Mutual information identifies spurious Hurst phenomena in resting state EEG and fMRI data

Frederic von Wegner,[1,*,†] Helmut Laufs,[1,2,†] and Enzo Tagliazucchi[1]

Department of Neurology and Brain Imaging Center, Goethe University, Schleusenweg 2-16, 60528 Frankfurt am Main, Germany
Department of Neurology, University Hospital Kiel, Arnold-Heller-Straße 3, 24105 Kiel, Germany

Long-range memory in time series is often quantified by the Hurst exponent H, a measure of the signal's variance across several time scales. We analyze neurophysiological time series from electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) resting state experiments with two standard Hurst exponent estimators and with the time-lagged mutual information function applied to discretized versions of the signals. A confidence interval for the mutual information function is obtained from surrogate Markov processes with equilibrium distribution and transition matrix identical to the underlying signal. For EEG signals, we construct an additional mutual information confidence interval from a short-range correlated, tenth-order autoregressive model. We reproduce the previously described Hurst phenomenon ($H > 0.5$) in the analytical amplitude of alpha frequency band oscillations, in EEG microstate sequences, and in fMRI signals, but we show that the Hurst phenomenon occurs without long-range memory in the information-theoretical sense. We find that the mutual information function of neurophysiological data behaves differently from fractional Gaussian noise (fGn), for which the Hurst phenomenon is a sufficient condition to prove long-range memory. Two other well-characterized, short-range correlated stochastic processes (Ornstein-Uhlenbeck, Cox-Ingersoll-Ross) also yield $H > 0.5$, whereas their mutual information functions lie within the Markovian confidence intervals, similar to neural signals. In these processes, which do not have long-range memory by construction, a spurious Hurst phenomenon occurs due to slow relaxation times and heteroscedasticity (time-varying conditional variance). In summary, we find that mutual information correctly distinguishes long-range from short-range dependence in the theoretical and experimental cases discussed. Our results also suggest that the stationary fGn process is not sufficient to describe neural data, which seem to belong to a more general class of stochastic processes, in which multiscale variance effects produce Hurst phenomena without long-range dependence. In our experimental data, the Hurst phenomenon and long-range memory appear as different system properties that should be estimated and interpreted independently.

I. INTRODUCTION

Complex physical and biological systems display a multitude of interesting features that are markedly different from standard equilibrium mechanics [1,2]. Long-range memory in physical and biological systems has been studied extensively. There is a close connection between physical systems at a critical point and long-range correlations, whose properties have been reviewed in detail [3]. A range of mechanisms leading to long-range correlated fluctuations in condensed matter, and conducting materials in particular, has been summarized in Ref. [1]. Mechanisms discussed there include transport effects, "smeared" or fractional kinetics, and the superposition of non-Gaussian processes. Around the same time, the concept of self-organized criticality has been introduced [4], establishing an important connection to living systems. This mechanism has been repeatedly applied to explain long-range dependence (LRD) in neural systems [5]. In neural systems we can find, for instance, non-Gaussian and self-similar distributions of certain quantities [6], long-range correlated time series [5],

fractal dimensions of function graphs in the plane [7,8], and the Hurst phenomenon, i.e., a scaling law for the local variance across different time scales. For each of these features, there is a clear mathematical definition and several numerical methods for estimation in empirical data are readily available [8–11]. Historically, the aim to find a parsimonious model that explains the Hurst phenomenon observed in river Nile levels gave rise to the notions of fractional Brownian motion (fBm) and its first-order differenced process, fractional Gaussian noise (fGn) [12,13]. This approach reduces model complexity drastically, summarizing all multi-scale features in a single parameter H, the Hurst exponent. Interestingly, for fBm and fGn, several of the above-mentioned properties coincide and can be parametrized by H. Thus, fBm is a nonstationary Gaussian process described by a self-similar probability density with scaling exponent H and fractal dimension $d = 2 - H$ [14]. Long-range correlations are an attribute of the fGn process, given by the increments of fBm. fGn is a stationary Gaussian long-range correlated process whose autocovariance function decays as a power law, parametrized by the Hurst exponent H.

For experimental data, the relations between the aforementioned features are more complicated. In the general case, where time-stationarity may not hold, a Hurst exponent of

$H > 0.5$ does not strictly imply long-range dependence or scaling laws [11,15–18]. Both features, the Hurst phenomenon on one side and long-range correlations on the other, should therefore be considered separately [10,11]. A highly interdisciplinary and critical exposition of true and spurious long-range memory mechanisms, ranging from elementary physics to the social sciences, is given in Ref. [19].

We here aim to analyze temporal dependencies in neurophysiological signals, namely electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) data recorded during resting state experiments of healthy subjects. The goal is to assess memory effects in these experimental signals using well-known Hurst exponent estimators, as well as using information theoretical measures. To this end, the state space of continuous, real-valued time series is partitioned into a set of discretized states, and the time-lagged mutual information function of the discretized process is compared to a Markovian null hypothesis. Comparing our results with well-characterized stochastic processes whose correlation properties are known analytically, we aim to distinguish Hurst phenomena due to long-range memory from spurious Hurst phenomena induced by slow exponential relaxation and non-stationarity.

The paper is structured as follows. First, we present three model systems whose correlation properties are well known. The models are the stationary, long-range correlated fractional Gaussian noise process (fGn, [12]), the stationary, exponentially correlated Ornstein-Uhlenbeck process (OU, [20]), and the heteroscedastic, short-range correlated Cox-Ingersoll-Ross process (CIR, [21]). We show that all three processes create the Hurst phenomenon even though two models (OU, CIR) are actually short-range correlated. We then analyze the shape of the mutual information function for the three model processes and contrast it with a first-order Markovian null hypothesis. We apply the same technique to resting state electroencephalographic (EEG) and to functional magnetic resonance imaging (fMRI) data, for which the Hurst phenomenon and long-range correlations have been postulated [5,22–27]. For the EEG alpha rhythm, we also compute a mutual information confidence interval from a more flexible tenth-order autoregressive model, also termed AR(10), fitted to the experimental data set [28,29].

II. MODELS AND DATA

To illustrate different paths leading to the Hurst phenomenon, we consider three well defined stochastic processes, prior to investigating EEG and fMRI data for which the Hurst phenomenon and long-range dependence (LRD) have been previously reported.

A. Fractional Gaussian noise

Fractional Gaussian noise is a stationary, long-range (power-law) correlated Gaussian process, its integral representing the famous fractional Brownian motion [12]. An excellent review of the developments that led to the concepts of long-range dependence, fractals, and self-similarity can be found in Ref. [13]. Long-range correlations are quantified by the parameter $H \in [0.5,1)$, where $H = 0.5$ yields an uncorrelated process and $0.5 < H < 1.0$ produces long-range
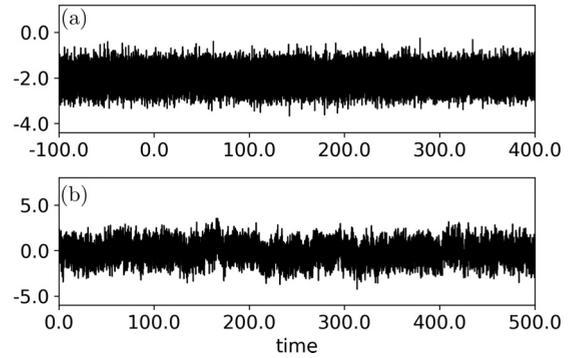

FIG. 1. Sample paths of the fractional Gaussian noise process. (a) Sample path for the uncorrelated, white noise-like case, $H = 0.5$. (b) Sample path showing long-range correlations, $H = 0.9$. Both samples are of length 50 000 and were synthesized using the covariance matrix method.

correlations. Fractional Gaussian noise can be defined via its integral (fractional Brownian motion, fBm) [12]:

$$B_H(t) = B_H(0) + \frac{1}{C} \int_{-\infty}^{\infty} \left[(t-s)_+^{H-\frac{1}{2}} - (-s)_+^{H-\frac{1}{2}} \right] dW_s,$$
(1)

where dW_s is a standard Brownian motion and $C = \Gamma(H + \frac{1}{2})$ is the normalization constant. Alternatively, fBm is defined via its autocorrelation function $\gamma(k)$ for $k > 0$ [11]:

$$\gamma(k) = \frac{1}{2}(\mid k+1 \mid^{2H} - 2 \mid k \mid^{2H} + \mid k-1 \mid^{2H}).$$
(2)

Sample paths were obtained numerically using circulant embedding of the covariance matrix [30]. We simulated fGn samples for Hurst exponents ranging from $H = 0.5$ to $H = 0.95$ using a step size of $\Delta H = 0.05$. Figure 1 shows fGn sample paths for $H = 0.5$ (a) and $H = 0.9$ (b). For $H = 0.5$, fGn is equivalent to uncorrelated Gaussian noise, whereas for $H = 0.9$, long-range correlations are generated by multi-scale fluctuations extending up to the signal length.

B. Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck (OU) process is the minimal representation of a stationary, Gaussian stochastic process with exponential, i.e., short-range autocorrelations. The autocorrelations decay exponentially with a time constant τ, the relaxation time of the process [31]. The process is mean-reverting with respect to the average value $\mathbb{E}(X) = \mu$. The OU process is integrated with respect to a standard Brownian motion dW_t and can be defined as

$$dX_t = -\frac{1}{\tau}(X_t - \mu) + \sigma dW_t.$$
(3)

As the autocorrelation coefficients of the process decay exponentially as $\gamma(k) = \frac{\tau}{2} \exp(-\frac{k}{\tau})$, the sequence of coefficients is integrable and the process is therefore short-range correlated. Sample paths were synthesized with Gillespie's method [31], using a time step of $dt = 0.01$. We used relaxation times $\tau = 0.1, 0.5, 1.0, 5.0, 10.0$ and constant variance $\sigma^2 = 1$. As the process is short-range correlated, the expected Hurst exponent for large samples is $H = 0.5$.
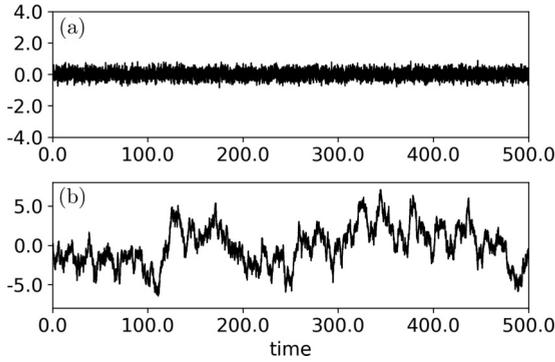
FIG. 2. Sample paths of stationary, exponentially correlated Gaussian noise, the Ornstein-Uhlenbeck process. (a) Sample path for the shortest relaxation time $\tau = 0.1$ resembling white noise. (b) A sample for the largest relaxation time $\tau = 10.0$ shows significant intermediate and low-frequency content, leading to variance contributions across many time scales in Hurst analyses.

Figure 2 shows two sample paths for different relaxation times τ. Figure 2(a) shows a sample with a short correlation time $\tau = 0.1$, similar to Gaussian white noise. The OU process sample shown in Fig. 2(b) has relaxation time $\tau = 10.0$ and shows large amplitude, slow fluctuations, increasing the autocorrelation time. For any parameter τ in the OU process, however, the autocorrelation function decays exponentially, and thus the process is short-range correlated by definition.

C. Cox-Ingersoll-Ross process

The Cox-Ingersoll-Ross (CIR) process generalizes the Ornstein-Uhlenbeck process by introducing heteroscedasticity, i.e., a time-varying conditional variance term. In the special case of the CIR process, the conditional variance term is proportional to $\sqrt{X_t}$. In contrast to OU sample paths, CIR samples show irregular bursts of large-amplitude activity separated by episodes of low amplitude. Figure 3(a) shows a sequence of short bursts, induced by a short relaxation time $\tau = 0.1$. In Fig. 3(b), using $\tau = 10.0$, we observe longer lasting bursts,
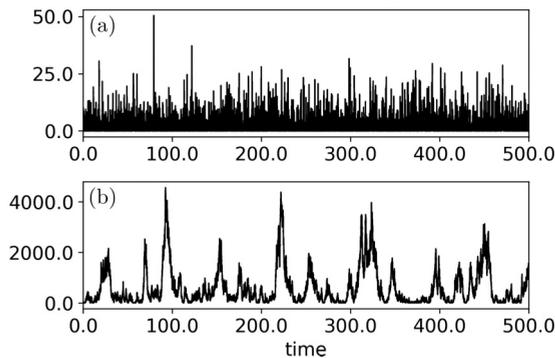

FIG. 3. Sample paths of the heteroscedastic, short-range correlated Cox-Ingersoll-Ross process. (a) Sample path with short relaxation time $\tau = 0.1$ showing a sequence of short lasting bursts. (b) For the largest relaxation time $\tau = 10.0$, the strictly positive bursts of activity show a longer duration and thus contribute to the global variance across many time scales in Hurst analysis.

leading to significant variance contributions at longer time scales in subsequent Hurst exponent analyses. The temporal autocorrelations of the CIR process, as for the OU process, also decay exponentially, i.e., the CIR process is short-range correlated [21].

We here include the CIR process as a simple heteroscedastic process that is short-range correlated on the one hand, but exhibits the Hurst phenomenon ($H > 0.5$) for finite samples on the other. While the CIR process does not capture the characteristic oscillations of EEG data, its amplitude dynamics are similar to the bursty nature of the EEG alpha oscillation envelope. Further below, we fit the more flexible AR(10) model to empirical EEG data. The CIR process is used as an example of intermediate complexity, showing that simple heteroscedastic models can produce variance dynamics leading to $H > 0.5$.

The Cox-Ingersoll-Ross process is defined as

$$dX_t = -\frac{1}{\tau}(X_t - \mu) + \sigma\sqrt{X_t}dW_t. \qquad (4)$$

Using a slightly different parametrization, $dX_t = (a - kX_t)dt + \sigma\sqrt{X_t}dW(t)$, sample paths X_t are assured to be nonnegative when the constraint $2a > \sigma^2$ holds [32]. We simulated sample paths for the same range of τ values as for the Ornstein-Uhlenbeck process and for $\sigma^2 = 1$ [32].

D. Resting state EEG and fMRI

We analyzed EEG and fMRI recordings from ten healthy subjects (age range, 19–27; mean age, 23 years) recorded during an eyes-closed wakeful rest condition. Each pair of EEG and fMRI data sets was acquired on the same day. We selected EEG data segments with optimum data quality, i.e., free of electrode artifacts, eye blinks or signs of drowsiness. EEG signals were recorded with a sampling rate of 5 kHz using the standard 10-10 electrode configuration. All channels were band-pass filtered to the alpha frequency range (8–12 Hz) using a zero-phase Butterworth filter with a slope of 24 dB/octave, down-sampled to 250 Hz and rereferenced to an average reference. EEG segments have a duration ranging from 100–303 s, corresponding to lengths of 25 000–75 750 samples. The amplitude modulation of two exemplary recordings is shown in Fig. 4. The traces show the analytical amplitude of the left occipital (O1) electrode's signal as computed from the Hilbert transform. In the two recordings shown in Fig. 4, as well as in the other subjects not shown, we observed irregular bursts of alpha activity, separated by low amplitude alpha intervals of variable duration. These dynamics suggest variance contributions across many time scales in Hurst exponent analysis.

We also analyzed EEG data using the so-called microstate segmentation algorithm [33]. The method applies a modified K-means clustering algorithm to the set of spatial EEG patterns selected at time points where the spatial standard deviation of the scalp potential has local maxima. Clustering is commonly performed to obtain a set of four representative EEG topographies, called microstates and labeled $\{A, B, C, D\}$, which maximize the global explained variance with respect to the original EEG time series [34]. The original EEG time series can then be represented by a symbolic time series of microstate labels, e.g., $\ldots, B, B, A, C, C, C, D, A, \ldots$, where
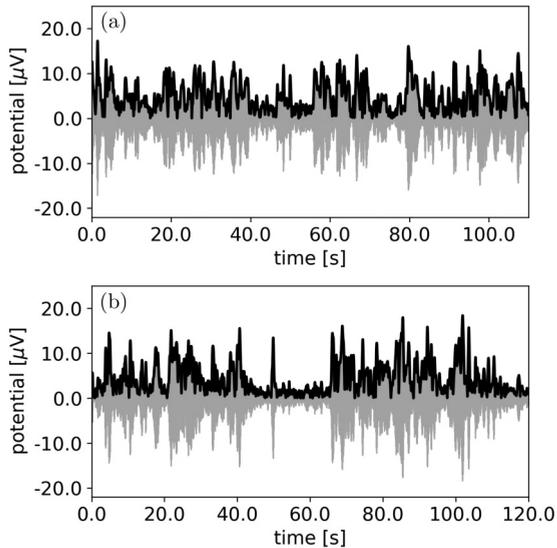
FIG. 4. Amplitude dynamics for two different resting state EEG recordings. (a) Subject 1, (b) Subject 6, as indexed in Table I. The respective left occipital EEG channel signals (O1, light gray) and their analytical amplitudes (black lines) are shown. The analytical amplitude shows irregular bursts of large amplitude activity, interlaced with periods of low-amplitude oscillations. In both subjects, we observe similar bursty dynamics, anticipating variance contributions across many time scales in subsequent Hurst analyses.

at each time point, the label indicates the microstate that has the maximum spatial correlation with the actual EEG topography. The algorithm modifies the classical K-means distance measure using the squared correlation coefficient as the metric between the microstate and the current EEG topography, i.e., the polarity of the potential is ignored. The microstate algorithm takes into account the temporal dynamics of the whole scalp surface potential, and not just of a single channel. For methodological details, we refer the reader to classical publications [34,35].

Functional MRI data were recorded in a 3 Tesla MR scanner (Siemens Trio, Erlangen, Germany) and consist of 1505 volumes of T2*-weighted echo planar images (TR/TE = 2080 ms/30 ms, matrix size 64 in x and y directions, voxel size 3 mm in x and y directions, and 2 mm in the z direction, distance factor 50%). All volumes were realigned, normalized to MNI space, and spatially smoothed with a Gaussian kernel (8 mm^3 full width at half maximum). Data were spatially downsampled to 4^3 mm^3 resolution for further processing. To allow for steady-state magnetization, the first five scans were ignored and time series of length 1500 were used for further computations. Cardiac-, respiration-, and motion-induced noise was regressed out as described in Ref. [36]. Finally, fMRI time series were band-pass filtered to the 0.01–0.10 Hz range using a sixth-order Butterworth filter. The preprocessing pipeline is equivalent to the procedures described in the literature on fMRI and LRD [36,37].

Written informed consent was obtained from all subjects, and the study was approved by the ethics committee of the Goethe University, Frankfurt, Germany.

