Statistical Signatures of Structural Organization: The Case of Long Memory in Renewal Processes

Sarah E. Marzen
James P. Crutchfield

SANTA FE INSTITUTE

Statistical Signatures of Structural Organization:
The case of long memory in renewal processes

Sarah E. Marzen[1, *] and James P. Crutchfield[2, †]

[1]*Department of Physics*
University of California at Berkeley, Berkeley, CA 94720-5800
[2]*Complexity Sciences Center and Department of Physics*
University of California at Davis, One Shields Avenue, Davis, CA 95616
(Dated: December 5, 2015)

Identifying and quantifying memory are often critical steps in developing a mechanistic understanding of stochastic processes. These are particularly challenging and necessary when exploring processes that exhibit long-range correlations. The most common signatures employed rely on second-order temporal statistics and lead, for example, to identifying long memory in processes with power-law autocorrelation function and Hurst exponent greater than $1/2$. However, most stochastic processes hide their memory in higher-order temporal correlations. Information measures— specifically, divergences in the mutual information between a process' past and future (excess entropy) and minimal predictive memory stored in a process' causal states (statistical complexity)— provide a different way to identify long memory in processes with higher-order temporal correlations. However, there are no ergodic stationary processes with infinite excess entropy for which information measures have been compared to autocorrelation functions and Hurst exponents. Here, we show that fractal renewal processes—those with interevent distribution tails $\propto t^{-\alpha}$—exhibit long memory via a phase transition at $\alpha = 1$. Excess entropy diverges only there and statistical complexity diverges there and for all $\alpha < 1$. When these processes do have power-law autocorrelation function and Hurst exponent greater than $1/2$, they do not have divergent excess entropy. This analysis breaks the intuitive association between these different quantifications of memory. We hope that the methods used here, based on causal states, provide some guide as to how to construct and analyze other long memory processes.

I. INTRODUCTION

Many time series of interest have "short memory", meaning (loosely speaking) that knowledge of the past confers exponentially diminishing returns for predicting the future. However, many other time series of interest— those with "long memory"—exhibit intrinsic timescales that grow without bound as the amount of available data increases [1–6]. Examples include the hydrological data first studied by Hurst [7] and modeled by Mandelbrot [8] and many others, e.g., see Refs. [9, 10].

These are qualitatively different processes that demand qualitatively different generative models. In other words, signatures of long memory imply a kind of structural organization of the underlying process that differs from one with short memory. This is the *inverse problem* of long memory: Which statistical signatures identify, uniquely or not, which intrinsic organizations? Sharp answers are critical to successful empirical analysis and often provide necessary first steps in predictive theory building. The complementary *forward problem*, an open question, is to identify the kinds of memoryful process structure that lead to one or another statistical signature. Answering this question requires defining statistical signatures that quantify memory in stochastic processes.

Many existing quantifications of long memory are based on second-order statistics; e.g., on using the autocorrelation function, power spectrum, or Hurst exponent. These approaches have had notable successes in analyzing hydrological data [7, 9], music [4], spin systems [2], astrophysical flicker noise [6], language [11, 12], natural scenery [13, 14], communication system error clustering [15], financial time series, and many other seemingly complex phenomena [5, 16].

However, there are at least two reasons to look to other statistics besides the Hurst exponent. First, second-order statistics alone can be misleading, as a process can "hide" signatures of long memory in higher-order statistics. For example, Fig. 1 shows a hidden Markov model (HMM) that, on the one hand, is patently quite structured, and, on the other, generates a process with a flat power spectrum [17]. Indeed, most stochastic processes seem to hide information about their temporal dependencies in higher-order statistics [18, 19]. Second, as suggested in Ref. [20], our determination of whether or not a pro-

———

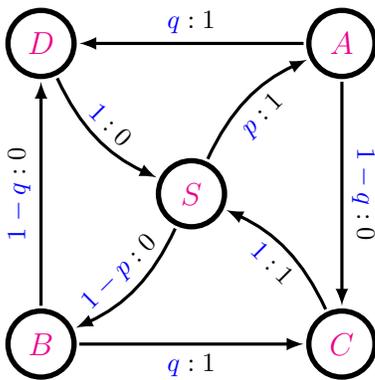* smarzen@berkeley.edu
† chaos@ucdavis.edu

FIG. 1. The Random-Random-XOR (RRXOR) Process is generated by the five-state (minimal unifilar) Hidden Markov model here. Labels $p|x$ denote that a state-to-state transition occurs with probability p and emits symbol x. If X_t is the random variable at time t, then the generated time series is $X_{t+2} = X_{t+1}$ XOR X_t, with X_{t+1} and X_t being Bernoulli(q) and Bernoulli(p), respectively, for $t = 0, 3, 6, \ldots$. With $p = q = 1/2$ and starting state probabilities $\Pr(S) = 1/3$ and $\Pr(A) = \Pr(B) = \Pr(C) = \Pr(D) = 1/6$ the output process is stationary white noise—a flat power spectrum [17].

cess has long memory ideally should be invariant under invertible transformations of one's measurement values. The challenge is not only to find a new statistic that addresses these two concerns, but to find a statistic that is also easy to operationalize.

References [21–23] suggested a process might be said to have long memory when the mutual information between its past and future (excess entropy) diverges, and Ref. [21] suggested that long memory is associated with divergent statistical complexity with the effective memory architecture given by a process' ϵ-machine. By construction, these statistics are invariant under invertible transformations of the data; and with sufficiently clever entropy estimation techniques, these statistics are also calculable directly from time series data.

Unfortunately, there is a paucity of concrete examples upon which to build intuition as to how these higher-order statistics and the more commonly used second-order statistics relate. In part, this lack of concrete examples might owe somewhat to the fact that it is nontrivial to construct ergodic stationary processes with divergent excess entropy, though see Refs. [24, 25]. (Note that the processes considered in Ref. [22] were nonergodic [26].)

To that end, we study a tractable class of processes that can have both divergent excess entropy and Hurst exponent greater than $1/2$: the *fractal renewal processes* [27–30] in which interevent intervals are drawn independently and identically (IID) from a probability distribution with tails $\propto t^{-\alpha}$. These processes are very widely used in the physical, biological, and social sciences to

model diverse long-memory phenomena, ranging from current fluctuations in electronic devices and neuronal spike trains to earthquakes and astrophysical time series [31–40].

Previous studies analyzed the second-order statistics of such processes in some detail [9, 41]. Here, we use techniques inspired by those in Refs. [25, 42] to calculate the excess entropy and statistical complexity of fractal renewal processes for the first time. We find that fractal renewal processes have divergent excess entropy only and exactly when $\alpha = 1$ and divergent statistical complexity as $\alpha \to 1$ from above and for all $0 < \alpha < 1$. However, fractal renewal processes have power-law power spectra for all $0 < \alpha < 2$ [41] and Hurst exponents greater than $1/2$ [9]—the latter being two of the conventional second-order statistical signatures of "long memory". Thus, even for these relatively straightforward processes, the excess entropy and statistical complexity encapsulate a different notion of long memory than one gleans using only second-order statistics. These results also add fractal renewal processes to a very short list of known stationary ergodic processes with divergent excess entropy [25, 42] and so, we hope, they pave the way for more general comparisons between different definitions of long memory.

Section II briefly reviews definitions of memory in stochastic processes. Section III calculates informational measures of memory for fractal renewal processes. Section IV then compares our findings to the second-order statistics calculated by Refs. [9, 41] and draws out the lessons for the above application examples. We close by reflecting on structural organization associated with long memory.

