Title

Exact complexity: The spectral decomposition of intrinsic computation

Permalink

Journal

ISSN

Authors

Crutchfield, James P
Ellison, Christopher J
Riechers, Paul M

Publication Date

DOI

Peer reviewed

Exact complexity: The spectral decomposition of intrinsic computation

James P. Crutchfield [a,*], Christopher J. Ellison [b], Paul M. Riechers [a]

[a] *Complexity Sciences Center and Department of Physics, University of California at Davis, One Shields Avenue, Davis, CA 95616, United States*
[b] *Center for Complexity and Collective Computation, University of Wisconsin-Madison, Madison, WI 53706, United States*

A B S T R A C T

We give exact formulae for a wide family of complexity measures that capture the organization of hidden nonlinear processes. The spectral decomposition of operator-valued functions leads to closed-form expressions involving the full eigenvalue spectrum of the mixed-state presentation of a process's ϵ-machine causal-state dynamic. Measures include correlation functions, power spectra, past-future mutual information, transient and synchronization informations, and many others. As a result, a direct and complete analysis of intrinsic computation is now available for the temporal organization of finitary hidden Markov models and nonlinear dynamical systems with generating partitions and for the spatial organization in one-dimensional systems, including spin systems, cellular automata, and complex materials via chaotic crystallography.

© 2016 Elsevier B.V. All rights reserved.

The emergence of organization in physical, engineered, and social systems is a fascinating and now, after half a century of active research, widely appreciated phenomenon [1–5]. Success in extending the long list of instances of emergent organization, however, is not equivalent to understanding what organization itself is. How do we say objectively that new organization has appeared? How do we measure quantitatively how organized a system has become?

Computational mechanics' answer to these questions is that a system's organization is captured in how it stores and processes information—how it computes [6]. *Intrinsic computation* was introduced two decades ago to analyze the inherent information processing in complex systems [7]: How much history does a system remember? In what architecture is that information stored? And, how does the system use it to generate future behavior?

Computational mechanics, though, is part of a long historical trajectory focused on developing a physics of information [8–10]. That nonlinear systems actively process information goes back to Kolmogorov [11], who adapted Shannon's communication theory [12] to measure the information production rate of chaotic dynamical systems. In this spirit, today computational mechanics is routinely used to determine physical and intrinsic computational properties in single-molecule dynamics [13], in complex materials

[14], and even in the formation of social structure [15], to mention several recent examples.

Thus, measures of complexity are important to quantifying how organized nonlinear systems are: their randomness and their structure. Moreover, we now know that randomness and structure are intimately intertwined. One cannot be properly defined or even practically measured without the other [16, and references therein].

Measuring complexity has been a challenge: Until recently, in understanding the varieties of organization to be captured; still practically, in terms of estimating metrics from experimental data. One major reason for these challenges is that systems with emergent properties are hidden: We do not have direct access to their internal, often high-dimensional state space; we do not know a priori what the emergent patterns are. Thus, we must "reconstruct" their state space and dynamics [17–20]. Even then, when successful, reconstruction does not lead easily or directly to measures of structural complexity and intrinsic computation [7]. It gives access to what is hidden, but does not say what the mechanisms are nor how they work.

Our view of the various kinds of complexity and their measures, though, has become markedly clearer of late. There is a natural semantics of complexity in which each measure answers a specific question about a system's organization. For example:

- How random is a process? Its *entropy rate* h_μ [11].
- How much state information must be stored for optimal prediction? Its *statistical complexity* C_μ [7].

* Corresponding author.
 E-mail addresses: chaos@ucdavis.edu (J.P. Crutchfield), cellison@wisc.edu (C.J. Ellison), pmriechers@ucdavis.edu (P.M. Riechers).

- How much of the future can be predicted? Its *past–future mutual information* or *excess entropy* **E** [16].
- How much information must an observer extract to know a process's hidden states? Its *transient information* **T** and *synchronization information* **S** [16].
- How much of the generated information (h_μ) affects future behavior? Its *bound information* b_μ [21].
- What's forgotten? Its *ephemeral information* ρ_μ [21].