E. Autoregressive AR(p) processes

Oscillations with burstlike amplitude dynamics are the most prominent pattern in resting state EEG signals. During the wakeful resting state, the main frequency component lies in the alpha frequency band (8–12 Hz). The analytical amplitude (the envelope) of alpha oscillations has been reported to exhibit the Hurst phenomenon, fGn-like dynamics, and long-range correlations [5,22,23]. As a first-order Markov process cannot capture the complex dynamics shown in Fig. 4, we add to our analyses a flexible short-range correlated EEG model that should be able to produce alpha oscillations with bursty, time-varying amplitude dynamics. The standard approach to model EEG oscillations stochastically is given by autoregressive (AR) models [28,29,38] whose general expression is given by

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + \epsilon_t. \tag{5}$$

Equation (5) can be rewritten in operator form as

$$\left(1 - \sum_{k=1}^{p} \phi_k L^{(k)}\right) X_t = \epsilon_t, \tag{6}$$

where $L^{(k)}$ is the kth order lag operator defined by $L^{(k)} X_t = X_{t-k}$. The roots of the complex polynomial $f(z) = 1 - \sum_{k=1}^{p} \phi_k z^k$ define the dynamics of the stochastic process X_t.

In Eqs. (5) and (6), the variable X_t represents the modeled EEG data, p denotes the AR model order and ϕ_i are the AR coefficients. The AR process is driven by the Gaussian white noise term ϵ_t. To model the full spectral content of EEG data, AR model orders of $p = 5$–15 are often used [29]. Here, we use AR(10) models which were fitted using Burg's method [29]. All fits converged and all roots of the autoregressive polynomials $f(z)$ were located outside the unit circle, assuring stable solutions. The fitted AR parameters ϕ_k are given in Table I.

Exemplary results of AR(10) fits are shown in Fig. 5. Figures 5(a) and 5(b) show resting state EEG data (alpha frequency band, O1 electrode) and Figs. 5(c) and 5(d) show a sample of the corresponding AR(10) fit, on two different time scales. Figure 5(a) illustrates the burstlike activity of alpha oscillations on a time scale of 100 s, and Fig. 5(b) on a shorter 8 s time scale. The analytical amplitude [black curve in Figs. 5(a), 5(c), dark gray curve in Figs. 5(b), 5(d)] is obtained as the modulus of the Hilbert transform. Figures 5(c) and 5(d) show the results of the corresponding AR(10) fit. It is observed that the model correctly reproduces burstlike alpha band activity, visually highly similar to the experimental data set. The quantitative properties (Hurst exponents, mutual information) of experimental and surrogate data are analyzed in the Results section.

III. METHODS

A. Estimation of the Hurst exponent

Hurst exponents were estimated using a method based on the discrete wavelet transform (DWT) [39]. We used a fifth-order Daubechies mother wavelet and fitted Hurst exponents to the scalogram across the wavelet scales $j = 6 - 12$ [40,41]. For two subjects, the sample size only allowed the computation

TABLE I. AR(10) coefficients for resting state EEG in ten healthy subjects S1–S10.

Subject no.	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}
1	5.06	−11.87	16.38	−13.34	4.03	4.47	−7.00	4.71	−1.73	0.29
2	5.27	−12.94	18.93	−16.90	7.06	3.19	−7.26	5.40	−2.13	0.38
3	5.26	−13.03	19.52	−18.23	8.61	2.27	−7.20	5.70	−2.33	0.42
4	4.86	−11.57	17.28	−17.16	10.70	−2.59	−2.16	2.63	−1.26	0.25
5	5.03	−11.73	16.16	−13.01	3.37	5.53	−8.06	5.32	−1.92	0.31
6	4.89	−11.06	14.64	−11.05	2.03	5.54	−7.10	4.34	−1.44	0.21
7	4.97	−11.58	16.06	−13.35	4.42	4.02	−6.74	4.60	−1.69	0.27
8	5.03	−11.77	16.28	−13.38	4.24	4.24	−6.89	4.71	−1.76	0.30
9	5.31	−13.26	19.92	−18.51	8.39	3.07	−8.17	6.39	−2.61	0.47
10	5.34	−13.39	20.24	−19.09	9.25	2.04	−7.28	5.87	−2.43	0.44

of DWT scales up to $j = 11$. The scale selection is based on values given in the literature, and the common scaling region is reproduced in our data as shown in Fig. 6. We observe that above $j = 6$, the scaling functions grow almost linearly toward the largest wavelet scale, defined by the sample size. At lower

scales, another previously observed scaling region is shown. That region contains filtering effects introducing short-range correlated components during preprocessing. Importantly, our results are in line with previously published scalograms and the Hurst exponents are highly similar to those presented in the literature. These results validate our data sets and the used estimation algorithms with respect to previously published results. Scale-dependent variance estimates at scale j are calculated as $\frac{1}{N_j} \sum_{k=1}^{N_j} | d_j(k) |^2$, where N_j is the number of wavelet coefficients at scale j.

For EEG microstate sequences, we used a recently published method that maps the symbolic microstate sequence to a metric time series based on a (2,2)-partition of the set of microstate labels [40]. There are three unique (2,2)-partitions $P_i = \{P_{i_1}, P_{i_2}\}, i = 1, 2, 3$, explicitly $P_1 = \{\{A,B\}, \{C,D\}\}$, $P_2 = \{\{A,C\}, \{B,D\}\}$ and $P_3 = \{\{A,D\}, \{B,C\}\}$. The values x_t of the microstate sequence are then mapped to the set $\{-1, +1\}$ by defining a function φ_i for each partition P_i, such that $\varphi_i(x_t) = +1$ if $x_t \in P_{i_1}$ and $\varphi_i(x_t) = -1$ if $x_t \in P_{i_2}$. Finally, the Hurst exponent of the random walk process $X_n = \sum_{i=0}^{n} \varphi_i(x_t)$
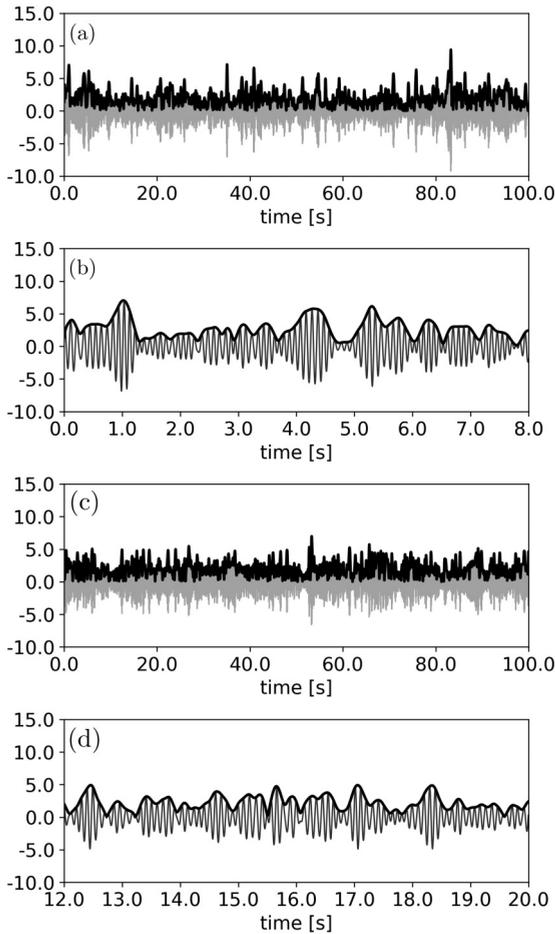

FIG. 5. Sample paths of experimental EEG alpha band (8–12 Hz) oscillations (a, b) and autoregressive AR(10) surrogates (c, d) on two different time scales, 100 s and 8 s. The AR(10) model data reproduce the alpha band dynamics on both time scales. Both, empirical EEG (a, b) and surrogate data (c, d) show irregular bursts of 10 Hz activity. The analytical amplitude (black) of the signal measures the instantaneous power of the alpha band oscillation (gray).
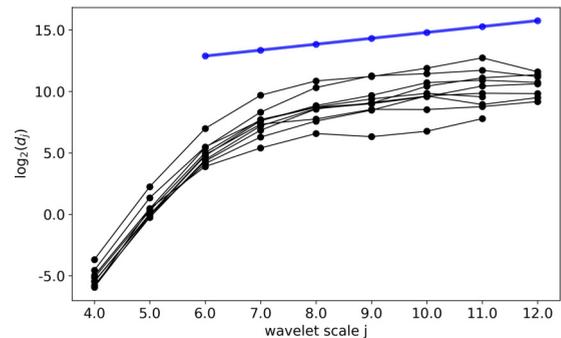

FIG. 6. Hurst exponent estimation using the discrete wavelet transform. The figure shows the scaling functions of all ten subjects, analyzing the analytical amplitude of the alpha envelope recorded at the left occipital electrode O1. The blue line above the scaling functions (black) shows the theoretical scaling function using the empirical average Hurst exponent across the ten subjects, $\bar{H} = 0.74$, restricted to the fitting interval $j_{\min} = 6$ to $j_{\max}$. For two subjects, the sample size determined $j_{\max} = 11$, for all other subjects $j_{\max} = 12$. The blue line is shifted vertically for visualization purposes. The smallest wavelet scales containing transient short-range correlations are not shown.

is estimated. From the three unique (2,2)-partitions of the state space $\{A,B,C,D\}$, we obtain three Hurst exponents for each time series.

