PHYSICAL REVIEW E VOLUME 47, NUMBER 6 JUNE 1993

Information and entropy

Carlton M. Caves
Santa Fe Institute, 1660 Old Pecos Trail, Suite A, Santa Fe, New Mexico 87501
and Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131-1156
(Received 4 January 1993)

The Landauer cost for erasing information demands that information about a physical system be included in the total entropy, as proposed by Zurek [Nature **341**, 119 (1989); Phys. Rev. A **40**, 4731 (1989)]. A consequence is that most system states—either classical phase-space distributions or quantum pure states—have total entropy much larger than thermal equilibrium. If total entropy is to be a useful concept, this must imply that work can be extracted in going from equilibrium to a typical system state. The work comes from randomization of a "memory" that holds a record of the system state.

PACS number(s): 89.70.+c, 03.65.−w

I. INTRODUCTION

To say that a system occupies a certain state implies that one has the information necessary to generate a complete description of that state. This apparently innocuous statement acquires physical significance from *Landauer's principle* [1, 2]: *to erase a bit of information in an environment at temperature T requires dissipation of energy $\geq k_B T \ln 2$.* Landauer's principle demands that information be granted a physical status as a negative contribution to *free* energy. This leads to a *total free energy*

$$\mathcal{F} = F - k_B T \ln 2\, I = E - k_B T \ln 2\,(H + I)\,, \qquad (1.1)$$

where E, H, and F are conventional energy, entropy (in bits), and free energy, and I is the information (in bits) required to describe the state of interest. The absolute information measure I, called *algorithmic information* [3], is defined as the length (in bits) of the shortest program on a universal computer that can generate a complete description of the state of interest. Defining total free energy is equivalent to defining a *total entropy*

$$\mathcal{S} = H + I\,, \qquad (1.2)$$

as proposed by Zurek [4, 5]. Throughout this paper I refer to the conventional entropy H as the *statistical entropy*, and I call the entity that stores information about the system a "memory."

This paper explores consequences of including information in free energy and entropy. The bottom line is that *it takes an enormous amount of information to describe most system states—either classical phase-space distributions or quantum pure states—much more information, in fact, than the statistical entropy of thermal equilibrium* [6]. As a result, most phase-space distributions and most pure states have total entropy much larger than equilibrium and total free energy much lower than equilibrium. If total entropy and total free energy are to be useful concepts, this means that work—indeed, an enormous amount of work—can be extracted in going from equilibrium to a typical phase-space distribution or a typical pure state.

This initially puzzling conclusion is illustrated by two *gedanken* examples—one classical and one quantum-mechanical—which are aimed at removing the puzzlement. They show that the work comes from randomization of the memory that holds a record of the system state. In the language of conventional statistical physics, the memory goes from a low-entropy, high-free-energy state, when it stores a record of the system in equilibrium, to a high-entropy, low-free-energy state, when it stores a record of a typical phase-space distribution or a typical pure state.

Far from just demystifying a theoretical puzzle, however, this paper establishes a framework for more ambitious investigations. Having established that typical phase-space distributions and typical pure states are algorithmically extremely complex and, furthermore, that it makes sense to attribute to them a high total entropy, one can proceed to the following question: can Hamiltonian evolution—either classical or quantum-mechanical—take an algorithmically simple initial state to one of the typical algorithmically complex states? Investigation of this question has begun [6–8], with interesting consequences for the second law of thermodynamics and the nature of irreversibility and for the connection between chaos and quantum mechanics.

Section II reviews pertinent elements of algorithmic information theory. Section III considers the total entropy of the conventional *distinct* states of a physical system and reconciles the memory's "inside view" [5], taken in this paper, with the "outside view" of conventional statistical physics. Section IV introduces the phase-space patterns and pure states that are the subject of this paper. Section V presents the gedanken examples that elucidate the meaning of total entropy and total free energy. Section VI concludes with a brief discussion.

II. ALGORITHMIC INFORMATION

Consider a set of $\mathcal{K}$ alternatives, labeled by an index k and occupied with probabilities p_k. In the remaining sections the alternatives become states of a physical system.

Three kinds of information arise in describing these alternatives. The first is the background information needed to give the statistical description. This background information, denoted by I_0, is the algorithmic information required to generate an overall description of the alternatives, including a list of all of them and their probabilities. The second kind of information, conventionally denoted $I_{k,0}$, but denoted just by I_k here, is the *joint* algorithmic information needed both to give the background information *and* to specify the particular alternative k. The third kind of information, conventionally denoted $I_{k|0}$, but denoted by ΔI_k here, is the *conditional* algorithmic information needed to specify alternative k, given the background information (more precisely, given the minimal program for the background information). The background information I_0, which is sufficient to generate a list of *all* the states, is often very much smaller than the the additional information ΔI_k needed to specify a particular *typical* alternative k.

