Santa Fe Institute Working Paper 13-09-030
arxiv.org:1309.5504 [nlin.CD]

Chaos Forgets and Remembers: Measuring Information Creation, Destruction, and Storage

Ryan G. James,[1,*] Korana Burke,[1,†] and James P. Crutchfield[1,2,‡]

[1] *Complexity Sciences Center and Physics Department, University of California at Davis, One Shields Avenue, Davis, CA 95616*
[2] *Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501*

(Dated: December 16, 2013)

The hallmark of deterministic chaos is that it creates information—the rate being given by the Kolmogorov-Sinai metric entropy. Since its introduction half a century ago, the metric entropy has been used as a unitary quantity to measure a system's intrinsic unpredictability. Here, we show that it naturally decomposes into two structurally meaningful components: A portion of the created information—the ephemeral information—is forgotten and a portion—the bound information—is remembered. The bound information is a new kind of intrinsic computation that differs fundamentally from information creation: it measures the rate of active information storage. We show that it can be directly and accurately calculated via symbolic dynamics, revealing a hitherto unknown richness in how dynamical systems compute.

Keywords: chaos, entropy rate, bound information, Shannon information measures, information diagram, Tent map, Logistic map, Lozi map

PACS numbers: 05.45.-a 89.75.Kd 89.70.+c 05.45.Tp

The world is replete with systems that generate information—information that is then encoded in a variety of ways: Erratic ant behavior eventually leads to intricate, structured colony nests [1, 2]; thermally fluctuating magnetic spins form complex domain structures [3]; music weaves theme, form, and melody with surprise and innovation [4]. We now appreciate that the underlying dynamics in such systems is frequently deterministic chaos [5, 6]. In others, the underlying dynamics appears to be fundamentally stochastic [7]. For continuous-state systems, at least, one operational distinction between deterministic chaos and stochasticity is found in whether or not information generation diverges with measurement resolution [8]. This result calls back to Kolmogorov's original use [9] of Shannon's mathematical theory of communication [10] to measure a system's rate of information generation in terms of the metric entropy. Since that time, metric entropy has been understood as a unitary quantity. Whether deterministic or stochastic, it is a system's degree of unpredictability. Here, we show that this is far too simple a picture—one that obscures much.

To ground this claim, consider two systems. The first, a fair coin: Each flip is independent of the others, leading to a simple uncorrelated randomness. As a result, no statistical fluctuation is predictively informative. For the second system consider a stock traded via a financial market: While its price is unpredictable, the direction and magnitude of fluctuations can hint at its future behavior. (This, at least, is the guiding assumption of the now-global financial engineering industry.) We make this distinction rigorous here, dividing a system's information generation into a component that is relevant to temporal structure and a component divorced from it. We show that the temporal component captures the system's internal information processing and, therefore, is of practical interest when harnessing the chaotic nature of physical systems to build novel machines and devices [11]. We first introduce the new measures, describe how to interpret and calculate them, and then apply them via a generating partition to analyze several dynamical systems—the Logistic, Tent, and Lozi maps—revealing a previously hidden form of active information storage.

We observe these systems via an optimal measuring instrument—called a generating partition—that encodes all of their behaviors in a *stationary process*: A distribution $\Pr(\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots)$ over a bi-infinite sequence of random variables with shift-invariant statistics. A contiguous block of observations $X_{t:t+\ell}$ begins at index t and extends for length ℓ. (The index is inclusive on the left and exclusive on the right.) If an index is infinite, we leave it blank. So, a process is compactly denoted $\Pr(X_{:})$. Our analysis splits $X_{:}$ into three segments: the *present* X_0, a single observation; the *past* $X_{:0}$, everything prior; and *future* $X_{1:}$, everything that follows.

