Information and Entropy in the Baker's Map

Rüdiger Schack[a] and Carlton M. Caves[a]
Center for Laser Studies, University of Southern California, Los Angeles, California 90089-1112
(Received 9 July 1992)

We analyze a random perturbation applied to the baker's map, a prototype for chaotic Hamiltonian evolution. We compare two strategies for following the perturbed evolution: (i) tracking the perturbed pattern in fine-grained detail; (ii) coarse graining by averaging over the perturbation. We find that the Landauer erasure cost associated with the information needed to implement the first strategy is overwhelmingly larger than the standard free-energy reduction associated with the second strategy. This finding provides a quantitative justification for coarse graining and thus an explanation of the second law of thermodynamics.

PACS numbers: 89.70.+c, 05.45.+b, 05.90.+m

Ordinary entropy measures the degree to which incomplete knowledge about a physical system reduces the ability to extract work from that system; it thus appears as a negative contribution to free energy. As a consequence of Landauer's principle [1, 2], which specifies the unavoidable energy cost $k_BT \ln 2$ connected with the erasure of a bit of information, the information (quantified by algorithmic information [3]) needed to give a complete description of the system state also reduces the amount of available work and thus should be added as a further negative contribution to free energy [4, 5].

This paper is motivated by the general question [6] of how available work decreases during chaotic Hamiltonian evolution subjected to random perturbations. We focus here on the baker's transformation [7], a prototype of a chaotic area-conserving map. General considerations [6] suggest that the results found here are not limited to this specific example. Moreover, a generalization to the large class of systems that can be modeled by symbolic dynamics [8] seems to be possible.

It is well known that, for mixing systems, a suitably defined coarse-grained entropy increases with time, approaching some equilibrium value. In Ref. [9], it is shown explicitly that the coarse-grained probability density for the baker's transformation approaches a constant in the long-time limit. It is tempting to regard this result as a proof of the second law of thermodynamics. The coarse graining, however, appears as an *ad hoc* assumption. We show that the presence of perturbations in any realistic system provides a justification for coarse graining. The famous example of a butterfly changing the pattern of density fluctuations in the atmosphere—and thereby changing the way a thunderstorm develops on the other side of the globe—illustrates the near impossibility of isolating a chaotic system.

It is intuitively obvious that it would be crazy to try to follow the detailed behavior of a perturbed chaotic system. In this paper, we quantify this intuition precisely. We show that coarse graining is the work-efficient strategy in the presence of perturbations. By this we mean that keeping track of the perturbed phase-space pattern leads to a much greater reduction in free energy than averaging over the small-scale perturbations. The reason is that the perturbation accesses algorithmically complex patterns of the sort discussed in Refs. [6, 10]. The information needed to specify a typical such pattern is so enormous that its Landauer erasure cost far outweighs the entropy increase due to averaging.

The concept of algorithmic information was introduced into physics as a way to characterize chaos [8, 11]. This characterization focuses on the algorithmic information needed to calculate a system trajectory. For discrete chaotic maps, the number of initial-condition digits needed to specify a trajectory to a given accuracy increases linearly with the number of steps one wishes to predict. Because we are interested in the question of available work, our approach is fundamentally different. We start with an initial condition given to some finite accuracy, corresponding to a finite initial area in phase space. Under the chaotic time evolution, this area is repeatedly stretched and folded, thereby evolving into an apparently complicated pattern. We ask for the algorithmic information needed to specify this pattern after n steps. We see below that this information is negligible for unperturbed evolution, in sharp contrast with the diverging amount of information needed to specify a single trajectory.