II. BACKGROUND

There are many definitions for a stochastic process to have long memory; Ref. [20] provides a particularly helpful survey. Consider a sequence of ℓ observations $x_0, x_1, \ldots, x_{\ell-1}$, realizations of discrete-valued random variables $X_0, X_1, \ldots, X_{\ell-1}$. For instance, if the *autocorrelation function* $\mathcal{C}(\tau)$ is asymptotically a power law multiplied by a slowly varying function $g(\tau)$, then a process can be said to have "long memory":

$$\mathcal{C}(\tau) = \sigma^{-2} \sum_{j=0}^{\ell} (x_j - \mu)(x_{j+\tau} - \mu)$$
$$\propto g(\tau)\tau^{-\gamma} \, ,$$

with $0 < \gamma < 1$, mean μ, and variance σ^2. Yet other definitions are based on the decay of the *spectral density*:

$$\mathcal{P}(f) = \ell^{-1} \left| \sum_{j=0}^{\ell} x_j e^{-ijf} \right|^2 \, .$$

The process has long memory when $\mathcal{P}(f) \propto f^{-\beta} L_1(f)$ as f approaches 0 (where $L_1(f)$ is a slowly varying function near $f = 0$) with $0 < \beta < 1$. Other definitions still are based on how variances deviate from time-local linear extrapolation. Starting with the variance of partial sums $S_j = X_1 + \cdots + X_j$, one uses the *rescaled-range* statistics:

$$RS(\ell) = \frac{\max_{0 \le j \le \ell}(S_j - \frac{j}{\ell}S_\ell) - \min_{0 \le j \le \ell}(S_j - \frac{j}{\ell}S_\ell)}{\sigma}$$
$$\propto \ell^{-H} ,$$

where $H \in (0,1)$ is the *Hurst index*. Processes with $H > 1/2$ are interpreted as having long memory. Unfortunately, even these second-order statistics are not always equivalent signatures of long memory. Section 5 of Ref. [20] provides examples of inconsistencies.

In a search for general principles from ergodic theory, Sec. 4 of Ref. [20] proposed that we require a definition of long memory independent of invertible transformations of the data. That is, if an invertible transformation is applied pointwise to each observation X_i, we would hope that the resulting process has long memory if and only if the original process had long memory. This desideratum is not satisfied by definitions based on the above second-order statistics.

Since strongly mixing processes have short memory and nonergodic processes could be said to have infinite memory [26], Ref. [20] proposed that one or another type of nonmixing property is a good candidate for long memory in ergodic stationary processes. This criterion satisfies the invariance desideratum above but can be rather difficult to evaluate.

Fortunately, the information-theoretic notions of memory we consider also satisfy the transformation-invariant desideratum and have been successfully deployed as quantifications for the "complexity" of stochastic processes [22, 43]. We study two: the *excess entropy* $\mathbf{E} = I[\overleftarrow{X}; \overrightarrow{X}]$, or the mutual information between a process' past $\overleftarrow{X} = \ldots X_{-3}X_{-2}X_{-1}$ and future $\overrightarrow{X} = X_0 X_1 X_2 \ldots$ [23]; and the *statistical complexity* C_μ, or the amount of information from the past $\overleftarrow{X}$ required to exactly predict the future $\overrightarrow{X}$ [43]. When the excess entropy diverges, we are interested in the asymptotic rate of divergence of finite-length excess entropy estimates $\mathbf{E}(\ell) = I[\overleftarrow{X}; \overrightarrow{X}^\ell]$ [22, 23]. This asymptotic rate of divergence is also invariant to temporally local convolutions and invertible transformations of the data [22].

To more precisely define and calculate the statistical complexity and the excess entropy, we need to recall the causal states of computational mechanics. Consider clustering pasts according to an equivalence relation $\sim$ in which two pasts are equivalent when they have the same conditional probability distribution over futures:

$\overleftarrow{x} \sim \overleftarrow{x}'$ if and only if $\Pr(\overrightarrow{X}|\overleftarrow{X} = \overleftarrow{x}) = \Pr(\overrightarrow{X}|\overleftarrow{X} = \overleftarrow{x}')$. The resulting clusters are *forward-time causal states* $\mathcal{S}^+$, which inherit a probability distribution from the probability distribution over pasts. The *forward-time statistical complexity* is the entropy of these causal states: $C_\mu^+ = H[\mathcal{S}^+]$. For more detail, see Refs. [44, 45].

We can similarly define the *reverse-time causal states* $\mathcal{S}^-$ by clustering futures with equivalent conditional probability distributions over pasts: $\overrightarrow{x} \sim \overrightarrow{x}'$ if and only if $\Pr(\overleftarrow{X}|\overrightarrow{X} = \overrightarrow{x}) = \Pr(\overleftarrow{X}|\overrightarrow{X} = \overrightarrow{x}')$. The *reverse-time statistical complexity* is the entropy of those reverse-time causal states: $C_\mu^- = H[\mathcal{S}^-]$. Renewal processes are time-reversal invariant [46], or *causally reversible*, so throughout the following we denote the statistical complexity as $C_\mu = C_\mu^+ = C_\mu^-$ without loss of precision.

Reverse-time causal states and forward-time causal states can be used to calculate the excess entropy [47, 48]:

$$\mathbf{E} = I[\mathcal{S}^+; \mathcal{S}^-] .$$

For discrete-time processes, $\mathbf{E}$ is a lower bound on C_μ:

$$\mathbf{E} \le C_\mu . \tag{1}$$

In other words, for discrete-time processes, if statistical complexity is finite, then so is excess entropy. Conversely, if excess entropy is infinite, then statistical complexity is infinite.

Often continuous-time processes have an uncountable set of causal states. For them, the statistical complexity is taken to be the differential entropy:

$$\widehat{C}_\mu = H[\mathcal{S}^+]$$
$$= - \int_\Delta d\mu(\sigma^+) \log \mu(\sigma^+) ,$$

where we have the simplex Δ of causal states and $\mu(\sigma^+)$ is their measure in Δ. In the continuous-time setting, the inequality analogous to Eq. (1) no longer necessarily holds [49]. We call the differential entropy $\widehat{C}_\mu$ the *continuous-time statistical complexity* to distinguish it from the discrete-time statistical complexity C_μ, but simply refer to it as the statistical complexity when context is clear.

One can also define finite-time reverse-time causal states, denoted $\mathcal{S}_\ell^-$, by clustering futures of finite-length ℓ with the same equivalence relation as above. From these, we obtain finite-length reverse-time statistical complexity $C_\mu^{-\ell} = H[\mathcal{S}_\ell^-]$, respectively. These can be used to calculate finite-future excess entropy estimates: $\mathbf{E}(\ell) = I[\mathcal{S}^+; \mathcal{S}_\ell^-]$ [47, 48].

For discrete-alphabet, discrete-time processes, the statistical complexity is invariant to relabelings of the mea-

surement alphabet. However, as just noted, when the causal states are uncountable, the statistical complexity involves a differential entropy, and differential entropies are not invariant to invertible transformations of the coordinate system of the distribution's support. A prosaic example of this is given in Ref. [50]. Modulo such factors, whether or not statistical complexity diverges, the rate of divergence of its finite-length estimates C_μ^ℓ is invariant to temporally local convolutions of the data.

Realizations from a renewal process consist of sequences of events separated by epochs of quiescence, the lengths of which are drawn independently from the same *interevent* distribution. Throughout, when discussing a discrete-time renewal process, we use the following notation [46]: $F(n)$ is the *interevent count* probability distribution function; $w(n) = \sum_{n'}^\infty F(n')$ is the *survival function*; and μ is its *mean interevent count*. We use the following notation for continuous-time renewal processes: $\phi(t)$ is the *waiting time* distribution; $\Phi(t)$ is its *survival function*; and $\overline{T}$ is the *mean interevent interval*. *Fractal renewal processes* have survival functions that have power-law tails, as introduced shortly.

III. INTRINSIC MEMORY IN FRACTAL RENEWAL PROCESSES

Fractal renewal processes—those with power-law interevent interval probability density functions—can have long memory in the sense of Ref. [51]. For instance, they can have Hurst index $H > 1/2$ [9] and their autocorrelation function can be (asymptotically) a power law [41]. Fractal renewal processes have been implicated in a variety of complex natural processes, to which the introduction alluded. Might these processes also have infinite statistical complexity or infinite excess entropy? To the best of our knowledge, the excess entropy and statistical complexity of fractal renewal processes have yet to be calculated.

Calculating statistical complexity and excess entropy can be challenging when going beyond finite causal-state processes [52]. To make progress with bounding the excess entropy of fractal renewal processes, we use two tools. The first tool is to coarse grain by time-binning. The Data Processing Inequality [53] then implies that the excess entropy of a discrete-time renewal process is always upper-bounded by the excess entropy of the corresponding continuous-time renewal process. See App. A. The second tool allows us to calculate excess entropy and statistical complexity even when the mean rate of events vanishes by conditioning on the presence of a proxy event. This tool was inspired by previous work [24] and is summarized in App. B.