For comparison, we also estimated Hurst exponents with detrended fluctuation analysis (DFA) [42] and obtained equivalent results. For better visibility, only wavelet-based results are shown in the remaining paper, DFA results are given in the text where necessary.

B. Discretization

To compute the discrete time-lagged mutual information function, we first have to partition the real-valued continuous process into a number of discrete classes. Thus, continuous probability densities are approximated by discrete distributions (histograms) and entropy calculations become straightforward [43,44]. Several choices for the optimum number of discretization steps exist in the literature. We here present results obtained with the Freedman-Diaconis rule [45]

$$N_{\mathrm{FD}} = \left\lceil \frac{R}{2\,\Delta\,n^3} \right\rceil, \qquad (7)$$

using the ceiling function $\lceil \cdot \rceil$. The rule does not apply assumptions about the underlying distribution.

We also tested two other discretization schemes, Scott's rule for normally distributed data [46], and Sturges' rule [47], the latter giving the lowest number of discretization steps. Scott's rule proposes the use of $N_{\mathrm{Sc}} = \lceil \frac{R}{3.5\,\sigma\,n^3} \rceil$, Sturges' rule the use of $N_{\mathrm{St}} = \lceil 1 + \log_2 n \rceil$. Sturges' rule may underestimate the optimum number but is taken to be a good approximation for large samples. In all expressions, $R = \max(X) - \min(X)$ is the range of the signal X, the standard deviation is abbreviated σ, and $\Delta = Q_{75} - Q_{25}$ is the interquartile range, i.e. the difference between the 75% and the 25% quantiles. The results for all discretization schemes were qualitatively similar, therefore the latter two schemes are not further discussed in this article. For the OU process, the Freedman-Diaconis rule gave an average of $\bar{N}_{\mathrm{FD}} = 98$ discretization steps and Scott's rule gave $\bar{N}_{\mathrm{Sc}} = 76$. For the CIR process, we obtained $\bar{N}_{\mathrm{FD}} = 211$ and $\bar{N}_{\mathrm{Sc}} = 102$. For fGn, we got $\bar{N}_{\mathrm{FD}} = 115$ and $\bar{N}_{\mathrm{Sc}} = 88$. Average values were rounded to the closest integer. For all processes, the sample size of $n_{\max} = 50\,000$ determined $N_{\mathrm{St}} = 17$.

C. Time-lagged mutual information

Time-lagged mutual information for discrete, positive time lags k is defined as [43]

$$I(k) = H(X_{t+k}) - H(X_{t+k} \mid X_t). \qquad (8)$$

Equation (8) defines mutual information as the difference between the entropies of two distributions, $P(X_{t+k})$, and the conditional probability of X_{t+k}, given X_t. Equivalently, mutual information is given as the Kullback-Leibler divergence [43] between the joint distribution $P(X_t, X_{t+1})$ and the product measure under the temporal independence assumption, $P(X_t, X_{t+1}) = P(X_t)P(X_{t+1})$. Thus, mutual information measures temporal dependence as the amount of information about X_{t+k} contained in X_t. Mutual information for the Markovian null hypothesis, with equilibrium distribution π

and transition matrix $P(X_{t+1} = j \mid X_t = i) = T_{ij}$ is given by

$$I(k) = -\sum_i \pi_i \log \pi_i + \sum_i \pi_i \sum_j T_{ij}^k \log T_{ij}^k, \qquad (9)$$

where T_{ij}^k denotes the (i,j)th element of the kth matrix potency of T.

To control for effects induced by choosing a specific discretization scheme, we also performed tests using the continuous (binless) Kozachenko-Leonenko estimate of mutual information [48]. We used the Kozachenko-Leonenko estimate as implemented in the Python machine learning package scikit-learn using the K-nearest neighbors approach with $K = 3$.

As EEG microstate sequences are discrete by construction, i.e., all values are taken from a finite set, mutual information can be computed exactly without discretization considerations.

Throughout the manuscript, we use the natural logarithm to the base e (Euler's number) in the computation of mutual information values, expressed in "nat" units.

D. Markov confidence intervals

To obtain Markovian surrogate data, the empirical transition matrix $\hat{T}_{ij} = P(X_{t+1} = j \mid X_t = i)$ and the empirical equilibrium distribution $\hat{\pi}$ are calculated for each discretized time series. As a first-order Markov chain is uniquely defined by T_{ij} and an initial distribution, a surrogate Markov data set with $\hat{T}$ and $\hat{\pi}$ identical to the empirical quantities can be synthesized [49]. For each time series X_t tested, we synthesized $n = 1000$ surrogate Markov sequences of the same length as X_t. The iterative algorithm is initialized according to $\hat{\pi}$ before iterating the remaining elements according to $\hat{T}_{ij}$. Thus, the value $X_{t+1} = j$ is a function of $X_t = i$ and the conditional transition probability $\hat{T}_{ij}$. Initialization and iteration are implemented with pseudorandom variables $r_t \sim U[0,1]$, uniformly and independently distributed on the unit interval. The initial state index j is given by $\sum_{i=0}^{j-1} \hat{\pi}_i \leqslant r_0 < \sum_{i=0}^{j} \hat{\pi}_i$. During iteration, the index j of the successor state $X_{t+1} = j$ is given by $\sum_{l=0}^{j-1} \hat{T}_{il} \leqslant r_t < \sum_{l=0}^{j} \hat{T}_{il}$.

Our aim is to test the neurobiologically derived hypothesis that the analytical amplitude of the alpha frequency band is a fGn-like, long-range correlated process [5,22,23]. As EEG data has to be preprocessed to extract the signal of interest (band-pass filter, Hilbert transform), we designed an additional null hypothesis for continuous EEG signals. This second null hypothesis for short-range memory takes into account preprocessing effects and the bursty nature of the instantaneous alpha power. To this end, we fitted an AR(10) model to each EEG data set and used the coefficients to synthesize $n = 100$ AR(10) surrogate samples for each EEG time series. The surrogate time series were preprocessed in the same way as experimental EEG (band-pass filter, Hilbert transform) and the mutual information estimates were obtained using the discretization method described above. From the set of AR(10) mutual information functions we obtain a confidence interval ($\alpha = 0.01$) under the assumption of autoregressive, short-range correlations. In case of long-range correlations, we expect the empirical (EEG) mutual information function to lie above the AR(10) confidence interval.
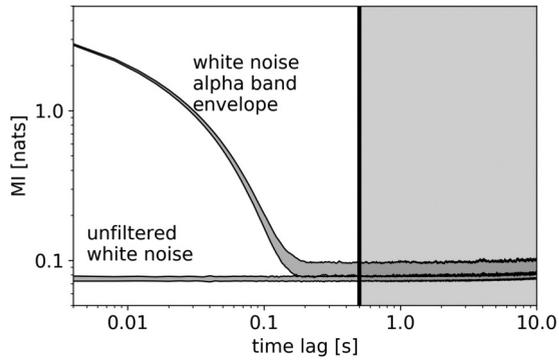
FIG. 7. Effects of preprocessing on mutual information estimation. Gaussian white noise yields low mutual information at all time lags as shown by the flat confidence interval ($\alpha = 0.01$) labeled "unfiltered white noise." Applying the EEG pre-processing algorithm to white noise, the confidence interval labeled "white noise alpha band envelope" shows short-term memory components up to approximately 200 ms. Memory effects in EEG time series are therefore considered for time lags >500 ms (gray shaded area).

E. Effects of preprocessing

To compare our results with previous studies assessing the instantaneous amplitude of individual EEG frequency bands, we follow the same preprocessing strategies as given in the literature [5,22,23,50]. As the occipital alpha rhythm (8–12 Hz) is the most prominent feature in wakeful resting state recordings, most studies consider this frequency band. Furthermore, as the amplitude of alpha oscillations is affected by vigilance and cognitive task engagement, among others, the instantaneous alpha amplitude is considered an indicator of the brain's functional state [51]. The neurobiological hypothesis behind the above mentioned investigations is that the brain's dynamic complexity is reflected in long-range temporal correlations of the alpha oscillation's envelope. The technical procedure leading to the envelope of a specific frequency band involves two linear and one nonlinear operator. In the first step, a (linear) zero-phase band-pass filter is used to extract the 8–12 Hz alpha frequency band. In the second step, the (linear) Hilbert transform of the band-pass filtered signal is computed. The band-pass filtered oscillation and its Hilbert transform define an analytical signal whose absolute value (nonlinear operation) is the alpha envelope. As the mutual information algorithm works on the preprocessed signal, we have to consider the possibility that the pre-processing strategy may affect the outcome. Intuitively, the envelope computation is expected to affect mutual information at short time lags since the envelope cannot change arbitrarily fast. Looking at the signals shown in Fig. 5, it is observed that the envelope smoothly connects subsequent peaks of the alpha band oscillation, i.e., it is a smooth curve between points approximately 100 ms apart. The lack of high-frequency components below this time scale suggests that our preprocessing may emphasize short-range correlations.

We therefore analyzed the effects of preprocessing on Gaussian distributed white noise for its uniform frequency content. Figure 7 illustrates the results. The horizontal gray band at the bottom of the figure is the first-order Markov confidence interval ($\alpha = 0.01$) for the mutual information function of

unprocessed white noise signals (labeled "unfiltered white noise"). As expected, there is a low, finite-memory content that does not depend on the time lag. When the white noise signal is processed in the same way as EEG data, we obtain the upper confidence interval labeled "white noise alpha band envelope." It is important to note that this represents the alpha band content of white noise, not a biological signal. In accordance with the arguments given above, the confidence interval for the filtered signal shows significant memory effects at time lags below 200 ms. Introducing a further safety margin, we only consider time lags >500 ms in all subsequent EEG analyses, indicated by the gray area to the right of the vertical black line in Fig. 7. These results are in line with previous studies assessing the range of scales where confounding effects due to filtering are expected [40,52,53].