And there are other useful measures ranging from degrees of irreversibility to quantifying model redundancy; see, for example, Ref. [22] and the proceedings in Refs. [23,24].

Unfortunately, except in the simplest cases where expressions are known for several, to date typically measures of intrinsic computation require extensive numerical simulation and estimation. Here we answer this challenge, providing exact expressions for a process's measures in terms of its ϵ-machine. In particular, we show that the spectral decomposition of this hidden dynamic leads to closed-form expressions for complexity measures. In this way, the remaining task in analyzing intrinsic computation reduces to mathematically constructing or reliably estimating a system's ϵ-machine in the first place.

Background

Our main object of study is a process $\mathcal{P}$, by which we mean the rather prosaic listing of all of a system's behaviors or realizations $\{\ldots x_{-2}, x_{-1}, x_0, x_1, \ldots\}$ and their probabilities: $\Pr(\ldots X_{-2}, X_{-1}, X_0, X_1, \ldots)$. We assume the process is stationary and ergodic and the measurement values range over a finite alphabet: $x \in \mathcal{A}$. This class describes a wide range of processes from statistical mechanical systems in equilibrium and in nonequilibrium steady states to nonlinear dynamical systems in discrete and continuous time on their attracting invariant sets.

Following Shannon and Kolmogorov, information theory gives a natural measure of a process's randomness as the uncertainty in measurement blocks: $H(L) = H[X_{0:L}]$, where H is the Shannon–Boltzmann entropy of the distribution governing the block $X_{0:L} = X_0, X_1, \ldots, X_{L-1}$. We monitor the *block entropy growth*—the average uncertainty in the next measurement X_{L-1} conditioned on knowing the preceding block $X_{0:L-1}$:

$$
\begin{aligned}
h_\mu(L) &= H(L) - H(L-1) \\
&= H[X_{L-1}|X_{0:L-1}] \\
&= -\left\langle \sum_{x_{L-1} \in \mathcal{A}} p(x_{L-1}) \log_2 p(x_{L-1}) \right\rangle_{\Pr(X_{0:L-1})},
\end{aligned}
\tag{1}
$$

where $p(x_{L-1}) = \Pr(x_{L-1}|x_{0:L-1})$. And when the limit exists, we say the process generates information at the *entropy rate*: $h_\mu = \lim_{L \to \infty} h_\mu(L)$.

Measurements, though, only indirectly reflect a system's internal organization. Computational mechanics extracts that hidden organization via the process's ϵ-machine [6], consisting of a set of recurrent *causal states* $\mathcal{S} = \{\sigma^0, \sigma^1, \sigma^2, \ldots\}$ and transition dynamic $\{T^{(x)} : T_{i,j}^{(x)} = \Pr(x, \sigma^j|\sigma^i)\}_{x \in \mathcal{A}}$. Each causal state represents a collection of "equivalent" histories—equivalent in the sense that each history belonging to an equivalence class yields the same prediction over futures. The ϵ-machine is a system's unique, minimal-size, optimal predictor from which two key complexity measures can be directly calculated.

The entropy rate follows immediately from the ϵ-machine as the causal-state averaged transition uncertainty:

$$
h_\mu = -\sum_{\sigma \in \mathcal{S}} \Pr(\sigma) \sum_{x \in \mathcal{A}} \Pr(x|\sigma) \log_2 \Pr(x|\sigma).
\tag{2}
$$

Here, the causal-state distribution $\Pr(\mathcal{S})$ is the stationary distribution $\langle \pi| = \langle \pi|T$ of the internal Markov chain governed by the row-stochastic matrix $T = \sum_{x \in \mathcal{A}} T^{(x)}$. The conditional probabilities $\Pr(x|\sigma)$ are the associated transition components in the labeled matrices $T_{\sigma,\sigma'}^{(x)}$. Note that the next state σ' is uniquely determined by knowing the current state σ and the measurement value x—a key property called *unifilarity*.