Fractal renewal processes are typically considered in continuous-time, with interevent intervals generated independently and identically distributed (IID) from the probability density function:

$$\phi(t) = \begin{cases} 0 & t < 1 \\ \alpha t^{-(\alpha+1)} & t \geq 1 \end{cases} . \tag{2}$$

The probability of seeing an interevent interval of length t or larger is the survival function:

$$\Phi(t) = \int_t^\infty \phi(t')dt'$$
$$= \begin{cases} 1 & t < 1 \\ t^{-\alpha} & t \geq 1 \end{cases} . \tag{3}$$

Time intervals are given in terms of the shortest possible interevent interval. When $\alpha > 1$, the mean interevent interval $\overline{T} = \frac{\alpha}{\alpha-1}$ is finite; when $0 < \alpha \leq 1$, the mean interevent interval is infinite, but one will always eventually see an event.

Appendix D describes how to manipulate the continuous-time analog of Eq. (B1) to obtain:

$$\widehat{\mathbf{E}} = \begin{cases} \log \frac{\alpha^2}{\alpha-1} - 1 & \alpha > 1 \\ \infty & \alpha = 1 \\ \frac{\alpha^2+\alpha-1}{\alpha(1-\alpha)} + \log \frac{\alpha}{1-\alpha} - (1-\alpha)K_\alpha & \alpha < 1 \end{cases} , \tag{4}$$

where $K_\alpha = \int_0^\infty (u^{-\alpha}-(1+u)^{-\alpha})\log(u^{-\alpha}-(1+u)^{-\alpha})du$. Note that at small values of α, K_α is difficult to evaluate numerically due to the integrand's long tails, even when $\widehat{\mathbf{E}}$ is quite small. For instance, when $\alpha = 1/4$, $\widehat{\mathbf{E}} \approx 0.089$ nats, but $\int_0^N (u^{-\alpha}-(1+u)^{-\alpha})\log(u^{-\alpha}-(1+u)^{-\alpha})du$ does not return positive estimates for the excess entropy until $N \geq 10^{11}$. A more obvious benefit of Eq. (4), then, is that we can study the excess entropy's asymptotic behavior near $\alpha = 1$, where $\widehat{\mathbf{E}}(\ell) \sim \log\log\ell$. This divergence is slower than any previously reported divergence [22, 24, 25], but is a divergence nonetheless.

When $\alpha > 1$ but close to its critical value, the excess entropy diverges as $\sim \log\frac{1}{\alpha-1}$. As $\alpha \to \infty$, $\widehat{\mathbf{E}}$ diverges as $\log\alpha$. This point is discussed more fully later on.

The discrete-time analog of fractal renewal processes has a survival function:

$$w(n) = \begin{cases} 1 & n = 0 \\ n^{-\alpha} & n \geq 1 \end{cases} . \tag{5}$$

The transient (small n) behavior of $w(n)$ may not match that in some applications, but only $w(n)$'s asymptotic behavior is relevant to $\mathbf{E}$'s divergence. Moreover, App. A guarantees that $\mathbf{E}$ is finite when $\alpha \neq 1$ and that at $\alpha = 1$

its divergence is at most $\log \log \ell$. Additional arguments in App. D, in turn, show that $\mathbf{E}(\ell)$ diverges at $\alpha = 1$ as $\log \log \ell$.

The excess entropy $\mathbf{E}$ captures the amount of predictable randomness of a stochastic process. As a comparison, we are also interested in the statistical complexity C_μ of discrete-time and continuous-time fractal renewal processes. The statistical complexity is the number of bits required to losslessly predict ($\mathbf{E}$ nats of) the process' future. Sometimes, C_μ is not much larger than $\mathbf{E}$; for discrete-time periodic processes, the two are equivalent and equal to the logarithm of the period. More often than not, C_μ is infinite while $\mathbf{E}$ is finite; e.g., for processes generated by most (nonunifilar) Hidden Markov Models.

Cryptic processes have large statistical complexity and small excess entropy [47]; the larger the crypticity, the more that a process' true structure is "hidden" from the observer. An open question is whether or not fractal renewal processes, with their statistical signatures of complexity, are highly cryptic. So, we focus some attention now on evaluating C_μ for fractal renewal processes.

We can calculate C_μ of time-binned continuous-time renewal processes in the infinitesimal-τ limit [49]:

$$C_{\mu_\tau} \sim \log \frac{1}{\tau} - \int_0^\infty \frac{\Phi(t)}{\overline{T}} \log \frac{\Phi(t)}{\overline{T}} dt \ .$$

As we will discuss elsewhere, the above expression is the differential entropy over continuous-time causal states—the expression given in Sec. II as the "continuous-time statistical complexity" $\widehat{C}_\mu$—plus the logarithm of our time-bin resolution. Thus, C_{μ_τ}'s $\log \frac{1}{\tau}$ divergence is an artifact of our failure to use the differential entropy when calculating memory storage requirements of continuous random variables [53]. As a result, we focus on C_{μ_τ}'s nondivergent component, $\widehat{C}_\mu = \lim_{\tau \to 0} \left(C_{\mu_\tau} + \log \tau \right)$, or what was earlier called the continuous-time statistical complexity. Straightforward algebra shows that:

$$\widehat{C}_\mu = \begin{cases} \frac{1}{\alpha-1} + \log \frac{\alpha}{\alpha-1} & \alpha > 1 \\ \infty & \alpha \leq 1 \end{cases} \ . \tag{6}$$

Again, we can say that the (continuous-time) C_μ diverges whenever the mean interevent interval $\overline{T}$ diverges. When $\alpha \leq 1$, finite-length statistical complexity estimates adapted to the continuous-time case from Eq. (B2) diverge as:

$$C_\mu^{+\ell} \sim \begin{cases} \log \ell & \alpha < 1 \\ \frac{1}{2} \log \ell & \alpha = 1 \end{cases} \ .$$

So, the special nature of $\alpha = 1$ is also revealed as a discontinuity in rates of divergence of the finite-length sta-

tistical complexity. In particular, the least cryptic fractal renewal process, among fractal renewal processes with divergent statistical complexity, is the process generated when $\alpha = 1$.

Equations (4) and (6) are plotted in Fig. 2. The divergences in $\widehat{\mathbf{E}}$ and $\widehat{C}_\mu$ at $\alpha = 1$ are apparent in the plot. If $\widehat{\mathbf{E}}$ and $\widehat{C}_\mu$ are taken to be systems-agnostic order parameters, then a fractal renewal process exhibits a nonequilibrium phase transition exactly when its mean interevent interval diverges.
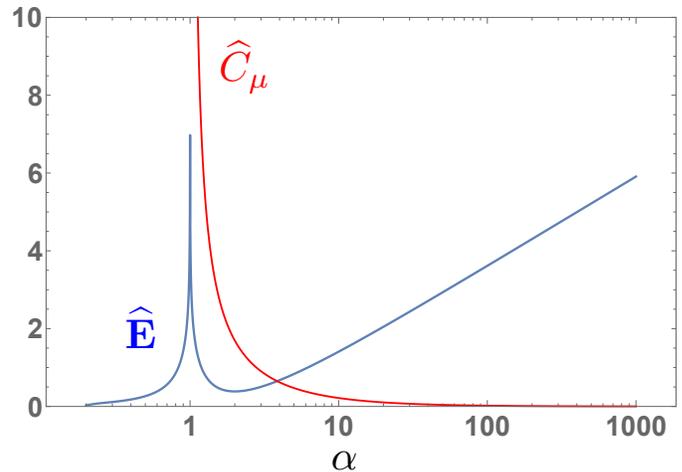
FIG. 2. **Excess entropy $\widehat{\mathbf{E}}$ and statistical complexity $\widehat{C}_\mu$ of continuous-time fractal renewal processes:** Process realizations are generated by drawing interevent intervals IID from the probability density function $\phi(t) = \alpha t^{-(\alpha+1)}$ for $t \geq 1$ and 0 otherwise. $\widehat{\mathbf{E}}$ in nats as a function of α, evaluated using Eq. (4). The nondivergent component of statistical complexity $\widehat{C}_\mu$ in nats as a function of α, evaluated using Eq. (6). Note that $\widehat{C}_\mu$ is a differential entropy and so not necessarily larger than the excess entropy $\widehat{\mathbf{E}}$; a subtlety when working with continuous-time processes.