The baker's transformation is most easily described in terms of its symbolic dynamics [8]. Each point in phase space is represented by a symbolic string $s = \cdots s_{-2}s_{-1}s_0.s_1s_2\cdots$ where $s_k = 0$ or 1. The string s is identified with a point (q,p) in the unit square by setting $q = \sum_{k=1}^{\infty} s_k 2^{-k}$ and $p = \sum_{k=0}^{\infty} s_{-k}2^{-k-1}$. *Sets* of points S are usually [8] represented by symbolic strings that do not extend indefinitely in both negative and positive directions. In those strings only a subset of symbols is specified; i.e., s_k is specified only for a subset of indices $k \in I \subset \{\ldots,-1,0,1,2,\ldots\}$. The set S consists of all the points (q,p) that have a binary expansion compatible with the specified symbols s_k, i.e., $S = \{(q,p)|q = \sum_{k=1}^{\infty} t_k 2^{-k},\ p = \sum_{k=0}^{\infty} t_{-k} 2^{-k-1},\ t_k \in \{0,1\},\ t_k = s_k \text{ for } k \in I\}$. For our purposes we find it convenient to write $s_k = x$ if s_k is not determined by I; for example, the set $\{(q,p)|0 \le q, p \le 1/2\}$ is denoted by $\cdots xx0.0xx\cdots$ or simply by $xx0.0xx$—i.e., $s_0 = s_1 = 0$ and $s_k = x$ for $k \notin \{0,1\}$. The position of the dot determines how a symbolic string is translated into a pattern on the unit square. We identify a set of strings with a uniform probability distribution on the union of the areas represented by the strings. The action of the baker's map on a symbolic string is given by the shift map U defined by $(Us)_k = s_{k+1}$, which means that, at each time step, the entire string is shifted one place to the left while the dot remains fixed. This corresponds to compressing the unit square in the p direction and stretching it in the q direction, while preserving the area, then cutting it vertically, and finally stacking the right part on top of the left part—in analogy to the way a baker kneads dough.

Suppose that the baker's map is applied n times to an algorithmically simple initial distribution. For simplicity, we assume that this initial condition corresponds to uniform probability over an area that is described by a single symbolic string containing a finite number q of *initial-condition* digits, i.e., a finite number of digits different from x. The algorithmic information (background information) needed to describe the unit square and the baker's map and to specify the initial condition we denote by I_0. We choose the (arbitrary) zero of the entropy H so that the string $xx.xx$, corresponding to a uniform probability over the entire square, has entropy $H = 0$. Hence the initial entropy is $H_0 = \log(2^{-q}/1) = -q$. (Throughout this paper information and entropy are measured in bits and log denotes the base-2 logarithm.)

After n steps, the simple initial pattern evolves into a complicated looking pattern of narrow horizontal stripes. To retain the ability to extract the work inherent in the initial state, i.e., to prevent the free energy from decreasing quickly, one must keep track of this pattern of stripes, thereby keeping the entropy constant: $H(n) = H_0$ or $\Delta H(n) = 0$. How much information must be supplied in addition to the initial I_0 to specify the pattern after n steps? Since the pattern after n steps is obtained simply by shifting the string n places to the left, the only additional information needed is the number n [5], requiring $\Delta I(n) \simeq \log n$ bits. The change in free energy is given by $\Delta F(n) = -k_B T \ln 2\,[\Delta H(n) + \Delta I(n)] \simeq -k_B T \ln 2 \log n$. The free-energy cost of keeping track of the evolved pattern grows very slowly. Unperturbed evolution of an algorithmically simple initial state does not lead to algorithmically complex states, at least for reasonable values of n. It is therefore possible in principle to retain the ability to extract work by keeping track of the evolving pattern.

The situation is dramatically different if one allows for perturbations. In order to model a perturbation, we divide the symbolic string into regions (see Fig. 1), which

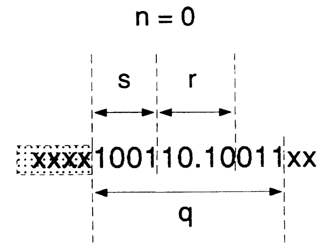

FIG. 1. A typical symbolic string at $n = 0$, just before the perturbation becomes effective. Application of the baker's map moves the leftmost of the q initial-condition digits into the shaded perturbation region. There are r decision digits, which partition the unit square into 2^r perturbation cells, and s digits separating the perturbation region from the decision digits.

remain fixed relative to the dot as the string moves to the left. The leftmost region, called the *perturbation region*, is separated by s digits ($s \ge 0$) from the *decision region*, r digits wide ($r \ge 1$), which partitions the unit square into 2^r congruent rectangles, which we call *perturbation cells.* Furthermore, we consider the set of *area-conserving perturbation maps*, which act on the digits in the perturbation region. A perturbed time step consists in first applying the unperturbed map U, then splitting the pattern into subpatterns defined by the perturbation cells (i.e., by the decision digits), and finally applying to each subpattern a perturbation map chosen at random. For an actual phase-space evolution, we would be primarily interested in energy-conserving perturbations, so as to separate our investigation of information and entropy from issues raised by dissipation. The analogous perturbations for two-dimensional maps, such as the baker's transformation, are area conserving.

