UC Davis
UC Davis Previously Published Works

Title

Fraudulent white noise: Flat power spectra belie arbitrarily complex processes

Permalink

Journal

ISSN

Authors

Riechers, Paul M
Crutchfield, James P

Publication Date

DOI

Peer reviewed

Fraudulent White Noise:
Flat power spectra belie arbitrarily complex processes

Paul M. Riechers[*]

Complexity Institute
Nanyang Technological University
639798 Singapore, Singapore and
School of Physical and Mathematical Sciences
Nanyang Technological University
637371 Singapore, Singapore

James P. Crutchfield[†]

Complexity Sciences Center
Department of Physics
University of California at Davis
One Shields Avenue, Davis, CA 95616
(Dated: September 2, 2019)

Power spectral densities are a common, convenient, and powerful way to analyze signals. So much so that they are now broadly deployed across the sciences and engineering—from quantum physics to cosmology, and from crystallography to neuroscience to speech recognition. The features they reveal not only identify prominent signal-frequencies but also hint at mechanisms that generate correlation and lead to resonance. Despite their near-centuries-long run of successes in signal analysis, here we show that flat power spectra can be generated by highly complex processes, effectively hiding all inherent structure in complex signals. Historically, this circumstance has been widely misinterpreted, being taken as the renowned signature of "structureless" *white noise*—the benchmark of randomness. We argue, in contrast, to the extent that most real-world complex systems exhibit correlations beyond pairwise statistics their structures evade power spectra and other pairwise statistical measures. To make these words of warning operational, we present constructive results that explore how this situation comes about and the high toll it takes in understanding complex mechanisms. First, we give the closed-form solution for the power spectrum of a very broad class of structurally-complex signal generators. Second, we demonstrate the close relationship between eigen-spectra of evolution operators and power spectra. Third, we characterize the minimal generative structure implied by *any* power spectrum. Fourth, we show how to construct arbitrarily complex processes with flat power spectra. Finally, leveraging this diagnosis of the problem, we point the way to developing more incisive tools for discovering structure in complex signals.

PACS numbers: 02.50.-r 05.45.Tp 02.50.Ey 02.50.Ga

We dare not neglect any of the tools that have proved useful in the past. But equally we dare not find ourselves confined to their use.

John Tukey, *The Future of Data Analysis*, 1962 [1, p. 60]

I. INTRODUCTION

Innovative science probes the unknown. Success in discovering the mechanisms that underlie the systems we seek to understand, though, requires distinguishing structure from noise. Often, this distinction falls to discretion: structure is that part of a signal we can predict, while noise stands in as a catch-all for everything else.

This conundrum holds especially in the analysis of signals from truly complex systems, as when analyzing data from multi-electrode arrays in brain tissue or social experiments. These systems are often said to be 'noisy' even though the so-called noise may be entirely functionally relevant, but in an unknown way. Such descriptions fall far short of a principled approach that explains all trends and correlational structure, which would claim success only when all that remains unexplained in the signal is structureless white noise. Even this principled approach ultimately begs the central question, though: how do we test if an apparently random signal is truly white noise?

The challenge of discovering structure in noisy signals is compounded manifold, as we demonstrate in the following, when our chosen observables hide arbitrary amounts of in-principle-predictable structure behind a familiar signature of white noise—the flat power spectrum. Said simply, observables can be completely de-

[*] pmriechers@gmail.com
[†] chaos@ucdavis.edu

void of pairwise correlation, while still embodying structure in higher-order correlations. More precisely, structure can be hidden beyond any arbitrarily-large order-N correlation—that not appearing in pairwise, three-way, nor any n-way statistics, up to some arbitrarily large N [2]. Moreover, the hidden structure can be arbitrarily sophisticated. It can be used, for example, to embed messages while shifting (and so hiding) the messages' content beyond N-way correlation. Here, we explore the structures conveyed and hidden by power spectra, revealing a novel perspective on the interplay between structure and noise in Fourier analysis.

Section II discusses temporal structure and provides closed-form expressions for the power spectra from autonomous signal generators. It highlights the intimate connection between power spectra and eigen-spectra of a system's time-evolution generator. Section III then introduces a suite of results on structure that is *hidden* by power spectra. Notably, it introduces a general condition for *fraudulent white noise* processes—structured processes with a flat power spectrum—which applies very broadly, including to input-dependent processes with nonstationary high-order statistics. Taken altogether the results emphasize the power spectrum's shortcomings for the task of structure detection. In response, Sec. IV considers more sophisticated measures of structure, introducing the *dependence function* which identifies the presence of novel finite-range dependencies that contribute to total correlation. Section V concludes the development. Appendices present detailed derivations, as well as several generalizations, of the main results.

II. STRUCTURE IN SPACE AND TIME?

Pairwise correlations are encountered throughout the sciences and engineering, especially in statistical physics. They are assumed, estimated, relied on, designed with, and used for interpretation widely. The following explores several specific examples of pairwise correlation that arise in different fields. These will set the context for our development, particularly for experts in the associated fields. However, our general results should be accessible and relevant across disciplines, as they rely primarily on basic probability theory and linear algebra.

A well-studied lesson from statistical physics is that diverging correlation length heralds the emergence of new types of order. Remarkably, mechanistically-distinct physical systems share many universal behaviors near a critical point of emergent order, including the scaling of spatial pairwise correlation length [3]. More broadly, pairwise correlations are indicators of fundamental physical processes. For example, the fluctuation–dissipation theorem says that pairwise temporal correlations in equilibrium determine the friction encountered in transport processes. The Green–Kubo relations [4] make this explicit. Far from equilibrium, say in computing devices and biological systems composed of excitable media, temporal correlations are signatures of richly coordinated state-trajectories.

Pairwise correlations are directly viewed in the frequency domain via power spectral densities. Indeed, power spectra are employed as a basic data analysis tool in many scientific domains and have been key to major scientific discoveries. For example, comparing alternative theoretical predictions for power spectra of incident electromagnetic radiation from locally-thermalized bodies, a unexpected discrepancy—the ultraviolet catastrophe—led to the acceptance of Planck's theory of quantized energies and the subsequent birth of quantum theory [5–7]. A contemporary example of the prominent role of power spectra is seen in the exquisitely detailed map of the cosmic microwave background (CMB)—a snapshot of the early universe's spatial correlations. In fact, models of the early universe are now benchmarked against their ability to replicate the CMB power spectrum [8].

In applied mathematics, power spectra played a key role in highlighting the defining features of the strange attractors of dynamical systems theory [9, 10]. This led to the discovery of Ruelle–Pollicott resonances, where mixing and the decay of correlations in chaotic systems were related to the point spectrum of the Ruelle–Perron–Frobenius operator [11–13]. Indeed, the power spectra of chaotic systems are still actively used to analyze the behavior of everything from open quantum systems [14, 15] to climate models [16].

Power spectra are regularly used to discover structure in materials science and biology, too. X-ray diffraction patterns—used to identify crystalline and molecular organization and central to discovering DNA's double-helix [17–20]—are power spectra of scatterer densities, as we explain in App. A. Power spectra have been used to identify temporal correlations in single-neuron spike trains, refuting the common Poissonian white-noise assumption common in theoretical and computational neuroscience [21–24]. This allows the possibility that temporal correlations in the spike train—rather than just the firing rate—can play an important role in the neural code [25, 26]. On a much larger (mean-field) scale, brain wave activity in different frequency bands gives signatures of normal brain functioning, as well as pathological conditions. Rhythmic brain-wave activity is clinically assessed through real-time power spectra of electroencephalography (EEG) signals [27–29].

From the smallest to the largest scales in the universe, when probing both the inanimate and the animate, power

spectra are a central diagnostic tool for structure and validating scientific models. Their use is so important that special-purpose spectral and network analyzers are standard laboratory test equipment; they can be readily purchased from dozens of major manufacturers.

Power spectra report pairwise correlations in a signal. But how much of a system's structure is faithfully represented by pairwise correlation? Are there important types of order that evade power spectra completely? To answer these questions, we first consider the problem of hidden structure concretely through the lens of autocorrelation and power spectra. Only then, once the strengths and weakness of power spectra are clear, do we move on along to more sophisticated measures of structure. Along the way we trace a path that begins to reveal what one can mean by "statistical dependency", "correlation", and "structure".

A.   Correlation and Power Spectra

To provide a common ground, consider discrete-time processes described by an interdependent sequence $\ldots X_0 X_1 X_2 \ldots$ of random variables X_t that at time t take on values $x \in \mathcal{A}$ within an alphabet assumed (for now) to be a subset of the complex numbers: $\mathcal{A} \subset \mathbb{C}$. (For other kinds of stochastic process, t may represent spatial or angular coordinates. For concreteness here, we interpret t as indexing time.) An observed process may have a discrete domain, as with a classical discrete-time communication channel or a series of quantum measurements, or otherwise may be a regularly-sampled process evolving in continuous time.

A signal's *power spectrum* or, more properly, its *power spectral density* quantifies how its power is distributed across frequency [30, 31]. For a discrete-domain process it is:

$$P(\omega) = \lim_{N \to \infty} \frac{1}{N} \left\langle \left| \sum_{t=1}^{N} X_t e^{-i\omega t} \right|^2 \right\rangle , \qquad (1)$$

where ω is the angular frequency and the angle brackets denote the expected value over the random variable chain $X_1 X_2 X_3 \ldots X_N$.

For *wide-sense stationary* stochastic processes the *autocorrelation function*:

$$\gamma(\tau) = \left\langle \overline{X}_t X_{t+\tau} \right\rangle , \qquad (2)$$

is independent of the global time shift t and depends only on the relative time-separation τ between observables [32]. The bar above X_t denotes its complex conjugate. Equation (2) makes plain the connection between pairwise statistics and the pairwise correlation function. For wide-sense stationary stochastic processes, the power spectrum is also determined by the signal's autocorrelation function $\gamma(\tau)$:

$$P(\omega) = \lim_{N \to \infty} \frac{1}{N} \sum_{\tau=-N}^{N} \left(N - |\tau| \right) \gamma(\tau) e^{-i\omega\tau} . \qquad (3)$$

The windowing function $N - |\tau|$ appearing in Eq. (3) is a direct consequence of Eq. (1); it is not imposed externally, as is common practice in signal analysis. (This factor is important for controlling convergence in our subsequent derivations.)

Equation (3) suggests that the power spectrum is very *nearly* the Fourier transform of the autocorrelation function, except for the $N - |\tau|$ term. In fact, the Wiener–Khinchin theorem proves that the power spectrum is indeed *equal* to the Fourier transform of the autocorrelation function for wide-sense stationary processes [33, 34]. Note, too, that the pairwise correlation function $\gamma(\tau)$ can be obtained via the inverse Fourier transform of the power spectrum $P(\omega)$.

B.   Temporal Structurelessness

Our goal is to understand temporal structure and to identify it in stochastic processes. To detect structure, even when hidden, we first must establish a baseline reference process that has no temporal structure: genuine white noise.

White noise processes are those for which each random variable X_t is statistically independent of all others $X_{t' \neq t}$, and each is identically distributed according to the same probability density function (PDF) over the alphabet. That is, the random variables in the sequence are *independent and identically distributed* (IID).

Familiar examples include a sequence of coin flips or the sequence of sums when rolling a pair of dice. As an example from contemporary physics, consider the (classical) process that results from observing a sequence of Bell-pair quantum states [35]. For each Bell pair, one of the entangled particles is sent to Alice and the other sent to Bob. Alice makes a sequence of measurements (along any measurement axis). The measurement output sequence she observes is pure white noise, with each measurement outcome having equal and independent probability of being up or down along the measurement axis. In fact, more sophisticated deployments of Bell pairs are being developed to provide certifiable random number generation [36]. Experiments now concentrate on increasing the rate of generating white noise [37, 38].

The most recognizable feature of all white noise pro-

cesses is their flat power spectrum. For any IID process, it follows directly from Eq. (2) that $\gamma(0) = \langle |X_t|^2 \rangle$, whereas $\gamma(\tau) = |\langle X_t \rangle|^2$ for $\tau \neq 0$. From Eq. (3), this immediately yields the familiar flat power spectrum of white noise, together with a δ-function at zero frequency, corresponding to the constant offset of the noise. For real-valued IID processes with zero mean, this simplifies further to $\gamma(\tau) = \sigma^2 \delta_{0,\tau}$ and so $P(\omega) = \sigma^2$. In fact, the flat power spectrum has height equal to the variance $\sigma^2 = \langle X_t^2 \rangle - \langle X_t \rangle^2$ of the white noise for any real-valued IID process. The flat power spectrum for IID processes indicates that any temporal structure in the generating source has such short memory that it vanishes within the short sampling time Δt between each observation.

Gaussian white noises tend to be the most commonly employed white noise processes and, usually, for good reason. By the central limit theorem, Gaussian white noise arises generically in systems whenever *many* events—with amplitude of finite variance and with rapidly decaying correlation (compared to the timescale between observations)—contribute additively to each individual observation. Suppose, for example, that the expected number of these contributions to each new observation is simply proportional to the time since the last observation. When sampled at interval Δt, the central limit theorem then tells us that each observation of the accumulated noise is IID and Gaussian distributed with variance $\sigma_\eta^2 \propto \Delta t$. This immediately leads to the familiar standard deviation of $\sigma_\eta \propto \sqrt{\Delta t}$ of the additive noise $\eta(t)$ that appears when numerically integrating stochastic differential equations (e.g., Langevin equations); this, in turn, produces the trajectories of slower random variables [39].

The memoryless nature of repetitive sampling from a distribution is apparent in the state machine shown in Fig. 1(a). The same Gaussian distribution is repeatedly sampled with probability 1 (as depicted by the self-transition probability there) for each observation, regardless of what happened previously.

Other "structureless" white noises are also possible. In fact, any of an uncountably infinite set of different IID processes—Gaussian, Poisson, Bernoulli, or any process that resamples a particular distribution at each timestep—all yield the flat power spectrum or white noise. Non-Gaussian noise can emerge from repetitive sampling of a system's (non-Gaussian) stationary distribution when the relaxation timescales are far shorter than the time elapsed between samples. Alternatively, non-Gaussian white noise can arise when only a few physical events contribute to each observation, in which case the non-Gaussianity may reveal features of the physical generative mechanism. Nevertheless, these processes possess no temporal structure on the timescale of observation
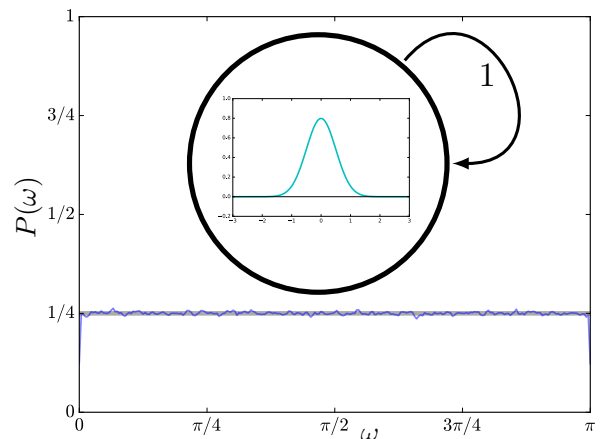
(a) Gaussian white noise process (inset) and its flat power spectrum.
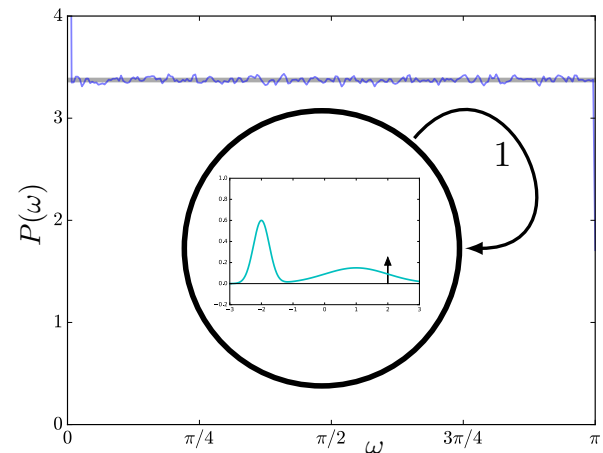
(b) Non-Gaussian white noise process (inset) and its flat power spectrum.

FIG. 1. Genuine white noise processes have no memory: This is represented graphically by a machine with a single state that is repeatedly visited with each observation. The same probability density function, inscribed in the state, is sampled at each timestep. (a) Gaussian white noise memoryless stochastic process. (b) Another white noise process, although non-Gaussian. For each (a) and (b), the flat power spectrum is given theoretically (thick gray), with height equal to the variance of the probability density function. We also display the numerically-obtained power spectrum (thin blue) for each. The class of all possible (not-necessarily Gaussian) memoryless white noises is identical with the class of processes generated by single-state machines. This class, in turn, is identical to that of all IID processes (spanning all possible probability density functions). These temporally structureless processes constitute all possible varieties of genuine white noise.

and, in particular, generate absolutely no correlations in the sequence of observations.

The hallmark of this structural paucity is the single state for the hidden Markov model (HMM) that describes all of these IID processes, as depicted in Fig. 1(b) [40]. The single states means that no influences from the past

can affect the next or future samples. These are the genuine white noises.

In sharp contrast, we will consider stochastic processes with arbitrarily sophisticated temporal structure on the timescale of observation. The much more general class we next consider allows for a thorough investigation of temporally structured stochastic processes. One surprising feature is that these very structured processes, described by arbitrarily complicated transition dynamics within memoryful collections of internal states, can have the flat power spectrum of white noise.
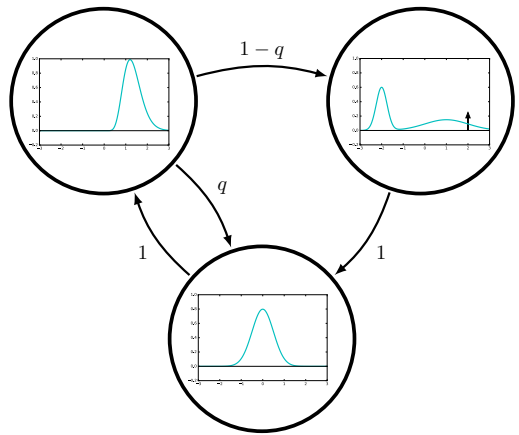

FIG. 2. Simple 3-state HMM that generates a stochastic process according to the state-to-state transition dynamic T and the probability density functions (PDFs) $\{p(X|s)\}_{s \in \boldsymbol{\mathcal{S}}}$ associated with each state. Theorem 1 asserts that its power spectrum is the same (modulo constant offset) as the power spectrum generated from an alternative process where the state's PDFs are solely concentrated at the average value $\langle x \rangle_{p(X|s)}$ of the original PDF associated with the state.

C. Models of Temporal Structure

Structure arises over time from the interdependence between observables. To explicitly address structure in a broad class of temporally structured processes, we use hidden Markov models (HMMs) as our preferred representation for autonomous signal generators [41–47]. Later sections will introduce yet more sophisticated models with input dependence.

Despite Markovian state-to-state transitions, HMMs can generate temporally-structured non-Markovian stochastic processes—those with infinite Markov order. Processes generated by even finite-state HMMs, in fact, *typically* have infinite-range statistical dependencies between observables since simple state-transition motifs guarantee this feature [48]. In addition to this richness and their ability to compactly generate the exact temporal statistics of nonlinear dynamical systems, HMMs are attractive since they are amenable to linear operator techniques [49–55].