The behavior of $\widehat{\mathbf{E}}$ and $\widehat{C}_\mu$ as α tends to infinity also deserves special mention, as the process appears to become infinitely predictable ($\widehat{\mathbf{E}} \to \infty$) while requiring less memory for prediction ($\widehat{C}_\mu \to 0$). As α tends to ∞, $\phi(t)$ becomes more and more sharply peaked at $t = 1$. In other words, the process moves closer and closer to that of a periodic process with period 1. Periodic processes are random enough, in that the phase of the process could be any real number between 0 and the period. In the language of computational mechanics, the causal state *is* the phase, and its differential entropy—the continuous-time statistical complexity $\widehat{C}_\mu$—is the logarithm of the process' period. As $\alpha \to \infty$, the mean interevent interval $\overline{T} = \frac{\alpha}{\alpha-1}$ tends to 1, and the continuous-time statistical complexity correspondingly tends to $\log 1 = 0$. However, periodic processes are also highly predictable, in that the time to next event is determined by the time

since last event; hence, the differential entropy of the time to next event *conditioned on* the time since last event tends towards negative infinity, resulting in an infinite $\widehat{\mathbf{E}} = \widehat{C}_\mu - H[\mathcal{S}^- | \mathcal{S}^+] \to \infty$. Similar behavior was seen in Ref. [49] as the noisiness of spike trains tended to zero, though. The least cryptic fractal renewal process, then, occurs in the limit that α tends to infinity.

IV. CONCLUSION

We showed that a fractal renewal process's excess entropy diverges precisely when its mean interevent interval diverges. This adds a relatively easily understood process and one of much broader applicability to the existing list of ergodic stationary processes with divergent excess entropy [24, 25].

Notably, the expected number of events observed in a finite time interval for a fractal renewal process with divergent excess entropy is zero. This brings in an interpretational challenge. A process that, on average, produces arbitrarily long silence is not often described as random. So, should not the excess entropy of a point process with infinite mean interevent interval be zero? However, the mutual information between finite-length pasts and futures, assuming that we do see an event, can diverge. And, we will almost surely see an event when we view a semi-infinite past.

Our calculations revealed that fractal renewal processes flip from finite to divergent statistical complexity and exhibit divergent excess entropy exactly when the mean interevent interval diverges. These information-theoretic measures of memory point to the power-law coefficient $\alpha = 1$ as being a "critical" parameter in this process family. When the mean interevent interval is finite, both excess entropy and continuous-time statistical complexity are finite, though excess entropy grows unbounded as α tends to infinity. When the mean interevent interval is infinite and the power-law coefficient is not $\alpha = 1$, excess entropy is finite, but continuous-time statistical complexity is infinite.

Employing signatures of long memory based on second-order statistics suggests, instead, that $\alpha = 2$ was a "critical point". Specifically, the power spectrum of a fractal renewal process exhibits power-law scaling when $\alpha < 2$ [41], and the Hurst index of the processes with $\alpha < 2$ is greater than $1/2$ and increases with decreasing α [9]. Therefore, at a minimum, drawing conclusions about a process' complex organization via such low-order statistics can be ambiguous.

Finally, our results suggest that certain previously studied experimental phenomenon are poised at a critical point between finite and infinite "memory", as suggested by many others using other definitions of criticality [54]. The stochastic process of neuron membrane ion channels opening and closing has divergent excess entropy when the kinetic rate adopts the form $k_{\text{eff}}(t) \approx t^{-1}$. This may be the case for some potassium-selective channels in cultured mouse hippocampal pyramidal cells near resting membrane voltage, $V = -60$ V [55, Fig. 10, bottom right]. Similarly, the phenomenological fit of the stopping probabilities used for Wikipedia edit-revert time series has divergent statistical complexity when $\alpha = 1$ and divergent excess entropy when $p = 1$ as well [56, 57]. This seems to suggest that increased cooperativity between editors drives Wikipedia towards increasing its social memory.

However, one lesson from our results is tantamount to a cautionary note on interpreting the implicated memory organization. To the extent that the estimated fractal renewal processes with divergent memory are good models, one cannot conclude that the content of that memory reflects sophisticated computational processing or highly organized storage of detailed information. Indeed, like all renewal processes, fractal renewal processes are simple: they count up to some threshold and reset. Surely these coarse statistics, while useful and even necessary as tools for a first-cut analysis, fall far short of fully describing the hierarchies of information processing in neurons and the rich social dynamics driving Wikipedia's accumulating human knowledge.

To close, let's return to our initial discussion of statistical signatures of structural organization. We drew a comparison of divergent memory in ergodic processes to that we previously identified in the so-called Bandit non-ergodic processes [26]. The mechanism underlying the latter was rather straightforward: from trial to trial the process remembers the operant ergodic component sub-process and so uses an infinite memory and exhibits an excess entropy that diverges as $\log \ell$. The case for ergodic process is more subtle. For renewal processes we showed that the divergence is $\log \log \ell$. What's the associated mechanism? Renewal processes track time between events and so, in computational model terms, it appears that the process somehow embeds a counter [21, Sec. 4.5.2]. An interesting contrast is the $\log \ell$ excess entropy divergence seen at the onset of chaos through period-doubling, associated with pushdown stack mechanism [21, Sec. 4.5.1], and seen in the branching copy process [24]. At this stage, though, the possibility of unique associations between the form of information measure divergence and mechanism is not sufficiently well explored. Nonetheless, with further extension and refinement information measures and their divergences will become increasingly more insightful diagnostics of nature's diverse forms of intrinsic computation.

Appendix A: Continuous- versus Discrete-time Excess Entropies

Often, integrals are easier to evaluate than the corresponding sums. One practical goal, leveraging this below, is to relate the excess entropy of time-binned continuous-time processes to that of corresponding discrete-time renewal processes.

Reference [46] found that the excess entropy of a discrete-time renewal process is:

$$
\mathbf{E} = \log(\mu + 1) - \frac{2}{\mu+1} \sum_{n=0}^{\infty} w(n) \log w(n) \\
+ \frac{1}{\mu+1} \sum_{n=0}^{\infty} (n+1) F(n) \log F(n) .
\tag{A1}
$$

While Ref. [49] showed that the excess entropy of a continuous-time renewal process $X(t)$ is:

$$
\widehat{\mathbf{E}} = \mathrm{I}[X(t)_{t<0}; X(t)_{t\geq0}] \\
= \log \overline{T} - \frac{2}{\overline{\overline{T}}} \int_0^\infty \Phi(t) \log \Phi(t) dt \\
+ \frac{1}{\overline{\overline{T}}} \int_0^\infty t\phi(t) \log \phi(t) dt , \tag{A2}
$$

which is in units of *nats* when the mean interevent interval $\overline{T}$ is finite.

Consider time-binning the continuous-time point process $X(t)$ by asking how many events are observed in an interval $[t, t+\tau)$. If at least one event is observed, then we record a 1; if no events are observed, then we record a 0. This data labeling technique is common; e.g., when studying neural spike trains. The probability of observing at least n counts between successive 1s is given by:

$$
w_\tau(n) = \Phi(n\tau) .
$$

When $\tau = 1$, then the survival function of the time-binned process is exactly that of the discrete-time renewal process with excess entropy given in Eq. (A1).

The excess entropy or estimates thereof for a discrete-time renewal process are upper bounded by the excess entropy of a corresponding continuous-time renewal process, as shown shortly. This is a special case of a more general statement: coarse-graining a time series always reduces its excess entropy, due to the Data Processing Inequality. This statement can be easily generalized to other discrete-alphabet, continuous-time processes. Despite its simplicity, it proves useful for the calculations to come in Sec. III.