IV. RESULTS

A. Hurst exponents

Hurst exponents were calculated using a discrete wavelet transform with Daubechies' Db5 wavelet. At each scale, the empirical fluctuation was computed and plotted against the wavelet scale j. For the three model stochastic processes (fGn, OU, CIR) and the parameter ranges of H and τ defined above, we synthesized and analyzed sample paths with a length of 50 000 samples each. For each process and each parameter value, we synthesized $n = 100$ sample paths. The results are summarized in Fig. 8. Figure 8(a) shows the estimated Hurst exponents $\hat{H}$ for fGn. The nominal Hurst exponent used in the simulations is shown on the abscissa and the wavelet estimate $\hat{H}$ on the ordinate. The blue line is the identity line indicating the correct H value. The mean value of the Hurst exponent estimates is indicated by black squares. We also calculated the standard error of the mean (SEM) for the $n = 100$ sample paths and added corresponding error bars to Fig. 8(a). However, the numerical values are so small (≈ 0.004–0.006) that the error bars can only be perceived under considerable magnification of the data. It is observed that the wavelet method gives precise estimates of the Hurst exponent in the case of fGn. There is a small negative bias of the wavelet estimate, which however is small compared to other methods presented in the literature [9,54]. Figure 8(b) shows the Hurst exponents for the OU process, with increasing relaxation time shown on the abscissa. It is observed that increasing τ is associated with large Hurst exponents $H \gg 0.5$. Figure 8(c) shows that also the CIR process yields Hurst exponents $H \gg 0.5$, especially for slow relaxation times, i.e., for large τ. As for fGn, black squares represent mean Hurst exponent estimates. Again, the SEM for the $n = 100$ estimates are so small that the error bars in Figs. 8(b) and 8(c) are invisible. All shown results represent wavelet estimated Hurst exponents, DFA estimates were practically identical and are therefore not visualized in Fig. 8.

B. Mutual information with Markovian confidence intervals

For the three stochastic processes considered above, we consistently found Hurst exponents $H > 0.5$, although two of the processes (OU, CIR) are actually short-range correlated. We will here assess the capacity of the mutual information function to measure the memory structure of the discretized
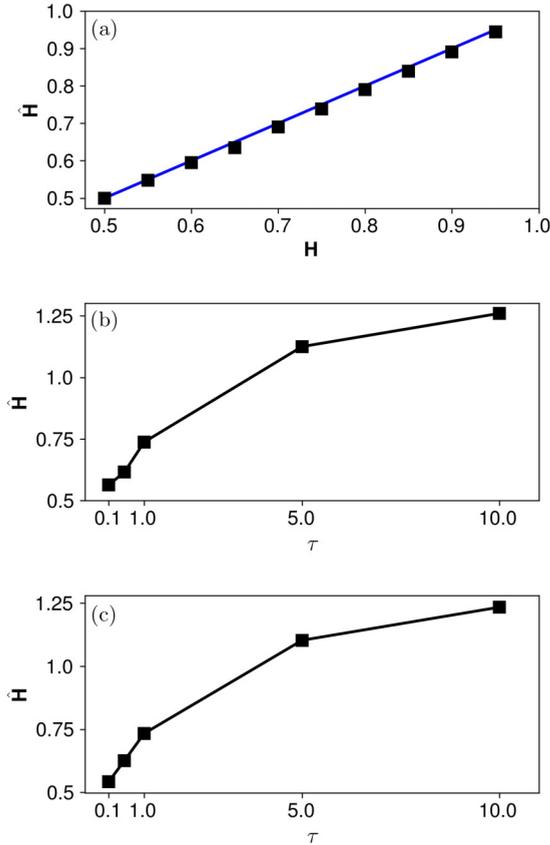
FIG. 8. Hurst exponents for fractional Gaussian noise, the Ornstein-Uhlenbeck process and the Cox-Ingersoll-Ross process. In (a–c), $n = 100$ sample paths of length 50 000 are used per data point. (a) For each nominal Hurst exponent H (abscissa), $\hat{H}$ is estimated by the wavelet method (ordinate). Average $\hat{H}$-values are shown as black squares, SEM error bars are so small that they are not visible. The blue line indicates the identity $\hat{H} = H$. It is observed that long-range correlations in fGn are correctly estimated. (b) For the OU process, the same estimation procedure shows increasing Hurst exponent estimates with increasing relaxation time τ, although the process is short-range correlated. (c) For the CIR process, the same behavior as in (b) is observed. It is concluded that for OU and CIR, the Hurst phenomenon occurs without long-range correlations.

FIG. 9. Time-lagged mutual information functions for Freedman-Diaconis discretized stochastic processes, LRD and relaxation time parameters (H, τ) increase from bottom to top. (a) fractional Gaussian noise (fGn), (b) Ornstein-Uhlenbeck (OU), (c) Cox-Ingersoll-Ross (CIR).

processes. Figure 9(a) shows the mutual information functions for fGn, with H increasing from $H = 0.5$ (bottom) to $H = 0.95$ (top). Figures 9(b) and 9(c) show the mutual information functions for the OU and CIR processes, with τ increasing from $\tau = 0.1$ (bottom) to $\tau = 10.0$ (top). The shape of the short-range correlated processes [Figs. 9(b) and 9(c)] is markedly different from the mutual information functions of fGn. Short-range correlated mutual information functions [Figs. 9(b) and 9(c)] show an inflection point whereas the long-range correlated mutual information functions [Fig. 9(a)] are convex. Figure 10 shows mutual information functions (bold black lines) with Markovian confidence intervals (gray shaded areas) for stochastic test signals as well as for EEG and AR(10) example data. Figure 10(a) shows the results for a fGn sample synthesized with $H = 0.85$, for which wavelet (DWT) and DFA estimates yield the correct value $\hat{H} = 0.85$. Compared with the first-order Markov confidence interval, the mutual
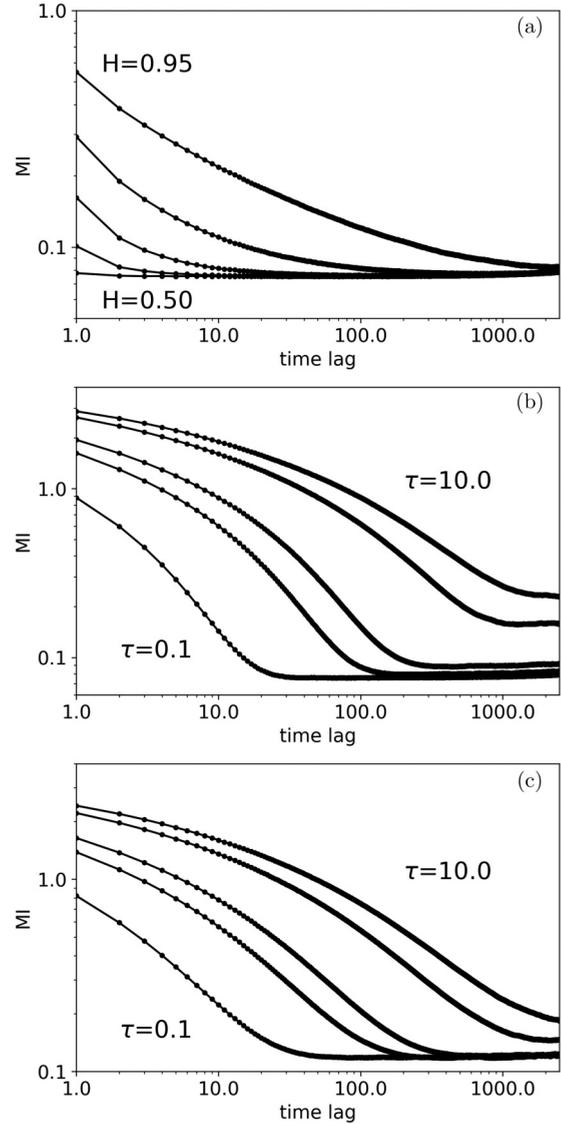
information function shows a much slower decay, indicating long-range memory. Figures 10(b) and 10(c) show the results for OU and CIR samples with relaxation time $\tau = 1.0$. The estimated Hurst exponents for the OU process are $\hat{H}_{\mathrm{DWT}} = 0.77$ and $\hat{H}_{\mathrm{DFA}} = 0.73$, for the CIR process $\hat{H}_{\mathrm{DWT}} = 0.80$ and $\hat{H}_{\mathrm{DFA}} = 0.74$. If these results are interpreted according to the rules valid for fGn, the estimated Hurst exponents seem to indicate LRD effects. The corresponding mutual information functions, however, lie within their first-order Markov confidence intervals across the range of measured time lags, demonstrating that the apparent Hurst phenomenon does not indicate long-range memory.