Let the 4-tuple $\mathcal{M} = \big(\boldsymbol{\mathcal{S}}, \mathcal{A}, \mathcal{P}, T\big)$ be a discrete-time HMM that generates the stationary stochastic process $\ldots X_{-2}X_{-1}X_0X_1X_2\ldots$ according to the following. $\boldsymbol{\mathcal{S}}$ is the (finite) set of hidden states of the hidden Markov chain and $\mathcal{A} \subseteq \mathbb{C}$ is the observable alphabet. $\mathcal{S}_t$ is the random variable for the hidden state at time t that takes on values $s \in \boldsymbol{\mathcal{S}}$. X_t is the random variable for the observation at time t that takes on values $x \in \mathcal{A}$.

Given the hidden state at time t, the possible observations are distributed according to the conditional probability density functions: $\mathcal{P} = \big\{p(X_t|\mathcal{S}_t = s)\big\}_{s \in \boldsymbol{\mathcal{S}}}$. For each $s \in \boldsymbol{\mathcal{S}}$, $p(X_t|\mathcal{S}_t = s)$ may be abbreviated as $p(X|s)$ since the probability density function in each state is assumed to not change over t. Similarly, we will write $p(x|s)$ for $p(X_t = x|\mathcal{S}_t = s)$. Finally, the hidden-state-to-state stochastic transition matrix T has elements $T_{s,s'} = \Pr(\mathcal{S}_{t+1} = s'|\mathcal{S}_t = s)$, which give the probability of transitioning from hidden state s to s' given that the system is in state s, where $s, s' \in \boldsymbol{\mathcal{S}}$. It is important for subsequent developments that $\Pr(\cdot)$ denotes a probability

in contrast to $p(\cdot)$ which denotes a probability *density*.

Epitomizing the class of processes considered, Fig. 2 presents a rather simple HMM with continuous observable alphabet $\mathcal{A} = \mathbb{R}$, whose samples are distributed according to the probability density function shown within each hidden state. As seen in the HMM's top-right state, both continuous probability density functions and discrete output probabilities can be accommodated in this framework: Finite probability of a particular observable is accomplished by an appropriately weighted Dirac δ-function in the probability density function. The memoryful structure in Fig. 2 should be contrasted with the completely memoryless process of sampled Gaussian white noise shown in Fig. 1.

Figure 3's Bayes network depicts the structure of conditional independence among the random variables for these memoryful signal generators. For example, for a generic HMM, $p(X_t|X_{t-N}\ldots X_{t-2}X_{t-1} = x_{t-N}\ldots x_{t-2}x_{t-1})$ cannot be simplified since the condition on even arbitrarily distant past observables can influence the probability of the current observable. However, when conditioning on hidden states, the situation can simplify markedly. For example:

$$\begin{aligned} p(X_t|&X_{t-N}\ldots X_{t-2}X_{t-1} = x_{t-N}\ldots x_{t-2}x_{t-1}, \\ &\mathcal{S}_{t-N}\ldots\mathcal{S}_{t-2}\mathcal{S}_{t-1} = s_{t-N}\ldots s_{t-2}s_{t-1}) \\ &= p(X_t|\mathcal{S}_{t-1} = s_{t-1}) \\ &= \sum_{s \in \boldsymbol{\mathcal{S}}} T_{s_{t-1},s}\, p(X|s) \ . \end{aligned}$$
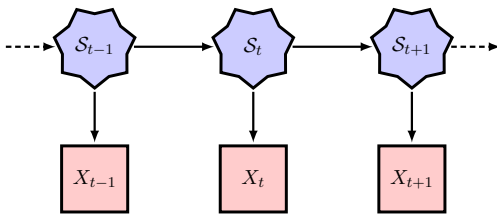
FIG. 3. Bayesian network for a state-emitting hidden Markov model graphically depicts the structure of conditional independence among random variables for the hidden state $\{\mathcal{S}_n\}_{n\in\mathbb{Z}}$ at each time n and the random variables $\{X_n\}_{n\in\mathbb{Z}}$ for the observation at each time n.

The general properties of HMMs allow one to calculate any statistic about the generated process from the hidden-state-to-state transition matrix T and set $\mathcal{P}$ of conditional probability density functions. For simplicity in the following, assume a finite set of hidden states and a single attracting component. Then every transition matrix T admits a unique stationary distribution $\boldsymbol{\pi}$. This is determined as T's left eigenvector associated with the eigenvalue of unity: $\langle\boldsymbol{\pi}|T = \langle\boldsymbol{\pi}|$. The eigenvector is normalized in probability: $\langle\boldsymbol{\pi}|\mathbf{1}\rangle = 1$, where $|\mathbf{1}\rangle$ is the column vector of all ones. Note also that $|\mathbf{1}\rangle$ is the right eigenvector of T associated with the eigenvalue of unity, $T|\mathbf{1}\rangle = |\mathbf{1}\rangle$. This property conserves state probability in hidden Markov chain evolution.

We can now provide the correlation functions and power spectral density in general and in closed form for this entire class of stochastic process generated by HMMs. Helpfully, for particular HMMs, the expressions become analytic in the model parameters.

Appendix B shows that the autocorrelation function is given by:

$$\gamma(\tau) = \begin{cases} \langle\boldsymbol{\pi}|\Omega T^{|\tau|}\overline{\Omega}|\mathbf{1}\rangle & \text{if } \tau \leq 1 \\ \langle|x|^2\rangle & \text{if } \tau = 0 \\ \langle\boldsymbol{\pi}|\overline{\Omega}T^{|\tau|}\Omega|\mathbf{1}\rangle & \text{if } \tau \geq 1 \end{cases}, \quad (4)$$

where Ω is the $|\mathcal{S}|$-by-$|\mathcal{S}|$ matrix defined by:

$$\Omega = \sum_{s\in\mathcal{S}} \langle x\rangle_{\mathrm{p}(X|s)} |s\rangle\langle s| . \quad (5)$$

We use the hidden-state basis in which $|s\rangle$ is the column vector of all 0s except for a 1 at the index corresponding to state s. $\langle s|$ is simply its transpose. This yields a natural decomposition of the identity operator: $I = \sum_{s\in\mathcal{S}} |s\rangle\langle s|$. In the hidden-state basis, then, the Ω matrix simply places state-conditioned average outputs along its diagonal.

The power spectrum is calculated starting from Eq. (3)

together with Eq. (4), using the spectral decomposition techniques developed for nonnormal and nondiagonalizable operators in Ref. [55]. In the derivation it is important to treat individual eigenspaces separately, as our generalized framework naturally accommodates. Appendix C gives the derivation's full details. Qualitatively, the power spectrum decomposes naturally into a discrete part $P_{\mathrm{d}}(\omega)$ (a weighted sum of Dirac δ-functions) and a continuous part $P_{\mathrm{c}}(\omega)$ (a collection of diffuse peaks):

$$P(\omega) = P_{\mathrm{c}}(\omega) + P_{\mathrm{d}}(\omega) .$$

For the power spectrum's continuous part the end result is:

$$P_{\mathrm{c}}(\omega) = \langle|x|^2\rangle + 2\,\mathrm{Re}\,\langle\boldsymbol{\pi}|\overline{\Omega}\,T\left(e^{i\omega}I - T\right)^{-1}\Omega|\mathbf{1}\rangle , \quad (6)$$

where $\mathrm{Re}(\cdot)$ denotes the real part of its argument.

Remarkably, *all* of the ω-dependence is in the apparently simple term $\left(e^{i\omega}I - T\right)^{-1}$. This is the *resolvent* of T along the unit circle in the complex plane. However, and central to our main results, this frequency dependence is *filtered* through $\langle\boldsymbol{\pi}|\overline{\Omega}$ and $\Omega|\mathbf{1}\rangle$. Notably, if the average-observation matrix was proportional to the identity, then all frequency dependence would be lost since $\mathrm{Re}\,\langle\boldsymbol{\pi}|\left(e^{i\omega}I - T\right)^{-1}|\mathbf{1}\rangle$ is independent of ω. Frequency dependence of the power spectrum thus requires that there are different averages associated with different states. Surprisingly though, none of the structure of the conditional probability density functions $\{\mathrm{p}(X|s)\}_s$ matters for the power spectrum, except for the average value observed in each state. Structure beyond averages is simply not captured.

D. Apparent Structure

To fully appreciate the structure that *is* captured by the power spectrum requires a spectral decomposition of the transition matrix. The set Λ_T of T's eigenvalues is calculated as usual. However, since transition matrices are generically nonnormal and often nondiagonalizable, the spectral projection operators associated with T deserve a brief review.

In particular, the spectral projection operator T_λ associated with eigenvalue λ can be defined as the residue of $(zI - T)^{-1}$ as $z \to \lambda$:

$$T_\lambda = \frac{1}{2\pi i} \oint_{C_\lambda} \left(zI - T\right)^{-1} dz , \quad (7)$$

where $z \in \mathbb{C}$ and C_λ is a small counterclockwise contour around the eigenvalue λ. Alternatively, the spectral projection operators can be constructed from all left

eigenvectors, generalized left eigenvectors, right eigenvectors, and generalized right eigenvectors associated with λ. The construction is given explicitly in Ref. [55]. In the simple case where the eigenvalue is nondegenerate, the eigenprojector takes on the simple form:

$$T_\lambda = \frac{|\lambda\rangle\langle\lambda|}{\langle\lambda|\lambda\rangle} \ .$$

However, the left and right eigenvectors are *not* simply complex-conjugate transposes of each other, as they would be in the normal-operator case familiar from closed quantum systems and undirected networks. For example, the spectral projection operator associated with stationarity—$T_1 = |\mathbf{1}\rangle\langle\boldsymbol{\pi}|$—can be interpreted as the classical version of a density matrix but, typically, the stationary distribution is not uniform and so $\langle\boldsymbol{\pi}|$ is not proportional to the transpose of $|\mathbf{1}\rangle$.

We will also use the broader class of spectral companion operators:

$$T_{\lambda,m} = T_\lambda(T-\lambda I)^m \ . \qquad (8)$$

They have the useful property that $T_{\lambda,m}T_{\zeta,n} = \delta_{\lambda,\zeta}T_{\lambda,m+n}$. Clearly, the spectral projection operator is contained in this set, as $T_\lambda = T_{\lambda,0}$. It should be noted that $T_{\lambda,m} = \mathbf{0}$ for $m \geq \nu_\lambda$, where ν_λ is the *index* of the eigenvalue λ—i.e., the size of the largest Jordan block associated with λ. One should keep in mind that the transition matrix can be represented as:

$$T = \sum_\lambda \big(\lambda T_{\lambda,0} + T_{\lambda,1}\big) \ .$$

While the resolvent has the general spectral decomposition:

$$(zI-T)^{-1} = \sum_{\lambda\in\Lambda_T}\sum_{m=0}^{\nu_\lambda-1}\frac{1}{(z-\lambda)^{m+1}}T_{\lambda,m} \ . \qquad (9)$$

The spectral expansion of the resolvent given by Eq. (9) allows us to better interpret the qualitative shape of the power spectrum:

$$P_{\mathrm{c}}(\omega) = \left\langle|x|^2\right\rangle + \sum_{\lambda\in\Lambda_T}\sum_{m=0}^{\nu_\lambda-1} 2\,\mathrm{Re}\frac{\langle\boldsymbol{\pi}|\,\overline{\Omega}\,T\,T_{\lambda,m}\Omega\,|\mathbf{1}\rangle}{(e^{i\omega}-\lambda)^{m+1}}. \qquad (10)$$

Note that $\langle\boldsymbol{\pi}|\,\overline{\Omega}\,T\,T_{\lambda,m}\Omega\,|\mathbf{1}\rangle$ is a complex-valued scalar and all of the frequency dependence now handily resides in the denominator. When T is diagonalizable, Eq. (10) reduces to:

$$P_{\mathrm{c}}(\omega) = \left\langle|x|^2\right\rangle + \sum_{\lambda\in\Lambda_T} 2\,\mathrm{Re}\left(\frac{\lambda\,\langle\boldsymbol{\pi}|\,\overline{\Omega}\,T_\lambda\Omega\,|\mathbf{1}\rangle}{e^{i\omega}-\lambda}\right) \ .$$

Appendix E gives the corresponding results for continuous-time processes.

The discrete (δ-function) portion of the power spectrum is:

$$P_{\mathrm{d}}(\omega) = \sum_{k=-\infty}^{\infty}\sum_{\substack{\lambda\in\Lambda_T\\|\lambda|=1}} 2\pi\,\delta(\omega-\omega_\lambda+2\pi k)\,\mathrm{Re}\,\langle\boldsymbol{\pi}|\,\overline{\Omega}\,T_\lambda\Omega\,|\mathbf{1}\rangle \ , \qquad (11)$$

where ω_λ is related to λ by $\lambda = e^{i\omega_\lambda}$. Equation (11) is valid even when T is nondiagonalizable: An extension of the Perron-Frobenius theorem guarantees that T's eigenvalues on the unit circle have index $\nu_\lambda = 1$. With $T_1 = |\mathbf{1}\rangle\langle\boldsymbol{\pi}|$, it is useful to note that $\langle\boldsymbol{\pi}|\,\overline{\Omega}\,T_1\Omega\,|\mathbf{1}\rangle = |\langle x\rangle|^2$, so that the δ-function at zero frequency appears whenever the average observation is nonzero.

When plotted as a function of the angular frequency ω around the unit circle, the power spectrum suggestively appears to emanate from the eigenvalues $\lambda \in \Lambda_T$ of the hidden linear dynamic. These are the coronal spectrograms displayed in Figs. 4(c) and (d); these will be discussed after the general phenomena is explained.

T's eigenvalues *on* the unit circle yield Dirac δ-functions in the power spectrum. T's eigenvalues *within* the unit circle yield more diffuse line profiles, increasingly diffuse as the magnitude of the eigenvalues retreats toward the origin. Moreover, the integrated magnitude of each contribution is determined from the amplitude $\langle\boldsymbol{\pi}|\,\overline{\Omega}\,T_\lambda\Omega\,|\mathbf{1}\rangle$. Finally, we note that nondiagonalizable eigenmodes yield qualitatively different line profiles.

The spectral decomposition of the power spectrum offers several insights into the minimal temporal structure required to generate the observed power spectrum. In particular, since (i) each local maxima in the power spectrum emanates from an eigenvalue of the hidden state-to-state transition matrix and (ii) since the number of unique eigenvalues is upper bounded by the number of hidden states (i.e., $|\Lambda_T| \leq |\boldsymbol{\mathcal{S}}|$), we have the following result: Counting both diffuse peaks and δ-functions, the number of observed peaks in the power spectrum (from $\omega \in (-\pi, \pi]$ in the discrete-time setting) puts a lower bound on the number of hidden states of *any* model capable of generating the observed stochastic process. Note further that all transition matrices must have an eigenvalue of unity and that this eigenvalue can only produce a δ-function at $\omega = 0$ with no other way to shape the power spectrum over other frequencies. This gives the immediate consequence that all single-state HMMs (i.e., all IID processes) have a flat power spectrum, as suggested earlier. In such cases, $\Lambda_T = \{1\}$, and there are no other eigenvalues to shape the power spectrum.

Figure 4 shows the power spectrum of a particular

(a) A b-parametrized HMM with mean values of each state's pdf $\langle x\rangle_{\mathrm{p}(X|s)}$ indicated as the number inside each state.

(b) Eigenvalue evolution for all $\lambda \in \Lambda_T$ sweeping transition parameter b from 1 (thick blue) to 0 (thin red).

(c) Power spectrum and eigenvalues at $b = 3/4$.
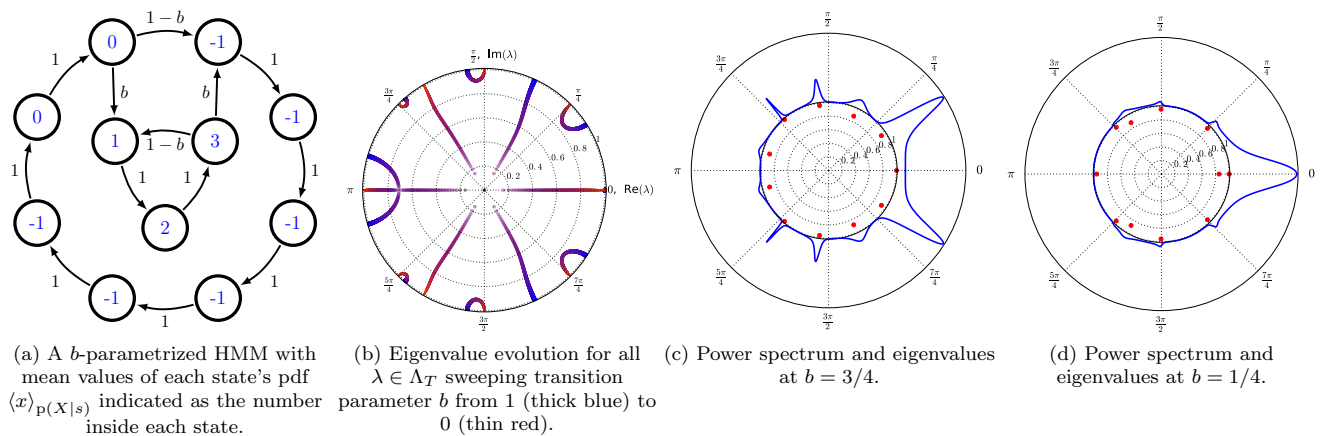
(d) Power spectrum and eigenvalues at $b = 1/4$.

FIG. 4. Parametrized HMM of a stochastic process, its eigenvalue evolution, and two coronal spectrograms showing power spectra emanating from eigen-spectra.

parametrized family of stochastic processes. Figure 4(a) displays the HMM's skeleton with state-to-state transition probabilities parametrized by b. The mean values $\langle x\rangle_{\mathrm{p}(X|s)}$ observed from each state are indicated as the blue number inside each state. The process generated depends on the actual PDFs and the transition parameter b. Although, and this is one of our main points, the power spectrum is ignorant to the PDFs' details.

The evolution of the eigenvalues Λ_T of the transition dynamic among hidden states is shown from thick blue to thin red markers in Fig. 4(b), as we sweep the transition parameter b from 1 to 0. A subset of the eigenvalues pass continuously but very quickly through the origin of the complex plane as b passes through $1/2$. The continuity of this is not immediately apparent numerically, but can be revealed with a finer increment of b near $b \approx 1/2$. Notice the persistent eigenvalue of $\lambda_T = 1$, which is guaranteed by the Perron–Frobenius theorem.

Using *coronal spectrograms*, introduced in Refs. [56] and [54], Figs. 4(c) and 4(d) illustrate how the observed power spectrum $P(\omega)$ emanates from the *eigen*-spectrum Λ_T of the hidden linear state-dynamic. Specifically, in Fig. 4(c) and again, at another parameter setting, in Fig. 4(d), we show the continuous part of the power spectrum $P_{\mathrm{c}}(\omega)$ (plotted around the unit circle in solid blue) and the eigen-spectrum Λ_T (plotted as red dots on and within the unit circle) of the state-to-state transition matrix for the 11-state hidden Markov chain (Fig. 4(a)) that generates it. As anticipated from Eq. (10), the power spectrum has sharper peaks when the eigenvalues are closer to the unit circle. The integrated magnitude of each peak depends on $\langle \boldsymbol{\pi}| \overline{\Omega} |\lambda\rangle \langle\lambda| \Omega |\mathbf{1}\rangle$.

Interestingly, our continuous spectrum (closely related to the continuous spectrum of unitary models of chaotic dynamics) is the shadow of the discrete spectrum of

nonunitary dynamics. This suggests that resonances in various physics domains concerned with a continuous spectrum can be modeled as consequences of simpler nonunitary dynamics. Indeed, hints of this appear in the literature already [57–59].