To simplify the analysis, we restrict ourselves to perturbations that do not affect the pattern on scales smaller than the smallest structures of the pattern itself. In symbolic language, this means that the symbol x must be invariant under any perturbation. The set of perturbation maps is thereby restricted to maps that act on each perturbation digit independently, either switching it or leaving it unchanged. Notice that the rightmost digit in the perturbation region characterizes the perturbation's "strength," whereas the linear dimensions of a perturbation cell give the perturbation's "correlation lengths."

Figure 2 shows a perturbed time step for the initial pattern $xx01.xx$. The perturbation region lies just to the left of the digit 0, as indicated by the vertical line. The $s = 2$ decision digits are marked by a box. Application of the unperturbed map leads to the pattern $x01x.xx$, shown in Fig. 2(b) together with the expansion into the four subpatterns given by the four possible choices for the decision digits. Since only one initial-condition digit of each subpattern is located in the perturbation region, there are just two different perturbation maps for each perturbation cell, the identity map and the *switch map*

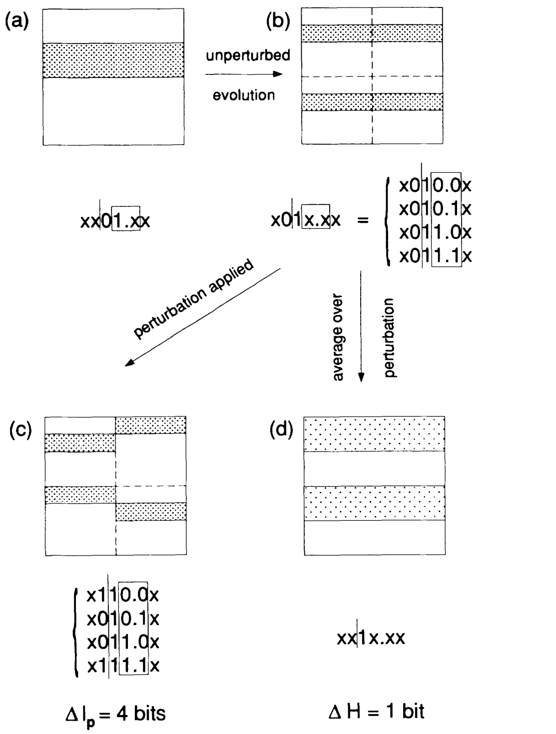

FIG. 2. An example of the action of the perturbed baker's map on the initial condition shown in (a). In (b), the pattern that results from application of the unperturbed baker's map is split into four subpatterns determined by the four perturbation cells. In the symbolic representation, these perturbation cells are distinguished by two decision digits, here enclosed in a box. In (c), a perturbation map is applied to each subpattern independently, affecting only the digits in the perturbation region to the left of the vertical line. The information needed to specify the perturbed pattern, given the initial pattern and the number of steps, is $\Delta I_p = 4$ bits. Averaging over all possible perturbation maps leads to the coarse-grained pattern in (d), with an entropy increase of $\Delta H = 1$ bit.

that interchanges 0 and 1. Figure 2(c) shows the result of applying one possible choice of perturbation maps to the pattern of Fig. 2(b).

In the following, we let $\Delta I_p(n)$ be the additional (conditional) algorithmic information needed to specify a *typical* perturbed pattern, given the background information I_0 and the number of steps n. Since the patterns are generated by a random mechanism that leads to equally likely alternatives, ΔI_p can be determined simply by counting [6, 12]: if N_p is the number of equally likely patterns, the information needed to specify a typical pattern is $\Delta I_p \simeq \log N_p$. In the example of Fig. 2(c), there are $N_p = 16$ patterns—two maps for each of four perturbation cells—leading to $\Delta I_p = 4$ bits. Averaging over the perturbation leads to the pattern $xx1x.xx$ shown in Fig. 2(d) and thus to an entropy increase of $\Delta H = 1$ bit.