Figures 10(d) and 10(e) illustrate the same analysis for an EEG example and a corresponding short-range correlated AR(10) surrogate. The EEG sample yields empirical Hurst exponents of $\hat{H}_{\mathrm{DWT}} = 0.72$ and $\hat{H}_{\mathrm{DFA}} = 0.75$, and the AR(10)
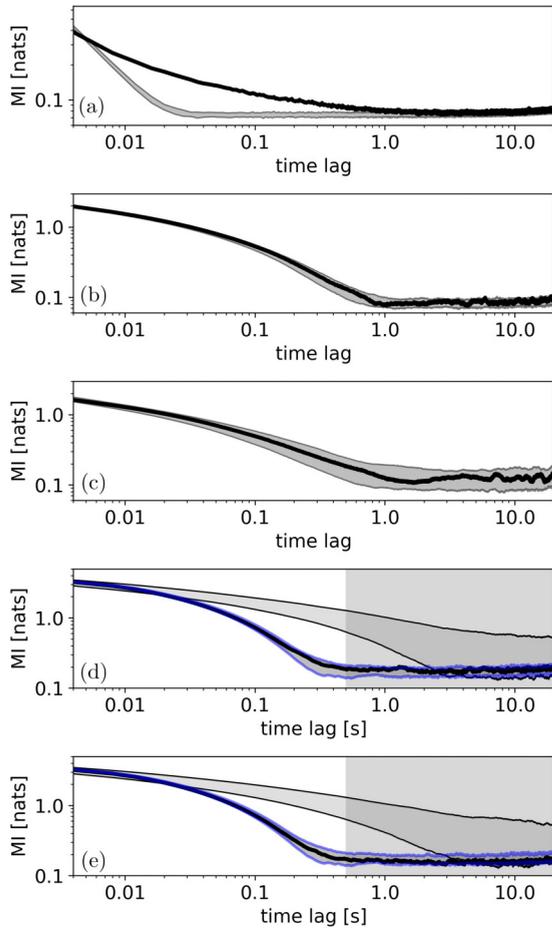
FIG. 10. Time-lagged mutual information functions (bold black lines) for fGn (a), OU (b), and CIR (c), and their first-order Markov confidence intervals (gray). fGn clearly deviates from the Markovian hypothesis, showing a significantly higher memory content and a much slower decay. For OU and CIR, the mutual information function lies within the Markov confidence interval, demonstrating their short-range correlated nature. For EEG data (d) and the AR(10) model (e), two types of confidence intervals are shown, the first-order Markov null hypothesis as in (a–c) (gray area with black borders) and the AR(10) confidence interval (gray area with blue borders). EEG and AR(10) data do not show a fGn-like mutual information function and are statistically indistinguishable from short-range correlated processes.

model gives estimates of $\hat{H}_{\mathrm{DWT}} = 0.62$ and $\hat{H}_{\mathrm{DFA}} = 0.78$. Again, these results suggest LRD effects. The mutual information functions, however, show inflection points similar to the short-range correlated processes shown in Figs. 10(b) and 10(c). As first-order Markov surrogates parametrize these processes insufficiently, we show both, the first-order Markov (gray area with black borders) and the AR(10) (gray area with blue borders) confidence intervals. The region for which pre-processing effects can be excluded (time lags >500 ms) is shown in light gray color. For EEG and AR(10) data, the mutual information functions lie within the corresponding AR(10) confidence intervals. This result can be formulated in two ways, (a) the memory structure of the EEG alpha envelope is fully captured by the AR(10) model, and (b) the EEG structure is
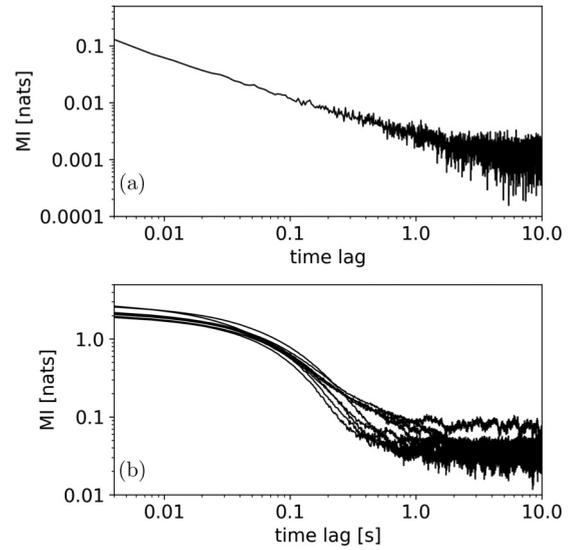

FIG. 11. Continuous Kozachenko-Leonenko estimates of mutual information for fGn and EEG. (a) Average mutual information function for fGn with Hurst exponent $H = 0.75$ ($n = 10$). (b) Individual mutual information functions for the ten EEG recordings. While fGn shows a power-law decay of memory effects, the information content of the alpha envelope decays like a short-range correlated process.

statistically not distinguishable from an autoregressive model, i.e., the EEG alpha envelope shows short-range memory. Moreover, the short-range correlated bursts of alpha activity produced by the AR(10) model reliably produce a spurious Hurst phenomenon.

C. Resting state EEG

The exemplary results for the analytical amplitude of resting state EEG alpha oscillations shown in Figs. 10(d) and 10(e) were further assessed for all ten subjects. For the ten data sets, the mean and SEM values of the estimated Hurst exponents were $\hat{H}_{\mathrm{DWT}} = 0.75 \pm 0.025$ and $\hat{H}_{\mathrm{DFA}} = 0.74 \pm 0.019$. We thus observe a consistent Hurst phenomenon ($H > 0.5$), with Hurst exponents identical to the values presented in the literature [5,22,23]. Next, we test if the difference in the mutual information functions of fGn and the EEG alpha envelope is caused by the discretization scheme. We therefore performed an additional analysis using the Kozachenko-Leonenko continuous estimate of mutual information for fractional Gaussian noise and EEG signals. The results are shown in Fig. 11. To compare fGn and EEG memory effects, Fig. 11(a) shows the average mutual information function of ten fGn samples with $H = 0.75$, to mimic the properties of the EEG data set. We observe that for fGn, the continuous Kozachenko-Leonenko estimate also indicates memory effects following a power-law, i.e. decaying linearly in log-log coordinates. Figure 11(b) shows the continuous mutual information estimates for the ten EEG samples. Although the EEG-derived mutual information functions show some variability, we observe the same shape as for the discretized processes shown in Figures 10(d), 10(e). Notably, the EEG alpha envelope and fGn show clearly dissimilar mutual information functions. To exclude spurious memory effects introduced by band-pass filtering and
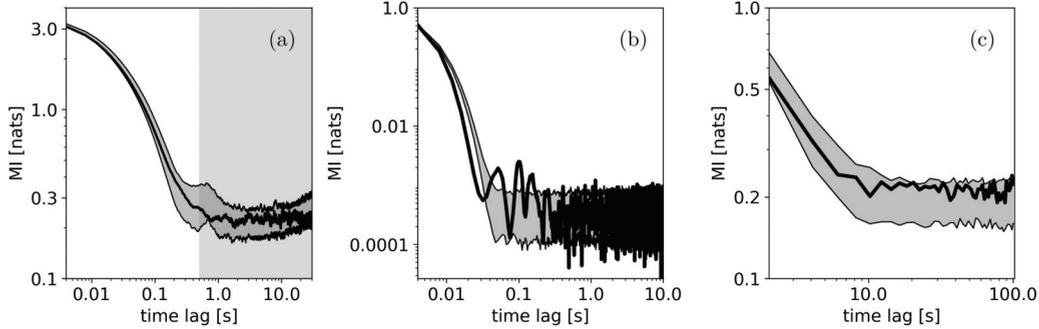
FIG. 12. Mutual information functions for neurophysiological data. (a) Analytical alpha amplitude of the left occipital EEG channel. Time scales not affected by pre-processing effects are indicated by the gray-shaded area. (b) Memory effects for the EEG microstate random walk. (c) fMRI signal recorded in the left occipital area, the same region where the signal in (a) was recorded using EEG. For (a–c), the empirical mutual information functions (black lines) lie within the Markovian confidence intervals (gray areas, $\alpha = 0.01$) supporting the short-range correlated null hypothesis.

envelope computations, we also analyze memory effects in EEG microstate time series computed from the same EEG data sets as those analyzed in Fig. 11(b), but using a completely different algorithmic approach including data from all EEG channels. Exemplary mutual information functions for Subject 2, as indexed in Table I, are shown in Fig. 12. On the left, Fig. 12(a) shows the discretized mutual information function for the analytical alpha amplitude recorded at electrode O1. The function lies within the Markovian confidence interval obtained from the AR(10) model (gray area with black borders), similar to the example shown in Fig. 10(d). Time scales not affected by pre-processing are again indicated by the light gray area.