In particular, let $\widehat{\mathbf{E}}$ denote the excess entropy of a continuous-time renewal process $X(t)$ with survival function $\Phi(t)$ and $\mathbf{E}$ the excess entropy of the discrete-time renewal process X_t with survival function $w(n) = \Phi(n)$ for all nonnegative integers n. Then, when $\widehat{\mathbf{E}} < \infty$:

$$
\mathbf{E} \leq \widehat{\mathbf{E}} .
$$

To see this, let $\mathbf{E}_\tau$ denote the excess entropy of the discrete-time process that comes from time-binning the continuous-time renewal process with discretization bin size τ. To obtain the above inequality, we apply the Data Processing Inequality:

$$
\mathbf{E}_{1/n} = I[\ldots, X(-2/n), X(-1/n); X(0), X(1/n), \ldots] \\
\geq I[\ldots, X_{-2}, X_{-1}; X_0, X_1, \ldots] \\
= \mathbf{E}_1 .
$$

If we take the limit of the left-hand side as $n \to \infty$, we obtain:

$$
\mathbf{E}_{\tau=1} \leq \lim_{n\to\infty} \mathbf{E}_{1/n} \\
= \lim_{\tau\to0} \mathbf{E}_\tau .
$$

Again by the Data Processing Inequality, $\mathbf{E}_{\tau=1}$ is lower-bounded by the mutual information between the counts since last event and counts to next event, as the former is a function of the past and the latter is a function of the future: $\mathbf{E} \leq \mathbf{E}_{\tau=1}$. By definition [58], $\lim_{\tau\to0} \mathbf{E}_\tau = \widehat{\mathbf{E}}$.

Appendix B: Renewal Processes with Infinite Mean Intervent Intervals

When the mean interevent interval $\overline{T}$ (or μ) is infinite, the formulae for excess entropy in Eqs. (A1) and (A2) no longer apply. Causal states, however, still provide a useful framework for calculating it. Using them we introduce an analysis method for discrete-time renewal processes in this case. The obvious extensions to continuous-time renewal processes follow when we replace $F(n)$ with $\phi(t)$,

$w(n)$ with $\Phi(t)$, and summations with integrals.

We calculate $\mathbf{E}(\ell)$ for renewal processes with infinite μ via an analysis technique inspired by Ref. [24] and then calculate $\mathbf{E}$ as a limit of $\mathbf{E}(\ell)$ as ℓ tends to infinity, seemingly valid for ergodic processes. First, we would like to directly calculate $\mathbf{E}(\ell)$ in terms of forward and reverse-time causal states [47]: $\mathbf{E}(\ell) = I[\overleftarrow{X}; \overrightarrow{X}^\ell] = I[\mathcal{S}^+; \mathcal{S}_\ell^-]$, where $\mathcal{S}_\ell^-$ are finite-time reverse-time causal states. Unfortunately, inspecting the corresponding joint probability distribution in App. II of Ref. [46] shows that while we can identify the joint probability distribution up to a normalization constant, this normalization constant is infinite when μ is infinite.

So, we define a "proxy" binary random variable U_ℓ which is 1 if there has been an event sometime in $\overrightarrow{X}^\ell$ and past $\overleftarrow{X}$, and 0 otherwise. A little reflection shows that $\Pr(U_\ell = 0) = \lim_{N\to\infty} w(N+\ell) = 0$. Even so, this auxiliary random variable is a surprisingly useful construct. A standard information-theoretic decomposition gives $\mathbf{E}(\ell) = I[\mathcal{S}^+; \mathcal{S}^{-\ell}|U_\ell] + I[\mathcal{S}^+; \mathcal{S}_\ell^-; U_\ell]$, but since $\Pr(U_\ell = 0) = 0$, we have that $I[\mathcal{S}^+; \mathcal{S}_\ell^-|U_\ell] = I[\mathcal{S}^+; \mathcal{S}_\ell^-|U_\ell = 1]$ and $I[\mathcal{S}^+; \mathcal{S}_\ell^-; U_\ell] = 0$. Altogether this yields:

$$\mathbf{E}(\ell) = I[\mathcal{S}^+; \mathcal{S}_\ell^-|U_\ell = 1] .$$

The conditional probability distribution $\Pr(\mathcal{S}^+, \mathcal{S}_\ell^-|U_\ell = 1)$ is normalizable and, as shown in App. C, leads to:

$$\begin{aligned}
\mathbf{E}(\ell) =& \log Z(\ell) - \frac{1}{Z(\ell)} \sum_{n=0}^\ell w(n) \log w(n) \\
& - \frac{1}{Z(\ell)} \Big(\sum_{n=0}^\infty (w(n) - w(n+\ell+1)) \\
& \times \log(w(n) - w(n+\ell+1)) \Big) \\
& + \frac{1}{Z(\ell)} \sum_{n=0}^\ell (n+1) F(n) \log F(n) \\
& + \frac{\ell+1}{Z(\ell)} \sum_{n=\ell+1}^\infty F(n) \log F(n) ,
\end{aligned} \tag{B1}$$

where $Z(\ell) = \sum_{n=0}^\ell w(n)$. If $\lim_{\ell\to\infty} \mathbf{E}(\ell)$ diverges, then we look for the asymptotic rate of divergence of $\mathbf{E}(\ell)$. Otherwise, the process' excess entropy can be defined as $\mathbf{E} = \lim_{\ell\to\infty} \mathbf{E}(\ell)$. We expect $\mathbf{E}$ will often be finite even when μ diverges.

A similar method allows us to calculate C_μ when mean interevent count is infinite. This time, we define U_ℓ as a proxy random variable that is 1 if there has been an event in $\overleftarrow{X}^\ell$ and 0 otherwise. Since U_ℓ is a function of $\mathcal{S}^+$, a standard information-theoretic identity implies that:

$$C_\mu = \mathrm{H}[\mathcal{S}^+|U_\ell] + \mathrm{H}[U_\ell]$$

and, in particular:

$$C_\mu = \lim_{\ell\to\infty} \left(\mathrm{H}[\mathcal{S}^+|U_\ell] + \mathrm{H}[U_\ell] \right) .$$

As before, $\lim_{\ell\to\infty} \Pr(U_\ell = 0) = \lim_{\ell\to\infty} w(\ell) = 0$, so $\lim_{\ell\to\infty} \mathrm{H}[U_\ell] = 0$. Also, $\mathrm{H}[\mathcal{S}^+|U_\ell] = \Pr(U_\ell = 0) \mathrm{H}[\mathcal{S}^+|U_\ell = 0] + \Pr(U_\ell = 1) \mathrm{H}[\mathcal{S}^+|U_\ell = 1]$ by definition. Since there is only one semi-infinite past without an event, $\lim_{\ell\to\infty} \mathrm{H}[\mathcal{S}^+|U_\ell = 0] = 0$. And, $\mathrm{H}[\mathcal{S}^+|U_\ell = 1] = -\sum_{n=0}^\ell \frac{w(n)}{Z(\ell)} \log \frac{w(n)}{Z(\ell)}$. Altogether, this implies:

$$C_\mu = \lim_{\ell\to\infty} \sum_{n=0}^\ell \frac{w(n)}{Z(\ell)} \log \left(1/\frac{w(n)}{Z(\ell)} \right) . \tag{B2}$$

One can also study the growth rate of finite-time statistical complexity estimates which are, after a moment's reflection, the $C_\mu^\ell = -\sum_{n=0}^\ell \frac{w(n)}{Z(\ell)} \log \frac{w(n)}{Z(\ell)}$ estimates above in Eq. (B2).

One comment, perhaps obvious from Eqs. (B1) and (B2), is that whether or not $\mathbf{E}$ and C_μ diverge depends entirely on the asymptotic form of $F(n)$. Another is that the sums in Eq. (B1) can be quite difficult to evaluate numerically when the renewal process has long-range temporal correlations, since then $F(n)$ decays slowly with n.