The amount of historical information the process stores also follows immediately: the *statistical complexity*, the Shannon–Boltzmann entropy of the causal-state distribution:

$$
C_\mu = -\sum_{\sigma \in \mathcal{S}} \Pr(\sigma) \log_2 \Pr(\sigma).
\tag{3}
$$

In this way, the ϵ-machine allows one to directly determine two important properties of a system's intrinsic computation: its information generation and its storage. Since it depends only on block entropies, however, h_μ can be calculated via other presentations; though not as efficiently. For example, h_μ can be determined from Eq. (2) using any unifilar predictor, which necessarily is always larger than the ϵ-machine. Only recently was a (rather more complicated) closed-form expression discovered for the excess entropy **E** using a representation closely related to the ϵ-machine [22]. Details aside, no analogous closed-form expressions for the other complexity measures are known, including and especially those for finite-L blocks, such as $h_\mu(L)$.

Mixed-state presentation

To develop these, we shift to consider how an observer represents its knowledge of a hidden system's current state and then introduce a spectral analysis of that representation. For our uses here, the observer has a correct model in the sense that it reproduces $\mathcal{P}$ exactly. (Any model that does we call a *presentation* of the process. There may be many.) Using this, the observer tracks a process's evolution using a distribution over the hidden states called a *mixed state* $\eta \equiv (\Pr(\sigma^0), \Pr(\sigma^1), \Pr(\sigma^2), \ldots)$. The associated random variable is denoted $\mathcal{R}$. The question is how does an observer update its knowledge (η) of the internal states as it makes measurements—$x_0, x_1, \ldots$?

If a system is in mixed state η, then the probability of seeing measurement x is: $\Pr(X = x|\mathcal{R} = \eta) = \langle \eta|T^{(x)}|\mathbf{1}\rangle$, where $\langle \eta|$ is the mixed state as a row vector and $|\mathbf{1}\rangle$ is the column vector of all 1s. This extends to measurement sequences $w = x_0 x_1 \ldots x_{L-1}$, so that if, for example, the process is in statistical equilibrium, $\Pr(w) = \langle \pi|T^{(w)}|\mathbf{1}\rangle = \langle \pi|T^{(x_0)} T^{(x_1)} \cdots T^{(x_{L-1})}|\mathbf{1}\rangle$. The mixed-state evolution induced by measurement sequence w is: $\langle \eta_{t+L}| = \langle \eta_t|T^{(w)}/\langle \eta_t|T^{(w)}|\mathbf{1}\rangle$. The set $\mathcal{R}$ of mixed states that we use here are those induced by all allowed words $w \in \mathcal{A}^*$ from initial mixed state $\eta_0 = \pi$. For each mixed state η_{t+1} induced by symbol $x \in \mathcal{A}$, the mixed-state-to-state transition probability is: $\Pr(\eta_{t+1}, x|\eta_t) = \Pr(x|\eta_t)$. And so, by construction, using mixed states gives a unifilar presentation. We denote the associated set of transition matrices $\{W^{(x)}\}$. They and the mixed states $\mathcal{R}$ define a process's *mixed-state presentation* (MSP), which describes how an observer's knowledge of the hidden process updates via measurements. The row-stochastic matrix $W = \sum_{x \in \mathcal{A}} W^{(x)}$ governs the evolution of the probability distribution over allowed mixed states.

The use of mixed states is originally due to Blackwell [25], who expressed the entropy rate h_μ as an integral of a (then uncomputable) measure over the mixed-state space $\mathcal{R}$. Although we focus here on the finite mixed-state case for simplicity, it is instructive to see in the general case the complicatedness revealed in a process using the mixed-state presentation: e.g., Figs. 17(a)–(c) of Ref. [26]. The Supplementary Materials give the detailed calculations for examples in the finite case.

Mixed states allow one to derive an efficient expression for the finite-L entropy-rate estimates of Eq. (1):

$$h_\mu(L) = H\left[X_{L-1} | \left(\mathcal{R}_{L-1} | \mathcal{R}_0 = \pi\right)\right] . \tag{4}$$

This says that one need only update the initial distribution over mixed states (with all probability density on $\eta_0 = \pi$) to the distribution at time L by tracking powers W^L of the internal transition dynamic of the MSP and not tracking, for that matter, an exponentially growing number of intervening sequences $\{x^L\}$. (This depends critically on the MSP's unifilarity.) That is, using the MSP reduces the original exponential computational complexity of estimating the entropy rate to polynomial time in L. Finally, and more to the present task, the mixed-state simplification is the main lead to an exact, closed-form analysis of complexity measures, achieved by combining the MSP with a spectral decomposition of the mixed-state evolution as governed by W^L.