The middle panel, Fig. 12(b), shows the results for EEG microstate analysis. The shape of the empirical mutual information function and the Markovian confidence interval are highly similar to the results obtained in Fig. 12(a). A difference is found at time lags around 100ms, corresponding to the EEG alpha frequency band (8–12 Hz), where the microstate mutual information function shows oscillatory peaks. Actually, the peaks are located at multiples of twice the main frequency component, here at multiples of approximately 50 ms. We explained this frequency doubling effect in a recent publication [55]. At time lags above 1–2 s, however, the mutual information functions lies within the confidence intervals defined by the first-order Markov null hypotheses, arguing against long-range memory effects. Although we find no long-range memory in the mutual information functions, the empirical Hurst exponents of all microstate sequences satisfy $H > 0.5$, in line with values reported in the literature [40,41]. Averaging across the ten subjects, we obtain $\hat{H}_{\mathrm{DWT}} = 0.63 \pm 0.013$ and $\hat{H}_{\mathrm{DFA}} = 0.64 \pm 0.013$ for partition P1. For P2, we get $\hat{H}_{\mathrm{DWT}} = 0.62 \pm 0.019$ and $\hat{H}_{\mathrm{DFA}} = 0.63 \pm 0.013$ and for P3, we find $\hat{H}_{\mathrm{DWT}} = 0.66 \pm 0.025$ and $\hat{H}_{\mathrm{DFA}} = 0.66 \pm 0.019$. As microstate analysis does not contain the preprocessing steps contained in Fig. 12(a), i.e., band-pass filtering to the alpha frequency band and analytical amplitude computation, we conclude that the observed short-range memory effects are a stable, inherent property of the neural processes underlying EEG data. We obtained the same kind of results for all other subjects that are not shown in Fig. 12. Figures 12(a) and 12(b) both show 99% confidence intervals computed from $n = 100$ surro-

gates. Interestingly, all recordings show a mutual information function with a reverse sigmoid shape, i.e., with an inflection point, similar to the short-range correlated stochastic processes shown in Figs. 9(b) and 9(c), and Figs. 10(b) and 10(c).

D. Resting state fMRI

To further corroborate our results, we also tested a different experimental technique (fMRI) for several reasons: (a) unlike the EEG signal, the fMRI signal does not suffer from electrical volume conduction through the skull, (b) no amplitude reconstruction by the Hilbert transform is required, and (c) much longer time series can be analyzed. We therefore analyzed fMRI recordings from the same set of subjects as considered in the EEG section. As EEG data were taken from the left occipital electrode (O1), we analyzed the Hurst exponents and the mutual information functions of fMRI time courses extracted from the left occipital cortex as indexed by the AAL anatomical MR atlas [56].

Hurst exponents were estimated using DWT (scales 1–4) and DFA (scales 10–50) and we obtained results of $\hat{H}_{\mathrm{DWT}} = 0.85 \pm 0.032$ and $\hat{H}_{\mathrm{DFA}} = 0.66 \pm 0.019$, in general accordance with the literature. For subject 2, the Freedman-Diaconis discretized mutual information function (black) and the first-order Markov 99% confidence interval (gray area) are shown in Fig. 12(c). Similar to EEG data, we observe that the mutual information function lies within the Markov confidence intervals. Analogous results were obtained for all other subjects. Note that due to the different sampling rates and different recording lengths of EEG and fMRI recordings, the fMRI analyses represent 50 min records with time lags up to 100 s.

V. DISCUSSION

We here present the time-lagged mutual information function as an effective tool to assess long-range dependence in finite length empirical data, especially in cases where Hurst exponent estimation yields false positive results. Our results can be summarized by the following points:

(1) Neural signals (EEG alpha envelope, EEG microstate sequences, fMRI) show a consistent Hurst phenomenon when analyzed with the wavelet and DFA algorithms but their

mutual information functions lie within the confidence intervals defined by short-range correlated surrogates.

(2) Neural data show mutual information functions similar to some short-range (exponentially) correlated stochastic processes (OU, CIR). Testing OU and CIR samples for slow relaxation time constants with the wavelet and DFA algorithms, we observe a stable Hurst phenomenon ($H > 0.5$). In these cases, however, the long-range dependence interpretation of the Hurst exponent is erroneous.

(3) Long-range correlations and the Hurst phenomenon, when present (fGn), are correctly quantified by all algorithms tested, i.e., the classical wavelet and DFA algorithms, as well as the information-theoretical approach investigated here.

(4) The time-lagged mutual information function with Markovian confidence intervals, applied to discretized signals, correctly distinguishes short-range from long-range correlated processes.

A. Slow relaxation and sample length

The wavelet and DFA algorithms both quantify variance contributions on different time scales for a given signal. The amount of variance produced at a given length scale is analyzed as a function of the scale parameter, i.e., the length of the time window considered. Due to the specially designed covariance structure of fGn, the Hurst exponent correctly parametrizes the long-range correlations existing in fGn. For the short-range correlated OU and CIR processes, if we use a relaxation time that is large relative to the analyzed sample size, we observe variance contributions at scales the size of the data sample, and this can result in $H > 0.5$ (Fig. 8). The problem can be partially resolved by considering larger samples. However, for any sample size we can encounter exponentially correlated processes with relaxation times that are too large for the chosen size. Also, in practice, sample size can be limited due to stationarity or data quality issues. In the case of resting state EEG data, vigilance changes are a common nonstationarity [57] and electrode or eye-blink artifacts can severely affect data quality. The duration of the EEG recordings used by us are within the range of those in previous studies reporting long-range correlations [24,52,53,58–63]. A notable exception is the study by Kantelhardt [50], where very long EEG recordings are reported and the scaling region was extended to the 50–500 s range. To address longer time scales, we analyzed 50 min fMRI recordings up to time scales of 100 s.

B. Mutual information with Markovian confidence intervals

The mutual information function analyzes temporal dependencies in time series by computing a dissimilarity measure between the independent and the dependent symbol distributions for X_t and X_{t+k}. The analytical form of the mutual information function for a first-order Markov process [Eq. (9)] and empirical confidence intervals are derived. True LRD data, i.e., fGn sample paths, show a linear (Fig. 11) or slightly convex shape (Fig. 10) of the mutual information function in log-log coordinates. Comparing the mutual information functions of the processes with known autocorrelation properties, as shown in Fig. 9, we observe that the short-range correlated OU and CIR processes show a (reverse) sigmoid shape markedly different from the fGn case.

To distinguish short-range from long-range memory, we construct confidence intervals based on a null hypothesis of short-range, Markovian correlations. In the case of fGn, the contrast between the long-range memory present in fGn and the Markovian memory decay of the confidence interval is clearly seen in Fig. 10(a). The figure also shows that OU and CIR processes are correctly characterized as short range correlated. For EEG data, we find that a simple first-order Markov process characterizes the real data only insufficiently. The more flexible AR(10) model, however, yields a confidence interval enclosing the empirical EEG mutual information function.

With Hurst exponent estimation, it was not possible to distinguish these cases as all data gave values indicating long-range correlations ($H > 0.5$). Analyzing the discrepancy between the empirical mutual information function and the Markov confidence interval, we find that it is possible to distinguish short-range from long-range dependence statistically, even in cases of slow relaxation times and finite sample size.

C. LRD in neural data

Applying the mutual information approach to experimental EEG and fMRI resting state data from healthy subjects, we consistently reproduce large Hurst exponents ($H > 0.5$), but at the same time find mutual information functions approaching those computed from Markovian surrogates. In the case of EEG, the tested signals were (a) the analytical amplitude of alpha frequency band oscillations, representing the most prominent EEG feature in the wakeful resting state condition, and (b) EEG microstate sequences computed from the same data sets, which capture the temporal dynamics of the spatial EEG profile after data compression to four representative topographies of the electrical potential. The fMRI signal is recorded from the same subjects and from the same brain regions where alpha rhythm generators are localized. For all modalities, long-range memory effects based on Hurst exponent estimates $H > 0.5$ have been reported. Our results confirm these previous findings of $H \approx 0.6$–0.8 [5,22,27,40], showing that large Hurst exponents seem to be a consistent phenomenon in neural data. On the other hand, our findings question the long-range memory interpretation when looking at the information content of the time series directly. The Shannon entropy of the neural signals decays over time, finally as fast as a Markovian process mimicking other signal properties such as underlying oscillations and bursty amplitude dynamics. Furthermore, our results still hold when restricted to time scales beyond those affected by preprocessing, as shown in Figs. 10 and 12. The fact that short-range correlated AR(10) models produce the same memory structure as experimental data, although these models have an integrable autocorrelation function, shows that Hurst exponents alone provide only partial evidence for LRD.

Analyzing the analytical amplitude of EEG signals visually (Fig. 4), we observe that resting state alpha activity occurs in irregular bursts which produce a complex, multi-scale variance structure of the signal envelope. This contribution of large variance terms at different time scales contributes to scaling exponents $H > 0.5$. Visually and quantitatively, neural signals behave differently from fGn samples, and therefore, fGn properties such as long-range correlations cannot always be inferred directly from empirical Hurst exponents of empirical data.

VI. CONCLUSIONS

Long-range memory is an exceptional feature generated by certain stochastic processes and by real-world complex systems. Stochastic processes such as fractional Gaussian noise can be constructed in such a way that Hurst exponents, power-law exponents and fractal dimension are connected via simple algebraic relations. While the resulting fGn process is stationary, i.e. the variance is distributed uniformly along the time axis, real-world signals often exhibit more irregular, bursty, and otherwise nonstationary properties. We present two elementary stochastic processes (OUP, CIR) whose Hurst exponents do not predict the long-range information content of the process correctly. This observation is not completely unexpected as whole families of nonstationary stochastic processes with Hurst exponents $H > 0.5$ but Markovian memory structure have been introduced in Ref. [17]. In those cases, $H > 0.5$ is not due to suboptimal estimation but represents an analytical property of the Markovian process.

The main message of our work is that the interpretation of memory effects in real-world signals may benefit from information-theoretical analyses, in addition to Hurst exponent estimation. We show that the time-lagged mutual information function is a useful method to measure memory effects. Moreover, the null hypotheses used to construct confidence intervals should capture as many features of the experimental time series

as possible, in the case of EEG including oscillations with bursty amplitude dynamics, while guaranteeing short-range correlations. To this end, we used the AR(10) model which is known to yield a satisfying parametrization of EEG data in other contexts [28,29].