III.   HIDDEN STRUCTURE

Remarkably, the power spectrum generated by any hidden-Markov process with continuous random variables for the observables is the same as the power spectrum generated by a potentially much simpler one—a process that is a function of the same underlying Markov chain but instead emits the state-dependent *expectation value* of the observable within each state.

Theorem 1. *Let $\mathcal{P} = \big\{p(X|s)\big\}_{s\in\boldsymbol{S}}$ specify any state-paired collection of probability density functions over the domain $\mathcal{A} \subseteq \mathbb{C}$. Let $\mathcal{B} = \big\{\langle x\rangle_{p(X|s)}\big\}_{s\in\boldsymbol{S}}$ and let $\mathcal{Q} = \big\{\delta(x - \langle x\rangle_{p(X|s)})\big\}_{s\in\boldsymbol{S}}$. Then, the power spectrum generated by any hidden Markov model $\mathcal{M} = \big(\boldsymbol{S}, \mathcal{A}, \mathcal{P}, T\big)$ differs at most by a constant offset from the power spectrum generated by the hidden Markov model $\mathcal{M}' = \big(\boldsymbol{S}, \mathcal{B}, \mathcal{Q}, T\big)$ that has the same hidden Markov chain but in any state $s \in \boldsymbol{S}$ emits, with probability one, the average value $\langle x\rangle_{p(X|s)}$ of the state-conditioned probability density function $p(X|s) \in \mathcal{P}$ of $\mathcal{M}$.*

Proof. *From Eqs. (6) and (11), we see that $P_c(\omega) + P_d(\omega) - \big\langle |x|^2 \big\rangle$ depends only on T and $\{\langle x\rangle_{p(X|s)}\}_{s\in\boldsymbol{S}}$. Thus, all HMMs sharing the same T and $\{\langle x\rangle_{p(X|s)}\}_{s\in\boldsymbol{S}}$ have the same power spectrum $P(\omega) = P_c(\omega) + P_d(\omega)$, modulo a constant offset determined by differences in $\big\langle |x|^2 \big\rangle$.*

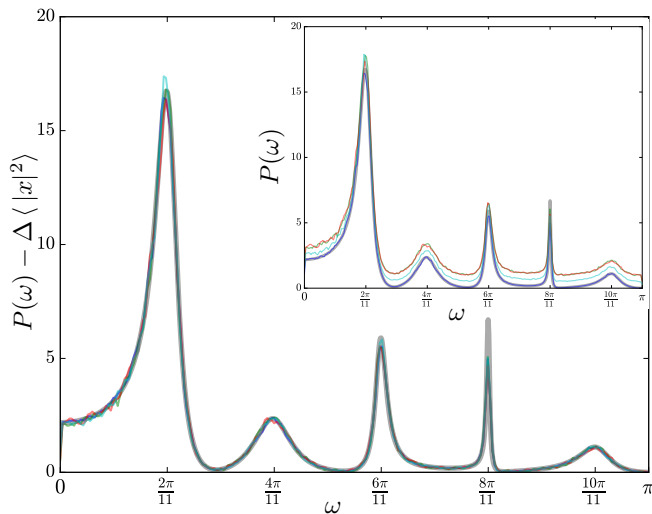Figure 5 demonstrates Thm. 1 for the power spectrum

FIG. 5. Demonstrating Thm. 1 for the processes generated by the HMM skeleton of Fig. 4(a), using transition parameter $b = 3/4$ as in Fig. 4(c). Besides an overall constant offset of $\langle|x|^2\rangle$, the power spectrum is insensitive to all details of the state-conditioned PDFs except for their averages. On top of the theoretical curve (thick gray) given by Eq. (6) we overlay horizontal offsets of the power spectra calculated numerically for stochastically-generated time series. The state-conditioned PDFs used to define the different stochastic processes are: (i) single δ-functions, (ii) single Gaussians, (iii) two symmetrically spaced δ-functions (with no support at the mean), and (iv) weighted δ-functions with asymmetric spacing. For each, a time series of length 2^{18} was generated. The Welch method was used to calculate the average power spectrum for each process using FFTs of segments of length 2^9. The inset shows the raw power spectrum for each process without the offset.

shown in Fig. 4(c).

One immediate consequence of Thm. 1 is the following:

Corollary 1. *Any hidden Markov chain with any arbitrary state-paired collection of zero-mean distributions, i.e.:*

$$\mathcal{P} \in \left\{ \{p(X|s)\}_{s \in \boldsymbol{\mathcal{S}}} : \langle x \rangle_{p(X|s)} = 0 \text{ for all } s \in \boldsymbol{\mathcal{S}} \right\} ,$$

generates a flat power spectrum indistinguishable from white noise.

Proof. *This follows immediately from Thm. 1 and the fact that the all-zero sequence has a power spectrum that is zero everywhere. Thus, the corresponding power spectrum of the actual process is a flat (nonzero) power spectrum of uniform height $\langle|x|^2\rangle$.*

We can relax the corollary to include cases where the state-conditioned PDFs are all equal to a potentially-nonzero constant. Although, a δ-function at zero frequency (of integrated magnitude equal to the square magnitude of the constant) will then be observed in ad-
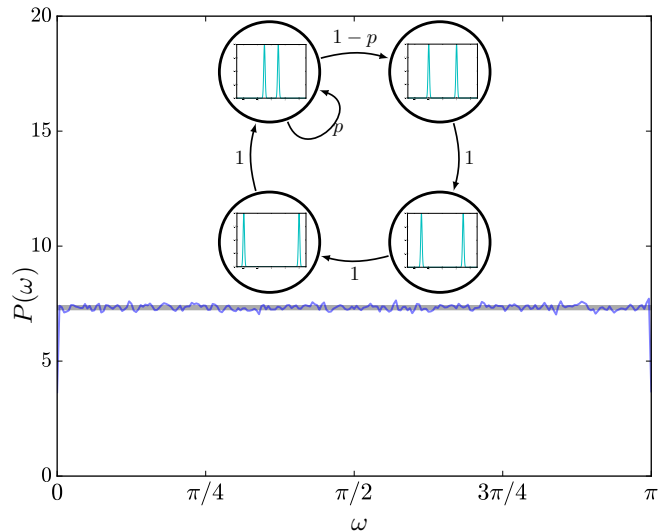

FIG. 6. Demonstrating Cor. 1: The overtly-structured stochastic process generated by the HMM (inset) has a flat power spectrum for all values of the transition parameter $p \in [0, 1]$. The flat power spectrum is shown analytically (thick gray) and numerically (thin blue) for $p = 1/10$. The numerical power spectrum was calculated from a simulated time series of length 2^{20} using the Welch method, performing FFTs on segments of length 2^9.

dition to the flat power spectrum.

The corollary's implications are striking. It is quite surprising, to consider one broad class of examples, that a power spectrum can be completely flat even when a ring of sequential states are visited that emit observables with probability density functions having no overlapping support. Figure 6 gives an example. In such a case, any cogent observer immediately detects the obvious structure in the mismatched supports—observed values are distinct—and forbidden realizations. Yet the power spectrum remains silent about this structure, reporting only the featureless signature of white noise.

In other more challenging settings, structure is not always so obvious without some reliable aid. Indeed, structure becomes increasingly difficult to detect (by any means) when the state-conditioned probability density functions have overlapping support. This is the generic case of non-Markovian processes. The hidden states cannot be detected via casual inspection.

While they give a concrete sense of missing structure, these cases fall far short of telling the full story of how power spectra mask structure. The following sections, culminating in Thm. 2, address additional ways white noise appears without needing to meet the requirements of Cor. 1. (Reference [2] goes further still, showing how the structure can indeed be hidden much deeper.)

A. Nonlinear Pairwise Correlation

In a way, the structure of the stochastic process in Fig. 6 was hidden as shallowly as possible to evade the power spectrum. As mentioned, the structure should be trivial to detect by other means. Indeed, while the *linear* pairwise correlation $\gamma(\tau)$ vanished for all $\tau > 0$, there is still pairwise dependence between the generated random variables, which is nonlinear. This pairwise dependence can be teased out using the pairwise mutual information $\mathrm{I}(X_0; X_\tau)$ between observables at different times. For the process of Fig. 6, if we take the limit of the narrow Gaussians in the state-conditioned PDFs to be pairs of δ-functions, then the pairwise mutual information can be calculated exactly as shown in App. F. In fact, $\mathrm{I}(X_0; X_\tau)$ will be unchanged for any set of four PDFs we could have chosen for the states of the example HMM, as long as the PDFs all have mutually exclusive support for the observable output. (This then makes the hidden state a function of the instantaneous observable.)

A concise summary of the pairwise mutual information is provided via Ref. [54]'s *power-of-pairwise-information* (POPI) spectrum:

$$\mathcal{I}(\omega) = -\mathrm{H}(X_0) + \lim_{N \to \infty} \sum_{\tau = -N}^{N} e^{-i\omega\tau}\, \mathrm{I}(X_0; X_\tau) \,,$$

where $\mathrm{H}(\cdot)$ is the Shannon entropy of its argument. We generated plots of both the pairwise mutual informations and the POPI spectrum for this example (shown in App. F) and find the decay of pairwise information to scale intuitively with the phase-slip-parameter p. While the example of Fig. 6 has no *linear* correlation, nevertheless it *does* have pairwise structure. Thus, the structure of the example process was hidden from power spectra, but not hidden from the POPI spectrum.

The following sections continue investigating temporally-structured processes, but focus on those with no linear pairwise correlation (and so a flat power spectrum) and no pairwise mutual information (and so a flat POPI spectrum). These will lead us to introduce a general condition for flat power spectra. And, since power spectra fail so often to detect structure, we turn from criticizing them to being constructive: introducing ways to detect hidden structure.

B. Sophisticated Fraudulent White Noise

Theorem 1 established that the power spectrum from processes with *continuous* observable random variables is the same as the power spectrum from much simpler corresponding processes with *discrete* observable random variables. Accordingly, Thm. 1 motivates studying the power spectra of processes with discrete observable random variables to determine if there are further ways to achieve a flat power spectrum, beyond Cor. 1's possibilities. For observables that are discrete random variables, it is sufficient to consider their probability *distributions* rather than their probability *density* functions.

We begin this next step of the development by establishing the following simple lemma:

Lemma 1. *Any stochastic process (not necessarily stationary) with the* Single-Condition-Independent Property *(SCIP):*

$$\Pr(X_t | X_{t'} = x) = \Pr(X_t)$$
$$= \Pr(X_{t'}) \,,$$

for all $x \in \mathcal{A}$ and all $t \neq t'$, generates a flat power spectrum, mimicking white noise.

Proof. *For any such process, $\Pr(X_t)$ is the stationary distribution $\boldsymbol{\mu}_X$ of the instantaneous observable under the stochastic dynamic. Moreover, SCIP means that the joint probability of any two observations decomposes:*

$$\Pr(X_t = x, X_{t+\tau} = x') = \Pr(X_{t+\tau} = x' | X_t = x) \Pr(X_t = x)$$
$$= \Pr(X_{t+\tau} = x') \Pr(X_t = x)$$
$$= \boldsymbol{\mu}_X(x') \boldsymbol{\mu}_X(x) \,.$$

Substituting $\boldsymbol{\mu}_X(x')\boldsymbol{\mu}_X(x)$ for $\Pr(X_t = x, X_{t+\tau} = x')$ in the autocorrelation definition of Eq. (2) immediately implies that SCIP processes have τ-independent pairwise correlation $\gamma(\tau) = |\langle x \rangle|^2$ for $\tau \neq 0$. The power spectrum is thus flat over all frequencies, except possibly with a δ-function at $\omega = 0$.

SCIP processes not only have a flat power spectrum but also a flat POPI spectrum. SCIP implies $\mathrm{I}(X_0; X_\tau) = 0$ for all $\tau \neq 0$ which, in turn, implies $\mathcal{I}(\omega) = 0$. These processes completely lack any pairwise correlation, whether linear or nonlinear.

Notably, Lem. 1 is not covered by Cor. 1; nor is Cor. 1 subsumed by Lem. 1. Accordingly, the following develops a single simple condition (culminating in Thm. 2) that covers all of these cases of fraudulent white noise.

Crucially, the class of potentially-fraudulent-white-noise processes suggested by Lem. 1 is nontrivial. In addition to IID processes, this class of processes includes non-Markovian processes that hide all of their structure beyond pairwise correlations.

The Random–Random–XOR process (RRXOR), discussed at length in Ref. [54], is an example. Over blocks of length 3, the first two bits are generated randomly from a uniform distribution and the third bit is then the

logical XOR operation of the last two. Explicitly:

$$X_{3n+\phi} = \text{XOR}(X_{3n-2+\phi}, X_{3n-1+\phi}) \text{ , whereas}$$
$$X_{3n-2+\phi} \sim \left(\tfrac{1}{2}, \tfrac{1}{2}\right) \text{ and}$$
$$X_{3n-1+\phi} \sim \left(\tfrac{1}{2}, \tfrac{1}{2}\right) \text{ ,}$$

for all $n \in \{1, 2, \dots\}$. As a SCIP process, the RRXOR process has a flat power spectrum although it does not fall under the purview of Cor. 1. Indeed, the RRXOR process has no pairwise correlation at all since $\text{I}(X_0; X_\tau) = 0$ for all $\tau > 0$. Accordingly, the POPI spectrum is zero over all frequencies. The structure in this process is strictly *three-way* correlation. In Ref. [54], the phase ϕ itself is a random variable, and synchronizing to the phase is a surprisingly difficult task [60]. No matter, whether or not the phase ϕ is given, the process has no pairwise correlation—resulting in a flat power spectrum and flat POPI spectrum—and only reveals correlation in its three-way structure.

It is interesting to note that the related RRX*NOR* process, where $X_{3n} = \text{XNOR}(X_{3n-2}, X_{3n-1})$, also has a flat power spectrum. In fact, this suggests a new method to hide structure: embed a correlated message into a sequence of RRXOR and RRXNOR 3-bit sequences that lifts all correlation beyond pairwise. Specifically, the original message is transformed into a sequence of choices about whether to use XOR or XNOR on the previous two random bits. As long as the read frame and the embedding mechanism is known, the message can be easily extracted. But, if it is not known that a message is embedded, it cannot be detected simply by looking for pairwise correlations.

Through similar construction, structure can be shifted up to arbitrarily-high orders of correlation [2]. Stochastic processes can be constructed with N-way correlation but no n-way correlation for all $n < N$. Moreover, an arbitrarily correlated message can be embedded within such a process, such that its structure is lifted beyond any desired order of correlation.

C. Content-preserving Whitening

Corollary 1 gave a method to construct an arbitrarily complex process with a truly flat power spectrum, so long as all latent states have the same average output. Here, we introduce an alternate method to construct arbitrarily complex processes with truly flat power spectra. These processes, in addition, are devoid of n-way correlation for all $n < N$.

1. Choose an embedding block length $N \geq 3$.
2. Choose any stochastic process ("Process A") with a binary output alphabet.

3. Construct "Process B" as follows:
   - Whenever Process A would produce a 0, Process B will sample a word uniformly from the set of all words of length N with an even number of 1s.
   - Whenever Process A would produce a 1, Process B will sample a word uniformly from the set of all words of length N with an odd number of 1s.

Any Process B constructed in this manner has a truly flat power spectrum. Process B will also be devoid of n-way correlation for all $n < N$. Moreover, if A is a stationary process such that its statistical complexity $C_\mu(\text{A})$ is well defined [61, 62], then Process B is also a stationary process with $C_\mu(\text{B}) \geq C_\mu(\text{A})$.

This also works for "infinity-structured" processes, those with divergent statistical complexity. Choose *any* binary Process-A family with $C_\mu \to \infty$. This can be, for example, Ref. [63]'s *Heavy-Tailed Periodic Mixture Process* that has infinite past–future mutual information: $\mathbf{E} \to \infty$. Then add some structure, via content-preserving whitening, to obtain a binary Process-B family with $C_\mu \to \infty$ and a truly flat power spectrum.

Similar constructions can also be developed for processes with larger alphabets. Further examples will be given in Ref. [2].

Through the lens of pairwise correlation, such structure is simply missed. However, before moving on to consider more advanced methods to detect such structure, we finish our investigation of flat power spectra from structured processes. The next section addresses a broad class of possibly-input-dependent process generators and we give a very general condition for when a flat power spectrum results.

D. Input-dependent Generators and Fraudulent White Noise

Probing fraudulent white noise more broadly, consider an arbitrarily correlated message $\vec{m}$ and an input-dependent generator $\mathcal{M}(\vec{m})$ of an observable output process $\{X_t\}_{t \in \mathcal{T}}$. The lengths of the inputs and outputs need not be commensurate, and the input and output alphabets may also be distinct. The generator is fully specified by the tuple $\mathcal{M}(\vec{m}) = \left(\boldsymbol{\mathcal{S}}, \mathcal{A}, \mathcal{P}, \{T_t(\vec{m})\}_t, \boldsymbol{\mu}_1\right)$. That is, the internal states $\boldsymbol{\mathcal{S}}$, output alphabet $\mathcal{A}$, and state-dependent PDFs $\mathcal{P}$ are static. However, the hidden-state-to-state transition matrix $T_t(\vec{m})$ at time t is potentially a function of the full input $\vec{m}$. Since stationarity is no longer assumed, the initial distribution $\boldsymbol{\mu}_1$ over hidden states must be specified for the statistics of the output process to be well defined.
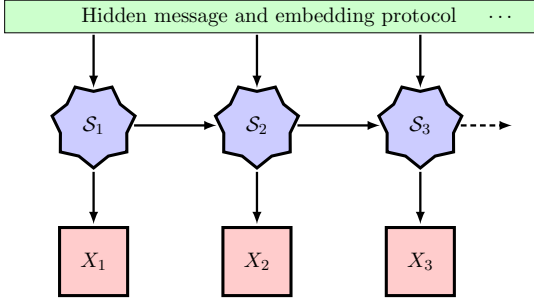
FIG. 7. Bayesian network for memoryful input-dependent generators.

Figure 7 shows the relevant Bayes network for this general type of input-dependent generator. Contrast this with Fig. 3, which showed the Bayes network of autonomous HMM generators. Autonomous HMMs can be seen as a special case of these possibly-input-dependent generators when the process $\mathcal{M}(\vec{m}) = \mathcal{M}$ is input-independent and the initial distribution $\boldsymbol{\mu}_1 = \boldsymbol{\pi}$ is taken to be the stationary distribution $\langle\boldsymbol{\pi}| = \langle\boldsymbol{\pi}| T$ of the time-independent transition matrix $T_t(\vec{m}) = T$.

The memoryful input-dependent generators we now consider also generalize the memoryful transducers considered in Ref. [64] to use continuous-variable outputs and allow the lengths of input and output to be incommensurate. Via any of the above models, very general message-embedding schemes can be developed that produce sophisticated fraudulent white noise.

Even with all the generalizations, we can determine autocorrelation and power spectra. Similar to the derivation for HMMs, we find that *if* the process is wide-sense stationary then (for $\tau \geq 1$):

$$\gamma(\tau) = \langle\boldsymbol{\mu}_t|\overline{\Omega}\, T_{t:t+\tau}(\vec{m})\,\Omega|\mathbf{1}\rangle \ , \qquad (12)$$

which must overall be t-independent (so long as $t \geq 1$). Here, $\langle\boldsymbol{\mu}_t| = \langle\boldsymbol{\mu}_1|\, T_{1:t}(\vec{m})$ and $T_{a:b}(\vec{m}) = \prod_{t=a}^{b-1} T_t(\vec{m})$, and Ω is again given by Eq. (5). (Notice that $T_{a:a+\tau}(\vec{m}) = T^{\tau}$ for the special case of autonomous HMMs.)