The information-theoretic relationships between these three random variable segments are graphically expressed in a Venn-like diagram, known as an I-diagram [12]; see Fig. 1. The rate h_μ of information generation is the amount of new information in an observation X_0 given

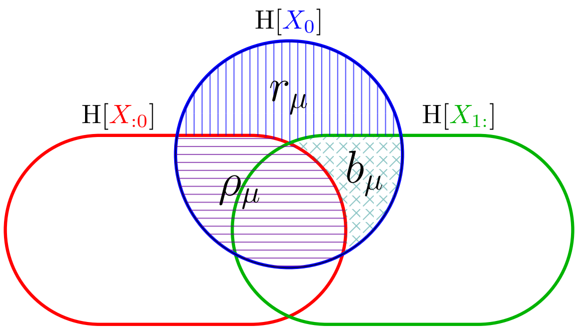

FIG. 1. A process's I-diagram showing how the past $X_{:0}$, present X_0, and future $X_{1:}$ partition each other into seven distinct information *atoms*. We focus only on the four regions contained in the *present information* $\mathrm{H}[X_0]$ (blue circle). That is, the present decomposes into three components: ρ_μ (horizontal lines), r_μ (vertical lines), and b_μ (diagonal crosshatching). The *redundant* information ρ_μ overlaps with the *past* $\mathrm{H}[X_{:0}]$; the *ephemeral* information r_μ falls outside both the *past* and the *future* $\mathrm{H}[X_{1:}]$. The *bound* information b_μ is that part of $\mathrm{H}[X_0]$ which is in the future yet not in the past.

all the prior observations $X_{:0}$:

$$h_\mu = \mathrm{H}[X_0|X_{:0}] \ , \tag{1}$$

where $\mathrm{H}[Y|Z]$ denotes the Shannon conditional entropy of random variable Y given variable Z. This quantity arises in various contexts and goes by many names: e.g., the Shannon entropy rate and the Kolmogorov-Sinai metric entropy, mentioned above [8]. The complement of the entropy rate is the *predicted information* ρ_μ:

$$\rho_\mu = \mathrm{I}[X_{:0} : X_0] \ , \tag{2}$$

where $\mathrm{I}[Y : Z]$ denotes the mutual information between random variables Y and Z [12]. Hence, ρ_μ is the information in the present that can be predicted from prior observations. Together, we have a decomposition of the information contained in the present: $\mathrm{H}[X_0] = h_\mu + \rho_\mu$.

A simple application of the entropy chain rule [12] to Eq. (1) leads us to a different view:

$$\begin{aligned} h_\mu &= \mathrm{I}[X_0 : X_{1:}|X_{:0}] + \mathrm{H}[X_0|X_{:0}, X_{1:}] \\ &= b_\mu + r_\mu \ . \end{aligned} \tag{3}$$

This introduces two new information measures:

$$b_\mu = \mathrm{I}[X_0 : X_{1:}|X_{:0}] \text{ and} \tag{4}$$

$$r_\mu = \mathrm{H}[X_0|X_{:0}, X_{1:}] \ . \tag{5}$$

That is, created information (h_μ) decomposes into two parts: information (b_μ) shared by the present and the future but not in the past and information (r_μ) in the present but in neither the past nor the future.

The r_μ component was first studied by Verdú and Weissman [13] as the *erasure entropy* (their H^-) to measure information loss in erasure channels. To emphasize that it is information existing only in a single moment—created and then immediately forgotten—we refer to r_μ as the *ephemeral information*. The second component b_μ we call the *bound information* since it is information created in the present that the system stores and that goes on to affect the future [14]. It was first studied as a measure of "interestingness" in computational musicology by Abdallah and Plumbley [15]. For a more complete analysis of this decomposition, as well as computation methods and related measures, see Ref. [16].

Isolating the information $\mathrm{H}[X_0]$ contained in the present and identifying its components provides the partitioning illustrated in Fig. 1. This is a particularly intuitive way of thinking about the information contained in an observation. While, some behavior (ρ_μ) can be predicted, the rest ($h_\mu = b_\mu + r_\mu$) cannot. Of that which cannot be predicted, some (b_μ) plays a role in the future behavior and some (r_μ) does not. As such, this is a natural decomposition of a time series; one that results in a semantic dissection of the entropy rate.