Turn now to a general analysis. We choose the zero of time ($n = 0$) so that the perturbation becomes effective at the first time step ($n = 1$)—i.e., so that the structure of the unperturbed pattern at the first step is on the scale set by the perturbation strength. In the symbolic representation, this corresponds to the situation shown in Fig. 1: at $n = 0$ the leftmost of the q initial-condition digits is situated just to the right of the perturbation region and is first perturbed at $n = 1$. For specificity, we assume that at $n = 1$ the unperturbed pattern is confined entirely to one perturbation cell. The corresponding condition $q > r + s$ (see Fig. 1) expresses the fact that all the decision digits at $n = 1$ are determined by the initial condition. At each step, one additional initial-condition digit moves into the perturbation region and is randomized by the perturbation. If one averages over the perturbation, the entropy increases by one bit per step (doubling of phase-space area), so $\Delta H(n) = n$, until the entropy reaches its maximum value, $\Delta H = q$, at which all information on the initial condition is lost.

The conditional algorithmic information $\Delta I_p(n)$ has quite different behavior, which we have calculated for $0 \le n \le q$ (see Table I). There are three regimes. For $0 \le n \le q - s - r$, the entire pattern remains inside a single perturbation cell—all decision digits are determined by the initial condition—so only one perturbation map is applied at each step. The n digits affected by the perturbation are specified by $\Delta I_p(n) = n$ bits. For $q-s-r < n \le q-s$, the pattern spreads over $2^{n-(q-s-r)}$ perturbation cells, in each of which n perturbed digits have to be specified, leading to $\Delta I_p(n) = n2^{n-(q-s-r)}$. For $q - s < n \le q$, the situation is more complicated, because now the subpattern in each perturbation cell consists of $2^{n-(q-s)}$ separate pieces. A perturbation map cannot change the correlations between the pieces inside

TABLE I. Algorithmic and entropic contributions to the decrease in free energy as a function of the number of steps n. The amount of information needed to keep track of the evolved pattern in the absence of perturbations, given by ΔI, grows as $\log n$. In the presence of perturbations, however, the information to keep track of the pattern, given by $\Delta I + \Delta I_p$, exceeds by far the entropy increase ΔH that results from averaging over the perturbation.

n	ΔI	ΔI_p	ΔH
$0 \le n \le q-s-r$	$\log n$	n	n
$q-s-r \le n \le q-s$	$\log n$	$n2^{n-(q-s-r)}$	n
$q-s \le n \le q$	$\log n$	$n2^r + 2^{r-2}(n-q+s)(n+q-s-1)$	n

one perturbation cell. Since the information needed to specify these correlations is equal to half the information to specify the preceding perturbation maps, back to and including the maps for $n = q - s$, we find that

$$\Delta I_p(n) = n2^r + \sum_{j=q-s}^{n-1} \frac{1}{2} j 2^r = n2^r + 2^{r-2}(n-q+s)(n+q-s-1).$$

For $n \geq 0$ the bottom two rows of Table I apply without change to the case $s < q \leq r + s$. The bottom row describes the $n \geq 0$ behavior of the case $q \leq s$, provided that one sets $q = s$ in the formula for ΔI_p. For arbitrary $q > 0$, we find that $\Delta I_p(n) \geq \Delta H(n) = n$ for $0 \leq n \leq q$ and, moreover, that $\Delta I_p(n = q) \geq q2^r$.

We thus arrive at our key result: For the perturbed baker's transformation, the information ΔI_p needed to keep track of the evolving pattern far exceeds the entropy ΔH that results from averaging over the perturbation. This key result expresses a hypersensitivity to perturbations. For the baker's map, it is a consequence of the pattern's spreading over an exponentially large number of perturbation cells—i.e., phase-space cells whose size is determined by the correlation lengths of the perturbation. Hence, we expect this key result to hold for all sensible perturbations of the baker's map. More generally, we expect all chaotic systems with positive KS (Kol'mogorov-Sinai) entropy (or metric entropy) to display a similar hypersensitivity to perturbations [6]. In contrast, classical regular systems should not display such hypersensitivity. There is evidence [6] that quantum systems, because of their phase freedom, display a hypersensitivity to perturbations similar to classical chaotic systems, thus suggesting a new connection between chaos and quantum mechanics.