Thus, autocorrelation can be calculated as $\langle\boldsymbol{\mu}_1|\overline{\Omega}\, T_{1:1+\tau}(\vec{m})\,\Omega|\mathbf{1}\rangle$, assuming that the pairwise statistics are stationary. This can also be written as:

$$\gamma(\tau) = \left\langle\, \langle\overline{x}\rangle_{\mathrm{p}(X|\mathcal{S}_t)}\, \langle x\rangle_{\mathrm{p}(X|\mathcal{S}_{t+\tau})} \,\right\rangle_{\Pr(\mathcal{S}_t, \mathcal{S}_{t+\tau})} , \qquad (13)$$

where we treat $\langle x\rangle_{\mathrm{p}(X|\mathcal{S}_t)}$ as a random variable that depends on $\mathcal{S}_t$ and the whole expression becomes t-independent assuming stationary pairwise statistics. Accordingly, the autocorrelation function is constant and

the power spectrum is flat whenever:

$$\Pr\big(\langle x\rangle_{\mathrm{p}(X|\mathcal{S}_{t+\tau})}\,|\mathcal{S}_t = s\big) = \Pr\big(\langle x\rangle_{\mathrm{p}(X|\mathcal{S}_{t+\tau})}\big)$$
$$= \Pr\big(\langle x\rangle_{\mathrm{p}(X|\mathcal{S}_t)}\big) \ ,$$

for all τ, for all $t \in \mathcal{T}$, and for all $s \in \boldsymbol{\mathcal{S}}$.

However, this requirement is too strict to cover all cases of interest. For example, it does *not* yet imply the flat power spectrum of the RRXOR process. More generally, constant autocorrelation and flat power spectra can be guaranteed by an even weaker condition.

To appreciate this, define the set Ξ of average outputs exhibited by the states: $\Xi \equiv \bigcup_{s \in \boldsymbol{\mathcal{S}}}\big\{\langle x\rangle_{\mathrm{p}(X|s)}\big\}$. Furthermore, we define $\boldsymbol{\mathcal{S}}_\xi \subset \boldsymbol{\mathcal{S}}$ as the set of states that all exhibit the same average output $\xi \in \Xi$. Explicitly, $\boldsymbol{\mathcal{S}}_\xi \equiv \{s \in \boldsymbol{\mathcal{S}} : \langle x\rangle_{\mathrm{p}(X|s)} = \xi\}$. Using these quantities, we can state our result more precisely as the following theorem.

Theorem 2. *Let $\{X_t\}_t$ be a stochastic process generated by any of the hidden-state models $\mathcal{M}(\vec{m})$ discussed above, including autonomous HMMs and input-dependent generators, X_t the random variable for the observable at time t, and $\mathcal{S}_t$ the random variable for the hidden state at time t. Such processes have constant autocorrelation and a flat power spectrum if:*

$$\Pr(\mathcal{S}_{t+\tau} \in \boldsymbol{\mathcal{S}}_{\xi'}|\mathcal{S}_t \in \boldsymbol{\mathcal{S}}_\xi) = \Pr(\mathcal{S}_{t+\tau} \in \boldsymbol{\mathcal{S}}_{\xi'})$$
$$= \Pr(\mathcal{S}_t \in \boldsymbol{\mathcal{S}}_{\xi'}) \ , \qquad (14)$$

for all separations $\tau > 0$, for all $t \in \mathcal{T}$, and for all $\xi, \xi' \in \Xi$.

Proof. *Starting from Eq. (13), we find the autocorrelation for all such processes:*

$$\gamma(\tau) = \left\langle\, \langle\overline{x}\rangle_{p(X|\mathcal{S}_t)}\, \langle x\rangle_{p(X|\mathcal{S}_{t+\tau})} \,\right\rangle_{\Pr(\mathcal{S}_t, \mathcal{S}_{t+\tau})}$$
$$= \sum_{s,s' \in \boldsymbol{\mathcal{S}}} \Pr(\mathcal{S}_t = s, \mathcal{S}_{t+\tau} = s')\, \langle\overline{x}\rangle_{p(X|s)}\, \langle x\rangle_{p(X|s')}$$
$$= \sum_{\xi,\xi' \in \Xi} \Pr(\mathcal{S}_t \in \boldsymbol{\mathcal{S}}_\xi, \mathcal{S}_{t+\tau} \in \boldsymbol{\mathcal{S}}_{\xi'})\, \overline{\xi}\,\xi'$$
$$= \sum_{\xi \in \Xi} \Pr(\mathcal{S}_t \in \boldsymbol{\mathcal{S}}_\xi)\overline{\xi}$$
$$\times \sum_{\xi' \in \Xi} \Pr(\mathcal{S}_{t+\tau} \in \boldsymbol{\mathcal{S}}_{\xi'}|\mathcal{S}_t \in \boldsymbol{\mathcal{S}}_\xi)\xi' \ . \qquad (15)$$

Combining Eq. (14) and Eq. (15), we see that:

$$\gamma(\tau) = \sum_{\xi \in \Xi} \Pr(\mathcal{S}_t \in \boldsymbol{\mathcal{S}}_\xi)\overline{\xi} \sum_{\xi' \in \Xi} \Pr(\mathcal{S}_{t+\tau} \in \boldsymbol{\mathcal{S}}_{\xi'})\xi'$$
$$= |\langle\xi\rangle|^2 \ ,$$

which is a constant. Thus, the power spectrum is flat, if

Eq. (14) holds.

Theorem 2 says that a flat power spectrum results whenever the average output of the future latent state is independent of the average output of the current latent state.

This generalized condition for flat power spectra covers the special case for HMMs as well as fraudulent white noise from message-embedding schemes with stationary pairwise statistics, but nonstationary high-order statistics. Appendix G shows that a modified version of Thm. 2 also applies to another class of generators that can be more natural for measured quantum systems and systems with computational dependencies. Theorem 2 subsumes Cor. 1 as well as Lem. 1. And, it offers the most general guarantee yet for constant autocorrelation and flat power spectrum.

By way of contrast consider the following. While zero pairwise mutual information is always a sufficient condition for flat power spectrum, it is not a necessary condition. Here, in Thm. 2, we find a very general condition for a flat power spectrum. Appendix H established a related theorem (Thm. 4) that further generalizes the condition for flat power spectra, allowing for time-dependent PDFs associated with each state. Moreover, Thm. 2 and Thm. 4 constructively suggest how to design such processes. Notably, these generalized conditions do not require a stationary dynamic over the hidden states of the observation-generating mechanism, which furthermore allows messages to ride undetected aboard fraudulent white noise.

This final result emphasizes the main argument's generality: power spectra are mute when detecting a broad range of observable structure. Whether observing physical, biological, or social systems, we seek structure that reveals mechanism and begets predictability. Through the lens of power spectra, or pairwise correlation more generally, much structure is simply missed. The challenge then is to look for structure beyond pairwise.

IV. STRUCTURE IN NOISE?

One systematic method for exploring beyond-pairwise correlations in stationary stochastic processes is through the sequence of *myopic entropy rates* [50, 53, 54, 65–67]:

$$h_L = \mathrm{H}(X_L | X_1 X_2 \ldots X_{L-1}) , \qquad (16)$$

with $h_1 = \mathrm{H}(X_1)$. For example, the RRXOR process has $h_1 = h_2 = \log |\mathcal{A}| = 1$ bit/symbol—it appears as random as possible when considering symbols individually or in pairs. Structure is unveiled, though, for $L \geq 3$ when $h_L < 1$. That is, progressively longer Markov-order-L

approximations of the infinite-Markov-order process reveal progressively more of its hidden structure.

In fact, h_L's convergence reflects how structure is hidden in the stochastic process [67]. As $L \to \infty$, h_L approaches the process' *Shannon entropy rate* h—the irreducible randomness per symbol after all orders of correlation have been taken into account. Notably, the accumulation of the excess myopic entropy $\sum_{L=1}^{\infty}(h_L - h) = \mathbf{E}$—the *excess entropy*—quantifies the total mutual information between the past and future of a process: $\mathbf{E} = \mathrm{I}(\ldots, X_{-1}, X_0 ; X_1, X_2, \ldots)$. So, while $\mathrm{I}(X_0; X_\tau) = 0$ for all $\tau > 0$ for the RRXOR process, the past and future are nevertheless correlated since $\mathbf{E} > 0$. And the convergence to predictability can be viewed in the frequency domain through the excess-entropy spectrum introduced in Ref. [54]. Taken together, this suggests that myopic entropy rates serve well to identify hidden structure beyond pairwise correlation. They show how predictability improves as progressively longer historical context is used.

However, correlations are not always restricted to contiguous blocks. Therefore, there can be pairwise correlations among distant observables while $h_2 = 0$. Moreover, the myopic entropy rates as defined above are restricted to stationary processes. Consequently, despite their utility, myopic entropies are not ideal for direct indication of L-way correlation in the most general setting.

A more direct indicator of L-way correlation is found in the *dependence function* D_L, which quantifies the maximal uniquely-L-way correlation that exists in a process. We say a set $\boldsymbol{\chi}$ of random variables is *fully correlated* if all constituent random variables inform all of the others; that is, if:

$$\mathrm{H}(X|\boldsymbol{\chi} \setminus \{X, X'\}) - \mathrm{H}(X|\boldsymbol{\chi} \setminus \{X\})$$
$$= \mathrm{I}\big(X ; X' | \boldsymbol{\chi} \setminus \{X, X'\}\big)$$
$$> 0 ,$$

for all $X, X' \in \boldsymbol{\chi}$. A process is then L-way correlated if it has a set of L random variables that are fully correlated. One way to quantify this L-way correlation is through the following dependence function:

$$D_L \equiv \sup_{\{\boldsymbol{\chi} \subset \{X_t\}_t : |\boldsymbol{\chi}|=L\}} \min_{X, X' \in \boldsymbol{\chi}} \mathrm{I}\big(X ; X' | \boldsymbol{\chi} \setminus \{X, X'\}\big) ;$$
$$(17)$$

defined here only for $L \geq 2$. L-way dependence is nonzero if and only if there are novel L-way contributions to a process' total correlation. Note that dependence can be applied to nonstationary processes and processes of finite duration.

Consider, as a simple example of noncontiguous depen-

dencies, a process consisting of two interlaced RRXOR processes with unambiguous phase. Explicitly:

$$X_{6n} = \text{XOR}(X_{6n-4}, X_{6n-2}) \text{ and}$$
$$X_{6n-1} = \text{XOR}(X_{6n-5}, X_{6n-3}) \,,$$

whereas X_{6n-5}, X_{6n-4}, X_{6n-3}, and X_{6n-2} are all generated from a uniform distribution for all $n \in \{1, 2, \dots\}$. Joint probabilities over contiguous variables are completely uncorrelated and as random as possible, up until a block-length of 5. Let's treat the example as a stationary process: Calculating probabilities from word frequencies in a single realization, with the implicit assumption of stationarity, effectively inducing random phase. Then, we find full randomness in the myopic entropy rates up to block length 5: $h_\ell = \log|\mathcal{A}| = 1$ bit for $1 \leq \ell < 5$. Then, finally, a reduction in apparent entropy occurs at h_5, after which $h_\ell < h_{\ell-1}$ for $\ell \geq 5$. Notably, h_3 reflects maximal randomness within its purview of inspection. Whereas, the process actually has three-way, but no lower-order dependencies. This yields $D_1 = D_2 = 0$ and $D_3 > 0$. With known phase, we would have $D_3 = 1$ bit.

However, when the process is unknown and only a single realization is available for analysis, probabilities can be inferred only from motifs of random-variable clusters. For example, estimating $\Pr(X_{t-2}, X_t, X_{t+2})$ *as if* the process were stationary, leads to finding $0 < \widetilde{D}_3 < 1$, where $\widetilde{D}_L$ denotes approximating the dependence function assuming stationarity and testing a limited set of motifs. Usefully, $\widetilde{D}_L$ sets a lower bound on D_L. So, nonzero $\widetilde{D}_L$ implies L-way dependence. Curiously, the assumption of stationarity induces $\widetilde{D}_L > 0$ for all $L \geq 3$; reminiscent of how $h_L - h_{L-1} > 0$ for all $L \geq 3$ for the RRXOR process with ambiguous phase. In each case, these higher-order correlations correspond to the observer's resolving phase ambiguity.

The dependence function seems to fulfill its desired role of identifying high-order correlations that cannot be explained by lower-order phenomena. Taking a step back, though, we might question the whole endeavor. Can a single model-free signal-analysis method ever reliably detect information processing and thus complex structure in the world around us? We have clearly ousted power spectra for this task. Nevertheless, our arguments here lend support to an affirmative answer, but at the cost of more nuanced and computationally intensive techniques. What is the range of validity of the informational measures discussed above? Can they be entrusted with finding structure in the noise?

First, it should be noted that Shannon entropy is only fully justifiable for alphabets $\mathcal{A}$ of countable cardinality. So, apparently continuous observables must be par-

titioned into measurable sets to apply the informational measures like the myopic entropy rates and the dependencies D_L. Nevertheless, quantum physics suggests that even very large and apparently continuous systems are, in principle, always represented in a countable basis. Practically too, measurement devices only have a finite precision, so observations are discretized in practice anyway. Therefore, Shannon entropies (like the myopic entropy rates and the dependencies) can be applied in principle.

However, a second and likely more-severe challenge arises from limited data: reliable estimates of probabilities are not always available. Model building offers the strongest response to this challenge. Generative models inferred from low-order statistics sometimes encapsulate predictions of rare events. And, at least, they give a prediction for high-order statistics. Testing these predictions against observation allows refining one's model and discovering new structure.

V. CONCLUSION

Our investigation began with the modest task of showing how to calculate the correlation function and power spectrum given a signal's generator. To this end, we briefly introduced hidden Markov models as signal generators and then used the linear-operator techniques of Ref. [55] to calculate their autocorrelation and power spectra in closed-form. This led to several lessons. First, we saw that the power spectrum is a direct fingerprint of the resolvent of the model's time-evolution operator, analyzed along the unit circle. Second, spectrally decomposing the not-necessarily-diagonalizable time evolution operator, we discovered the range of qualitative behaviors that can be exhibited by autocorrelation functions and power spectra. Third, contributions from eigenvalues on the unit circle had to be extracted and dealt with separately. Contributions from eigenvalues on the unit circle correspond to Dirac δ-functions—the analog of Bragg peaks in diffraction. Whereas, eigen-contributions from inside the unit circle correspond to diffuse peaks, which become sharper for eigenvalues closer to the unit circle. Finally, we found that nondiagonalizable eigenmodes yield qualitatively different line profiles than their diagonalizable counterparts.

These first results incisively answer the challenges raised by Ruelle–Pollicott resonance theory about the possible relationship between complex eigenvalues of time-evolution operators and the correlation and power spectra of observables [11–13]. In short, we provided the exact relationship between the time-evolution operator and the correlation functions and power spectra, as well as the possible behavior modes of each. The result is

a deeper theoretical understanding and constructive calculational methods that complement early investigations that experimentally delivered meromorphic power spectra from chaotic dynamical systems [9, 10].

Accordingly, our findings are relevant to modern applications of Ruelle–Pollicott resonance theory. These applications are leading, for example, to better understanding of sensitivities in climate models [16] and the dynamics of open quantum systems via their correspondence to classical chaotic dynamical systems [14, 15]. Our results provide full analytical correspondence between observed correlation and the spectral properties of nonunitary models. Our approach also bears on Koopman operator theory and its applications, which has received a new wave of attention due to the success of recent data-driven algorithms [68]. However, our results also clarify that resonances discovered via pairwise correlation are generically an insufficient representation of the spectral features of such nonnormal dynamics. This emphasizes that the full spectral representation of the effective nonnormal dynamics [55], generically inaccessible via pairwise correlation, is worth pursuing. Success in this will immediately yield predictions about many complex systems of interest.

The most surprising and more immediate finding, though, is that temporal structure can fully evade detection by power spectra. Arbitrarily sophisticated processes can have exactly flat power spectra and so masquerade as white noise. Accordingly, we called such processes *fraudulent white noise* processes. Theorem 1 and Cor. 1 characterized the many ways that structure can be hidden from power spectra. And, ultimately, Thm. 2 addressed the more general condition for fraudulent white noise, in which the generated time-series could be input-dependent and nonstationary.

We started out noting that, on the one hand, divergent correlation length often heralds the emergence of new types of order. And, on the other, that pairwise correlation is generically identified as the structure in random systems. However, we showed that there is often rich structure even in the absence of pairwise correlations. What types of order are we *failing* to predict due to an historical emphasis on pairwise correlations? Complex systems surely exhibit emergent structure beyond the reach of pairwise statistics. There is almost surely more functionally-relevant brain activity available in EEGs beyond what is reported in their power spectra. Perhaps, however, we should consider beyond-pairwise structure for even simple generators of structure. For example, cosmological models could be more thoroughly tested against structure in the CMB beyond what is contained in the two-point angular correlation functions.

Having diagnosed the structures inaccessible via power

spectra, we briefly discussed how to detect beyond-pairwise structure in general, introducing the dependence function to detect any L-way correlations for any L. We also stressed the importance of model building whenever possible. In particular, it can help anticipate and perhaps avoid never-yet encountered catastrophes, which are often a byproduct of the high interconnectivity of complex socio-economic systems [69]. Model building also allows us to discover new mechanisms in nature.

This all said, nature can still keep us in the dark. We showed that the correlations in a message can be shifted to arbitrarily-high orders of correlation. The result is that, for finite length messages, statistical inference can be made effectively impossible regardless of one's sophistication [2]. Nature herself employs this technique whenever we observe an increase in entropy—giving the impression of randomness generated, when it is only ever structure hidden in inaccessibly-obscure high-order correlations. In a sense, having pulled the wool over our eyes, Mother Nature lulls us into complacency with a soothing hiss of fraudulent white noise. Waking up to the true hues of reality—prying open the black box, dispelling apparent white noise—continues to require new theory and new experimentation.

... it is clearly wise to learn what a procedure really seems to be telling us about.

John Tukey, *The Future of Data Analysis*, 1962 [1, p. 60]

Appendix A: Diffraction patterns as power spectra

Diffraction patterns are used extensively to infer material structure from the scattering of, for example, an incident X-ray beam [70–74]. Generally, consider $\vec{r} \in \mathbb{R}^n$ to be a vector in n-dimensional real space. The spatial arrangement of elastic scatterers is given by the scatterers' density $f(\vec{r})$. Ideally, we wish to recover $f(\vec{r})$ from our diffraction experiments, which provide measured intensities. However, far-field patterns of diffracted intensity yield only $I_{\text{diff}}(\vec{\omega}) = c|F(\vec{\omega})|^2$, where $F(\vec{\omega}) =$

$\int_{\mathbb{R}^n} f(\vec{r}) e^{-i\vec{\omega} \cdot \vec{r}} \, d^n \vec{r}$ is the n-dimensional Fourier transform of $f(\vec{r})$ and c is some constant. In other words, $F(\vec{\omega})$'s phase information is lost when only intensity is measured. The X-ray beam's expected diffracted intensity is proportional to $\left\langle |F(\vec{\omega})|^2 \right\rangle$, which is the n-dimensional generalization of a power spectrum. However, it is also interesting to relate the n-dimensional diffraction pattern, along a curve in reciprocal space, to the more familiar one-dimensional power spectrum.