Two results from algorithmic information theory are needed. The first result is the sensible one that

$$I_k = I_0 + \Delta I_k + O(1) \; , \tag{2.1}$$

where $O(1)$ denotes a computer-dependent constant, which arises from defining algorithmic information in terms of computer programs and which is bounded in absolute value for any particular universal computer. Equation (2.1) reveals the reason for my unconventional notation for algorithmic information: it is aimed at harmonizing with the conventional thermodynamic notation for differences between states—a reasonable aim when H and I are included on an equal footing in the total entropy.

The second result from algorithmic information theory is a double inequality [5, 9–11],

$$H \le \sum_{k=1}^{\mathcal{K}} p_k \Delta I_k \le H + O(1) \; , \tag{2.2}$$

which relates the *average* conditional information to the usual statistical information [12]

$$H = -\sum_{k=1}^{\mathcal{K}} p_k \log_2 p_k \; , \tag{2.3}$$

which is stored in $\mathcal{K}$ alternatives that have probabilities p_k ($\log_2$ denotes a base-2 logarithm, so H is in bits). The left-hand inequality in Eq. (2.2) is strict, whereas the right-hand inequality is soft because of the computer-dependent constant. For *typical* alternatives, ΔI_k is very close to the code-word length for alternative k in an optimal instantaneous code [12] for all $\mathcal{K}$ alternatives.

For the remainder of this paper, I specialize to equally likely alternatives ($p_k = 1/\mathcal{K}$), as in the microcanonical ensemble. In this situation the double inequality (2.2) becomes

$$\log_2 \mathcal{K} \le \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \Delta I_k \le \log_2 \mathcal{K} + O(1) \; . \tag{2.4}$$

It reduces the remainder of this paper to counting alternatives: the conditional information ΔI_k to specify a *typical* alternative k is very nearly $\log_2 \mathcal{K}$; there are simple alternatives for which $\Delta I_k \ll \log_2 \mathcal{K}$, but they are atypical.

III. CONVENTIONAL STATES AND THE INSIDE AND OUTSIDE VIEWS

Consider now an isolated physical system that has $\mathcal{J}$ *distinct* states, labeled by an index j and all of energy E_0. Classically these states are *nonoverlapping* cells in phase space, defined by a phase-space coarse graining; quantum mechanically they are *orthonormal* basis states in a $\mathcal{J}$-dimensional Hilbert space. These states are distinguished from the phase-space patterns and pure states introduced in Sec. IV precisely because they are distinct in the usual sense. To make a semantic distinction, I refer to a set of distinct states as conventional states.

Thermal equilibrium for this system is described by the microcanonical ensemble, in which all states are equally likely. Equilibrium has energy E_0, statistical entropy

$$H_0 = \log_2 \mathcal{J} \; , \tag{3.1}$$

and free energy

$$F_0 = E_0 - k_B T \ln 2\, H_0 = E_0 - k_B T \ln 2\, \log_2 \mathcal{J} \; . \tag{3.2}$$

The background information I_0 needed to describe equilibrium is small ($I_0 \simeq \log_2 H_0$ [6, 11]) and can be neglected. Hence, the total entropy and total free energy of equilibrium are essentially the same as their conventional counterparts.

Each conventional state j has statistical entropy $H_j = 0$, corresponding to a statistical-entropy reduction

$$\Delta H_j = H_j - H_0 = -\log_2 \mathcal{J} \tag{3.3}$$

and a free-energy increase

$$\Delta F_j = -k_B T \ln 2\, \Delta H_j = k_B T \ln 2\, \log_2 \mathcal{J} \; , \tag{3.4}$$

relative to equilibrium. The additional information required to describe a *typical* conventional state j, however, is

$$\Delta I_j \simeq \log_2 \mathcal{J} \; . \tag{3.5}$$

Thus a typical conventional state has the same total entropy as equilibrium, i.e.,

$$\Delta \mathcal{S}_j = \mathcal{S}_j - \mathcal{S}_0 = \Delta H_j + \Delta I_j \simeq 0 \tag{3.6}$$

and, likewise, the same total free energy, i.e.,

$$\Delta\mathcal{F}_j = \mathcal{F}_j - \mathcal{F}_0 = \Delta F_j - k_B T \ln 2\, \Delta I_j = -k_B T \ln 2\, (\Delta H_j + \Delta I_j) \simeq 0 \,. \quad (3.7)$$

Physically, this means that *once the Landauer erasure cost is recognized, no work is available in going from a typical conventional state to equilibrium.*