Spectral decomposition

State distribution evolution involves iterating the transition dynamic W^L—that is, taking powers of a row-stochastic square matrix. As is well known, functions of a diagonalizable matrix can often be carried out efficiently by operating on its eigenvalues. More generally, using the Cauchy integral formula for operator-valued functions [27] and given W's eigenvalues $\Lambda_W \equiv \{\lambda \in \mathbb{C} : \det(\lambda I - W) = 0\}$, we find that W^L's spectral decomposition is:

$$W^L = \sum_{\substack{\lambda \in \Lambda_W \\ \lambda \neq 0}} \lambda^L W_\lambda \left\{ I + \sum_{N=1}^{\nu_\lambda - 1} \binom{L}{N} \left(\lambda^{-1} W - I\right)^N \right\}$$
$$+ [0 \in \Lambda_W] \left\{ \delta_{L,0} W_0 + \sum_{N=1}^{\nu_0 - 1} \delta_{L,N} W_0 W^N \right\}, \tag{5}$$

where $[0 \in \Lambda_W]$ is the Iverson bracket (unity when $\lambda = 0$ is an eigenvalue, vanishing otherwise), $\delta_{i,j}$ is the Kronecker delta function, and ν_λ is the size of the largest Jordan block associated with λ: $\nu_\lambda \leq 1 + a_\lambda - g_\lambda$, where g_λ and a_λ are λ's geometric (subspace dimension) and algebraic (order in the characteristic polynomial) multiplicities, respectively. The matrices $\{W_\lambda\}$ are a mutually orthogonal set of projection operators given by the residues of W's resolvent:

$$W_\lambda = \frac{1}{2\pi i} \oint_{C_\lambda} (zI - W)^{-1} dz , \tag{6}$$

a counterclockwise integral around singular point λ.

For simplicity here, consider only Ws that are diagonalizable. In this case: $g_\lambda = a_\lambda$ and Eq. (5) simplifies to $W^L = \sum_{\lambda \in \Lambda_W} \lambda^L W_\lambda$, where the projection operators reduce to $W_\lambda = \prod_{\substack{\zeta \in \Lambda_W \\ \zeta \neq \lambda}} (W - \zeta I) / (\lambda - \zeta)$. Thus, the only L-dependent operation in forming W^L is simply exponentiating its eigenvalues. The powers determine all of a process's properties, both transient (finite-L) and asymptotic. Moreover, when $a_\lambda = 1$, there are further simplifications. In particular, $W_\lambda = |\lambda\rangle\langle\lambda| / \langle\lambda|\lambda\rangle$, where $|\lambda\rangle$ and $\langle\lambda|$ are the right and left eigenvectors, respectively, of W associated with λ. Contrary to the Hermitian case of quantum mechanics, the left eigenvectors are not simply the conjugate transpose of the right eigenvectors.

Closed-form complexities

Forming the mixed-state presentation of process's ϵ-machine, its spectral decomposition leads directly to analytic, closed-form expressions for many complexity measures—here we present formulae only for $h_\mu(L)$, **E**, **S**, and **T**. Similar expressions for correlation functions and power spectra, partition functions, b_μ, r_μ, and others are presented elsewhere.