By way of an example, consider a few simple processes and how their present information decomposes into these three components. A periodic process of alternating 0s and 1s ($\ldots 01010101 \ldots$) has $\mathrm{H}[X_0] = 1$ bit since 0s and 1s occur equally often. Given a prior observation, one can accurately predict exactly which symbol will occur next and so $\mathrm{H}[X_0] = \rho_\mu = 1$ bit, while $r_\mu = b_\mu = 0$ bits. On the other extreme is a fair coin flip. Again, each outcome is equally likely and so $\mathrm{H}[X_0] = 1$ bit. However, each flip is independent of all others and so $\mathrm{H}[X_0] = r_\mu = 1$ bit, while $\rho_\mu = b_\mu = 0$ bits.

Between these two extrema lie interesting processes: those with *stochastic structure*. Processes expressing a fixed template, like the periodic process above, contain a finite amount of information. Those with stochastic structure, however, constantly generate information and store it in the form of patterns. Being neither purely predictable nor independently random, these patterns are captured by b_μ. The more intricate the organization, the larger b_μ. More to the point, generating these patterns requires intrinsic computation in a system—information creation, storage, and transformation [17]. We propose b_μ as a simple method of discovering this type of physical computation: Where there are intricate patterns, there is sophisticated processing.

How useful is the proposed decomposition and its mea-

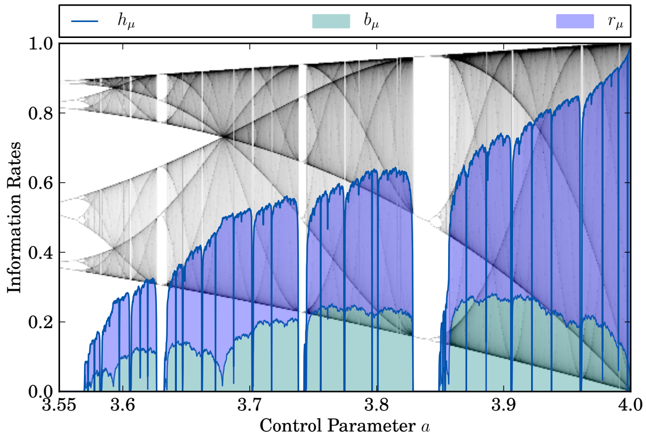

FIG. 2. Logistic map information anatomy as a function of control parameter a: Bound information b_μ is the lower (green shaded) component; ephemeral information r_μ is the upper (blue shaded) component. Entropy rate is the top (blue) line: $h_\mu = b_\mu + r_\mu$. As reference to the dynamical behavior, the map's bifurcation diagram is displayed in the background.

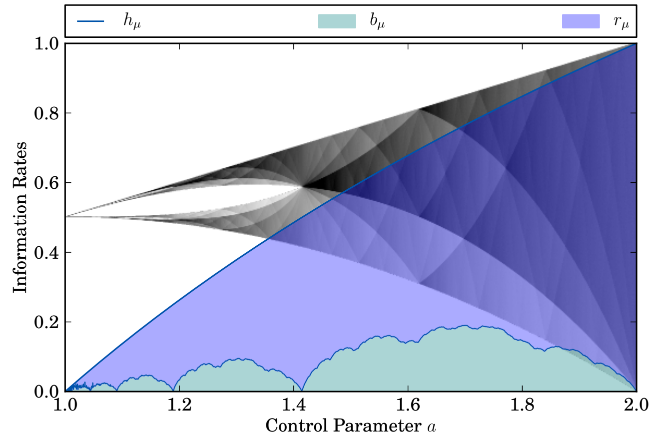

FIG. 3. Tent map information anatomy: Although $h_\mu = b_\mu + r_\mu$ is a smooth function of control—$h_\mu = \log_2 a$—the decomposition into bound and ephemeral informations is not. Graphics layout as in previous figure.