For a given directional frequency $\vec{\omega}$, decompose $\vec{r} = \vec{r}_\parallel + \vec{r}_\perp$, where $\vec{r}_\parallel \equiv (\vec{r} \cdot \widehat{\omega})\widehat{\omega}$ and $\widehat{\omega} = \vec{\omega}/|\vec{\omega}|$. Then, let $\mu_\perp(\vec{r}_\parallel)$ be the accumulated density within the entire cross-sectional plane perpendicular to and uniquely identified by $\vec{r}_\parallel$; i.e., $\mu_\perp(\vec{r}_\parallel) \equiv \int_{\mathbb{R}^{n-1}} f(\vec{r}_\parallel + \vec{r}_\perp) \, d^{n-1}\vec{r}_\perp$. We then find that in general:

$$I_{\text{diff}}(\vec{\omega}) = c \left| \int_{\mathbb{R}} \mu_\perp(\vec{r}_\parallel) e^{-i\omega r_\parallel} \, dr_\parallel \right|^2 . \qquad (A1)$$

In particular, we see that the diffraction pattern along any line $\vec{\omega} = \omega\widehat{\omega}$ (with varying ω but fixed $\widehat{\omega}$) is the power spectrum of the net magnitude of scatterers within sequential cross sections of real space perpendicular to $\widehat{\omega}$.

For molecular or crystalline structures, when the net scatterer density is well-approximated by a superposition of more elementary densities $f(\vec{r}) = \sum_j f_j(\vec{r} - \vec{r}_j)$, we obtain the alternative expression:

$$I_{\text{diff}}(\vec{\omega}) = c \left| \sum_{\ell=-\infty}^{\infty} F^{(\ell)}(\vec{\omega}) e^{-i\omega\ell d} \right|^2 . \qquad (A2)$$

Here, the layer-scattering factors are defined:

$$F^{(\ell)}(\vec{\omega}) \equiv \sum_{j \in \{m : \ell d \leq r_{m_\parallel} < \ell d + d\}} F_j(\vec{\omega}) e^{-i\omega(r_{j_\parallel} - \ell d)} ,$$

for cross-sectional layers of depth d, where $r_{j_\parallel} = \vec{r}_j \cdot \widehat{\omega}$. And:

$$F_j(\vec{\omega}) = \int_{\mathbb{R}^n} f_j(\vec{r}) e^{-i\vec{\omega} \cdot \vec{r}} \, d^n \vec{r}$$

is the n-dimensional Fourier transform of $f_j(\vec{r})$. Again, the diffraction pattern (along some line in reciprocal space passing through the origin) appears as a one-dimensional power spectrum. However, this time the diffraction pattern is the power spectrum of the complex-valued layer-scattering factors $F^{(\ell)}(\vec{\omega})$ over a discrete spatial domain. The frequency-dependence of $F^{(\ell)}(\vec{\omega})$ is often factored out to 'correct' the diffraction pattern, so that only the structure of interest remains [75, 76]. The corrected diffraction pattern is thus a standard power spectrum, with the form of Eq. (1). Equation (A2) is an especially useful expression for layered structures, as demonstrated in Ref. [56].

Appendix B: Autocorrelation for processes generated by autonomous HMMs

Let's derive the autocorrelation function in general and in closed form for the class of autonomous HMMs introduced in the main text. Helpfully, for particular models, the expressions become analytic in terms of the model parameters.

Directly calculating, we find that the autocorrelation function, for $\tau > 0$, for any such HMM is:

$$\begin{aligned}
\gamma(\tau) &= \left\langle \overline{X_t} X_{t+\tau} \right\rangle \\
&= \int_{x \in \mathcal{A}} \int_{x' \in \mathcal{A}} \overline{x} x' \, \mathrm{p}(X_0 = x, X_\tau = x') \, dx \, dx' \\
&= \sum_{s \in \mathcal{S}} \sum_{s' \in \mathcal{S}} \int_{x \in \mathcal{A}} \int_{x' \in \mathcal{A}} \overline{x} x' \, \mathrm{p}(X_0 = x, X_\tau = x', \mathcal{S}_0 = s, \mathcal{S}_\tau = s') \, dx \, dx' \\
&= \sum_{s \in \mathcal{S}} \sum_{s' \in \mathcal{S}} \int_{x \in \mathcal{A}} \int_{x' \in \mathcal{A}} \overline{x} x' \, \mathrm{Pr}(\mathcal{S}_0 = s, \mathcal{S}_\tau = s') \, \mathrm{p}(X_0 = x | \mathcal{S}_0 = s) \, \mathrm{p}(X_\tau = x' | \mathcal{S}_\tau = s') \, dx \, dx' \\
&= \sum_{s \in \mathcal{S}} \sum_{s' \in \mathcal{S}} \langle \boldsymbol{\pi} | s \rangle \langle s | T^\tau | s' \rangle \langle s' | \mathbf{1} \rangle \left(\int_{x \in \mathcal{A}} \overline{x} \, \mathrm{p}(x|s) \, dx \right) \left(\int_{x' \in \mathcal{A}} x' \, \mathrm{p}(x'|s') \, dx' \right) \\
&= \langle \boldsymbol{\pi} | \left(\sum_{s \in \mathcal{S}} \langle \overline{x} \rangle_{\mathrm{p}(X|s)} | s \rangle \langle s | \right) T^\tau \left(\sum_{s' \in \mathcal{S}} \langle x \rangle_{\mathrm{p}(X|s')} | s' \rangle \langle s' | \right) | \mathbf{1} \rangle ,
\end{aligned}$$

where the integrals are written in a form meant to be easily accessible but should generally be interpreted as Lebesgue

integrals. In the above derivation, note that:

$$p(X_0 = x, X_\tau = x', \mathcal{S}_0 = s, \mathcal{S}_\tau = s') = \Pr(\mathcal{S}_0 = s, \mathcal{S}_\tau = s')p(X_0 = x, X_\tau = x'|\mathcal{S}_0 = s, \mathcal{S}_\tau = s')$$

holds by definition of conditional probability. The decomposition of:

$$p(X_0 = x, X_\tau = x'|\mathcal{S}_0 = s, \mathcal{S}_\tau = s') = p(X_0 = x|\mathcal{S}_0 = s)p(X_\tau = x'|\mathcal{S}_\tau = s')$$

for $\tau \neq 0$ follows from the conditional independence in the relevant Bayesian network shown in Fig. 3. Moreover, the equality:

$$\Pr(\mathcal{S}_0 = s, \mathcal{S}_\tau = s') = \langle \boldsymbol{\pi}|s\rangle \langle s| T^\tau |s'\rangle \langle s'|\mathbf{1}\rangle$$

can be derived by marginalizing over all possible intervening state sequences. We can use the hidden-state basis, where $|s\rangle$ is the column vector of all 0s except for a 1 at the index corresponding to state s, while $\langle s|$ is simply its transpose. This yields a natural decomposition of the identity operator: $I = \sum_{s \in \mathcal{S}} |s\rangle \langle s|$.

Since the autocorrelation is a Hermitian function—i.e., $\gamma(-\tau) = \overline{\gamma}(\tau)$—and:

$$\gamma(0) = \left\langle |x|^2 \right\rangle_{\boldsymbol{\pi}(X)}$$
$$= \langle \boldsymbol{\pi}| \sum_{s \in \mathcal{S}} \left\langle |x|^2 \right\rangle_{p(X|s)} |s\rangle \ ,$$

we find the full autocorrelation function is given by:

$$\gamma(\tau) = \begin{cases} \langle \boldsymbol{\pi}| \Omega\, T^{|\tau|}\, \overline{\Omega} |\mathbf{1}\rangle & \text{if } \tau \leq 1 \\ \left\langle |x|^2 \right\rangle & \text{if } \tau = 0 \ , \\ \langle \boldsymbol{\pi}| \overline{\Omega}\, T^{|\tau|}\, \Omega |\mathbf{1}\rangle & \text{if } \tau \geq 1 \end{cases} \qquad \text{(B1)}$$

where Ω is the $|\mathcal{S}|$-by-$|\mathcal{S}|$ matrix defined by:

$$\Omega = \sum_{s \in \mathcal{S}} \langle x \rangle_{p(X|s)} |s\rangle \langle s| \ .$$

The Ω matrix simply places state-conditioned average outputs along its diagonal.

To better understand the range of possible behaviors of autocorrelation, we can go a step further. In particular, we employ the general spectral decomposition of T^τ derived in Ref. [55] for nonnormal and potentially non-diagonalizable operators:

$$T^\tau = \Big[\sum_{m=0}^{\nu_0 - 1} \delta_{\tau,m} T_{0,m} \Big] + \sum_{\lambda \in \Lambda_T \setminus \{0\}} \sum_{m=0}^{\nu_\lambda - 1} \binom{\tau}{m} \lambda^{\tau-m} T_{\lambda,m},$$

$$\text{(B2)}$$

where $\binom{\tau}{m}$ is the generalized binomial coefficient:

$$\binom{\tau}{m} = \frac{1}{m!} \prod_{n=1}^{m} (\tau - n + 1) \ ,$$

with $\binom{\tau}{0} = 1$. As briefly summarized in Sec. II D, Λ_T is the set of T's eigenvalues while T_λ is the spectral projection operator associated with the eigenvalue λ. Recall that ν_λ is the *index* of the eigenvalue λ, i.e., the size of the largest Jordan block associated with λ, and $T_{\lambda,m} = T_\lambda (T - \lambda I)^m$. Substituting Eq. (B2) into Eq. (B1) yields:

$$\gamma(\tau) = \Big[\sum_{m=1}^{\nu_0 - 1} \delta_{\tau,m} \langle \boldsymbol{\pi}| \overline{\Omega}\, T_{0,m}\, \Omega |\mathbf{1}\rangle \Big]$$
$$+ \sum_{\lambda \in \Lambda_T \setminus \{0\}} \sum_{m=0}^{\nu_\lambda - 1} \binom{\tau}{m} \lambda^{\tau-m} \langle \boldsymbol{\pi}| \overline{\Omega}\, T_{\lambda,m}\, \Omega |\mathbf{1}\rangle \ ,$$

for $\tau > 0$.

It is significant that the zero eigenvalue contributes a qualitatively distinct ephemeral behavior to the autocorrelation while $|\tau| < \nu_0$. All other eigenmodes contribute products of polynomials times decaying exponentials in τ. When T is diagonalizable, the autocorrelation is simply a sum of decaying exponentials.

Appendix C: Analytical power spectra

The following derives both the continuous and discrete part of the power spectrum for HMM-generated processes. The development parallels that in Ref. [56], although that derivation was restricted to the special case of diffraction patterns from Mealy (i.e., edge-emitting) HMMs with countable alphabets. In contrast, the following derives analytical expressions for the power spectrum of any stochastic process generated by an HMM. Notably, it also allows uncountably infinite alphabets. Also, it is developed for Moore (i.e., state-emitting) HMMs—although Mealy and Moore HMMs are class-equivalent and can be easily transformed from one to the other.

1. Diffuse Spectra

Recall Eq. (3):

$$P(\omega) = \lim_{N \to \infty} \tfrac{1}{N} \sum_{\tau=-N}^{N} \left(N - |\tau|\right)\gamma(\tau)e^{-i\omega\tau} \,,$$

and Eq. (4)'s explicit expression for the correlation function:

$$\gamma(\tau) = \begin{cases} \langle \boldsymbol{\pi}| \,\Omega\, T^{|\tau|}\, \overline{\Omega}\, |\mathbf{1}\rangle & \text{if } \tau \leq 1 \\ \langle |x|^2\rangle & \text{if } \tau = 0 \\ \langle \boldsymbol{\pi}| \,\overline{\Omega}\, T^{|\tau|}\, \Omega\, |\mathbf{1}\rangle & \text{if } \tau \geq 1 \end{cases} \,.$$

From these we can rewrite the power spectrum directly in terms of the generating HMM's transition matrix:

$$P(\omega) = \langle |x|^2\rangle +$$
$$\lim_{N \to \infty} \frac{2}{N} \,\mathrm{Re} \sum_{\tau=1}^{N} (N - \tau)\, \langle \boldsymbol{\pi}| \,\overline{\Omega}\, T^{\tau}\, \Omega\, |\mathbf{1}\rangle\, e^{-i\omega\tau}$$
$$= \langle |x|^2\rangle +$$
$$\lim_{N \to \infty} \frac{2}{N} \,\mathrm{Re} \,\langle \boldsymbol{\pi}| \,\overline{\Omega} \left(\sum_{\tau=1}^{N}(N - \tau)T^{\tau}e^{-i\omega\tau}\right)\Omega\, |\mathbf{1}\rangle \,.$$

(C1)

We used the fact that $z + \overline{z} = 2\mathrm{Re}(z)$ for any $z \in \mathbb{C}$. For convenience, we introduce the variable $\mathfrak{z} \equiv e^{-i\omega}$. We then note that the summation splits:

$$\sum_{\tau=1}^{N}(N - \tau)T^{\tau}e^{-i\omega\tau} = N \sum_{\tau=1}^{N}(\mathfrak{z}T)^{\tau} - \sum_{\tau=1}^{N}\tau(\mathfrak{z}T)^{\tau} \,.$$

For positive integer N, it is always true that:

$$(I - \mathfrak{z}T) \sum_{\tau=1}^{N}(\mathfrak{z}T)^{\tau} = \mathfrak{z}T - \mathfrak{z}^{N+1}T^{N+1} \,,$$

and:

$$(I - \mathfrak{z}T) \sum_{\tau=1}^{N}\tau(\mathfrak{z}T)^{\tau} = -N\mathfrak{z}^{N+1}T^{N+1} + \sum_{\tau=1}^{N}(\mathfrak{z}T)^{\tau} \,.$$

Hence, whenever $I - \mathfrak{z}T$ is invertible (i.e., whenever $e^{i\omega} \notin \Lambda_T$), we have:

$$\sum_{\tau=1}^{N}(\mathfrak{z}T)^{\tau} = (I - \mathfrak{z}T)^{-1}\left(\mathfrak{z}T - \mathfrak{z}^{N+1}T^{N+1}\right) \,,$$

and:

$$\sum_{\tau=1}^{N}\tau(\mathfrak{z}T)^{\tau} = (I - \mathfrak{z}T)^{-1}\left(-N\mathfrak{z}^{N+1}T^{N+1} + (I - \mathfrak{z}T)^{-1}\left(\mathfrak{z}T - \mathfrak{z}^{N+1}T^{N+1}\right)\right) \,.$$

Together, this yields:

$$\sum_{\tau=1}^{N}(N - \tau)T^{\tau}e^{-i\omega\tau} = N \sum_{\tau=1}^{N}(\mathfrak{z}T)^{\tau} - \sum_{\tau=1}^{N}\tau(\mathfrak{z}T)^{\tau}$$
$$= N(I - \mathfrak{z}T)^{-1}\left(\mathfrak{z}T - \mathfrak{z}^{N+1}T^{N+1} + \mathfrak{z}^{N+1}T^{N+1}\right) - (I - \mathfrak{z}T)^{-2}\left(\mathfrak{z}T - \mathfrak{z}^{N+1}T^{N+1}\right)$$
$$= NT(\mathfrak{z}^{-1}I - T)^{-1} - (I - \mathfrak{z}T)^{-2}\left(\mathfrak{z}T - \mathfrak{z}^{N+1}T^{N+1}\right) \,.$$

Noting that $(\mathfrak{z}^{-1}I - T)^{-1} = (e^{i\omega}I - T)^{-1}$, this implies that the continuous (i.e., diffuse) part of the power spectrum becomes:

$$P_{\mathrm{c}}(\omega) = \langle |x|^2\rangle + \lim_{N \to \infty} \frac{2}{N} \,\mathrm{Re} \,\langle \boldsymbol{\pi}| \,\overline{\Omega} \left(\sum_{\tau=1}^{N}(N - \tau)T^{\tau}e^{-i\omega\tau}\right)\Omega\, |\mathbf{1}\rangle$$
$$= \langle |x|^2\rangle + \lim_{N \to \infty} \frac{2}{N} \,\mathrm{Re} \,\langle \boldsymbol{\pi}| \,\overline{\Omega} \left(NT(\mathfrak{z}^{-1}I - T)^{-1} - (I - \mathfrak{z}T)^{-2}\left(\mathfrak{z}T - \mathfrak{z}^{N+1}T^{N+1}\right)\right)\Omega\, |\mathbf{1}\rangle$$
$$= \langle |x|^2\rangle + 2\,\mathrm{Re} \,\langle \boldsymbol{\pi}| \,\overline{\Omega}\, T(\mathfrak{z}^{-1}I - T)^{-1}\Omega\, |\mathbf{1}\rangle - \lim_{N \to \infty} \frac{2}{N}\,\mathrm{Re}\,\langle \boldsymbol{\pi}| \,\overline{\Omega}\,(I - \mathfrak{z}T)^{-2}\left(\mathfrak{z}T - \mathfrak{z}^{N+1}T^{N+1}\right)\Omega\, |\mathbf{1}\rangle \qquad \text{(C2)}$$
$$= \langle |x|^2\rangle + 2\,\mathrm{Re} \,\langle \boldsymbol{\pi}| \,\overline{\Omega}\, T(e^{i\omega}I - T)^{-1}\Omega\, |\mathbf{1}\rangle \,. \qquad \text{(C3)}$$

Equation (C3) is the principle result, yielding the continuous part of the power spectrum in closed form. However, it is also worth noting that Eq. (C2) (without the $N \to \infty$ limit yet being taken) provides the exact result for the expected periodogram from finite-length N samples.

2. Discrete Spectra

The transition dynamic's eigenvalues $\Lambda_{\rho(T)} = \{\lambda \in \Lambda_T : |\lambda| = 1\}$ on the unit circle are responsible for a power spectrum's Dirac δ-functions. In the physical context of diffraction patterns, these δ-functions are the familiar Bragg peaks. For finite length-N samples, eigenvalues on the unit circle give rise to Dirichlet kernels. As $N \to \infty$, the analysis simplifies since the Dirichlet kernels converge to δ-functions.

The following derives the exact form of the δ-function contributions, showing how their presence and integrated magnitude can be calculated directly from the stochastic transition dynamic. Recall that the spectral projection operator $T_{\lambda,0}$ associated with the eigenvalue λ can be defined as the residue of $(zI - T)^{-1}$ as $z \to \lambda$:

$$T_{\lambda,0} = \frac{1}{2\pi i} \oint_{C_\lambda} \left(zI - T \right)^{-1} dz \ .$$

The spectral companion operators are:

$$T_{\lambda,m} = T_{\lambda,0}(T - \lambda I)^m \ ,$$

with the useful property that $T_{\lambda,m}T_{\zeta,n} = \delta_{\lambda,\zeta}T_{\lambda,m+n}$ and $T_{\lambda,m} = \mathbf{0}$ for $m \geq \nu_\lambda$. The index ν_λ of the eigenvalue λ is the size of the largest Jordan block associated with λ.

The Perron-Frobenius theorem guarantees that all eigenvalues on the unit circle have an index of one: i.e., $\nu_\lambda = 1$ for all $\lambda \in \Lambda_{\rho(T)}$. This means that the algebraic and geometric multiplicities of these eigenvalues coincide and they are all associated with diagonalizable subspaces.

Taking advantage of the index-one nature of the eigenvalues on the unit circle, and using the shorthand $T_\lambda \equiv T_{\lambda,0}$ for the spectral projection operators, we define:

$$\Theta \equiv \sum_{\lambda \in \Lambda_{\rho(T)}} \lambda T_\lambda$$

and

$$F \equiv T - \Theta \ .$$

We then consider how the spectral decomposition of T^τ splits into contributions from these two independent components: From Ref. [55], and employing the simplifying notation that $0^{\tau-m} = \delta_{\tau-m,0}$, we find:

$$T^\tau = \sum_{\lambda \in \Lambda_T} \sum_{m=0}^{\nu_\lambda - 1} \lambda^{\tau-m} \binom{\tau}{m} T_{\lambda,m}$$

$$= \left(\sum_{\lambda \in \Lambda_{\rho(T)}} \lambda^\tau T_\lambda \right) + \left(\sum_{\lambda \in \Lambda_T \backslash \Lambda_{\rho(T)}} \sum_{m=0}^{\nu_\lambda - 1} \lambda^{\tau-m} \binom{\tau}{m} T_{\lambda,m} \right)$$

$$= \Theta^\tau + F^\tau \ ,$$

where $\binom{\tau}{m} = \frac{1}{m!} \prod_{n=1}^{m} (\tau - n + 1)$ is the generalized binomial coefficient.

