



# Partial Autoinformation to Characterize Symbolic Sequences

Frederic von Wegner<sup>1,2\*</sup>

<sup>1</sup> Epilepsy Center Rhein-Main, Goethe University Frankfurt, Frankfurt am Main, Germany, <sup>2</sup> Department of Neurology and Brain Imaging Center, Goethe University Frankfurt, Frankfurt am Main, Germany

## OPEN ACCESS

### Edited by:

Srdjan Kesic,  
University of Belgrade, Serbia

### Reviewed by:

Hussein M. Yahia,  
Institut National de Recherche en  
Informatique et en Automatique  
(INRIA), France  
Luca Faes,  
Università degli Studi di Palermo, Italy

### \*Correspondence:

Frederic von Wegner  
vonWegner@med.uni-frankfurt.de

### Specialty section:

This article was submitted to  
Fractal Physiology,  
a section of the journal  
Frontiers in Physiology

**Received:** 30 May 2018

**Accepted:** 11 September 2018

**Published:** 11 October 2018

### Citation:

von Wegner F (2018) Partial  
Autoinformation to Characterize  
Symbolic Sequences.  
*Front. Physiol.* 9:1382.  
doi: 10.3389/fphys.2018.01382

An information-theoretic approach to numerically determine the Markov order of discrete stochastic processes defined over a finite state space is introduced. To measure statistical dependencies between different time points of symbolic time series, two information-theoretic measures are proposed. The first measure is time-lagged mutual information between the random variables  $X_n$  and  $X_{n+k}$ , representing the values of the process at time points  $n$  and  $n + k$ , respectively. The measure will be termed autoinformation, in analogy to the autocorrelation function for metric time series, but using Shannon entropy rather than linear correlation. This measure is complemented by the conditional mutual information between  $X_n$  and  $X_{n+k}$ , removing the influence of the intermediate values  $X_{n+k-1}, \dots, X_{n+1}$ . The second measure is termed partial autoinformation, in analogy to the partial autocorrelation function (PACF) in metric time series analysis. Mathematical relations with known quantities such as the entropy rate and active information storage are established. Both measures are applied to a number of examples, ranging from theoretical Markov and non-Markov processes with known stochastic properties, to models from statistical physics, and finally, to a discrete transform of an EEG data set. The combination of autoinformation and partial autoinformation yields important insights into the temporal structure of the data in all test cases. For first- and higher-order Markov processes, partial autoinformation correctly identifies the order parameter, but also suggests extended, non-Markovian effects in the examples that lack the Markov property. For three hidden Markov models (HMMs), the underlying Markov order is found. The combination of both quantities may be used as an early step in the analysis of experimental, non-metric time series and can be employed to discover higher-order Markov dependencies, non-Markovianity and periodicities in symbolic time series.

**Keywords:** EEG microstates, information theory, entropy, mutual information, Markovianity, stationarity

## 1. INTRODUCTION AND BACKGROUND

Information theory occupies a central role in time series analysis. The concept of entropy provides numerous important connections to statistical physics and thermodynamics, often useful in the interpretation of the results (Kullback, 1959; Cover and Thomas, 2006). Despite the large number of available measures, there is no generally accepted systematic procedure for the analysis of

symbolic time series, although collections of theory and methods are readily available (Daw et al., 2003; Mézard and Montanari, 2009). In metric time series analysis however, a hierarchical set of analyses and tests has been established by Box and Jenkins (Box and Jenkins, 1976). The seminal work by these authors deals with autoregressive and moving average processes, some of the most prominent Markov processes across many fields of science (Häggström, 2002). The result is a standardized procedure for analyzing continuous valued, discrete time stochastic processes (Box and Jenkins, 1976). The procedure addresses the impressive complexity of possible stochastic processes by combining semi-quantitative, visual analysis steps with a number of rigorous statistical test procedures. The first step in Box-Jenkins analysis is the visual and statistical assessment of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the data. In particular, the order of purely autoregressive processes can be directly deduced from the PACF coefficients. For a  $p$ -th order autoregressive process, it can be shown that PACF coefficients for time lags larger than  $p$  are equal to zero, within statistical limits.