sures? To answer this we analyze several discrete-time chaotic dynamical systems—the Logistic and Tent maps of the interval and the Lozi map of the plane—uncovering a number of novel properties embedded in these familiar and oft-studied systems. As an independent calibration for the measures, we employ Pesin's theorem [18]: h_μ is the sum of the positive Lyapunov characteristic exponents (LCEs). The maps here have at most one positive LCE λ, so $h_\mu = \max\{0, \lambda\}$. The symbols $s_0, s_1, s_2, \ldots, s_N$ for each process we analyze come from a generating partition. We produce a long sample of $N \approx 10^{10}$ symbols, extracting subsequence statistics via a sliding window [19]. Each window consists of a past, present, and future symbol sequence and we estimate r_μ and b_μ using truncated forms of Eqs. (4) and (5).

Consider first the Logistic map, perhaps one of the most studied chaotic systems:

$$x_{n+1} = a x_n (1 - x_n) \ , \tag{6}$$

where $a \in [0, 4]$ is the control parameter and the initial condition is $x_0 \in [0, 1]$. Its generating partition is defined by:

$$s_n = \begin{cases} 0 & \text{if } x_n < \frac{1}{2} \\ 1 & \text{if } x_n \geq \frac{1}{2} \end{cases} \ . \tag{7}$$

Figure 2 shows the resulting measures as a function of control a, with the map's bifurcation diagram displayed in the background for reference.

The first point of interest is that the system's information generation is, in fact, a mixture of ephemeral (r_μ) and bound (b_μ) informations at nearly all chaotic ($h_\mu > 0$) parameter values. The second is that the division into the two components varies in a nontrivial way as a function of the control parameter a. Moreover, the boundary between the two appears nondifferentiable. At first blush, this is not surprising given that their sum h_μ ($= \lambda$) is known to be nondifferentiable. Finally, b_μ vanishes nontrivially only at parameters that coincide with the merging of the chaotic bands (e.g., $a = 4.0, 3.67857\ldots, 3.59257\ldots, \ldots$). Thus, the information generated by the Logistic map at these parameters is entirely forgotten.

Is the complex and nondifferentiable boundary between r_μ and b_μ simply a consequence of the entropy rate's complicated behavior or due a dynamical mechanism distinct from information creation? We answer this by analyzing the Tent map:

$$x_{n+1} = \frac{a}{2} \left(1 - 2 \left| x_n - \frac{1}{2} \right| \right) \ , \tag{8}$$

where $a \in [0, 2]$ is the control parameter. The generating partition for the Tent map is the same as for the Logistic map. Since the Tent map is piecewise linear, its Lyapunov exponent is simply $\lambda = \log_2 a$ and, by Pesin's theorem, so is the information generation $h_\mu = \log_2 a$; a rather smooth parameter dependence. As a result, the intricate structures exhibited in the Tent map's bifurcation diagram cannot be resolved by studying solely the behavior of the Lyapunov exponent (or h_μ) itself. Figure 3 demonstrates that, despite the entropy rate's simple logarithmic dependence on control, its decomposition

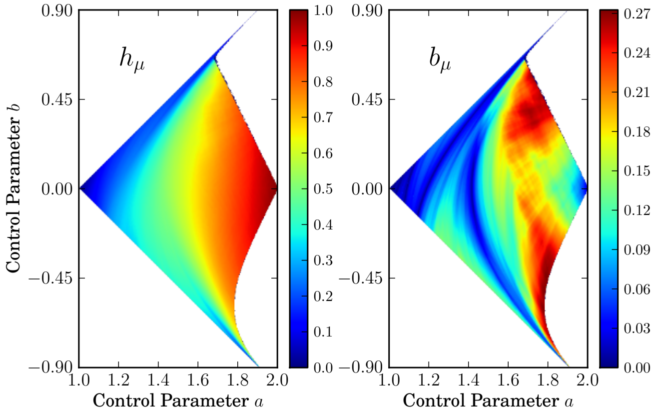

FIG. 4. Lozi map information anatomy: (Left) h_μ as a function of controls a and b. (Right) b_μ similarly. b_μ is maximized on the upper-right and lower-right edges of the a-b region that supports an attractor near the origin.