Starting from its mixed-state expression in Eq. (4) for the length-L entropy-rate approximates $h_\mu(L)$, we find the closed-form expression:

$$h_\mu(L) = \langle\delta_\pi | W^{L-1} | H(W^{\mathcal{A}})\rangle$$
$$= \sum_{\lambda \in \Lambda_W} \lambda^{L-1} \langle\delta_\pi | W_\lambda | H(W^{\mathcal{A}})\rangle , \tag{7}$$

where δ_π is the distribution with all probability density on the MSP's unique start state—the mixed state corresponding to the ϵ-machine's equilibrium distribution π.[1] In addition, $|H(W^{\mathcal{A}})\rangle$ is a column vector of transition uncertainties from each allowed mixed state η:

$$|H(W^{\mathcal{A}})\rangle = -\sum_{\eta \in \mathcal{R}} |\delta_\eta\rangle \sum_{x \in \mathcal{A}} \langle\delta_\eta | W^{(x)} | \mathbf{1}\rangle \log_2 \langle\delta_\eta | W^{(x)} | \mathbf{1}\rangle .$$

Taking $L \to \infty$, one finds the entropy rate (cf. Eq. (2)):

$$h_\mu = \langle\delta_\pi | W_1 | H(W^{\mathcal{A}})\rangle = \langle\pi_W | H(W^{\mathcal{A}})\rangle ,$$

where π_W is the stationary distribution over the MSP.

Let's turn to analyze the past-future mutual information, the excess entropy $\mathbf{E} = I[X_{-\infty:0} ; X_{0:\infty}]$: the information from the past that reduces uncertainty in the future. In general, $\mathbf{E}$ is not the statistical complexity C_μ, which is the information from the past that must be stored in order make optimal predictions about the future. Although Eq. (3) makes it clear that the stored information C_μ is immediately calculable from the ϵ-machine, $\mathbf{E}$ is substantially less direct. To see this, recall that the excess entropy has an equivalent definition—$\mathbf{E} = \sum_{L=1}^{\infty} \left[h_\mu(L) - h_\mu\right]$—to which we can apply Eq. (7), obtaining:

$$\mathbf{E} = \sum_{\substack{\lambda \in \Lambda_W \\ |\lambda| < 1}} \frac{1}{1 - \lambda} \langle\delta_\pi | W_\lambda | H(W^{\mathcal{A}})\rangle . \tag{8}$$

This should be compared to the only previously known general closed-form, which uses a process and its time-reversal [22,28]: $\mathbf{E} = I[\mathcal{S}^- ; \mathcal{S}^+]$, where $\mathcal{S}^-$ and $\mathcal{S}^+$ are the causal states of the reverse-time and forward-time ϵ-machines, respectively. Intriguingly, it appears that Eq. (8)—the spectral expression from the forward process—captures aspects of the reverse-time process. In short, the MSP's transient structure encapsulates the relevant information in the recurrent causal states of the reverse process. This suggests that there is a transformation between transient states of a forward process and recurrent states of its reverse process. (We explore this in a sequel.)

How does an observer come to know a hidden process's internal state? We monitor this via the *average state uncertainty* having seen all length-L words [16]:

$$\mathcal{H}(L) = -\sum_{w \in \mathcal{A}^L} \Pr(w) \sum_{\sigma \in \mathcal{S}} \Pr(\sigma|w) \log_2 \Pr(\sigma|w)$$
$$= \sum_{\eta \in \mathcal{R}} \Pr(\mathcal{R}_L = \eta | \mathcal{R}_0 = \pi) H[\eta] ,$$

where the last line is the mixed-state version with $H[\eta]$ being the presentation-state uncertainty specified by the mixed state η.

---

[1] To be clear, let's review: π is the stationary distribution over the ϵ-machine's states. It also happens to be—as a distribution over $\mathcal{S}$—a mixed state. π does not denote the initial distribution over the mixed states, δ_π does, as it is the distribution peaked at the mixed-state π. The use of δ_π is consistent with our use of δ_η for other distributions over the MSP that are peaked at the mixed-state η.

(Although recurrent causal states are induced by infinite histories, a partial history w, as above, starting from the stationary distribution π, can induce a nonpeaked distribution over causal states: $0 \leq \Pr(\sigma|w) \leq 1$.) Applying the spectral decomposition yields, for diagonalizable W:

$$\mathcal{H}(L) = \langle \delta_\pi | W^L | H[\eta] \rangle = \sum_{\lambda \in \Lambda_W} \lambda^L \langle \delta_\pi | W_\lambda | H[\eta] \rangle \,,$$

where $|H[\eta]\rangle$ is the column vector of state-distribution uncertainties for each allowed mixed state $\eta \in \mathcal{R}$.

