aspirations? Asked whether he wouldn't have liked to live in Pericles' Athens, Isaac Asimov answered, "I doubt it—statistically, I would have been a slave!"

Such are the productive pleasures of studying invertibility in the context of our own world.

[16] Of course I agree with Don Lancaster (www.tinaja.com/glib/bashpseu.pdf) that "finding a source of unlimited free energy would be the most unimaginably heinous crime possible against humanity. For it would inevitably turn the planet into a cinder."

[17] So when I die my first question to God will be "How does quantum mechanics really work?"—unless, of course, I get the Nobel prize first, and my first question will be the present one.

# References


[1] Aladyev, Viktor, "Computability in Homogeneous Structures," *Izv. Akad. Nauk. Estonian SSR, Fiz.-Mat.* **21** (1972), 80–83.

[2] Barenco, Adriano, Charles Bennett, Richard Cleve, David DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John Smolin, and Harald Weinfurter, "Report on new gate constructions for quantum computation", *Phys. Rev. A* **52** (1995), 3457–3467.

[3] Bennett, Charles H, "Logical reversibility of computation," *IBM J Res. Develop.* **6** (1973), 525–532.

[4] Bennett, Charles H, "Logical depth and physical complexity," in *The Universal Turing Machine—a Half-Century Survey* (Rolf Herken ed.), Oxford U Press 1988, 227–257.

[5] Bennett, Charles H, "Notes on the history of reversible computation," *IBM J Res. Develop.* **32** (1988), 16–23.

[6] Bohr, Mark, "The high-$k$ solution," *IEEE Spectrum* **44**:10 (Oct 2007), 29–35. Semiconductors: October 2007

[7] Burks, Arthur (ed.), *Essays on Cellular Automata*, U Illinois Press 1970.

[8] de Finetti, Bruno, *Theory of Probability—A critical introductory treatment*, vol. 1 and 2, Wiley & Sons 1974, 1975, translated by A Machí and A Smith from *Teoria della Probabilità*, vol. 1 and 2, Einaudi 1970.

[9] Dobzhansky, Theodosius, "Nothing makes sense in biology except in the light of evolution." *American Biology Teacher* **35** (1973), 125–129.

[10] Fredkin, Edward, and Tommaso Toffoli, "Conservative logic," *Int. J Theor. Phys.* **21** (1982), 219–253.

[11] Gardner, Martin, "The Fantastic Combinations of John Conway's New Solitaire Game 'Life'," *Sc. Am.* **223**:4 (April 1970), 120–123.

[12] Geroch, Robert, *Mathematical Physics*, U Chicago Press 1985.

[13] Jaynes, Edwin, *Probability Theory—The logic of science,* Cambridge U Press 2003.

[14] Kari, Jarkko, "Theory of cellular automata: A survey," *Theor. Comp. Sci.* **334** (2005), 3–33.

[15] Landauer, Rolf, "Irreversibility and heat generation in the computing process," *IBM J.* **5** (1961), 183–191.

[16] Papanikolaou, Nick, "An introduction to quantum cryptography," *ACM Crossroads* **11**:3 (Spring 2004).

[17] Toffoli, Tommaso, "Computation and construction universality of reversible cellular automata," *J Comp. Syst. Sci.* **15** (1977), 213–231.

[18] Toffoli, Tommaso, "Reversible Computing," *Automata, Languages and Programming* (de Bakker and van Leeuwen eds.), Springer-Verlag 1980, 632–644.

[19] Toffoli, Tommaso, and Norman Margolus, "Invertible cellular automata: A review," *Physica D* **45** (1990), 229–253.

[20] Toffoli, Tommaso, "Computation: our 'theoretical physics' kit," *TRG: Transient Realities and their Generators*, FoAM, Brussels 2006, in close cooperation with Time's Up, Linz. See pm1.bu.edu/~tt/linz.

[21] Toffoli, Tommaso, Silvio Capobianco, and Patrizia Mentrasti, "How to turn a second-order cellular automaton into a lattice gas: a new inversion scheme," *Theor. Comp. Sci.* **325** (2004), 329–344.

[22] Toffoli, Tommaso, Silvio Capobianco, and Patrizia Mentrasti, "When—and how—can a cellular automaton be rewritten as a lattice gas?" *Theor. Comp. Sci.* **403** (2008), 71–88.

[23] von Neumann, John, *Theory of Self-Reproducing Automata* (edited and completed by Arthur Burks), Univ. of Illinois Press 1966.

[24] Wood, Roger, "The feasibility of magnetic recording at 1 Terabit per square inch, *IEEE Trans. Magnetics* **36** (2000), 36–42.

[25] Zuse, Konrad, *Rechnender Raum*, Vieweg, Braunschweig 1969; translated as "Calculating Space," *Tech. Transl. AZT-70-164-GEMIT*, MIT Project MAC 1970.