As the sequence length $N \to \infty$, the summation over τ in Eq. (C1) divided by the sequence length becomes:

$$\lim_{N \to \infty} \sum_{\tau=1}^{N} \frac{N - \tau}{N} T^\tau e^{-i\omega\tau}$$

$$= \sum_{\tau=1}^{\infty} T^\tau e^{-i\omega\tau}$$

$$= \left(\sum_{\tau=1}^{\infty} \Theta^\tau e^{-i\omega\tau} \right) + \left(\sum_{\tau=1}^{\infty} F^\tau e^{-i\omega\tau} \right) \ . \quad \text{(C4)}$$

In Eq. (C4), only the summation involving Θ is capable of contributing δ-functions. Expanding that sum yields:

$$\sum_{\tau=1}^{\infty} \Theta^\tau e^{-i\omega\tau}$$

$$= \sum_{\lambda \in \Lambda_{\rho(T)}} T_\lambda \sum_{\tau=1}^{\infty} (\lambda e^{-i\omega})^\tau$$

$$= \sum_{\lambda \in \Lambda_{\rho(T)}} T_\lambda \left(-1 + \sum_{\tau=0}^{\infty} e^{i(\omega_\lambda - \omega)\tau} \right)$$

$$= \sum_{\lambda \in \Lambda_{\rho(T)}} T_\lambda \left(\frac{-1}{1 - e^{i(\omega - \omega_\lambda)}} + \sum_{k=-\infty}^{\infty} \pi \, \delta(\omega - \omega_\lambda + 2\pi k) \right),$$

$$\text{(C5)}$$

where ω_λ is related to λ by $\lambda = e^{i\omega_\lambda}$. The last line is obtained using well-known properties of the discrete-time Fourier transform [77].

From Eqs. (C1), (C4), and (C5), we find that the potential δ-function at ω_λ (and its 2π-periodic offsets) has integrated magnitude:

$$\Delta_\lambda \equiv \lim_{\epsilon \to 0} \int_{\omega_\lambda - \epsilon}^{\omega_\lambda + \epsilon} P(\omega)\, d\omega$$

$$= \lim_{\epsilon \to 0} \int_{\omega_\lambda - \epsilon}^{\omega_\lambda + \epsilon} 2\, \mathrm{Re}\, \langle \boldsymbol{\pi} | \, \overline{\Omega} \left(\lim_{N \to \infty} \sum_{\tau=1}^{N} \frac{N - \tau}{N} T^\tau e^{-i\omega\tau} \right) \Omega \, | \mathbf{1} \rangle \, d\omega$$

$$= \lim_{\epsilon \to 0} \int_{\omega_\lambda - \epsilon}^{\omega_\lambda + \epsilon} 2\, \mathrm{Re}\, \langle \boldsymbol{\pi} | \, \overline{\Omega} \left(\sum_{\tau=1}^{\infty} \Theta^\tau e^{-i\omega\tau} \right) \Omega \, | \mathbf{1} \rangle \, d\omega$$

$$= \lim_{\epsilon \to 0} \int_{\omega_\lambda - \epsilon}^{\omega_\lambda + \epsilon} 2\, \mathrm{Re}\, \langle \boldsymbol{\pi} | \, \overline{\Omega} \sum_{\zeta \in \Lambda_{\rho(T)}} T_\zeta \left(\frac{-1}{1 - e^{i(\omega - \omega_\zeta)}} + \sum_{k=-\infty}^{\infty} \pi\, \delta(\omega - \omega_\zeta + 2\pi k) \right) \Omega \, | \mathbf{1} \rangle \, d\omega$$

$$= 2\pi\, \mathrm{Re}\, \langle \boldsymbol{\pi} | \, \overline{\Omega}\, T_\lambda \, \Omega \, | \mathbf{1} \rangle \lim_{\epsilon \to 0} \int_{\omega_\lambda - \epsilon}^{\omega_\lambda + \epsilon} \delta(\omega - \omega_\lambda)\, d\omega$$

$$= 2\pi\, \mathrm{Re}\, \langle \boldsymbol{\pi} | \, \overline{\Omega}\, T_\lambda \, \Omega \, | \mathbf{1} \rangle \; . \tag{C6}$$

Finally, from Eq. (C6) and the 2π-periodicity of the power spectrum, we obtain the full discrete (i.e., δ-function) contribution to the power spectrum:

$$P_\mathrm{d}(\omega) = \sum_{\lambda \in \Lambda_{\rho(T)}} 2\pi\, \mathrm{Re}\, \langle \boldsymbol{\pi} | \, \overline{\Omega}\, T_\lambda \Omega \, | \mathbf{1} \rangle \sum_{k=-\infty}^{\infty} \delta(\omega - \omega_\lambda + 2\pi k). \tag{C7}$$

Appendix D: Cross-correlation and spectral densities

Cross-correlation and cross-spectral densities are often important in applications [78, 79]. These may be especially useful when analyzing input–output processes, to characterize the correlation of input and output, or to characterize the correlation between different aspects of the output. Our results can be easily extended to address these quantities.

Using an HMM that describes the *joint* stochastic process of two observables $(x, y) \in \mathcal{A}$, it is straightforward to generalize our developments to *cross-correlation* $\gamma_{XY}(\tau)$:

$$\gamma_{XY}(\tau) = \langle \overline{X_t} Y_{t+\tau} \rangle$$

(rather than necessarily *auto*correlation $\gamma = \gamma_{XX}$) and the associated *cross-spectral densities* $P_{XY}(\omega)$:

$$P_{XY}(\omega) = \lim_{N \to \infty} \frac{1}{N} \left\langle \left(\sum_{t=1}^{N} \overline{X_t} e^{i\omega t} \right) \left(\sum_{t=1}^{N} Y_t e^{-i\omega t} \right) \right\rangle$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{\tau=-N}^{N} (N - |\tau|) \gamma_{XY}(\tau) e^{-i\omega\tau}$$

of distinct observables $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. The individual stochastic processes for each observable by itself

can simply be obtained by marginalizing over the other observable.

Explicitly, the expressions take the form:

$$\gamma_{XY}(\tau) = \begin{cases} \langle \boldsymbol{\pi} | \, \Omega_Y\, T^{|\tau|}\, \overline{\Omega}_X \, | \mathbf{1} \rangle & \text{if } \tau \le 1 \\ \langle \overline{X_t} Y_t \rangle & \text{if } \tau = 0 \\ \langle \boldsymbol{\pi} | \, \overline{\Omega}_X\, T^{|\tau|}\, \Omega_Y \, | \mathbf{1} \rangle & \text{if } \tau \ge 1 \end{cases} ,$$

where:

$$\Omega_Y = \sum_{s \in \boldsymbol{\mathcal{S}}} \langle y \rangle_{\mathrm{p}(X,Y|s)} |s\rangle \langle s| \; ,$$

and:

$$\langle \overline{X_t} Y_t \rangle = \sum_{s \in \boldsymbol{\mathcal{S}}} \langle \boldsymbol{\pi} | s \rangle \, \langle \overline{x} y \rangle_{\mathrm{p}(X,Y|s)} \; .$$

Moreover, the continuous part of the cross-spectral density is given by:

$$P_{XY\mathrm{c}}(\omega) = \langle \overline{X_t} Y_t \rangle + \langle \boldsymbol{\pi} | \, \overline{\Omega}_X\, T \left(e^{i\omega} I - T \right)^{-1} \Omega_Y \, | \mathbf{1} \rangle$$
$$+ \langle \boldsymbol{\pi} | \, \Omega_Y\, T \left(e^{-i\omega} I - T \right)^{-1} \overline{\Omega}_X \, | \mathbf{1} \rangle \; .$$

And so on.

Appendix E: Continuous-time processes

For simplicity and generality, the main development addressed discrete-time dynamics. (Indeed, discrete-time dynamics are, in a sense, more general than continuous-time dynamics, while continuous-time dynamics can be obtained as the limiting behavior of discrete-time dynamics.) However, correlation and power spectra are often applied to continuous-time processes. And so, the

continuous-time setting is often of direct interest. The following makes a more explicit connection to continuous-time processes.

First, we note that a continuous-time process is typically observed only intermittently at some sampling frequency f_0. The duration $\tau_0 = 1/f_0$ between observations thus induces a discrete-time transition operator T_{τ_0} between states in that time interval. In such cases, the discrete-time transition matrix is related to the continuous-time generator G of time evolution by $T_{\tau_0} = e^{\tau_0 G}$. Accordingly, the continuous-time generator can be obtained from the discrete-time dynamic as: $G = f_0 \ln T_{\tau_0}$ [80]. The relationship between discrete and continuous time is the same relationship that yields the well-known conformal mapping of the interior of the unit circle in the complex plane to the left-half of the complex plane, which also relates z-transforms and Laplace transforms. Of most direct relevance here, the eigenvalues of T_{τ_0} and G are simply related by $\Lambda_{T_{\tau_0}} = \bigcup_{\zeta \in \Lambda_G} \{e^{\tau_0 \zeta}\}$.

Continuous-time representations can be analyzed directly. Consider the generic case of a continuous-time dynamic over a hidden state-space, with two types of example in mind:

1. The system evolves through a continuous state-space and describes both typical linear systems and typical nonlinear systems, including chaotic dynamical systems. Then G is the generator that induces the finite-time Ruelle–Frobenius–Perron operator.

2. Or, observations are functions of a finite state space with continuous-time transition rates. For example, we observe current flowing or not, depending on the conformation of a biomolecular ion channel.

These different settings all have the same expression for the autocorrelation and power spectrum. We now give them in closed-form.

For $\tau > 0$, we find the autocorrelation to be:

$$\gamma(\tau) = \langle \boldsymbol{\pi} | \overline{\Omega} \, e^{|\tau| G} \, \Omega \, | \mathbf{1} \rangle \ .$$

From this, we derive the continuous part of the power spectrum $P_c(\omega)$ with respect to angular frequency $\omega = 2\pi f \in \mathbb{R}$, with the result that:

$$P_c(\omega) = 2 \operatorname{Re} \langle \boldsymbol{\pi} | \overline{\Omega} \left(i\omega I - G \right)^{-1} \Omega | \mathbf{1} \rangle \ .$$

Appealing to the resolvent's spectral expansion again allows us to better understand the possible shapes of their power spectrum:

$$P_c(\omega) = \sum_{\lambda \in \Lambda_G} \sum_{m=0}^{\nu_\lambda - 1} 2 \operatorname{Re} \frac{\langle \boldsymbol{\pi} | \overline{\Omega} \, G_{\lambda, m} \Omega | \mathbf{1} \rangle}{(i\omega - \lambda)^{m+1}} \ . \tag{E1}$$

Since all of the frequency-dependence is isolated in the denominator and since $\langle \boldsymbol{\pi} | \overline{\Omega} \, G_{\lambda, m} \Omega | \mathbf{1} \rangle$ is a frequency-independent complex-valued constant, peaks in $P_c(\omega)$ arise only via contributions of the form $\operatorname{Re} \frac{c}{(i\omega - \lambda)^n}$ for $c \in \mathbb{C}$, $\omega \in \mathbb{R}$, $\lambda \in \Lambda_G$, and $n \in \mathbb{Z}_+$.

This provides a rich starting point for applications and further theoretical investigation. For example, Eq. (E1) helps explain the shapes of power spectra of nonlinear dynamical systems, as have appeared, e.g., in Ref. [10]. Furthermore, it suggests an approach to the inverse problem of inferring the spectrum of the hidden linear dynamic via power spectra.

Appendix F: Pairwise mutual information example

For the process generated by the HMM given in Fig. 6, taking the limit of ever-narrower Gaussians in the state-conditioned PDFs, so that we work with pairs of δ-functions, then the process becomes Markovian and the pairwise mutual information can be calculated exactly:

$$
\begin{aligned}
\mathrm{I}(X_0; X_\tau) &= H(X_0) - H(X_\tau | X_0) \\
&= H(X_0, \mathcal{S}_0) - H(X_\tau, \mathcal{S}_\tau | X_0, \mathcal{S}_0) \\
&= H(\mathcal{S}_0) + \mathrm{H}(X_0 | \mathcal{S}_0) - H(X_\tau, \mathcal{S}_\tau | \mathcal{S}_0) \\
&= H(\mathcal{S}_0) + \mathrm{H}(X_0 | \mathcal{S}_0) - H(\mathcal{S}_\tau | \mathcal{S}_0) - H(X_\tau | \mathcal{S}_\tau) \\
&= H(\mathcal{S}_0) - H(\mathcal{S}_\tau | \mathcal{S}_0) \\
&= H(\boldsymbol{\pi}) - \sum_{s \in \mathcal{S}} \boldsymbol{\pi}(s) H(\mathcal{S}_\tau | \mathcal{S}_0 = s) \\
&= H(\boldsymbol{\pi}) - \sum_{s \in \mathcal{S}} \boldsymbol{\pi}(s) H\big(\langle s | T^\tau \big) \\
&= H(\boldsymbol{\pi}) + \sum_{s, s' \in \mathcal{S}} \boldsymbol{\pi}(s) \langle s | T^\tau | s' \rangle \log \langle s | T^\tau | s' \rangle, \tag{F1}
\end{aligned}
$$

where $\boldsymbol{\pi} = [1, 1-p, 1-p, 1-p]/(4 - 3p)$.

Continuing, $\langle s | T^\tau | s' \rangle$ can be calculated via T's spectral decomposition. Since T is diagonalizable and nondegenerate for all values of the transition parameter p, we find:

$$\langle s | T^\tau | s' \rangle = \sum_{\lambda \in \Lambda_T} \lambda^\tau \langle s | T_\lambda | s' \rangle \ .$$

Moreover:

$$
\begin{aligned}
\langle s | T_1 | s' \rangle &= \langle s | \mathbf{1} \rangle \langle \boldsymbol{\pi} | s' \rangle \\
&= \boldsymbol{\pi}(s') \ ,
\end{aligned}
$$

so $\langle s | T^\tau | s' \rangle$ simplifies somewhat to:

$$\langle s | T^\tau | s' \rangle = \boldsymbol{\pi}(s') + \sum_{\lambda \in \Lambda_T \setminus \{1\}} \lambda^\tau \langle s | T_\lambda | s' \rangle$$

In fact, Eq. (F1) is valid for any set of four PDFs we could have chosen for the example HMM's states, as long as the PDFs all have mutually exclusive support for the observable output, since this then makes the hidden state a function of the instantaneous observable.

Using the linear algebra of Eq. (F1), we calculate the pairwise mutual information and POPI spectrum numerically. The pairwise mutual informations are shown for $p \in \{0.1, 0.5, 0.9\}$ in Fig. 8. Reasonably, the loss of information is monotonic over temporal distance. More surprisingly, the decay of pairwise mutual information is very-nearly exponential as made clear in the inset logarithmic plot.

The POPI spectrum, which can be rewritten for a wide-sense stationary process as:

$$\mathcal{I}(\omega) = \lim_{N \to \infty} 2 \sum_{\tau=1}^{N} \cos(\omega\tau) \, \mathrm{I}(X_0; X_\tau) \, ,$$

is shown for these same p-values in Fig. 9. The POPI spectrum was approximated by truncating the summation of modulated pairwise mutual informations at a sufficiently large separation of $\tau = 2000$.
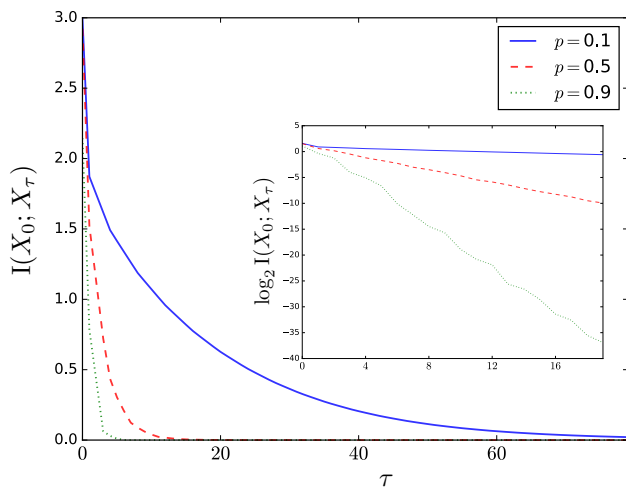

FIG. 8. Nontrivial pairwise mutual information for the process from Fig. 6 with a flat power spectrum.

Appendix G: Measurement Feedback Models

Let's now turn to describe an alternative set of possibly-input-dependent models, which may be more convenient for describing certain phenomena. For example, they are more natural for describing measured quantum systems. They also reduce to the canonical models used in *computational mechanics* [62, 64] after a number of simplifying assumptions.
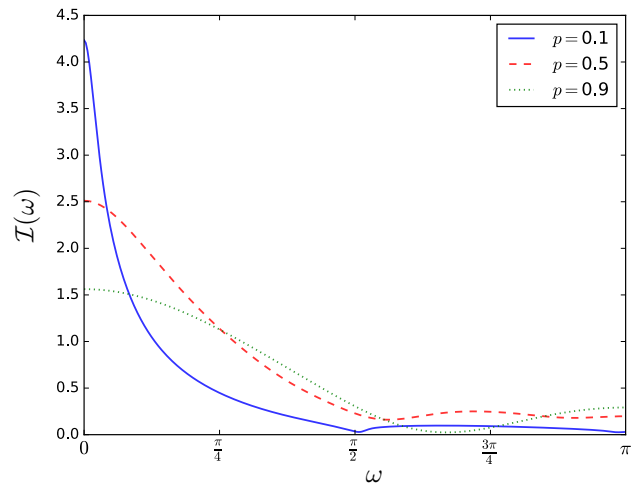

FIG. 9. Power-of-Pairwise-Information (POPI) spectrum for the process from Fig. 6.

After introducing them, we show that Thm. 2 applies to them as well as to the other model types discussed in the main body. In this way, we extend the theory of fraudulent white noise to these models as well.

The models we consider generate observable behavior *during transitions* between states, rather than in the states themselves. This is a natural approach in the quantum setting since measurement feedback changes the state of the quantum system with dependence on the measurement outcome. For projective measurements, measurement fully defines the new state, but for the much more general class of quantum measurements described by *positive operator valued measures* (POVMs), the measurement outcome plays a more nuanced role in updating the state. More generally, edge-emitting models can be natural descriptors of complex systems with control and feedback. And, fittingly, edge-emitting models have been used elsewhere as well. For instance, they appear extensively in computer science and computational mechanics—the latter of which spans the study of natural computation in physical systems and the minimal resources required for prediction.