Equation (3.7) is the contemporary formula for exorcising Maxwell demons [4–6, 10, 11]. A demon-memory can operate an engine cycle in which it first observes the system in equilibrium, finding it in state j (with probability $1/\mathcal{J}$), then extracts work ΔF_j as the system returns to equilibrium, and finally pays a Landauer erasure cost $k_B T \ln 2\, \Delta I_j$ to return to its background state of knowledge. The *net* work extracted *on the average*,

$$\frac{1}{\mathcal{J}}\sum_{j=1}^{\mathcal{J}}(\Delta F_j - k_B T \ln 2\, \Delta I_j) = \frac{1}{\mathcal{J}}\sum_{j=1}^{\mathcal{J}}\Delta\mathcal{F}_j = k_B T \ln 2 \left(\log_2 \mathcal{J} - \frac{1}{\mathcal{J}}\sum_{j=1}^{\mathcal{J}}\Delta I_j\right) \le 0 \,, \quad (3.8)$$

is guaranteed not to be positive by the strict left-hand inequality of the double inequality (2.4). Though the demon-memory cannot win, the soft right-hand inequality implies that it can come close to breaking even, at least in principle. Equation (3.8) illustrates how the Landauer erasure cost leads naturally to the notion of total free energy.

This analysis—and this paper—use the memory's own inside view [5], which assigns to the memory the specific, *minimal* amount of information (relative to a particular universal computer) needed to describe the system state—an amount of information that varies from one system state to another. The memory is also a physical system. Viewed from the outside, the memory should be treated by conventional statistical physics, like any other physical system. Once the demon-memory has observed the system's state, the outside view assigns equal probabilities to the $\mathcal{J}$ different memory configurations that record the system states and thus attributes a statistical entropy $\log_2 \mathcal{J}$ to the memory. Thus, from the outside view, the demon-memory cannot win because its conventional statistical entropy increases after it observes the system and must be reduced to return to its background state of knowledge. The double inequality (2.4) is a consistency condition: it ensures that an average over the inside view is equivalent to the conventional outside view.

The import of the first sentence of this paper is that a system state *implies* the existence of a memory that contains its description. Aside from the computer dependence, *the amount of information I contained in the memory and, hence, the total entropy $\mathcal{S}$ are properties of the system state.* From the inside view, it is a convenient shorthand to say that a system state *has* total entropy $\mathcal{S}$, even though from the outside view the information part of the total entropy is the conventional statistical entropy of another physical system, the memory.

In contrast to typical conventional states, a *simple* conventional state j has $\Delta I_j \ll H_0$, so the information contribution to the total entropy and total free energy can be neglected. A simple conventional state has lower total entropy and higher total free energy than equilibrium by the conventional amounts. Work

$$\Delta\mathcal{F}_j \simeq \Delta F_j = k_B T \ln 2\, \log_2 \mathcal{J} \quad (3.9)$$

must be supplied to transform the system reliably from equilibrium to a simple state and can be extracted as the system returns to equilibrium from a simple state. In conventional statistical physics, simple conventional states and equilibrium constitute the entire subject, precisely because the information needed to describe them can be neglected; typical conventional states are dealt with not individually, but only statistically within the equilibrium ensemble.

IV. TYPICAL PHASE-SPACE PATTERNS AND TYPICAL PURE STATES

Typical conventional states are already too complex algorithmically to be dealt with individually in conventional statistical physics, but they are by no means the most algorithmically complex states that have a statistical-entropy reduction $\Delta H = -\log_2 \mathcal{J}$ relative to equilibrium. Classically, any *pattern* of *fine-grained* phase-space cells whose total phase-space volume is the same as the volume of a coarse-grained cell has $\Delta H = -\log_2 \mathcal{J}$. Quantum mechanically, any pure state—i.e., any linear combination of the orthonormal basis states—has $\Delta H = -\log_2 \mathcal{J}$. The number of fine-grained phase-space patterns is much greater than the number $\mathcal{J}$ of coarse-grained cells and is limited only by the scale of the fine graining; the number of pure states is much greater than the dimension $\mathcal{J}$ of Hilbert space and is limited only by the number of significant digits in the basis-state amplitudes—i.e., in $\mathcal{J}$ probabilities and $\mathcal{J}$ phases [6, 13]. Hence, the information ΔI needed to specify a *typical* fine-grained phase-space pattern or a *typical* pure state, beyond the background information (which must be supplemented by a small amount to specify the scale of the fine graining or the number of digits in the quantum amplitudes), is much greater than the equilibrium entropy $\log_2 \mathcal{J}$ [6]. Thus emerges the key result of this paper: *a typical fine-grained phase-space pattern or a typical pure state has total entropy much greater than equilibrium.*

Although this key result has far-reaching consequences, it is by no means difficult to understand. It is a consequence of dropping the usual insistence that states be distinct. The fine-grained phase-space patterns can overlap, so one is freed from the constraint set by the number of nonoverlapping patterns that can be fitted into the available volume of phase space. The pure states need not be orthogonal, so one is freed from the constraint set by the number of Hilbert-space dimensions.