The observer is *synchronized* when the state uncertainty vanishes: $\mathcal{H}(L) = 0$. The total amount of state information, then, that an observer must extract to become synchronized is the *synchronization information* [16] $\mathbf{S} \equiv \sum_{L=0}^\infty \mathcal{H}(L)$. Applying the above spectral decomposition results in the following closed-form expression:

$$\mathbf{S} = \sum_{\substack{\lambda \in \Lambda_W \\ |\lambda| < 1}} \frac{1}{1 - \lambda} \langle \delta_\pi | W_\lambda | H[\eta] \rangle \,. \tag{9}$$

This form makes it clear that mixed states and the transitions between them capture fundamental properties of the underlying process. For example, rewriting C_μ as the entropy of the initial mixed state—$\langle \delta_\pi | H[\eta] \rangle$—reinforces this observation. A related measure, the presentation-dependent synchronization information [29], diverges when the presentation is not synchronizable. Note also the close similarity between the excess entropy formula Eq. (8) above and Eq. (9). The only difference is that they average different informational quantities—the transition or the state uncertainties, respectively.

Although there are a number of additional complexity measures, as we discussed above, the final example we present is the *transient information* $\mathbf{T}$ [16]. It measures the amount of information one must extract from observations so that the block entropy converges to its linear asymptote: $\mathbf{T} \equiv \sum_{L=1}^\infty L \left[h_\mu(L) - h_\mu \right]$. The spectral decomposition readily yields the following closed-form expression:

$$\mathbf{T} = \sum_{\substack{\lambda \in \Lambda_W \\ |\lambda| < 1}} \frac{1}{(1 - \lambda)^2} \langle \delta_\pi | W_\lambda | H(W^{\mathcal{A}}) \rangle \,.$$

This form reveals the close relationship between transient information and excess entropy: they differ only in the eigenvalue weighting.

Discussion

There are a number of comments to make at this point to draw out the results' usefulness and importance. A process's structural complexity is not controlled by only the first spectral gap—the difference between the maximum and next eigenvalue. Rather, the *entire* spectrum is implicated in calculating the complexity measures. In terms of the temporal dynamics, all subspaces of the underlying causal-state process contribute. Naturally, there will be cases in which only some subspaces dominate, but as the expressions show this is not the general case. In addition, there is much structural information to be extracted from the projection operators W_λ, such as the dimension (geometric multiplicity) of the associated subspaces on which they act. This, in turn, gives the number of active degrees of freedom for the constituent subprocesses. As a result, we see that complexity measures capture inherently different properties, far beyond pairwise correlations and power spectra. Specifically, they measure transient, finite-time, and time asymptotic information processing, properties that rely on all-way correlations. The result gives one of the most thorough quantitative views of a process' intrinsic computation [7].

Although their derivations have not been laid out, the formulae as given are immediately usable. The Supplementary Materials provide calculations of complexity measures for several examples, including those for processes generated by typical unifilar hidden Markov models, nonlinear dynamical systems, cellular automata, and materials that form close-packed structures, as determined experimentally from X-ray diffraction spectra. And, a sequel demonstrates how to analyze the intrinsic computation in the stochastic switching dynamics of neural ion channels [30,31].

One of the more direct physical consequences of computational mechanics is that a system's organization is synonymous with how it stores and transforms information—a view complementary to that physics takes in terms of energy storage and transduction. In short, how a system is organized is how it computes. The theory just introduced grounds this information processing view practically and mathematically, giving quantitative and exact analysis of how hidden processes are organized. And, it should be contrasted with Kolmogorov–Chaitin complexity analysis [32]. For both stochastic and deterministic chaotic processes, the Kolmogorov–Chaitin complexity is (i) dominated by randomness and (ii) uncomputable. The theory here could not be more different: A wide variety of distinct kinds of information storage and processing are identified and they are exactly calculable.

Acknowledgements

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.physleta.2016.01.008.