Appendix C: Finite-time Excess Entropy Estimates with Infinite Mean Interevent Interval

From App. II of Ref. [46]:

$$\begin{aligned}
&\Pr(\mathcal{S}^+ = \sigma^+, \mathcal{S}_\ell^- = \sigma^-|U_\ell = 1) \\
&= \frac{1}{Z} \begin{cases} F(\sigma^+ + \sigma^-) & \sigma^- \le \ell \\ 0 & \sigma^- = \ell + 1 \end{cases} ,
\end{aligned}$$

where the normalization constant is:

$$\begin{aligned}
Z &= \sum_{\sigma^-=0}^\ell \sum_{\sigma^+=0}^\infty F(\sigma^+ + \sigma^-) \\
&= \sum_{\sigma^-=0}^\ell w(\sigma^-) .
\end{aligned}$$

The marginals are easily calculated:

$$\Pr(\mathcal{S}^+ = \sigma^+|U_\ell = 1) = \frac{1}{Z}(w(\sigma^+) - w(\sigma^+ + \ell + 1))$$

and:

$$\Pr(\mathcal{S}_\ell^- = \sigma^- | U_\ell = 1) = \frac{1}{Z} \begin{cases} w(\sigma^-) & \sigma^- \leq \ell \\ 0 & \sigma^- = \ell + 1 \end{cases}.$$

From this, we calculate finite-length excess entropy in nats:

$$\mathbf{E}(\ell) = H[\mathcal{S}_\ell^- | U_\ell = 1] + H[\mathcal{S}^+ | U_\ell = 1] \\ - H[\mathcal{S}^+, \mathcal{S}_\ell^- | U_\ell = 1]$$

$$= \log Z - \frac{1}{Z} \sum_{n=0}^{\ell} w(n) \log w(n)$$

$$- \frac{1}{Z} \Big(\sum_{n=0}^{\infty} (w(n) - w(n+\ell+1))$$

$$\times \log(w(n) - w(n+\ell+1)) \Big)$$

$$+ \frac{1}{Z} \sum_{n=0}^{\infty} \sum_{m=0}^{\ell} F(n+m) \log F(n+m) .$$

This simplifies to:

$$\mathbf{E}(\ell) = \log Z - \frac{1}{Z} \sum_{n=0}^{\ell} w(n) \log w(n)$$

$$- \frac{1}{Z} \Big(\sum_{n=0}^{\infty} (w(n) - w(n+\ell+1))$$

$$\times \log(w(n) - w(n+\ell+1)) \Big)$$

$$+ \frac{1}{Z} \sum_{n=0} \ell(n+1) F(n) \log F(n)$$

$$+ \frac{\ell+1}{Z} \sum_{n=\ell+1}^{\infty} F(n) \log F(n) .$$

Similar manipulations hold for continuous-time processes. Briefly, the time since last event t and time to next event t' have a joint probability distribution proportional to $\phi(t+t')$, since the time since last event plus the time to next event is an interevent interval.

Appendix D: Fractal Renewal Processes

The $\alpha > 1$ case simply requires substituting $\phi(t)$ and $\Phi(t)$ from Eqs. (2)-(3) into Eq. (A2) and solving:

$$\widehat{\mathbf{E}} = \log \overline{T} - \frac{2}{\overline{T}} \int_0^{\infty} \Phi(t) \log \Phi(t) dt$$

$$+ \frac{1}{\overline{T}} \int_0^{\infty} t\phi(t) \log \phi(t) dt. \qquad (D1)$$

After straightforward calculations, we find that:

$$\overline{T} = \frac{\alpha}{\alpha - 1} ,$$

$$\frac{1}{\overline{T}} \int_0^{\infty} \Phi(t) \log \Phi(t) dt = -\frac{1}{\alpha - 1} , \text{ and}$$

$$\frac{1}{\overline{T}} \int_0^{\infty} t\phi(t) \log \phi(t) dt = \log \alpha - \frac{\alpha+1}{\alpha-1} .$$

These together yield:

$$\widehat{\mathbf{E}} = \log \frac{\alpha^2}{\alpha - 1} - 1 .$$

Now, we turn our attention to the case of $0 < \alpha \leq 1$. There are two possibilities for $\widehat{\mathbf{E}}$ when $0 < \alpha \leq 1$. One is that $\widehat{\mathbf{E}}$ diverges, in which case, we only care about the asymptotic rate of divergence of $\widehat{\mathbf{E}}(\ell)$. The other possibility is that $\widehat{\mathbf{E}}$ does not diverge, in which case, we only care about contributions Q to $\widehat{\mathbf{E}}(\ell)$ that are not $o(1)$; i.e., that satisfy $\lim_{\ell \to \infty} Q \neq 0$. Our strategy in evaluating $\widehat{\mathbf{E}}(\ell)$ from Eq. (D1) is to systematically find closed-form expressions for all components that are not $o(1)$.

Direct solution gives:

$$Z = \begin{cases} \frac{\ell^{1-\alpha}}{1-\alpha} & \alpha < 1 \\ \log \ell & \alpha = 1 \end{cases} , \qquad (D2)$$

plus components of $o(1)$:

$$-\frac{1}{Z} \int_0^{\ell} \Phi(t) \log \Phi(t) dt = \begin{cases} -\frac{\alpha}{1-\alpha} + \alpha \log \ell & \alpha < 1 \\ \frac{1}{2} \log \ell & \alpha = 1 \end{cases} \qquad (D3)$$

plus components of $o(1)$; and:

$$\frac{1}{Z} \int_1^{\ell} t\phi(t) \log \phi(t) dt + \frac{\ell}{Z} \int_{\ell}^{\infty} \phi(t) \log \phi(t) dt$$

$$= \begin{cases} -\frac{1-\alpha-2\alpha^2}{\alpha(1-\alpha)} + \log \alpha - (1+\alpha) \log \ell & \alpha < 1 \\ -2 - \log \ell & \alpha = 1 \end{cases} , \qquad (D4)$$

plus components of $o(1)$.

Finally, we address the only component with no simple closed-form expression:

$$\frac{1}{Z} \int_0^{\infty} (\Phi(t) - \Phi(t+\ell)) \log(\Phi(t) - \Phi(t+\ell)) dt$$

$$= \frac{1}{Z} \int_1^{\infty} (t^{-\alpha} - (t+\ell)^{-\alpha}) \log(t^{-\alpha} - (t+\ell)^{-\alpha}) dt$$

$$+ \frac{1}{Z} \int_0^1 (1 - (t+\ell)^{-\alpha}) \log(1 - (t+\ell)^{-\alpha}) dt .$$

Since:

$$\lim_{\ell\to\infty}\frac{1}{Z}\int_0^1(1-(t+\ell)^{-\alpha})\log(1-(t+\ell)^{-\alpha})dt = 0 \ ,$$

we ignore that term as a correction of $o(1)$. The case for $\alpha = 1$ can actually be evaluated explicitly since $\frac{1}{t}-\frac{1}{t+\ell} = \frac{\ell}{t(t+\ell)}$:

$$\lim_{\ell\to\infty}\frac{1}{Z}\int_1^\infty\frac{\ell}{t(t+\ell)}\log(\frac{\ell}{t(t+\ell)})dt = -\frac{1}{2}\log\ell \ .$$

Now, consider the case of $\alpha < 1$. We extract the asymptotic scaling in ℓ of the first term by the change of vari-

ables $u = \ell t$, giving:

$$\frac{1}{Z}\int_1^\infty(t^{-\alpha}-(t+\ell)^{-\alpha})\log(t^{-\alpha}-(t+\ell)^{-\alpha})dt$$

$$=\frac{\ell^{1-\alpha}}{Z}\int_{1/\ell}^\infty(u^{-\alpha}-(1+u)^{-\alpha})\log(\ell^{-\alpha}(u^{-\alpha}-(1+u)^{-\alpha}))du$$

$$=-\alpha\frac{\ell^{1-\alpha}\log\ell}{Z}\int_{1/\ell}^\infty u^{-\alpha}-(1+u)^{-\alpha}du$$

$$+\frac{\ell^{1-\alpha}}{Z}\int_{1/\ell}^\infty(u^{-\alpha}-(1+u)^{-\alpha})\log(u^{-\alpha}-(1+u)^{-\alpha})du \ .$$

The first of the two integrals can be evaluated explicitly as:

$$\int_{1/\ell}^\infty u^{-\alpha}-(1+u)^{-\alpha}du = -\frac{\ell^{\alpha-1}}{1-\alpha}+\frac{\ell^{\alpha-1}}{1-\alpha}(\ell+1)^{1-\alpha} \ .$$

So, that we find the first term's asymptotic behavior to be:

$$-\alpha\frac{\ell^{1-\alpha}\log\ell}{Z}\int_{1/\ell}^\infty u^{-\alpha}-(1+u)^{-\alpha}du \sim -\alpha\log\ell \ ,$$

plus corrections of $o(1)$. One of the more notable corrections of $o(1)$ is proportional to $\frac{\log\ell}{Z}$, which is $o(1)$ for $\alpha < 1$ and otherwise has a nonzero limiting value when $\ell\to\infty$.