Figure 1 depicts the resulting entropy and free-energy relationships. The two systems of Fig. 1—one classical and one quantum-mechanical—focus the remainder

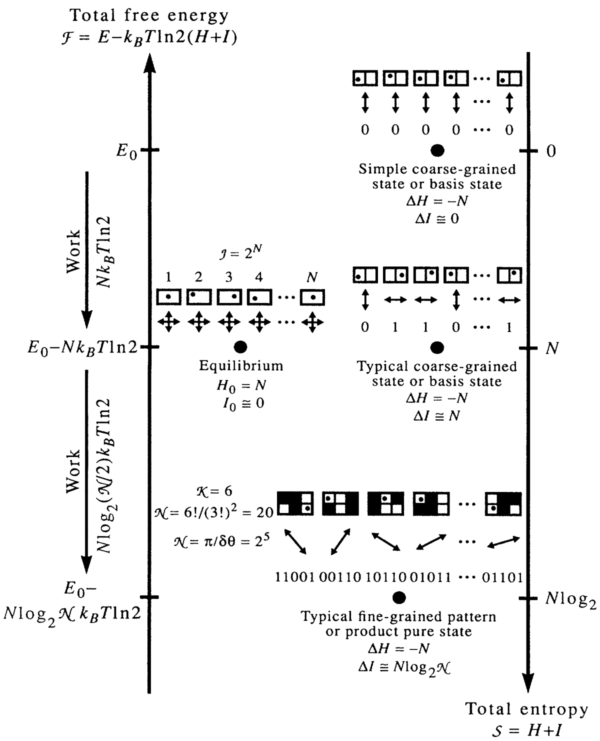

FIG. 1. Total free energy (upward) and total entropy (downward) for a classical system—N molecules enclosed in separate containers—and a quantum system—N linear polarizations (double-ended arrows). There are $\mathcal{J} = 2^N$ *distinct* states: coarse graining divides each container into two equal volumes; basis states are vertical and horizontal polarizations. In equilibrium each molecule explores its entire container, and both polarizations are equally likely. A classical coarse-grained state or quantum basis state has $\Delta H = -N$ and is represented by an N-digit binary string—0 for left or vertical and 1 for right or horizontal. *Typical* classical coarse-grained states or *typical* quantum basis states are represented by *random* strings, whose specification requires $\Delta I \simeq N$ bits; typical coarse-grained states or typical basis states have the same total entropy as equilibrium. *Simple* classical coarse-grained states or *simple* quantum basis states, illustrated by extreme cases in which all molecules are on the left or all polarizations are vertical, are represented by strings that can be specified by $\Delta I \ll N$ bits—in the cases here, essentially one bit, for 0 or 1; simple coarse-grained states or simple basis states have total entropy lower than equilibrium by nearly N bits. Classical fine-grained patterns and quantum pure states, both of which have $\Delta H = -N$, are depicted at the bottom. Fine graining divides each container into $\mathcal{K}$ identical boxes; for each container there are $\mathcal{N} = \mathcal{K}!/[(\mathcal{K}/2)!]^2$ fine-grained patterns, each made up of half the boxes (white in the drawing) and each represented by a $(\log_2 \mathcal{N})$-digit binary code word. For each photon a pure linear-polarization state is represented by an angle θ, defined relative to vertical and given modulo π to $\log_2 \mathcal{N}$ binary digits. A *typical* fine-grained pattern or a *typical* product pure state for all N constituents requires $\Delta I \simeq N \log_2 \mathcal{N}$ bits for its description and thus has total entropy much larger than equilibrium. Work can be extracted in going from equilibrium to a typical fine-grained pattern or a typical pure state.

of the paper. The classical system has N containers, each enclosing a single molecule; the choice of coarse graining divides each container into two equal volumes, which can be separated by a partition. The quantum-mechanical system is made up of the polarizations of N photons; the chosen orthogonal basis states are vertical and horizontal linear polarization. For both systems there are $\mathcal{J} = 2^N$ distinct states.

For the classical system, suppose the fine graining divides each container into $\mathcal{K}$ boxes of equal volume. A fine-grained pattern is formed by selecting half the boxes in each container. These boxes (white in Fig. 1) can be separated physically from the other half of the boxes (black in Fig. 1) by a suitable partition. The number of patterns is $\mathcal{N}^N$, where

$$\mathcal{N} = \frac{\mathcal{K}!}{[(\mathcal{K}/2)!]^2} \tag{4.1}$$

is the number of patterns per container. Suppose the quantum system is restricted to *product* states of *linear* polarizations [14]. Then a pure state is a product of N linear polarizations, each specified by an angle θ relative to vertical. If this polarization angle is given to accuracy $\delta\theta$, the number of pure states is $\mathcal{N}^N$, where

$$\mathcal{N} = \pi/\delta\theta \tag{4.2}$$

is the number of linear polarizations per photon. A fine-grained pattern or a pure state has a statistical-entropy reduction

$$\Delta H = -\log_2 \mathcal{J} = -N \,, \tag{4.3}$$

but the information needed to specify a *typical* pattern or a *typical* pure state is

$$\Delta I \simeq \log_2 \mathcal{N}^N = N \log_2 \mathcal{N} \,. \tag{4.4}$$

For the classical system $\Delta I \sim N\mathcal{K}$ for $\mathcal{K} \gg 1$.