References

[1] C. Domb, M.S. Green, J.L. Lebowitz (Eds.), Phase Transitions and Critical Phenomena, vols. 1–20, Academic Press, 1972–2001.
[2] M.C. Cross, P.C. Hohenberg, Rev. Mod. Phys. 65 (3) (1993) 851–1112.
[3] S. Camazine, J.-L. Deneubourg, N.R. Franks, J. Sneyd, G. Theraulas, E. Bonabeau, Self-Organization in Biological Systems, Princeton University Press, New York, 2003.
[4] M. Newman, A.-L. Barabasi, D.J. Watts, The Structure and Dynamics of Networks, Princeton University Press, Princeton, New Jersey, 2006.
[5] D. Helbing (Ed.), Social Self-Organization, Springer, New York, 2012.
[6] J.P. Crutchfield, Nat. Phys. 8 (2012) 17–24.
[7] J.P. Crutchfield, K. Young, Phys. Rev. Lett. 63 (1989) 105–108.
[8] W. Weaver, Am. Sci. 36 (1948) 536.
[9] L. Brillouin, Science and Information Theory, Academic Press, New York, 1956.
[10] R. Landauer, IBM J. Res. Dev. 5 (3) (1961) 183–191.
[11] A.N. Kolmogorov, Dokl. Akad. Nauk SSSR 124 (1959) 754 (in Russian), Math. Rev. 21 (2035b).
[12] C.E. Shannon, Bell Syst. Tech. J. 27 (1948) 379–423, 623–656.
[13] C.-B. Li, H. Yang, T. Komatsuzaki, Proc. Natl. Acad. Sci. USA 105 (2008) 536–541.
[14] D.P. Varn, G.S. Canright, J.P. Crutchfield, Acta Crystallogr., Sect. A 69 (2) (2013) 197–206.
[15] D. Darmon, J. Sylvester, M. Girvan, W. Rand, arXiv:1306.6111, 2013.
[16] J.P. Crutchfield, D.P. Feldman, Chaos 13 (1) (2003) 25–54.
[17] N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, Phys. Rev. Lett. 45 (1980) 712.
[18] F. Takens, in: D.A. Rand, L.S. Young (Eds.), Symposium on Dynamical Systems and Turbulence, vol. 898, Springer-Verlag, Berlin, 1981, p. 366.
[19] J.P. Crutchfield, B.S. McNamara, Complex Syst. 1 (1987) 417–452.
[20] J.D. Farmer, J. Sidorowitch, Phys. Rev. Lett. 59 (1987) 366.
[21] R.G. James, C.J. Ellison, J.P. Crutchfield, Chaos 21 (3) (2011) 037109.

[22] J.P. Crutchfield, C.J. Ellison, J.R. Mahoney, Phys. Rev. Lett. 103 (9) (2009) 094101.
[23] B. Marcus, K. Petersen, T. Weissman (Eds.), Entropy of Hidden Markov Process and Connections to Dynamical Systems, Lecture Notes Series, vol. 385, London Mathematical Society, 2011.
[24] J.P. Crutchfield, J. Machta, Chaos 21 (3) (2011) 037101.
[25] D. Blackwell, Publishing House of the Czechoslovak Academy of Sciences, Prague, 1957, pp. 13–20, vol. 28.
[26] J.P. Crutchfield, Physica D 75 (1994) 11–54.
[27] S. Hassani, Mathematical Physics, Springer, New York, 1999.
[28] C.J. Ellison, J.R. Mahoney, J.P. Crutchfield, J. Stat. Phys. 136 (6) (2009) 1005–1034.
[29] J.P. Crutchfield, C.J. Ellison, J.R. Mahoney, R.G. James, Chaos 20 (3) (2010) 037105.
[30] P. Kienker, Proc. R. Soc. Lond. B 236 (1989) 269–309.
[31] W.J. Bruno, J. Yang, J.E. Pearson, Proc. Natl. Acad. Sci. USA 102 (18) (2005) 6326–6331.
[32] M. Li, P.M.B. Vitanyi, An Introduction to Kolmogorov Complexity and Its Applications, Springer-Verlag, New York, 1993.