$h_\mu = b_\mu + r_\mu$ is not a smooth function of a. To emphasize, in sharp contrast with h_μ's simplicity, r_μ and b_μ again appear nondifferentiable—a complexity masked by the smooth h_μ. Thus, the two informational components capture a property in the chaotic system's behavior that is both quantitatively and qualitatively new. As with the Logistic map, we once again find that the bound information vanishes and that all of the information the Tent map generates is forgotten ($h_\mu = r_\mu$) at parameters corresponding to merging of chaotic bands ($a = 2^{1/2^k}, k = 0, 1, 2, \ldots$). In the Supplementary Materials we show how to calculate b_μ and r_μ in closed form for the Tent map at Misiurewicz parameters.

To explore how these measures apply more generally, we extend information anatomy to two dimensions by analyzing the Lozi map:

$$\begin{aligned} x_{n+1} &= 1 - a\,|x_n| + y_n \\ y_{n+1} &= b x_n \ . \end{aligned} \tag{9}$$

The map exhibits an attractor near the origin within a diamond-shaped parameter region inside $(a, b) \in [1, 2] \times [-0.9, 0.9]$. Note that when $b = 0$ the map becomes isomorphic to the Tent map. The generating partition is given by:

$$s_n = \begin{cases} 0 & \text{if } x_n < 0 \\ 1 & \text{if } x_n \geq 0 \end{cases} . \tag{10}$$

Figure 4 shows h_μ (left) and b_μ (right) in the attracting parameter region. Mirroring the Tent map, the Lozi map's entropy rate varies smoothly over the attractor region, whereas b_μ varies in a more complicated manner. There are swaths of low b_μ corresponding to "fuzzy" mergings of chaotic bands. Notably, while the maximal h_μ occurs along the line $b = 0$, maximal b_μ occurs far from $b = 0$. Hence, large h_μ does not necessarily imply large bound information b_μ.

To sum up, we showed that a process's information creation rate decomposes, via a chain rule, into two structurally meaningful components. The components, the ephemeral information r_μ and the bound information b_μ, provide direct insights into a system's behavior without detailed modeling or appealing to domain-specific knowledge. That is to say, they are relatively easily defined measures that can be straightforwardly estimated. More to the point, however, b_μ is a strong indicator of intrinsic computation. While related to information generation, we demonstrated that it captures a different kind of informational processing—a mechanism that actively stores information.

Concretely, decomposing information creation in the symbolic dynamics of the Logistic, Tent, and Lozi systems delineated the topography of their intrinsic-computation landscape. Awareness of this rich (and previously hidden) landscape will lead to improved engineering of natural systems as substrates for information processing [11]. And, it will lead to an expanded understanding of evolved information processing systems, such as the linguistic processes comprising human natural languages. A sequel will develop the decomposition further, including a geometric interpretation of active information storage that parallels the geometric view of information creation expressed in the Lyapunov exponents.

KB is supported by a UC Davis Chancellor's Post-Doctoral Fellowship. This work was partially supported by ARO grant W911NF-12-1-0288.

* rgjames@ucdavis.edu
† kburke@ucdavis.edu
‡ chaos@ucdavis.edu

[1] E. Bonabeau, G. Theraulaz, and M. Dorigo, editors. *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, New York, 1999.

[2] S. Camazine, J.-L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulas, and E. Bonabeau. *Self-Organization in Biological Systems*. Princeton University Press, New York, 2003.