It is worth spelling out the physical meaning of our key result. Whenever $\Delta I_p \gg \Delta H$, averaging over the perturbation, which amounts to a coarse graining matched to the strength of the perturbation, is a far better strategy for preserving available work than is keeping track of the perturbed system state in fine-grained detail. This is a compelling justification for coarse graining in the presence of a perturbation and for the accompanying increase in entropy. There is a way around this conclusion: the excess information needed to specify the fine-grained pattern can be used to extract an equivalent amount of work from the perturbing system. Given a split between a system of interest and its surroundings, however, we have provided an explanation of the second law.

R.S. acknowledges the support of Deutsche Forschungsgemeinschaft. We both thank the Aspen Center for Physics for its hospitality while this paper was completed.

(a) Present address: Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131-1156.

[1] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).

[2] R. Landauer, Nature (London) **355**, 779 (1988).

[3] G. J. Chaitin, *Information, Randomness, and Incompleteness* (World Scientific, Singapore, 1987).

[4] W. H. Zurek, Nature (London) **341**, 119 (1989).

[5] W. H. Zurek, Phys. Rev. A **40**, 4731 (1989).

[6] C. M. Caves, in "Physical Origins of Time Asymmetry," edited by J. J. Halliwell, J. Pérez-Mercader, and W. H. Zurek (Cambridge Univ. Press, Cambridge, England, to be published).

[7] V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968).

[8] V. M. Alekseev and M. V. Yakobson, Phys. Rep. **75**, 287 (1981).

[9] L. E. Reichl, *A Modern Course in Statistical Physics* (University of Texas, Austin, 1980).

[10] C. M. Caves, "Information and Entropy" (to be published).

[11] J. Ford, Phys. Today **36**, No. 4, 40 (1983).

[12] C. M. Caves, in *Complexity, Entropy, and the Physics of Information*, edited by W. H. Zurek (Addison-Wesley, Redwood City, CA, 1990), p. 91.

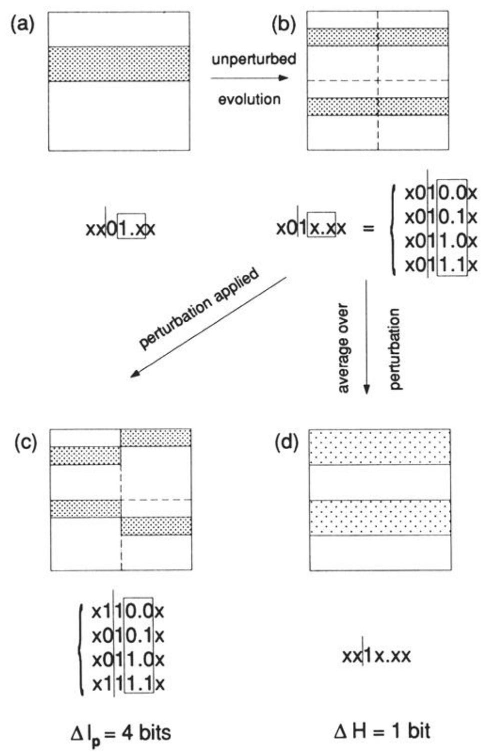

FIG. 2. An example of the action of the perturbed baker's map on the initial condition shown in (a). In (b), the pattern that results from application of the unperturbed baker's map is split into four subpatterns determined by the four perturbation cells. In the symbolic representation, these perturbation cells are distinguished by two decision digits, here enclosed in a box. In (c), a perturbation map is applied to each subpattern independently, affecting only the digits in the perturbation region to the left of the vertical line. The information needed to specify the perturbed pattern, given the initial pattern and the number of steps, is $\Delta I_p = 4$ bits. Averaging over all possible perturbation maps leads to the coarse-grained pattern in (d), with an entropy increase of $\Delta H = 1$ bit.