To put our results into context, it is interesting to observe that our results are similar to previous information-theoretical analyses of EEG data using the diffusion entropy method, where short-range correlations and spurious LRD results were observed [64,65]. In two recent studies, we could show that Hurst exponents and memory effects of EEG microstate sequences describe two different aspects of the time series, and that nonstationarity and slow relaxation affect Hurst exponents [41,55].

The fact that several experimental modalities representing the same neural process (resting state activity in parieto-occipital regions) show converging results considering memory effects, suggests that the patterns observed describe some biologically robust phenomenon. While Hurst exponents are a useful measure to characterize different physiological states, as shown multiple times [24,27,37,50,58,66], we propose to add complementary information about the time series' memory structure using information theoretical measures. The additional approach should complement and not substitute existing methodologies, and thus add to our understanding of the complex dependencies apparent in neural signals.

[1] M. B. Weissmann, Rev. Mod. Phys. **60**, 537 (1988).

[2] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008).

[3] P. C. Hohenberg and B. I. Halperin, Rev. Mod. Phys. **49**, 435 (1977).

[4] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).

[5] K. Linkenkaer-Hansen, V. V. Nikouline, J. M. Palva, and R. J. Ilmoniemi, J. Neurosci. **21**, 1370 (2001).

[6] F. Freyer, K. Aquino, P. A. Robinson, P. Ritter, and M. Breakspear, J. Neurosci. **29**, 8512 (2009).

[7] P. Paramanathan and R. Uthayakumar, Comput. Biol. Med. **38**, 372 (2008).

[8] T. Gneiting, H. Ševčíková, and D. B. Percival, Statist. Sci. **27**, 247 (2012).

[9] J. Beran, Statist. Sci. **7**, 404 (1992).

[10] Rangarajan and Ding, Phys. Rev. E **61**, 4991 (2000).

[11] J. Gao, J. Hu, W.-W. Tung, Y. Cao, N. Sarshar, and V. P. Roychowdhury, Phys. Rev. E **73**, 016117 (2006).

[12] B. B. Mandelbrot and J. W. Van Ness, SIAM Rev. **10**, 422 (1968).

[13] T. Graves, R. B. Gramacy, N. Watkins, and C. L. E. Franzke, Entropy **19**(9), 437 (2017).

[14] B. Mandelbrot, Phys. Scr. **32**, 257 (1985).

[15] T. Karagiannis, M. Faloutsos, and R. H. Riedi, IEEE Globecom **3**, 2165 (2002).

[16] D. Maraun, H. W. Rust, and J. Timmer, Nonlin. Process. Geophys. **11**, 495 (2004).

[17] K. E. Bassler, G. H. Gunaratne, and J. L. McCauley, Physica A **369**, 343 (2006).

[18] J. L. McCauley, G. H. Gunaratne, and K. E. Bassler, Physica A **379**, 1 (2007).

[19] D. Sornette, *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization, and Disorder: Concepts and Tools*, Physics and Astronomy Online Library (Springer-Verlag, Berlin, Heidelberg, 2004).

[20] G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930).

[21] J. C. Cox, J. E. Ingersoll, Jr., and S. A. Ross, Econometrica **53**, 385 (1985).

[22] V. V. Nikulin and T. Brismar, Clin. Neurophysiol. **115**, 1896 (2004).

[23] V. V. Nikulin and T. Brismar, Neuroscience **130**, 549 (2005).

[24] L. Berthouze, L. M. James, and S. F. Farmer, Clin. Neurophysiol. **121**, 1187 (2010).

[25] E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. A. Carpenter, and M. Brammer, Hum. Brain Mapp **12**, 61 (2001).

[26] S. Thurner, C. Windischberger, E. Moser, P. Walla, and M. Barth, Physica A **326**, 511 (2003).

[27] P. Ciuciu, G. Varoquaux, P. Abry, S. Sadaghiani, and A. Kleinschmidt, Front. Physiol. **3**, 186 (2012).

[28] J. J. Wright, R. R. Kydd, and A. A. Sergejew, Biol. Cybern. **62**, 201 (1990).

[29] J. Pardey, S. Roberts, and L. Tarassenko, Med. Eng. Phys. **18**, 2 (1996).

[30] J. Beran, *Statistics for Long-memory Processes*, Monographs on statistics and applied probability (Chapman & Hall, New York, 1994).

[31] Gillespie, Phys. Rev. E **54**, 2084 (1996).

[32] A. Alfonsi, Monte Carlo Methods Appl. **11**, 355 (2018).

[33] D. Lehmann, H. Ozaki, and I. Pal, Electroencephalogr. Clin. Neurophysiol. **67**, 271 (1987).

[34] M. M. Murray, D. Brunet, and C. M. Michel, Brain Topogr. **20**, 249 (2008).

[35] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, IEEE Trans. Biomed. Eng. **42**, 658 (1995).

[36] E. Tagliazucchi, F. von Wegner, A. Morzelewski, V. Brodbeck, S. Borisov, K. Jahnke, and H. Laufs, NeuroImage **70**, 327 (2013).

[37] E. Tagliazucchi, F. von Wegner, A. Morzelewski, V. Brodbeck, K. Jahnke, and H. Laufs, Proc. Natl. Acad. Sci. U.S.A. **110**, 15419 (2013).

[38] K. J. Blinowska and M. Malinowski, Biol. Cybern. **66**, 159 (1991).

[39] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch, Wavelets for the analysis, estimation, and synthesis of scaling data, in *Self-Similar Network Traffic and Performance Evaluation* (John Wiley & Sons, Inc., New York, 2002), pp. 39–88.

[40] D. Van de Ville, J. Britz, and C. M. Michel, Proc. Natl. Acad. Sci. U.S.A. **107**, 18179 (2010).

[41] F. von Wegner, E. Tagliazucchi, V. Brodbeck, and H. Laufs, NeuroImage **141**, 442 (2016).

[42] C. K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, Nature **356**, 168 (1992).

[43] S. Kullback, *Information Theory and Statistics* (Dover Publications, Inc., Mineola, NY, 1959).

[44] S. Kullback, M. Kupperman, and H. H. Ku, Technometrics **4**, 573 (1962).

[45] D. Freedman and P. Diaconis, Z. Wahrscheinlichkeitstheorie verw Gebiete **57**, 453 (1981).

[46] D. W. Scott, Biometrika **66**, 605 (1979).

[47] H. A. Sturges, J. Am. Stat. Assoc. **21**, 65 (1926).

[48] A. Kraskov, H. Stögbauer, and P. Grassberger, Phys. Rev. E **69**, 066138 (2004).

[49] O. Häggström, *London Mathematical Society Student Texts* (Cambridge University Press, Cambridge, 2002).

[50] J. W. Kantelhardt, S. Tismer, F. Gans, A. Y. Schumann, and T. Penzel, Europhys. Lett. **112**, 18001 (2015).

[51] H. Laufs, J. L. Holt, R. Elfont, M. Krams, J. S. Paul, K. Krakow, and A. Kleinschmidt, NeuroImage **31**, 1408 (2006).

[52] J.-S. Lee, B.-H. Yang, J.-H. Lee, J.-H. Choi, I.-G. Choi, and S.-B. Kim, Clin. Neurophysiol. **118**, 2489 (2007).

[53] S. Karkare, G. Saha, and J. Bhattacharya, Chaos, Solitons Fractals **42**, 2067 (2009).

[54] P. Abry and D. Veitch, IEEE Trans. Inf. Theor. **44**, 2 (1998).

[55] F. von Wegner, E. Tagliazucchi, and H. Laufs, NeuroImage **158**, 99 (2017).

[56] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, NeuroImage **15**, 273 (2002).

[57] E. Tagliazucchi and H. Laufs, Neuron **82**, 695 (2014).

[58] T. Bojic, A. Vuckovic, and A. Kalauzi, J. Theor. Biol. **262**, 214 (2010).

[59] M. Buiatti, D. Papo, P.-M. Baudonnière, and C. van Vreeswijk, Neuroscience **146**, 1400 (2007).

[60] M. A. Colombo, Y. Wei, J. R. Ramautar, K. Linkenkaer-Hansen, E. Tagliazucchi, and E. J. W. V. Someren, Frontiers Physiol. **7**, 576 (2016).

[61] R. C. Hwa and T. C. Ferree, Phys. Rev. E Stat. Nonlin Soft Matter. Phys. **66**, 021901 (2002).

[62] T. Inouye, S. Ukai, K. Shinosaki, A. Iyama, Y. Matsumoto, and S. Toi, Neurosci. Lett. **174**, 105 (1994).

[63] T. Montez, S.-S. Poil, B. F. Jones, I. Manshanden, J. P. A. Verbunt, B. W. van Dijk, A. B. Brussaard, A. van Ooyen, C. J. Stam, P. Scheltens, and K. Linkenkaer-Hansen, Proc. Natl. Acad. Sci. U.S.A. **106**, 1614 (2009).

[64] M. Ignaccolo, M. Latka, W. Jernajczyk, P. Grigolini, and B. J. West, Phys. Rev. E **81**, 031909 (2010).

[65] M. Ignaccolo, M. Latka, W. Jernajczyk, P. Grigolini, and B. J. West, J. Biol. Phys. **36**, 185 (2010).

[66] N. W. Churchill, R. Spring, C. Grady, B. Cimprich, M. K. Askren, P. A. Reuter-Lorenz, M. S. Jung, S. Peltier, S. C. Strother, and M. G. Berman, Sci. Rep. **6**, 30895 (2016).