[3] J. J. Binney, N. J. Dowrick, A. J. Fisher, and M. E. J. Newman. *The Theory of Critical Phenomena*. Oxford University Press, Oxford, 1992.

[4] B. R. Simms. *Music of the Twentieth Century: Style and Structure*. Schirmer Books, New York, second edition,

1996.

[5] S. H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering.* Westview Press, 1994.

[6] J. B. Jose and E. J. Saletan. *Classical Dynamics: A Contemporary Approach.* Cambridge University Press, New York, 1998.

[7] R. F. Streater. *Statistical Dynamics: A Stochastic Approach to Nonequilibrium Thermodynamics.* Imperial College Press, London, second edition, 2009.

[8] P. Gaspard and X.-J. Wang. Noise, chaos, and (epsilon,tau)-entropy per unit time. *Physics Reports*, 235(6):291–343, 1993.

[9] A. N. Kolmogorov. Entropy per unit time as a metric invariant of automorphisms. *Dokl. Akad. Nauk. SSSR*, 124:754, 1959. (Russian) Math. Rev. vol. 21, no. 2035b.

[10] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948.

[11] W. L. Ditto, A. Miliotis, K. Murali, S. Sinha, and M. L. Spano. Chaogates: Morphing logic gates that exploit dynamical patterns. *Chaos*, 20(3):037107, 2010.

[12] R. W. Yeung. *Information Theory and Network Coding.* Springer, New York, 2008.

[13] S. Verdú and T. Weissman. The Information Lost in Erasures. *IEEE Trans. Info. Th.*, 54(11):5030–5058, 2008.

[14] Our terminology avoids the misleading use of the phrase "predictive information" for b_μ. The latter is not the amount of information needed to predict the future. Rather, it is part of the *predictable* information—that portion of the future which *can* be predicted.

[15] S. A. Abdallah and M. Plumbley. Information dynamics: Patterns of expectation and surprise in the perception of music. *Connection Science*, 21(2):89–117, June 2009.

[16] R. G. James, C. J. Ellison, and J. P. Crutchfield. Anatomy of a Bit: Information in a Time Series Observation. *Chaos*, 21(3):1–15, 2011.

[17] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Let.*, 63:105–108, 1989.

[18] Y. B. Pesin. Characteristic Lyapunov exponents and smooth ergodic theory. *Russ. Math. Surveys*, 32(4):55–114, 1977.

[19] Window width is adaptively chosen in inverse proportion to the LCE. When the latter is low we use a longer window than when the system is fully chaotic. The minimum window width of $L = 31$ and adaptive widths were chosen so that numerical estimates varied by less than 0.01% when the width is incremented.

Chaos Forgets and Remembers: Measuring Information Creation, Destruction, and Storage

Supplementary Material

Ryan G. James, Korana Burke, and James P. Crutchfield

COMPUTING BOUND AND EPHEMERAL INFORMATIONS ANALYTICALLY

To obviate the data requirements for accurately estimating b_μ from a time series, we present a method for computing it analytically, in closed form. An analytic expression is possible if one can construct forward-time and reverse-time models of the system. These models are sufficient statistics of the past about the present and future, and the future about the present and past, respectively. Here, we use ϵ-machines [S1, S2], which are the minimal sufficient statistics. From these, b_μ can be computed via:

$$b_\mu = \mathrm{I}[X_0 : \mathcal{S}_0^+ | \mathcal{S}_1^-] \ , \tag{11}$$

where $\mathcal{S}_0^+$ is the forward ϵ-machine's state random variable at time 0—the minimal sufficient statistic of the past about the present and future—and $\mathcal{S}_1^-$ is the reverse-time ϵ-machine's state random variable at time 1—the minimal sufficient statistic of the future about the present and future. Due to their standing as sufficient statistics, these states stand in for the future $X_{1:}$ and the past $X_{:0}$ of Eq. (4).