1. Measurement Feedback Models

Here we introduce Measurement Feedback Models (MFMs) $\mathcal{M}_{\mathrm{MFM}}(\vec{m})$, which are input-dependent generators of an observable output process $\{X_t\}_{t \in \mathcal{T}}$. As before, the lengths and alphabets of the inputs and outputs need not be commensurate. The output is generated via $\mathcal{M}_{\mathrm{MFM}}(\vec{m}) = \big(\mathcal{S}, \mathcal{A}, \{T_t^{(x)}(\vec{m})\}_{t \in \mathcal{T}, x \in \mathcal{A}}, \boldsymbol{\mu}_1\big)$, where $\mathcal{S}$ is the countable set of hidden states, $\mathcal{A}$ is the alphabet of observables, and $\boldsymbol{\mu}_1$ is the initial distribution over hid-

den states. For a given t and x, the matrix elements $\langle s| T_t^{(x)}(\vec{m}) |s'\rangle$ provide the probability density of transitioning from state s to s' while emitting the observable x; that is:

$$\langle s| T_t^{(x)}(\vec{m}) |s'\rangle = \mathrm{p}_{\vec{m}}(X_{t+1} = x, \mathcal{S}_{t+1} = s'|\mathcal{S}_t = s) .$$

where $\mathrm{p}_{\vec{m}}$ is the probability density (*induced* by $\vec{m}$) of the labeled transition. The symbol-labeled transition matrices $\{T_t^{(x)}(\vec{m})\}_{t \in \mathcal{T}, x \in \mathcal{A}}$ yield the net state-to-state transition probabilities when marginalizing over all possible observations:

$$\int_{x \in \mathcal{A}} T_t^{(x)}(\vec{m}) \, dx = T_t(\vec{m})$$

where $\langle s| T_t(\vec{m}) |s'\rangle = \Pr_{\vec{m}}(\mathcal{S}_{t+1} = s'|\mathcal{S}_t = s)$.
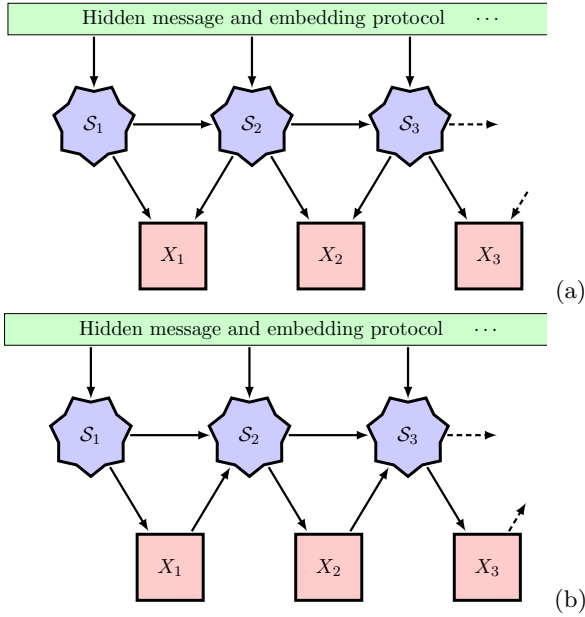

FIG. 10. Alternative Bayesian networks for measurement feedback models.

Figure 10 displays two different (but equally valid) Bayesian networks for the decomposition of conditional dependencies among observables and latent states of a MFM. Each decomposition suggests a preferred interpretation. The decomposition of the top panel (a) allows identifying a PDF with each directed edge between latent states of a measurement feedback model $\mathcal{M}_{\mathrm{MFM}}(\vec{m})$. Accordingly, panel (a) suggests that the transited edge determines the probability of the observable. Whereas, the decomposition of the bottom panel (b) suggests that the observation determines the probability of the latent state transition. The fact that both decompositions are valid insists, perhaps surprisingly, that the interpreta-

tions have no physical distinction. The interpretation of causality is ambiguous although each calculus of conditional dependencies is reliable.

The measurement feedback models may initially appear rather restrictive when considering the possibilities of, say, measuring a quantum system in different bases and with different instruments. However, in principle, the different measurement choices are incorporated through the different transformations $T_t(\vec{m})$, both through any pre-determined measurement choices in $\vec{m}$ and through dynamic-determination via feedback of the measurement outcomes themselves.

Reference [81]'s *process tensors* can also be used to model classical observable processes generated by general quantum dynamics. Although unnecessarily elaborate for most purposes, process tensors are appealing since they rigorously incorporate general quantum measurements. Ultimately though, they, together with a set of "experiments" $\vec{m}$, could be mapped onto the alternative rather-simpler models proposed here, if the goal is only to model the observable classical output process.

2. Theorem 2 for Measurement Feedback

The MFM's average-observation matrices are:

$$\Omega_t = \int_{x \in \mathcal{A}} x \, T_t^{(x)}(\vec{m}) \, dx .$$

Notably, they are no longer diagonal in the hidden-state basis. Rather, they assign to each matrix element the average observation associated with that transition, multiplied by the probability of the edge being traversed when conditioned on occupying the outgoing state. That is:

$$\begin{aligned}
\langle s|\Omega_t |s'\rangle &= \int_{x \in \mathcal{A}} x \, \mathrm{p}_{\vec{m}}(\mathcal{S}_{t+1} = s', X_t = x|\mathcal{S}_t = s) \, dx \\
&= \Pr_{\vec{m}}(\mathcal{S}_{t+1} = s'|\mathcal{S}_t = s) \\
&\quad \times \int_{x \in \mathcal{A}} x \, \mathrm{p}_{\vec{m}}(X_t = x|\mathcal{S}_t = s, \mathcal{S}_{t+1} = s') \, dx \\
&= \langle s| T_t(\vec{m}) |s'\rangle \, \langle x\rangle_{\mathrm{p}_{\vec{m}}(X_t|\mathcal{S}_t=s, \mathcal{S}_{t+1}=s')} .
\end{aligned} \tag{G1}$$

If the process is wide-sense stationary, then for $\tau > 0$:

$$\gamma(\tau) = \langle \boldsymbol{\mu}_t|\overline{\Omega}_t \, T_{t+1:t+\tau}(\vec{m}) \, \Omega_{t+\tau}|\mathbf{1}\rangle , \tag{G2}$$

which must be t-independent.

For input-independent processes with time-independent transition dynamics—where $T_t^{(x)}(\vec{m}) = T^{(x)}$ and $\boldsymbol{\mu}_1 = \boldsymbol{\pi}$—this simplifies to the autonomous Mealy-type HMMs with continuous PDFs for the observable associated with each hidden-state-to-state transition.

The autocorrelation function (for $\tau \geq 1$) then reduces to:

$$\gamma(\tau) = \langle \boldsymbol{\pi} | \overline{\Omega} \, T^{\tau-1} \, \Omega | \mathbf{1} \rangle \ ,$$

while the power spectrum's continuous part is:

$$P_{\mathrm{c}}(\omega) = \langle |x|^2 \rangle + 2 \operatorname{Re} \langle \boldsymbol{\pi} | \, \overline{\Omega} \left(e^{i\omega} I - T \right)^{-1} \Omega | \mathbf{1} \rangle \ . \quad \text{(G3)}$$

Note that this expression lacks T, the transition dynamic, when compared to Eq. (6). This follows since Ω induces a transition for these Mealy-type HMMs, reducing the number of subsequent transitions by one.

Let's return to the general setting for autocorrelation given by Eq. (G2) for processes generated by possibly-input-dependent models. Developing the analog of Thm. 2 requires recognizing that the average observation *on each edge* matters, rather than previously, where the average observation from each state mattered. For MFMs, constant autocorrelation and flat power spectrum can again be guaranteed by a rather weak condition: The average output of the current edge does not by itself influence the average output of a future edge.

More explicitly, consider the set of all edges:

$$\boldsymbol{\mathcal{E}}^{(t)} \equiv \left\{ (s, s') \in \boldsymbol{\mathcal{S}} \times \boldsymbol{\mathcal{S}} : \langle s | \, T_t(\vec{m}) \, | s' \rangle \neq 0 \right\} \ ,$$

which are transitions between hidden states that can be traversed at time t with positive probability. Since outputs occur during edge transitions, we redefine Ξ as the set of average outputs exhibited by the edges. Equa-

tion (G1) indicates that the desired definition is:

$$\Xi \equiv \bigcup_{t \in \mathcal{T}} \bigcup_{(s,s') \in \boldsymbol{\mathcal{E}}^{(t)}} \left\{ \frac{\langle s | \, \Omega_t \, | s' \rangle}{\langle s | \, T_t(\vec{m}) \, | s' \rangle} \right\} \ .$$

Furthermore, we define $\mathcal{E}_t$ to be the random variable for the edge traversed at time t; i.e., $\mathcal{E}_t$ is the joint random variable: $\mathcal{E}_t = (\mathcal{S}_t, \mathcal{S}_{t+1})$. And we define $\boldsymbol{\mathcal{E}}_\xi^{(t)} \subset \boldsymbol{\mathcal{E}}^{(t)}$ as the set of edges (at time t) with average output $\xi \in \Xi$:

$$\boldsymbol{\mathcal{E}}_\xi^{(t)} \equiv \left\{ (s, s') \in \boldsymbol{\mathcal{E}}^{(t)} : \frac{\langle s | \Omega_t | s' \rangle}{\langle s | T_t(\vec{m}) | s' \rangle} = \xi \right\} \ . \quad \text{(G4)}$$

With these in hand, we can state the theorem analogous to Thm. 2.

Theorem 3. *Let $\{X_t\}_t$ be a stochastic process generated by any measurement feedback model $\mathcal{M}_{MFM}(\vec{m})$, including autonomous Mealy-type HMMs and input-dependent generators. Such processes have* constant autocorrelation *and a flat power spectrum* if:

$$\Pr(\mathcal{E}_{t+\tau} \in \boldsymbol{\mathcal{E}}_{\xi'}^{(t+\tau)} | \mathcal{E}_t \in \boldsymbol{\mathcal{E}}_\xi^{(t)}) = \Pr(\mathcal{E}_{t+\tau} \in \boldsymbol{\mathcal{E}}_{\xi'}^{(t+\tau)})$$

and there exists a constant $c \in \mathbb{C}$ such that:

$$\sum_{\xi \in \Xi} \xi \, \Pr(\mathcal{E}_t \in \boldsymbol{\mathcal{E}}_\xi^{(t)}) = c \ ,$$

for all separations $\tau > 0$, $t \in \mathcal{T}$, and $\xi, \xi' \in \Xi$.

Proof. *Starting from Eq. (G2), we find the autocorrelation for all such processes by calculating:*

$$
\begin{aligned}
\gamma(\tau) &= \langle \boldsymbol{\mu}_t | \overline{\Omega}_t \, T_{t+1:t+\tau}(\vec{m}) \, \Omega_{t+\tau} | \mathbf{1} \rangle \\
&= \sum_{s,s',s'',s''' \in \boldsymbol{\mathcal{S}}} \langle \boldsymbol{\mu}_t | s \rangle \langle s | \overline{\Omega}_t | s' \rangle \langle s' | \, T_{t+1:t+\tau}(\vec{m}) \, | s'' \rangle \langle s'' | \Omega_{t+\tau} | s''' \rangle \langle s''' | \mathbf{1} \rangle \\
&= \sum_{\substack{(s,s') \in \boldsymbol{\mathcal{E}}^{(t)} \\ (s',s''') \in \boldsymbol{\mathcal{E}}^{(t+\tau)}}} \left(\frac{\langle s | \overline{\Omega}_t | s' \rangle}{\langle s | T_t(\vec{m}) | s' \rangle} \right) \left(\frac{\langle s'' | \Omega_{t+\tau} | s''' \rangle}{\langle s'' | T_{t+\tau}(\vec{m}) | s''' \rangle} \right) \langle \boldsymbol{\mu}_t | s \rangle \langle s | T_t(\vec{m}) | s' \rangle \langle s' | T_{t+1:t+\tau}(\vec{m}) | s'' \rangle \langle s'' | T_{t+\tau}(\vec{m}) | s''' \rangle \langle s''' | \mathbf{1} \rangle \\
&= \sum_{\substack{(s,s') \in \boldsymbol{\mathcal{E}}^{(t)} \\ (s',s''') \in \boldsymbol{\mathcal{E}}^{(t+\tau)}}} \left(\frac{\langle s | \overline{\Omega}_t | s' \rangle}{\langle s | T_t(\vec{m}) | s' \rangle} \right) \left(\frac{\langle s'' | \Omega_{t+\tau} | s''' \rangle}{\langle s'' | T_{t+\tau}(\vec{m}) | s''' \rangle} \right) \Pr(\mathcal{S}_t = s, \mathcal{S}_{t+1} = s', \mathcal{S}_{t+\tau} = s'', \mathcal{S}_{t+\tau+1} = s''') \\
&= \sum_{\substack{(s,s') \in \boldsymbol{\mathcal{E}}^{(t)} \\ (s',s''') \in \boldsymbol{\mathcal{E}}^{(t+\tau)}}} \left(\frac{\langle s | \overline{\Omega}_t | s' \rangle}{\langle s | T_t(\vec{m}) | s' \rangle} \right) \left(\frac{\langle s'' | \Omega_{t+\tau} | s''' \rangle}{\langle s'' | T_{t+\tau}(\vec{m}) | s''' \rangle} \right) \Pr\big(\mathcal{E}_t = (s, s'), \mathcal{E}_{t+\tau} = (s'', s''')\big) \\
&= \sum_{\xi,\xi' \in \Xi} \overline{\xi} \xi' \, \Pr\big(\mathcal{E}_t \in \boldsymbol{\mathcal{E}}_\xi^{(t)}, \mathcal{E}_{t+\tau} \in \boldsymbol{\mathcal{E}}_{\xi'}^{(t+\tau)}\big) \\
&= \sum_{\xi \in \Xi} \overline{\xi} \, \Pr\big(\mathcal{E}_t \in \boldsymbol{\mathcal{E}}_\xi^{(t)}\big) \Big[\sum_{\xi' \in \Xi} \xi' \, \Pr\big(\mathcal{E}_{t+\tau} \in \boldsymbol{\mathcal{E}}_{\xi'}^{(t+\tau)} | \mathcal{E}_t \in \boldsymbol{\mathcal{E}}_\xi^{(t)}\big) \Big] \ .
\end{aligned}
$$

Now, suppose that:

$$\Pr(\mathcal{E}_{t+\tau} \in \boldsymbol{\mathcal{E}}_{\xi'}^{(t+\tau)}|\mathcal{E}_t \in \boldsymbol{\mathcal{E}}_\xi^{(t)}) = \Pr(\mathcal{E}_{t+\tau} \in \boldsymbol{\mathcal{E}}_{\xi'}^{(t+\tau)})$$

and there exists some constant $c \in \mathbb{C}$ such that:

$$\sum_{\xi \in \Xi} \xi \, \Pr(\mathcal{E}_t \in \boldsymbol{\mathcal{E}}_\xi^{(t)}) = c \; ,$$

for all separations $\tau > 0$, $t \in \mathcal{T}$, and $\xi, \xi' \in \Xi$. Then, we find:

$$\gamma(\tau) = \sum_{\xi \in \Xi} \overline{\xi} \Pr(\mathcal{E}_t \in \boldsymbol{\mathcal{E}}_\xi^{(t)})$$
$$\times \left[\sum_{\xi' \in \Xi} \xi' \Pr(\mathcal{E}_{t+\tau} \in \boldsymbol{\mathcal{E}}_{\xi'}^{(t+\tau)}|\mathcal{E}_t \in \boldsymbol{\mathcal{E}}_\xi^{(t)}) \right]$$
$$= \left(\sum_{\xi \in \Xi} \overline{\xi} \Pr(\mathcal{E}_t \in \boldsymbol{\mathcal{E}}_\xi^{(t)}) \right)\left(\sum_{\xi' \in \Xi} \xi' \Pr(\mathcal{E}_{t+\tau} \in \boldsymbol{\mathcal{E}}_{\xi'}^{(t+\tau)}) \right)$$
$$= |c|^2 \; ,$$

which is a constant for all separations $\tau > 0$, $t \in \mathcal{T}$, and $\xi, \xi' \in \Xi$. Finally, a process with stationary low-order statistics and a flat autocorrelation has a flat power spectrum, as an immediate consequence of Eq. (3). This proves Thm. 3.

For the special case of an autonomous HMM that generates observations during hidden-state-to-state transitions, this condition simplifies significantly. Specifically, $\Omega_t \to \Omega$ and $T_t(\vec{m}) \to T$ become t-independent, which furthermore means that $\boldsymbol{\mathcal{E}}_\xi^{(t)} \to \boldsymbol{\mathcal{E}}_\xi$ becomes t-independent. For autonomous wide-sense stationary processes, we have $\Pr(\mathcal{E}_t) = \Pr(\mathcal{E}_{t+\tau})$ for all separations $\tau > 0$ and for all $t \in \mathcal{T}$. It then trivially follows that $\sum_{\xi \in \Xi} \xi \Pr(\mathcal{E}_t \in \boldsymbol{\mathcal{E}}_\xi)$ is constant for all $t \in \mathcal{T}$. So, the only requirement for an autonomous edge-emitting HMM to produce fraudulent white noise is that it satisfies the condition:

$$\Pr(\mathcal{E}_{t+\tau} \in \boldsymbol{\mathcal{E}}_{\xi'}|\mathcal{E}_t \in \boldsymbol{\mathcal{E}}_\xi) = \Pr(\mathcal{E}_{t+\tau} \in \boldsymbol{\mathcal{E}}_{\xi'})$$

for all separations $\tau > 0$, $t \in \mathcal{T}$, and $\xi, \xi' \in \Xi$.

Theorem 3 provides a very general condition for flat power spectra from measurement feedback models.

Appendix H: Theorem 2 for time-dependent PDFs

Moreover, Thm. 3 suggests how Thm. 2 generalizes even further to possibly-input-dependent hidden-state models with time-dependent PDFs associated with each state. We will call these *morphing hidden models* (MHMs) $\mathcal{M}_{\mathrm{MHM}}(\vec{m})$. MHMs include, as special cases, all

models (Moore-type HMMs and input-dependent generators) considered in the main text. We employ methods similar to those used in § G 2.

A MHM is a possibly-input-dependent generator of an observable output process $\{X_t\}_{t \in \mathcal{T}}$. The output is generated via $\mathcal{M}_{\mathrm{MHM}}(\vec{m}) = \left(\boldsymbol{\mathcal{S}}, \mathcal{A}, \{\mathcal{P}_t(\vec{m})\}_t, \{T_t(\vec{m})\}_t, \boldsymbol{\mu}_1 \right)$. Here, again, the lengths and alphabets of the inputs and outputs need not be commensurate. That is, the internal states $\boldsymbol{\mathcal{S}}$ and output alphabet $\mathcal{A}$ are static. However, the hidden-state-to-state transition matrix $T_t(\vec{m})$—as well as the state-dependent PDFs $\mathcal{P}_t(\vec{m})$—are time-dependent such that their values at time t are potentially a function of the full input vector $\vec{m}$. More specifically, $\mathcal{P}_t(\vec{m})$ is the set of hidden-state-dependent probability density functions $\mathrm{p}_{\vec{m}}(X_t|s)$ at time t. As before, $\boldsymbol{\mu}_1$ specifies the initial distribution over hidden states: $\mathcal{S}_1 \sim \boldsymbol{\mu}_1$.

For such cases, set:

$$\Omega_t = \sum_{s \in \boldsymbol{\mathcal{S}}} \langle x \rangle_{\mathrm{p}_{\vec{m}}(X_t|s)} |s\rangle \langle s| \; .$$

The Ω_t matrix is time-dependent with the state-conditioned expected outputs along its diagonal.

Since the average state output now varies in time, we must generalize Ξ from its more restricted use in the main text. Specifically, redefine Ξ as the set of state-dependent average outputs generated throughout time:

$$\Xi \equiv \bigcup_{t \in \mathcal{T}} \bigcup_{s \in \boldsymbol{\mathcal{S}}} \left\{ \langle x \rangle_{\mathrm{p}_{\vec{m}}(X_t|s)} \right\} \; .$$

Furthermore, we define $\boldsymbol{\mathcal{S}}_\xi^{(t)} \subset \boldsymbol{\mathcal{S}}$ as the set of states (at time t) with average output $\xi \in \Xi$:

$$\boldsymbol{\mathcal{S}}_\xi^{(t)} \equiv \left\{ s \in \boldsymbol{\mathcal{S}} : \langle x \rangle_{\mathrm{p}_{\vec{m}}(X_t|s)} = \xi \right\} \; .$$

Using these, we can state the following theorem, which generalizes Thm. 2.