A typical fine-grained pattern or typical pure state has total entropy higher than equilibrium by

$$\Delta \mathcal{S} = \Delta H + \Delta I \simeq N \log_2(\mathcal{N}/2) \tag{4.5}$$

and total free energy lower by

$$\Delta \mathcal{F} = -k_BT \ln 2\, \Delta \mathcal{S} \simeq -k_BT \ln 2\, N \log_2(\mathcal{N}/2) \,. \tag{4.6}$$

To erase the memory's record of a typical fine-grained pattern or a typical pure state, energy $k_BT \ln 2\, N \log_2 \mathcal{N}$ must be dissipated, of which $k_BT \ln 2\, N$ can be recovered in work as the system "expands" to equilibrium. Thus, to go from a typical fine-grained pattern or a typical pure state to equilibrium, energy $k_BT \ln 2\, N \log_2(\mathcal{N}/2) = -\Delta\mathcal{F}$ must be dissipated. There should be a reverse transformation, in which work $-\Delta\mathcal{F}$ is extracted in going from equilibrium to a typical fine-grained pattern or a typical pure state. Section V spells out such a reverse transformation.

V. GEDANKEN EXAMPLES

To describe the transformation requires just one molecule or one photon, so henceforth I specialize to the case $N = 1$. I aim for clarity by presenting the gedanken examples in terms of typical fine-grained patterns or typical pure states, thereby neglecting the possibility of compressing records for simple states. A more rigorous account, particularly for the engine cycles discussed below, uses the double inequality (2.4) to constrain averages, but reaches the same conclusions.

Consider first the classical system, a single molecule enclosed in a container (Fig. 2). The objective is to transform the molecule from equilibrium, where it roams the entire container, to occupation of a fine-grained pattern that is recorded in the memory, while extracting work $-\Delta\mathcal{F} = k_BT\ln 2\,\log_2(\mathcal{N}/2)$. To store the $(\log_2\mathcal{N})$-bit record for which pattern—i.e., the additional information beyond the background information—the "memory" uses $\log_2\mathcal{N}$ binary registers, which are initially in a "standard" state, storing no information. The fine-grained patterns come in complementary pairs (white and black in Fig. 2), for each of which the memory has a partition that separates white and black boxes. There being $\mathcal{N}/2$ pattern pairs, it takes $\log_2(\mathcal{N}/2)$ bits to specify a pair or its partition. Figure 2 shows a three-step transformation.

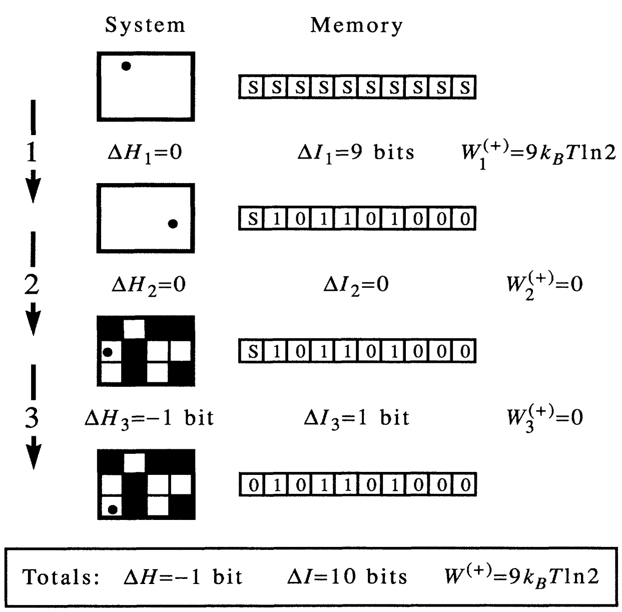

FIG. 2. A single molecule, initially free to explore its entire container, is transformed in three steps to occupy a fine-grained pattern consisting of 6 out of $\mathcal{K} = 12$ boxes. A memory with 10 binary registers, each initially in a standard state (denoted by "S"), stores the 10 bits ($\log_2\mathcal{N} = \log_2[12!/(6!)^2] = \log_2 924 = 9.85$) needed to specify the pattern. Given at each step are the change in the statistical entropy, the change in the number of bits of memory used, and the work extracted. After the transformation, the statistical entropy has changed by $\Delta H = -1$ bit, the memory stores $\Delta I = 10$ bits, and work $W^{(+)} = k_BT\ln 2\,(\Delta H + \Delta I) = 9k_BT\ln 2$ has been extracted.