We explicitly implement this calculation for one parameter value of the Tent map. In particular, consider the Misiurewicz point where $f^4(\frac{1}{2}) = f^5(\frac{1}{2})$. Solving this constraint gives parameter value:

$$\begin{aligned} a &= \alpha + \frac{2}{3\alpha} \\ &= 1.76929235\ldots \end{aligned} \tag{12}$$

where $\alpha = \sqrt[3]{\sqrt{\frac{19}{27}} + 1}$. There, the Tent map admits a Markov partition [S3], as Fig. 5 demonstrates. From this, a Markov chain is constructed and the generating partition overlaid. The result is the hidden Markov model of Fig. 6(right) that exactly describes the map's symbolic dynamics stochastic process.

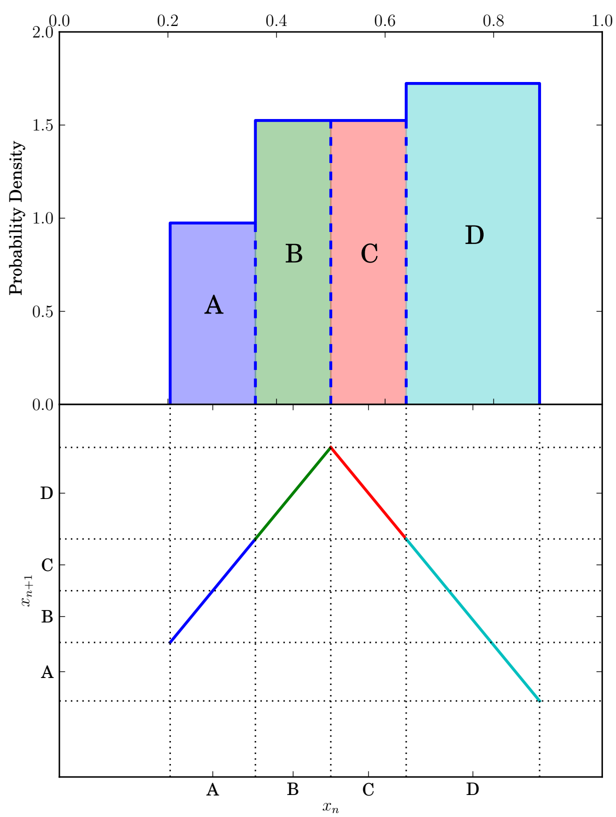

FIG. 5. (Above) Tent map's invariant distribution at parameter a as defined in Eq. (12), consisting of three contiguous, uniformly distributed parts, each differently colored for clarity. (Below) The same colors superimposed on the map itself along with (dotted line) guides to show that the uniform components of the invariant distribution do indeed form a Markov partition.

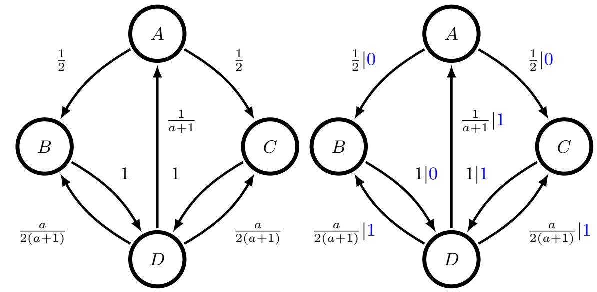

FIG. 6. (Left) Markov chain induced by the Markov partition. (Right) The generating partition applied to the transitions, resulting in a hidden Markov model that describes the Tent map's symbolic stochastic process.

Equation (11) requires the model to be a sufficient statistic to calculate b_μ. And so, we transform the hidden Markov model of Fig. 6(right) to one that is *unifilar* and, in particular, to the *ϵ-machine* of Fig. 7.