Theorem 4. *Let $\{X_t\}_t$ be a stochastic process generated by any morphing hidden model $\mathcal{M}_{MHM}(\vec{m})$. Such processes have* constant *autocorrelation and a flat power spectrum if:*

$$\Pr(\mathcal{S}_{t+\tau} \in \boldsymbol{\mathcal{S}}_{\xi'}^{(t+\tau)}|\mathcal{S}_t \in \boldsymbol{\mathcal{S}}_\xi^{(t)}) = \Pr(\mathcal{S}_{t+\tau} \in \boldsymbol{\mathcal{S}}_{\xi'}^{(t+\tau)})$$

and there exists a constant $c \in \mathbb{C}$ such that:

$$\sum_{\xi \in \Xi} \xi \, \Pr(\mathcal{S}_t \in \boldsymbol{\mathcal{S}}_\xi^{(t)}) = c \; ,$$

for all separations $\tau > 0$, $t \in \mathcal{T}$, and $\xi, \xi' \in \Xi$.

Proof. *For the processes under consideration, we find the linear pairwise correlation (for $\tau \geq 1$) to be:*

$$\langle \overline{X_t} X_{t+\tau} \rangle_{p_{\vec{m}}(X_t, X_{t+\tau})} = \langle \boldsymbol{\mu}_t | \overline{\Omega}_t \, T_{t:t+\tau}(\vec{m}) \, \Omega_{t+\tau} | \mathbf{1} \rangle$$

$$= \sum_{s,s' \in \boldsymbol{S}} \langle \boldsymbol{\mu}_t | s \rangle \langle s | \overline{\Omega}_t | s \rangle \langle s | T_{t:t+\tau}(\vec{m}) | s' \rangle \langle s' | \Omega_{t+\tau} | s' \rangle \langle s' | \mathbf{1} \rangle$$

$$= \sum_{\xi,\xi' \in \Xi} \overline{\xi} \xi' \sum_{\substack{s \in \boldsymbol{S}_\xi^{(t)} \\ s' \in \boldsymbol{S}_{\xi'}^{(t+\tau)}}} \Pr_{\vec{m}}(\mathcal{S}_t = s, \mathcal{S}_{t+\tau} = s')$$

$$= \sum_{\xi,\xi' \in \Xi} \overline{\xi} \xi' \Pr(\mathcal{S}_t \in \boldsymbol{S}_\xi^{(t)}, \mathcal{S}_{t+\tau} \in \boldsymbol{S}_{\xi'}^{(t+\tau)})$$

$$= \sum_{\xi \in \Xi} \overline{\xi} \Pr(\mathcal{S}_t \in \boldsymbol{S}_\xi^{(t)}) \Big(\sum_{\xi' \in \Xi} \xi' \Pr(\mathcal{S}_{t+\tau} \in \boldsymbol{S}_{\xi'}^{(t+\tau)} | \mathcal{S}_t \in \boldsymbol{S}_\xi^{(t)}) \Big) \,. \tag{H1}$$

Now, suppose that:

$$\Pr(\mathcal{S}_{t+\tau} \in \boldsymbol{S}_{\xi'}^{(t+\tau)} | \mathcal{S}_t \in \boldsymbol{S}_\xi^{(t)}) = \Pr(\mathcal{S}_{t+\tau} \in \boldsymbol{S}_{\xi'}^{(t+\tau)})$$

and there exists some constant $c \in \mathbb{C}$ such that:

$$\sum_{\xi \in \Xi} \xi \, \Pr(\mathcal{S}_t \in \boldsymbol{S}_\xi^{(t)}) = c$$

for all separations $\tau > 0$, for all $t \in \mathcal{T}$, and for all $\xi, \xi' \in \Xi$. Then, we find:

$$\langle \overline{X_t} X_{t+\tau} \rangle_{p_{\vec{m}}(X_t, X_{t+\tau})}$$
$$= \sum_{\xi \in \Xi} \overline{\xi} \Pr(\mathcal{S}_t \in \boldsymbol{S}_\xi^{(t)}) \Big(\sum_{\xi' \in \Xi} \xi' \Pr(\mathcal{S}_{t+\tau} \in \boldsymbol{S}_{\xi'}^{(t+\tau)} | \mathcal{S}_t \in \boldsymbol{S}_\xi^{(t)}) \Big)$$
$$= \Big(\sum_{\xi \in \Xi} \overline{\xi} \Pr(\mathcal{S}_t \in \boldsymbol{S}_\xi^{(t)}) \Big) \Big(\sum_{\xi' \in \Xi} \xi' \Pr(\mathcal{S}_{t+\tau} \in \boldsymbol{S}_{\xi'}^{(t+\tau)}) \Big)$$
$$= |c|^2 \,.$$

is constant for all $t \in \mathcal{T}$, and $\forall \xi, \xi' \in \Xi$.

That $\langle \overline{X_t} X_{t+\tau} \rangle_{p_{\vec{m}}(X_t, X_{t+\tau})}$ is constant verifies that the autocorrelation does not depend on the overall time shift of the process, so $\langle \overline{X_t} X_{t+\tau} \rangle_{p_{\vec{m}}(X_t, X_{t+\tau})} = \gamma(\tau)$. Moreover, $\gamma(\tau)$ is constant. Finally, a process with constant autocorrelation has a flat power spectrum, as an immediate consequence of Eq. (3). This proves Thm. 4.

[1] J. W. Tukey. The future of data analysis. *Ann. Math. Stat.*, 33, 1962. 1, 15

[2] P. M. Riechers, J. P. Crutchfield, and M. Gu. Lifting correlation to hide all structure beyond any level of scrutiny. (in preparation), 2019. 2, 9, 11, 15

[3] R.K. Pathria and P.D. Beale. *Statistical Mechanics*. Elsevier Science, 2011. 2

[4] R. Kubo. The fluctuation–dissipation theorem. *Reports on progress in physics*, 29(1):255, 1966. 2

[5] M. Planck. On the law of distribution of energy in the normal spectrum. *Verhandl. Dtsch. Phys. Ges.*, 2:237–245, 1900. 2

[6] A. Einstein. Über einen die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt. *Annalen der physik*, 322(6):132–148, 1905.

[7] M. J. Klein. Max planck and the beginnings of the quantum theory. *Archive for History of Exact Sciences*, 1(5):459–479, Oct 1961. 2

[8] Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, and S. Basak. Planck 2018 results. I. Overview and the cosmological legacy of planck. *arXiv:1807.06205*. 2

[9] P. Holmes. 'Strange' phenomena in dynamical systems and their physical implications. *Applied Mathematical Modelling*, 1(7):362–366, 1977. 2, 15

[10] J. D. Farmer, J. P. Crutchfield, H. Froehling, N. H. Packard, and R. S. Shaw. Power spectra and mixing properties of strange attractors. *Ann. New York Acad. Sci.*, 357:453, 1980. 2, 15, 21

[11] M. Pollicott. On the rate of mixing of Axiom A flows. *Inventiones mathematicae*, 81(3):413–426, 1985. 2, 14

[12] D. Ruelle. Resonances of chaotic dynamical systems. *Phys. Rev. Let.*, 56(5):405, 1986.

[13] P. Gaspard. *Chaos, scattering and statistical mechanics*, volume 9. Cambridge University Press, 2005. 2, 14

[14] K. Pance, W. Lu, and S. Sridhar. Quantum fingerprints of classical Ruelle-Pollicott resonances. *Phys. Rev. Lett.*, 85:2737–2740, Sep 2000. 2, 15

[15] I. García-Mata and M. Saraceno. Spectral properties and classical decays in quantum open systems. *Physical Review E*, 69(5):056211, 2004. 2, 15

[16] M. D. Chekroun, J. D. Neelin, D. Kondrashov, J. C. McWilliams, and M. Ghil. Rough parameter dependence in climate models and the role of Ruelle-Pollicott resonances. *Proc. Natl. Acad. Sci. USA*, 111(5):1684–1690, 2014. 2, 15

[17] J. D. Watson and F. H. C. Crick. A structure for deoxyribose nucleic acid. *Nature*, 171(3):737–738, 1953. 2

[18] M. H. F. Wilkins, H. R. Wilson, and A. R. Stokes. Molecular structure of deoxypentose nucleic acids. *Nature*, 171(3):738–740, 1953.

[19] R. E. Franklin and R. G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171(3):740–741, 1953.

[20] J. D. Watson and F. H. C. Crick. Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171(4361):964–967, 1953. 2

[21] W. Bair, C. Koch, W. Newsome, and K. Britten. Power spectrum analysis of bursting cells in area mt in the behaving monkey. *Journal of Neuroscience*, 14(5):2870–2892, 1994. 2

[22] F. Farkhooi, M. F. Strube-Bloss, and M. P. Nawrot. Serial correlation in neural spike trains: Experimental evidence, stochastic modeling, and single neuron variability. *Physical Review E*, 79(2):021905, 2009.

[23] B. Dummer, S. Wieland, and B. Lindner. Self-consistent determination of the spike-train power spectrum in a neural network with sparse connectivity. *Frontiers in Computational Neuroscience*, 8:104, 2014.

[24] R. F. O. Pena, S. Vellmer, D. Bernardi, A. C. Roque, and B. Lindner. Self-consistent scheme for spike-train power spectra in heterogeneous sparse networks. *Frontiers in Computational Neuroscience*, 12:9, 2018. 2

[25] M. J. Chacron, L. Maler, and J. Bastian. Electroreceptor neuron dynamics shape information transmission. *Nature neuroscience*, 8(5):673, 2005. 2

[26] T. Gollisch and M. Meister. Rapid neural coding in the retina with relative spike latencies. *Science*, 319(5866):1108–1111, 2008. 2

[27] P. Mellado, J. Diedler, and T. Steiner. Continuous EEG monitoring in the ICU. In *Intensive Care Medicine*, pages 693–704. Springer, 2007. 2

[28] J. Claassen, F. S. Taccone, P. Horn, M. Holtkamp, N. Stocchetti, and M. Oddo. Recommendations on the use of EEG monitoring in critically ill patients: Consensus statement from the neurointensive care section of the ESICM. *Intensive Care Medicine*, 39(8):1337–1351, 2013.

[29] W. O. Tatum. Ellen R. Grass lecture: Extraordinary EEG. *The Neurodiagnostic Journal*, 54(1):3–21, 2014. 2

[30] R. B. Blackman and J. W. Tukey. *The Measurement of Power Spectra—From the Point of View of Communications Engineering*. Dover, New York, New York, 1958. 3

[31] P. Stoica and R. L. Moses. *Spectral Analysis of Signals*. Pearson Prentice Hall, Upper Saddle River, New Jersey, 2005. 3

[32] For wide-sense stationary processes that are also ergodic, the expected value over realizations $\gamma(\tau) = \left\langle \overline{X_t} X_{t+\tau} \right\rangle_{\mathrm{p}(X_t, X_{t+\tau})}$ coincides with the time-average $\left\langle \overline{x}_t x_{t+\tau} \right\rangle_t$ over a single realization. In practice, scientists and engineers typically perform this latter time-average to obtain the autocorrelation function $\gamma(\tau) = \left\langle \overline{X_t} X_{t+\tau} \right\rangle_{\mathrm{p}(X_t, X_{t+\tau})} = \left\langle \overline{x}_t x_{t+\tau} \right\rangle_t$ from experimental data. 3

[33] N. Wiener. Generalized harmonic analysis. *Acta Math.*, 55:117–258, 1930. 3

[34] A. Khintchine. Korrelationstheorie der stationären stochastischen prozesse. *Mathematische Annalen*, 109(1):604–615, 1934. 3

[35] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, United Kingdom, tenth anniversary edition, 2011. 3

[36] A. Acín and L. Masanes. Certified randomness in quantum physics. *Nature*, 540(7632):213, 2016. 3

[37] L. Shen, J. Lee, L. P. Thinh, J.D. Bancal, A. Cerè, A. Lamas-Linares, A. Lita, T. Gerrits, S. W. Nam, V. Scarani, and C. Kurtsiefer. Randomness extraction from Bell violation with continuous parametric down-conversion. *Phys. Rev. Lett.*, 121:150402, Oct 2018. 3

[38] Y. Liu, Q. Zhao, M.H. Li, J.Y. Guan, Y. Zhang, B. Bai, W. Zhang, W.Z. Liu, C. Wu, X. Yuan, et al. Device-independent quantum random-number generation. *Nature*, page 1, 2018. 3

[39] The noise standard deviation being proportional to the square root of the timestep—$\sigma_\eta \propto \sqrt{\Delta t}$—rather than proportional to the timestep itself is familiar in numerically integrating Langevin and other stochastic differential equations. Though sometimes met with confusion, the dependence is a direct consequence of the central limit theorem since many independent noise contributions accumulate within each timestep and the variance of the accumulated value is proportional to the number of contributions. The simplest stochastic integral—integrating Gaussian white noise—produces the famous Wiener process W_t which is the canonical model for Brownian motion. Conversely, regularly sampling changes in the Wiener state produces Gaussian white noise: $W_{t+\Delta t} - W_t \sim \mathcal{N}(0, \Delta t)$. 4

[40] We choose to ignore the power spectrum's δ-function at zero frequency when classifying white noise, since it merely corresponds to a constant (DC) offset of the signal. Naturally, we make an exception for *completely* deterministic DC signals (with no noise at all), since these should not be considered among the white noise processes. 4

[41] A. Paz. *Introduction to Probabilistic Automata*. Academic Press, New York, 1971. 5

[42] L. R. Rabiner and B. H. Juang. An introduction to hidden Markov models. *IEEE ASSP Magazine*, January, 1986.

[43] L. R. Rabiner. A tutorial on hidden Markov models and selected applications. *IEEE Proc.*, 77:257, 1989.

[44] R. J. Elliot, L. Aggoun, and J. B. Moore. *Hidden Markov Models: Estimation and Control*, volume 29 of *Applications of Mathematics*. Springer, New York, 1995.

[45] Y. Ephraim and N. Merhav. Hidden Markov processes. *IEEE Trans. Info. Th.*, 48(6):1518–1569, 2002.

[46] 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.

[47] J. Bechhoefer. Hidden Markov models for stochastic thermodynamics. *New. J. Phys.*, 17:075003, 2015. 5

[48] 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. 5

[49] B. O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931. 5

[50] J. P. Crutchfield and N. H. Packard. Symbolic dynamics of noisy chaos. *Physica*, 7D(1-3):201 – 223, 1983. 13

[51] H. Jaeger. Observable operator models for discrete stochastic time series. *Neural computation*, 12(6):1371–1398, 2000.

[52] A. M. Fraser. *Hidden Markov models and dynamical systems*, volume 107. Siam, 2008.

[53] P. M. Riechers and J. P. Crutchfield. Spectral simplicity of apparent complexity, Part I: The nondiagonalizable metadynamics of prediction. *Chaos*, 28:033115, 2018. 13

[54] P. M. Riechers and J. P. Crutchfield. Spectral simplicity of apparent complexity, Part II: Exact complexities and complexity spectra. *Chaos*, 28:033116, 2018. 8, 10, 11, 13

[55] P. M. Riechers and J. P. Crutchfield. Beyond the spectral theorem: spectrally decomposing arbitrary functions of nondiagonalizable operators. *AIP Advances*, 8(6):065305, 2018. 5, 6, 7, 14, 15, 17, 19

[56] P. M. Riechers, D. P. Varn, and J. P. Crutchfield. Diffraction patterns of layered close-packed structures from hidden Markov models. arxiv.org:1410.5028. 8, 16, 17

[57] E. Narevicius, P. Serra, and N. Moiseyev. Critical phenomena associated with self-orthogonality in non-Hermitian quantum mechanics. *EPL (Europhys. Lett.)*, 62(6):789, 2003. 8

[58] A. V. Sokolov, A. A. Andrianov, and F. Cannata. Non-Hermitian quantum mechanics of non-diagonalizable Hamiltonians: Puzzles with self-orthogonal states. *J. Physics A*, 39(32):10207, 2006.

[59] A. Mostafazadeh. Spectral singularities of complex scattering potentials and infinite reflection and transmission coefficients at real energies. *Phys. Rev. Lett.*, 102:220402, 2009. 8

[60] Indeed, the burden of resolving this phase ambiguity ren-

ders the RRXOR process infinite Markov order. 11

[61] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Let.*, 63:105–108, 1989. 11

[62] J. P. Crutchfield. Between order and chaos. *Nature Physics*, 8(January):17–24, 2012. 11, 22

[63] N. Travers and J. P. Crutchfield. Infinite excess entropy processes with countable-state generators. *Entropy*, 16:1396–1413, 2014. 11

[64] N. Barnett and J. P. Crutchfield. Computational mechanics of input–output processes: Structured transformations and the ϵ-transducer. *J. Stat. Physics*, 161(2):404–451, 2015. 12, 22

[65] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948. 13

[66] A. del Junco and M. Rahe. Finitary codings and weak bernoulli partitions. *Proc. AMS*, 75:259, 1979.

[67] J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: Levels of entropy convergence. *CHAOS*, 13(1):25–54, 2003. 13

[68] M. Budii, R. Mohr, and I. Mezic. Applied Koopmanism. *Chaos*, 22(4), 2012. 15

[69] J. P. Crutchfield. The hidden fragility of complex systems: Consequences of change, changing consequences. In P. Alsina and J. Perello, editors, *Cultures of Change — Changing Cultures*, pages 98–111, Barcelona, Spain, 2009. ACTAR Publishers. 15

[70] A. Guinier. *X-Ray Diffraction in Crystals, Imperfect Crystals and Amorphous Bodies*. W. H. Freeman and Company, San Francisco, 1963. 15

[71] B. E. Warren. *X-Ray Diffraction*. Addison-Wesley, New York, 1969.

[72] T. R. Welberry. Diffuse x-ray scattering and models of disorder. *Rep. Prog. Phys.*, 48:1543, 1985.

[73] N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. Sanders College Publishing, New York, 1976.

[74] T. R. Welberry. One hundred years of diffuse X-ray scattering. *Metallurgical and Materials Transactions A*, 45, 2014. 15

[75] E. Prince, editor. *International Tables for Crystallography Vol. C, Mathematical, Physical and Chemical Tables*. Springer, 2004. 16

[76] M. M. Woolfson. *An introduction to X-ray crystallography*. Cambridge University Press, 1997. 16

[77] A. V. Oppenheim and R. W. Schafer. *Discrete-time signal processing*. Pearson Education, 2014. 19

[78] J. S. Bendat and A. G. Piersol. Engineering applications of correlation and spectral analysis. *New York, Wiley-Interscience, 1980. 315 p.*, 1980. 20

[79] J. S. Bendat and A. G. Piersol. *Random data: analysis and measurement procedures*, volume 729. John Wiley & Sons, 2011. 20

[80] One can choose the simplest branch of the complex logarithm since the apparent freedom offered through its different branches correspond to unobserved cyclic behaviors at inaccessibly small timescales that, if they exist, would be aliased by the sampling frequency f_0. In other words, the branch freedom is a gauge freedom supporting all dynamics that could exist but is suppressed by the

mapping $G \mapsto e^{\tau_0 G}$. 21

[81] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi. Non-Markovian quantum processes: Complete framework and efficient characterization. *Phys. Rev. A*, 97:012127, Jan 2018. 23