Step 1: The $\log_2(\mathcal{N}/2)$ memory registers after the first are allowed to randomize, with extraction of work (inverse of erasing [10])

$$W_1^{(+)} = k_BT\ln 2\,\log_2(\mathcal{N}/2)\,, \tag{5.1}$$

after which the memory stores

$$\Delta I_1 = \log_2(\mathcal{N}/2)\text{ bits}\,. \tag{5.2}$$

An explicit way for the memory to perform this step is to use an auxiliary system that has $\mathcal{N}/2$ distinct states, one for each record that can be stored in the $\log_2(\mathcal{N}/2)$ memory registers after the first. This auxiliary system is initially in equilibrium, with equal probability to occupy each of its distinct states. The memory first observes the state of the auxiliary system, storing a record of the observed state in its $\log_2(\mathcal{N}/2)$ registers after the first and thereby reducing the statistical entropy of the auxiliary system by $\log_2(\mathcal{N}/2)$ bits. The memory then uses its record to extract work $W_1^{(+)}$ as the auxiliary system returns to equilibrium, after which the auxiliary system is irrelevant to the further discussion.

Step 2: The memory uses its $[\log_2(\mathcal{N}/2)]$-bit record to select a partition and applies it to the container.

Step 3: The memory observes whether the molecule is in the white or the black part of the container and records

$$\Delta I_3 = 1\text{ bit} \tag{5.3}$$

for white (0) or black (1) in its first register, thereby changing the molecule's statistical entropy by

$$\Delta H_3 = -1\text{ bit}\,. \tag{5.4}$$

After step 3 the molecule occupies a particular pattern, with statistical-entropy reduction

$$\Delta H = \Delta H_3 = -1\text{ bit}\,; \tag{5.5}$$

the memory stores

$$\Delta I = \Delta I_1 + \Delta I_3 = \log_2\mathcal{N}\text{ bits}\,, \tag{5.6}$$

which record which pattern the molecule occupies; and work

$$W^{(+)} = W_1^{(+)} = k_BT\ln 2\,\log_2(\mathcal{N}/2) \tag{5.7}$$

has been extracted, as promised. An alternative transformation lets all $\log_2\mathcal{N}$ memory registers randomize, with extraction of work $k_BT\ln 2\,\log_2\mathcal{N}$, of which $k_BT\ln 2$ is used to "compress" the system into the pattern stored in the memory's record.

It is instructive to consider two engine cycles based on this example. The two cycles differ in what is regarded as the memory's background state of knowledge. The first cycle has the same initial state (situation before step 1): the molecule is in equilibrium, and the memory stores no information. This cycle proceeds through steps 1–3 and adds two further steps to get back to the initial state.

Step 4: The molecule returns to equilibrium with extraction of work

$$W_4^{(+)} = k_B T \ln 2 \,. \tag{5.8}$$

One way to extract this work is to move all the black boxes to one side of the container and all the white boxes to the other side, separating the two by a partition down the middle; then, knowing which side the molecule is on, the memory inserts a piston into the empty side and extracts work $k_B T \ln 2$ as the molecule pushes the piston out of the container.

Step 5: The memory returns to its standard state (all registers empty), paying a Landauer erasure cost

$$W_5^{(-)} = k_B T \ln 2\, \Delta I = k_B T \ln 2\, \log_2 \mathcal{N} \,. \tag{5.9}$$

The net work extracted, $W_1^{(+)} + W_4^{(+)} - W_5^{(-)}$, is zero. In this cycle the background information is the information needed to generate a description of the system at the level of division into boxes. The information gathered by the memory during the cycle includes both the $\log_2(\mathcal{N}/2)$ bits for choosing a partition and the 1 black-or-white bit from observing the molecule.

The second kind of cycle uses the result of step 1 as the initial state: the molecule is in equilibrium, the memory's first register is in the standard state, and the remaining $\log_2(\mathcal{N}/2)$ memory registers store which-partition information. Step 4 is the same, but in step 5 only the first memory register needs to be erased—at cost $k_B T \ln 2$—to return the memory to its background state of knowledge. In this cycle the $\log_2(\mathcal{N}/2)$ bits for choosing a partition are background information, which tells the memory how to partition the container into two equal volumes. The cycle is just a fancy Szilard [15] engine, with the container partitioned in an unusual way instead of by inserting a partition down the middle.

These two cycles emphasize that the change in total entropy must be defined relative to some initial system state *and* to some initial background information. The difference between the two cycles is precisely whether the $\log_2(\mathcal{N}/2)$ which-partition bits are background information, to be carried forward from cycle to cycle, or foreground information, collected afresh during each cycle and erased to return to the initial background information. The total-entropy increase when the molecule is confined to a typical pattern is real, but it must be understood as an increase relative to physical equilibrium *and* to equilibrium background information.

Turn now to photon polarization. The objective is to take an initially unpolarized photon to linear polarization at some angle θ (within $\delta\theta = \pi/\mathcal{N}$) that is recorded in the memory, while extracting work $k_B T \ln 2\, \log_2(\mathcal{N}/2)$. The memory has $\log_2 \mathcal{N}$ binary registers, initially in a standard state, to store the $(\log_2 \mathcal{N})$-bit polarization angle. The analog of the classical partitions is a polarizing beam splitter that, when set at angle θ_p, separates orthogonal polarizations at angles θ_p and $\theta_p + \pi/2$. Figure 3 shows a transformation that mimics the three steps in the classical example.