As the final step, we construct the process's *bidirectional machine* [S1, S2] from the ϵ-machine; the result is shown in Fig. 8. Then, from it we calculate the joint distribution $\Pr(\mathcal{S}_0^+, \mathcal{S}_0^-, X_0, \mathcal{S}_1^+, \mathcal{S}_1^-)$. This, in turn, allows one to calculate $b_\mu = \mathrm{H}[X_0 | \mathcal{S}_0^+, \mathcal{S}_1^-]$ and $r_\mu = \mathrm{I}[X_0 : \mathcal{S}_1^- | \mathcal{S}_0^+]$. We find that the Tent map at the

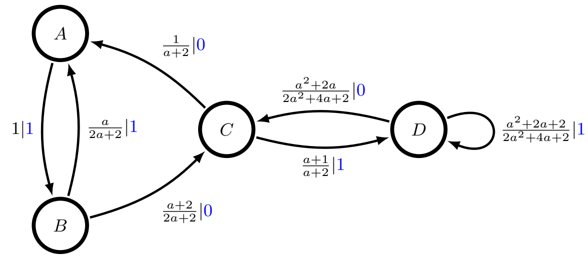

FIG. 7. The (unifilar) ϵ-machine for the hidden Markov model of Fig. 6(right). That is, for each state there is at most a single outgoing transition labeled with each symbol. This makes the states a function of the past $X_{:0}$ and so allows for the required calculation.

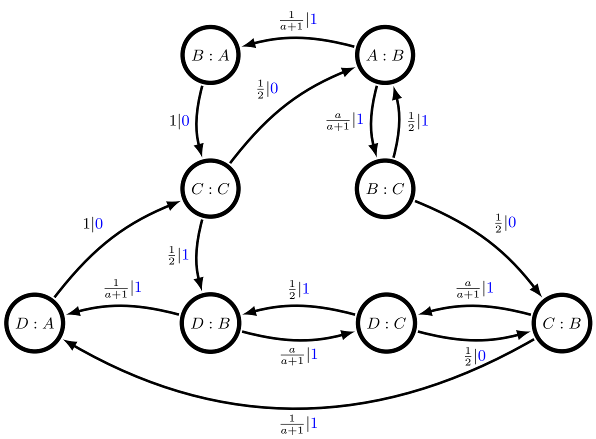

FIG. 8. Bidirectional machine of the stochastic process generated by the Tent map's symbolic dynamics at the Misiurewicz parameter a. Its states are pairs $\mathcal{S}_0^+ : \mathcal{S}_0^-$. By utilizing the dynamic (the edges that connect states), it allows one to directly calculate $\mathrm{H}[X_0|\mathcal{S}_0^+, \mathcal{S}_1^-]$ and, thus, r_μ and b_μ.

Misiurewicz parameter a has the following information measures (in bits per step):

$$
\begin{aligned}
h_\mu &= \log_2 a = \log_2 \left(\frac{\sqrt[3]{9+\sqrt{57}} + \sqrt[3]{9-\sqrt{57}}}{3^{\frac{2}{3}}} \right) \\
&= 0.823172\ldots \\
r_\mu &= \frac{1}{4}\left(3 - \frac{2}{a+1} - \frac{4}{a+2} + \frac{9}{2a+3}\right) \\
&= \frac{1}{9}\left(\frac{\sqrt[3]{207\sqrt{57}-1349}}{19^{2/3}} - \frac{32}{\sqrt[3]{19\left(207\sqrt{57}-1349\right)}} + 7 \right) \\
&= 0.648258\ldots \\
b_\mu &= h_\mu - r_\mu \\
&= 0.174915\ldots\ .
\end{aligned}
$$

Supplementary References

S1. J. P. Crutchfield, C. J. Ellison, and J. R. Mahoney, "Time's Barbed Arrow: Irreversibility, Crypticity, and Stored Information", Phys. Rev. Lett. **103**:9 (2009) 094101.

S2. C. J. Ellison, J. R. Mahoney, and J. P. Crutchfield, "Prediction, Retrodiction, and the Amount of Information Stored in the Present", J. Stat. Phys. **136**:6 (2009) 1005–1034.

S3. D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1999.