Surprisingly, the latter of the two integrals limits to a finite value for $\alpha < 1$:

$$\lim_{\ell\to\infty}\frac{\ell^{1-\alpha}}{Z}\int_{1/\ell}^\infty(u^{-\alpha}-(1+u)^{-\alpha})\log(u^{-\alpha}-(1+u)^{-\alpha})du$$

$$=(1-\alpha)\int_0^\infty(u^{-\alpha}-(1+u)^{-\alpha})\log(u^{-\alpha}-(1+u)^{-\alpha})du \ ,$$

where we used $\lim_{\ell\to\infty}\frac{\ell^{1-\alpha}}{Z} = 1-\alpha$ for $\alpha < 1$. As a result, we find that:

$$\frac{1}{Z}\int_0^\infty(\Phi(t)-\Phi(t+\ell))\log(\Phi(t)-\Phi(t+\ell))dt$$

$$=\begin{cases}-\frac{1}{2}\log\ell & \alpha = 1\\ -\alpha\log\ell+(1-\alpha)\int_0^\infty(u^{-\alpha}-(1+u)^{-\alpha})\log(u^{-\alpha}-(1+u)^{-\alpha})du & 0<\alpha<1\end{cases} \ , \quad (D5)$$

plus corrections of $o(1)$. Altogether, combining Eqs. (D2)-(D4) and (D5) into Eq. (D1), we recover Eq. (4) of the main text.

As discussed there, we still must evaluate $\mathbf{E}(\ell)$ at $\alpha = 1$. We focus again on asymptotic expansions in ℓ and drop corrections to expressions that do not contribute to

E. When $\alpha = 1$:

$$Z(\ell) = 1 + \sum_{n=1}^\ell\frac{1}{n} = \log\ell \ ,$$

plus corrections of $O(1)$. Next, we evaluate:

$$-\sum_{n=0}^{\ell} w(n) \log w(n) = \sum_{n=1}^{\ell} \frac{\log n}{n}$$
$$= \sum_{n=2}^{\ell} \frac{\log n}{n} .$$

Since $\frac{\log n}{n}$ is a monotone decreasing function with n, we lower- and upper-bound this sum using integrals: $\int_2^{\ell+1} \frac{\log n}{n} dn \leq \sum_{n=2}^{\ell} \frac{\log n}{n} \leq \frac{\log 2}{2} + \int_2^{\ell} \frac{\log n}{n} dn$. These are easily evaluated, giving:

$$-\sum_{n=0}^{\ell} w(n) \log w(n) = -\frac{1}{2} \log^2 \ell ,$$

plus corrections of $O(1)$. For other sums, we need an expression for $F(n)$:

$$F(n) = w(n) - w(n+1)$$
$$= \begin{cases} 0 & n = 0 \\ \frac{1}{n(n+1)} & n \geq 1 \end{cases} .$$

Then, we evaluate:

$$\sum_{n=0}^{\ell} (n+1) F(n) \log F(n) = -2 \sum_{n=1}^{\ell} \frac{\log n}{n} + \sum_{n=1}^{\ell} \frac{\log(1 + \frac{1}{n})}{n}$$
$$= \log^2 \ell ,$$

plus corrections of $O(1)$, where we have noted that $\sum_{n=1}^{\infty} \frac{\log(1+\frac{1}{n})}{n}$ converges since $\int_1^{\infty} \frac{\log(1+\frac{1}{x})}{x} dx$ converges. The next term takes the form:

$$(\ell+1) \sum_{\ell+1}^{\infty} F(n) \log F(n) = -(\ell+1) \sum_{\ell+1}^{\infty} \frac{\log(n(n+1))}{n(n+1)} .$$

We can bound the sum using $\int_{\ell+1}^{\infty} \frac{\log(n(n+1))}{n(n+1)} dn \leq \sum_{\ell+1}^{\infty} \frac{\log(n(n+1))}{n(n+1)} \leq \frac{\log(\ell^2+\ell)}{\ell^2+\ell} + \int_{\ell+1}^{\infty} \frac{\log(n(n+1))}{n(n+1)} dn$. These integrals are both easily evaluated, revealing an asymptotic form of:

$$(\ell+1) \sum_{\ell+1}^{\infty} F(n) \log F(n) = -2 \log \ell ,$$

plus corrections of $O(1)$. Finally, to evaluate the last

term in the sum, we note that:

$$w(n) - w(n+\ell+1) = \frac{1}{n(1 + \frac{n}{\ell+1})}$$
$$= \frac{1/\ell + 1}{\frac{n}{\ell+1}(1 + \frac{n}{\ell+1})} ,$$

when $n \geq 1$. We define $x_n = \frac{n}{\ell+1}$ with $dx_n = \frac{1}{\ell+1}$ and write:

$$w(n) - w(n+\ell+1) = \frac{dx_n}{x_n(1 + x_n)} .$$

Then:

$$\sum_{n=0}^{\infty} (w(n) - w(n+\ell+1)) \log(w(n) - w(n+\ell+1))$$
$$= (1 - w(\ell+1)) \log(1 - w(\ell+1))$$
$$+ \log dx_n \sum_{n=1}^{\infty} \frac{dx_n}{x_n(1 + x_n)} + \sum_{n=1}^{\infty} \frac{\log(x_n(1 + x_n))}{x_n(1 + x_n)} dx_n .$$

The first term is $o(1)$, since $\lim_{\ell \to \infty} (1 - w(\ell+1)) \log(1 - w(\ell+1)) = 0$. We can view the other two sums as Riemann sums for integrals $\int_{1/\ell}^{\infty} \frac{dx}{x(1+x)}$ and $\int_{1/\ell}^{\infty} \frac{\log(x(1+x))}{x(1+x)} dx$ respectively, giving:

$$\sum_{n=1}^{\infty} \frac{dx_n}{x_n(1 + x_n)} = \log \ell ,$$

plus corrections of $o(1)$ and:

$$\sum_{n=1}^{\infty} \frac{\log(x_n(1 + x_n))}{x_n(1 + x_n)} dx_n = -\frac{1}{2} \log^2 \ell .$$

plus corrections of $o(1)$. Altogether, substituting the above expressions into Eq. (B1) yields:

$$\mathbf{E}(\ell) = \log \log \ell - 2 ,$$

plus corrections of $o(1)$. The various divergences of order $\log \ell$ all cancel one another, but the divergence of $\log \log \ell$ due to the $\log \ell$ divergence in $Z(\ell)$ remains, just as for the continuous-time case. When $F(n)$ is monotone decreasing at some finite N sufficiently rapidly, manipulations similar to those above imply that divergence in $\widehat{\mathbf{E}}$ is a sufficient condition for divergence in $\mathbf{E}$.

[1] G. M. Shim and M. Y. Choi. Algebraic decay of correlations in neural networks. *Phys. Rev. A*, 46(8):5292–5295, 1992.

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

[3] J. F. Alves and V. Pinheiro S. Luzzatto. Markov structures and decay of correlations for non-uniformly expanding dynamical systems. *Ann. I. H. Poincare – Anal. Nonlin.*, 22:817–839, 2005.

[4] R. F. Voss and J. Clarke. '1/f noise' in music and speech. *Nature*, 258:317–318, 27 November 1975.

[5] A. d'Amico and P. Mazzetti. *Noise in Physical Systems and 1/f Noise.* Elsevier Science Publishers, Amsterdam, The Netherlands, 1986.

[6] W. H. Press. Flicker noises in astronomy and elsewhere. *Comments on Astrophysics*, 7(4):103–119, 1978.

[7] H. E. Hurst. Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Engineers*, 116:770–799, 1951.