Step 1: The $\log_2(\mathcal{N}/2)$ memory registers after the first are allowed to randomize, with extraction of work

$$W_1^{(+)} = k_B T \ln 2\, \log_2(\mathcal{N}/2) \,; \tag{5.10}$$

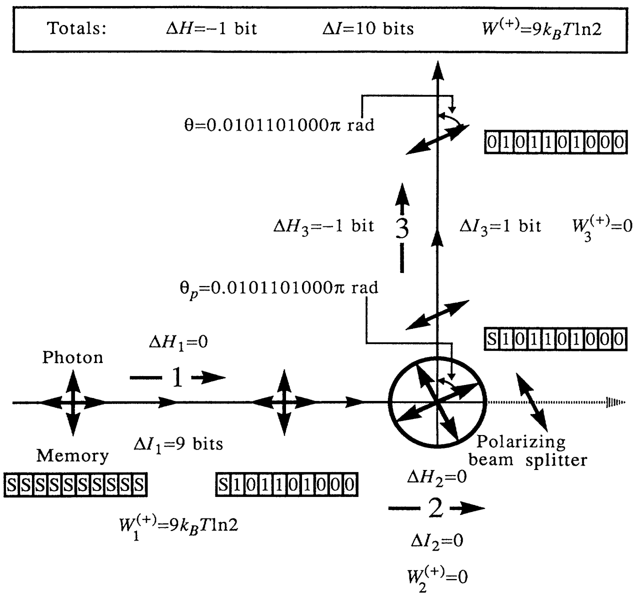

FIG. 3. An unpolarized photon is transformed in three steps to linear polarization at angle θ (relative to the vertical) given to 10 binary digits (accuracy $\delta\theta = 2^{-10}\pi$). The state of linear polarization is depicted in a plane rotated by 90° so that it lies in the plane of the paper. A memory with 10 binary registers, each initially in a standard state (denoted by "S"), stores the 10 angle bits. Given at each step are the change in the statistical entropy, the change in the number of bits of memory used, and the work extracted. After the transformation, the statistical entropy has changed by $\Delta H = -1$ bit, the memory stores $\Delta I = 10$ bits, and work $W^{(+)} = k_B T \ln 2\,(\Delta H + \Delta I) = 9 k_B T \ln 2$ has been extracted.

after this step the memory stores a binary string r of length

$$\Delta I_1 = \log_2(\mathcal{N}/2) \text{ bits} \,. \tag{5.11}$$

Step 2: The memory sets the polarizing beam splitter at angle $\theta_p = 0.0r\,\pi$ ($0.0r$ is the binary representation of θ_p/π), after which the photon passes through the beam splitter.

Step 3: The memory observes the photon's output direction, recording in its first register the

$$\Delta I_3 = 1 \text{ bit} \tag{5.12}$$

for which orthogonal polarization—0 for angle θ_p and 1 for angle $\theta_p + \pi/2$; the observation changes the statistical entropy by

$$\Delta H_3 = -1 \text{ bit} \,. \tag{5.13}$$

After step 3 the photon has linear polarization at angle $\theta = \theta_p = 0.0r\,\pi$ or at angle $\theta = \theta_p + \pi/2 = 0.1r\,\pi$ and thus has statistical-entropy reduction

$$\Delta H = \Delta H_3 = -1 \text{ bit} \,; \tag{5.14}$$

the memory stores the polarization angle modulo π as the binary string $0r$ or $1r$ of length

$$\Delta I = \Delta I_1 + \Delta I_3 = \log_2 \mathcal{N} \text{ bits ;} \tag{5.15}$$

and work

$$W^{(+)} = W_1^{(+)} = k_B T \ln 2 \, \log_2(\mathcal{N}/2) \tag{5.16}$$

has been extracted.

In both examples the memory's record splits naturally into two parts: $\log_2(\mathcal{N}/2)$ bits that tell the memory *how* to observe the system—which partition or which beam splitter angle—and 1 bit from the subsequent observation. The $\log_2(\mathcal{N}/2)$ bits of which-observation information are not stored in the system and cannot be obtained by observing it. The $\log_2(\mathcal{N}/2)$ which-observation registers are "called into existence" in a standard state when one contemplates storing a record of a typical fine-grained pattern or a typical pure state. The work $W^{(+)}$ comes wholly from randomizing these which-observation registers.

From the outside view of conventional statistical physics, the standard memory state is a low-entropy, high-free-energy state. Supplying memory registers in the standard state is the same as supplying fuel that can be turned into useful work [10]. Indeed, from the outside view, the work is extracted as the which-observation registers go from the low-entropy standard state, far from equilibrium, to the high-entropy equilibrium state, in which they store a record of how to observe the system.

The inside and outside views agree on the *average* work extracted or dissipated in steps 1–5. Thus both views agree on the *average* free-energy decrease in going from the system in equilibrium, with a record of equilibrium stored in the memory, to the system's occupying a fine-grained pattern or pure state, with a record of the pattern or pure state stored in the memory. The outside view achieves all this using only the conventional statistical entropies of the system and the memory. The inside view agrees with the conventional outside view *on the average*, but it has the advantage that it gives an account—indeed, it must give an account—of each memory state. This advantage is of little consequence in analyzing an engine cycle, since only average behavior is relevant for questions of thermodynamic efficiency, but it becomes important when one applies the concept of total entropy to system dynamics that begin with a particular initial state.

VI. DISCUSSION

This paper provides background and motivation for addressing the following question: can Hamiltonian dynamics, either classical or quantum-mechanical, take a simple initial state—i.e., an algorithmically simple phase-space cell or an algorithmically simple quantum basis state—to one of the typical fine-grained phase-space patterns or typical quantum pure states, which are algorithmically extremely complex and thus have high total entropy? This question, foreshadowed by Zurek's [5] analysis of classical ergodic, but nonmixing systems, is only beginning to be investigated. Initial work [6] indicates that the answer is no, at least for reasonable times. The negative answer holds both for classical systems, either regular or chaotic, and for quantum systems.

Things get more interesting, however, when one realizes that the state that evolves from a simple initial state under classical chaotic—but not regular—evolution or under quantum-mechanical evolution, although itself algorithmically simple, can be easily perturbed into one of the typical algorithmically complex phase-space patterns or pure states [6]. Detailed analyses of a stochastically perturbed version of the baker's map [7], a prototype for classical chaos, and of a stochastically perturbed version of a quantized baker's map [8], indicate that the algorithmic information needed to keep track of the perturbed state grows extremely fast. Indeed, the growth of algorithmic information is much faster than the increase in statistical entropy that follows from averaging over the stochastic perturbation. These detailed analyses, together with the heuristic analyses [6] that motivated them, have deep implications for the second law of thermodynamics and for the nature of irreversibility and, moreover, hint at a previously unrecognized connection between classical chaos and quantum mechanics.

One further question can be addressed profitably to the gedanken examples in Sec. V: how are classical and quantum systems different? Successive observations of the classical molecule, using different partitions, eventually isolate the molecule in a single box, thus uncovering the fine-grained structure beneath the initial coarse graining. Successive observations of the photon, using different beam splitter angles, yield fresh information as long as there is memory space to store it, never revealing any structure beneath the pure states (no hidden variables). This ability of quantum systems to manufacture new information is the central mystery of quantum mechanics, stated in information-theoretic language.

ACKNOWLEDGMENTS

I thank W. G. Unruh for challenging me to come up with the examples in Figs. 2 and 3, and I thank S. Lloyd, R. Schack, and W. H. Zurek for advising me how to revise this paper from its original, shorter form. This work was supported in part by the Complexity, Entropy, and Physics of Information program at the Santa Fe Institute.

[1] R. Landauer, IBM J. Res. Develop. **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 University, Cambridge, England, 1993).
[7] R. Schack and C. M. Caves, Phys. Rev. Lett. **69**, 3413 (1992).

[8] R. Schack and C. M. Caves (unpublished).
[9] A. K. Zvonkin and L. A. Levin, Usp. Mat. Nauk **25** (6), 85 (1970) [Russ. Math. Surveys **25** (6), 83 (1970)].
[10] C. H. Bennett, Int. J. Theor. Phys. **21**, 905 (1982).
[11] C. M. Caves, in *Complexity, Entropy and the Physics of Information*, edited by W. H. Zurek (Addison-Wesley, Redwood City, CA, 1990), p. 91.
[12] R. G. Gallager, *Information Theory and Reliable Communication* (Wiley, New York, 1968).
[13] I. C. Percival, in *Quantum Chaos, Quantum Measurement*, edited by P. Cvitanović, I. Percival, and A. Wirzba, (Kluwer, Dordrecht, Holland, 1992), p. 199. Percival emphasizes that so much information is required to specify a typical pure state that it is impossible in practice to prepare one.
[14] If one allows arbitrary pure states within the ($\mathcal{J} = 2^N$)-dimensional Hilbert space of the N photon polarizations, then the information to specify a typical state is $\Delta I \simeq \mathcal{J} \times$(number of bits to specify a real amplitude and a phase for each Hilbert-space dimension). The number of pure states, at the Hilbert-space resolution defined by the number of digits in the amplitudes and phases, is $\mathcal{N} \simeq 2^{\Delta I}$—an enormous number even for Hilbert spaces with a modest number of dimensions. A more rigorous discussion of this point can be found in Ref. [6].
[15] L. Szilard, Z. Phys. **53**, 840 (1929) [English translation in *Quantum Theory and Measurement*, edited by J. A. Wheeler and W. H. Zurek (Princeton University, Princeton, NJ, 1983), p. 539].