[8] B. B. Mandelbrot and J. R. Wallis. Noah, Joseph, and operational hydrology. *Water Resources Research*, 4(5):909–918, 1968.

[9] D. J. Daley. The Hurst index of long-range dependent renewal processes. *Ann. Prob.*, 27(4):2035–2041, 1999.

[10] J. Beran, Y. Feng, S. Ghosh, and R. Kulik. *Long-memory Processes: Probabilistic Properties and Statistical Methods.* Springer, London, United Kingdom, 2013.

[11] G. K. Zipf. *The Psycho-Biology of Language: An Introduction to Dynamic Philology.* Houghton Mifflin Company, Boston, Massachusetts, 1935.

[12] B. Mandelbrot. An informational theory of the statistical structure of languages. In W. Jackson, editor, *Communication Theory*, pages 486–502. Butterworths, London, 1953.

[13] D. C. Knill, D. Field, and D. Kersten. Human discrimination of fractal images. *J. Opt. Soc. Am. A*, 7(6):1113–1123, 1990.

[14] T. Kumar, P. Zhou, and D. A. Glaser. Comparison of human performance with algorithms for estimating fractal dimension. *J. Opt. Soc. Am. A*, 10(6):1136–46, 1993.

[15] B. B. Mandelbrot. Self-similar error clusters in communication systems and the concept of conditional stationarity. *IEEE Trans. Commun. Technol.*, 13:71–90, 1965.

[16] B. B. Mandelbrot. *Multifractals and 1/f Noise: Wild Self-Affinity in Physics (1963- 1976).* Springer, New York, first edition, 1999.

[17] P. Riechers. Private communication, 2015.

[18] B. D. Johnson, J. P. Crutchfield, C. J. Ellison, and C. S. McTague. Enumerating finitary processes. page submitted, 2012. SFI Working Paper 10-11-027; arxiv.org:1011.0036 [cs.FL].

[19] R. G. James, J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield. Many roads to synchrony: Natural time scales and their algorithms. *Phys. Rev. E*, 89:042135, 2014.

[20] G. Samorodnitsky. Long range dependence. *Foundations and Trends in Stochastic Systems*, 1(3):163–257, 2007.

[21] J. P. Crutchfield. The calculi of emergence: Computation, dynamics, and induction. *Physica D*, 75:11–54, 1994.

[22] W. Bialek, I. Nemenman, and N. Tishby. Predictability, complexity, and learning. *Neural Comp.*, 13:2409–2463, 2001.

[23] J. P. Crutchfield and D. P. Feldman. Regularities un-

[24] N. Travers and J. P. Crutchfield. Infinite excess entropy processes with countable-state generators. *Entropy*, 16:1396–1413, 2014.

[25] L. Debowski. On hidden Markov processes with infinite excess entropy. *J. Theo. Prob.*, pages 1–13, 2012.

[26] J. P. Crutchfield and S. Marzen. Signatures of infinity: Nonergodicity and resource scaling in prediction, complexity, and learning. *Phys. Rev. E*, 91(5):050106, 2015.

[27] W. L. Smith. Renewal theory and its ramifications. *J. Roy. Stat. Soc. B*, 20(2):243–302, 1958.

[28] W. Gerstner and Kistler. Statistics of spike trains. In *Spiking Neuron Models*. Cambridge University Press, Cambridge, United Kingdom, 2002.

[29] F. Beichelt. *Stochastic Processes in Science, Engineering and Finance.* Chapman and Hall, New York, 2006.

[30] V. S. Barbu and N. Limnios. *Semi-Markov chains and hidden semi-Markov models toward applications: Their Use in Reliability and DNA Analysis*, volume 191. Springer, New York, 2008.

[31] S. B. Lowen and M. C. Teich. Fractal renewal processes. *IEEE Trans. Info. Th.*, 39(5):1669–1671, 1993.

[32] S. Thurner, S. B. Lowen, M. C. Feurstein, and C. Heneghan. Analysis, synthesis and estimation of fractal-rate stochastic point processes. *Fractals*, 5(4):565–595, 1997.

[33] R. Cakir, P. Grigolini, and A. A. Krokhin. Dynamical origin of memory and renewal. *Phys. Rev. E*, 74:021108, 2006.

[34] S. Bianco, M. Ignaccolo, M. S. Rider, M. J. Ross, P. Winsor, and P. Grigolini. Brain, music, and non-Poisson renewal processes. *Phys. Rev. E*, 75:061911, 2007.

[35] C.-B. Li, H. Yang, and T. Komatsuzaki. Multiscale complex network of protein conformational fluctuations in single-molecule time series. *Proc. Natl. Acad. Sci. USA*, 105:536–541, 2008.

[36] T. Akimoto, T. Hasumi, and Y. Aizawa. Characterization of intermittency in renewal processes: Application to earthquakes. *Phys. Rev. E*, 81:031133, 2010.

[37] D. Kelly, M. Dillingham, A. Hudson, and K. Wiesner. A new method for inferring hidden Markov models from noisy time sequences. *PLoS One*, 7(1):e29703, 01 2012.

[38] M. Montero and J. Villarroel. Monotonic continuous-time random walks with drift and stochastic reset events. *Phys. Rev. E*, 87:012116, 2013.

[39] M. Bologna, B. J. West, and P. Grigolini. Renewal and memory origin of anomalous diffusion: A discussion of their joint action. *Phys. Rev. E*, 88:062106, 2013.

[40] T. Onaga and S. Shinomoto. Bursting transition in a linear self-exciting point process. *Phys. Rev. E*, 89:042817, 2014.

[41] S. B. Lowen and M. C. Teich. Fractal renewal processes generate 1/f noise. *Phys. Rev. E*, 47(2):992–1001, 1993.

[42] N. Travers and J. P. Crutchfield. Equivalence of history and generator ϵ-machines. 2014. SFI Working Paper 11-11-051; arxiv.org:1111.4500 [math.PR].

[43] J. P. Crutchfield and K. Young. Inferring statistical com-

plexity. *Phys. Rev. Let.*, 63:105–108, 1989.

[44] C. R. Shalizi and J. P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *J. Stat. Phys.*, 104:817–879, 2001.

[45] W. Lohr. Properties of the statistical complexity functional and partially deterministic HMMs. *Entropy*, 11(3):385–401, 2009.

[46] S. Marzen and J. P. Crutchfield. Informational and causal architecture of discrete-time renewal processes. *Entropy*, 17(7):4891–4917, 2015.

[47] J. P. Crutchfield, C. J. Ellison, and J. R. Mahoney. Time's barbed arrow: Irreversibility, crypticity, and stored information. *Phys. Rev. Lett.*, 103(9):094101, 2009.

[48] 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):1005–1034, 2009.

[49] S. Marzen, M. R. DeWeese, and J. P. Crutchfield. Time resolution dependence of information measures for spiking neurons: Scaling and universality. *Front. Comput. Neurosci.*, 9:109, 2015.

[50] S. Marzen and J. P. Crutchfield. Information anatomy of stochastic equilibria. *Entropy*, 16(9):4713–4748, 2014.

[51] T. Graves, R. B. Gramacy, C. Franzke, and N. Watkins. A brief history of long memory. *arXiv preprint arXiv:1406.6018*, 2014.

[52] J. P. Crutchfield, P. Riechers, and C. J. Ellison. Exact complexity: Spectral decomposition of intrinsic computation. submitted. Santa Fe Institute Working Paper 13-09-028; arXiv:1309.3792 [cond-mat.stat-mech].

[53] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, second edition, 2006.

[54] T. Mora and W. Bialek. Are biological systems poised at criticality? *J. Stat. Phys.*, 144(2):268–302, 2011.

[55] L. S. Liebovitch and J. M. Sullivan. Fractal analysis of a voltage-dependent potassium channel from cultured mouse hippocampal neurons. *Biophys. J.*, 52(6):979–988, 1987.

[56] S. Dedeo. Collective phenomena and non-finite state computation in a human social system. *PLoS One*, 8(10):e75818, 2013.

[57] S. Dedeo. Group minds and the case of Wikipedia. *Human Computation*, 1(1):5–29, 2014.

[58] M. S. Pinsker. *Information and information stability of random variables and processes*. Holden-Day, San Francisco, California, 1964.