Information-theoretical time series analysis is closely linked to the theory of Markov processes, as Markov processes are defined via their temporal dependencies. The elemental case is a first-order Markov process  $(X_n)_{n \in \mathbb{Z}}$ , for which the transition  $X_n \rightarrow X_{n+1}$  depends on the current state  $X_n$  only. Due to this property, first-order Markov processes are termed memory-less, as their past does not influence transitions to future states. Markov processes of order  $M$  generalize this property and have transition probabilities defined by  $M$  past states  $(X_n, \dots, X_{n-M+1})$ , thus representing finite memory effects of the time series. Using information-theoretical methods, memory effects and temporal dependencies can be quantified. A more formal treatment of Markov processes follows in the Materials and Methods section, after introducing the necessary notation.

To assess the Markov property of a time series, and to identify the Markov order of an empirical symbol sequence, classical statistics derives a number of tests as detailed in the Materials and Methods section (Hoel, 1954; Anderson and Goodman, 1957; Goodman, 1958; Billingsley, 1961; Kullback et al., 1962). We recently used the tests developed in (Kullback et al., 1962) to characterize electroencephalographic (EEG) data transformed into a symbolic time series termed microstate sequences (von Wegner and Laufs, 2018). Using time-lagged mutual information, we could show that the symbolic four-state sequences still retain periodic features of the underlying continuous EEG signal (von Wegner et al., 2017, 2018). The aim of the present article is to introduce partial autoinformation as a measure that is complementary to the time-lagged mutual information function, in the same sense that the PACF complements the autocorrelation function in classical time series analysis of continuous random variables. In the past, we have used the term autoinformation function (AIF) for time-lagged mutual information, in analogy with the autocorrelation function (ACF) of classical time series analysis (von Wegner et al., 2017). To continue the analogy, the newly introduced measure will

be termed partial autoinformation function (PAIF), because it answers the same question about the information content of a symbolic time series as the partial autocorrelation function (PACF) does about correlations. The new measure is derived based on the analogy with the PACF and theoretical connections with well-known functionals such as the entropy rate and active information storage are established. Next, we apply the AIF/PAIF approach to a number of symbolic time series ranging from Markov and non-Markov model data with known properties to simulated data representing physical systems (Ising model, abstract ion channel model) and experimental EEG microstate data. Finally, limitations and possible applications are discussed for larger state spaces and finite samples.

## 2. MATERIAL AND METHODS

### 2.1. Autoregressive Processes

To illustrate the motivation for this study, an exemplary autoregressive process is used to explain the principles of time series analysis with the (partial) autocorrelation approach. Autoregressive (AR) processes model time series of continuous random variables in discrete time (Box and Jenkins, 1976). The  $p$ -th order or AR( $p$ ) process models the dependency of  $X_n$  on its past via a linear combination of the  $p$  values preceding  $X_n$ :

$$X_n = \phi_1 X_{n-1} + \dots + \phi_p X_{n-p} + \varepsilon_n \tag{1}$$

where  $\phi_1, \dots, \phi_p$  are called the autoregression coefficients and  $\varepsilon_n$  represents identically and independently distributed (iid) Gaussian noise.

The linear dependencies created by Equation (1) can be quantified by the time autocorrelation function (ACF). The ACF coefficients  $\rho_k$  of a stationary stochastic process  $X_n$  are defined as:

$$\rho_k = C(X_{n+k}, X_n) \tag{2}$$

where  $C(X, Y) = \sum_i \frac{(X_i - \mu_X)(Y_i - \mu_Y)}{\sigma_X \sigma_Y}$  denotes Pearson's correlation coefficient. The ACF coefficients describe the linear correlation between process values at two different time steps  $X_n$  and  $X_{n+k}$ , without taking into account the effect of the intermediate time steps  $X_{n+k-1}, \dots, X_{n+1}$ . However,  $X_n$  could be correlated with  $X_{n+k}$  directly, independent of the intermediate values, or the correlation between  $X_n$  and  $X_{n+k}$  could be conveyed via the intermediate values and vanish when conditioned on these intermediates. To distinguish these cases, the PACF performs a multivariate regression of  $X_{n+k}$  on all values  $X_{n+k-1}, \dots, X_n$  and finally records the conditioned or partial correlation between  $X_n$  and  $X_{n+k}$ , removing the effect of the intermediate values:

$$\varphi_{kk} = C(X_{n+k}, X_n | X_{n+k-1}, \dots, X_{n+1}). \tag{3}$$

Continuous-valued, discrete time AR processes can be systematically assessed using the combination of the autocorrelation function and the PACF (Box and Jenkins, 1976).

We here present an example using a third-order autoregressive process that is parametrized by:

$$X_n = 0.85X_{n-1} - 0.2X_{n-2} + 0.1X_{n-3} + \varepsilon_n \quad (4)$$

with AR coefficients  $\phi_1 = 0.85$ ,  $\phi_2 = -0.2$  and  $\phi_3 = 0.1$ . **Figure 1** shows the ACF/PACF analysis of a simulated sample path ( $N = 10^5$  samples). The left panel, **Figure 1A** shows the exponentially decaying ACF. Though the analytical form of the ACF can be expressed in terms of the three AR coefficients, visual analysis does not allow to deduce the order of the AR process or the magnitude of the coefficients. The right panel, **Figure 1B** shows the PACF whose zero-lag coefficient  $\varphi_{00} = 1$ , by definition. The following three values  $\varphi_{11-33}$  directly reflect the relative magnitude and sign of the AR coefficients  $\phi_{1-3}$ . Thus, through PACF analysis, the AR(3) structure of the process can be conjectured from visual analysis already. In practice, the statistical significance of each coefficient can also be assessed quantitatively. In **Figure 1**, confidence intervals ( $\alpha = 0.05$ ) are shown in blue.

The classical Box-Jenkins approach to time series analysis considers the magnitude of ACF and PACF coefficients to guess the statistical structure of the data (Box and Jenkins, 1976). For a pure autoregressive process of order  $p$ , the PACF coefficients  $\varphi_{kk}$  vanish for  $k > p$ . In case of a pure moving average process, the expected value of the ACF coefficients  $\rho_k$  are zero for  $k > p$ . For mixed (ARMA) processes, the model orders cannot be determined visually. Although the pure AR model order can be deduced from the decay of the PACF, and the PACF coefficients  $\varphi_{kk}$  can be expressed in terms of the AR coefficients  $\phi_k$ , the exact value of the AR coefficients  $\phi_k$  cannot be derived visually, with the exception of a few simple low-order cases.

## 2.2. Information Theory

Information theory is rooted in mathematical statistics and uses entropy as one of its main concepts (Kullback, 1959). Entropy characterizes the shape of probability distributions and thereby, the amount of uncertainty or surprise associated with samples generated from the distribution. This section summarizes the concepts and definitions needed to derive the PAIF, more extensive treatments can be found in classical and more recent monographs (Kullback, 1959; Cover and Thomas, 2006; Mézard and Montanari, 2009). Connections of the PAIF with other information-theoretical quantities are derived in the first paragraph of the Results section. Logarithms are computed as  $\log_2$ , such that all information-theoretical quantities are measured in bits.

We here consider stochastic processes  $(X_n)_{n \in \mathbb{Z}}$ , i.e., sequences of random variables  $X_n$ , where each  $X_n$  takes values in some finite alphabet of  $L$  different symbols  $S = \{s_0, \dots, s_{L-1}\}$ . In practice, we deal with finite samples of the underlying process,  $(X_n)_{n=0, \dots, N-1}$ . In the following, contiguous data blocks starting from index  $n$ , and covering the past  $k$  values of the process,  $(X_n, X_{n-1}, \dots, X_{n-k+1})$  will be termed  $k$ -histories and are written as

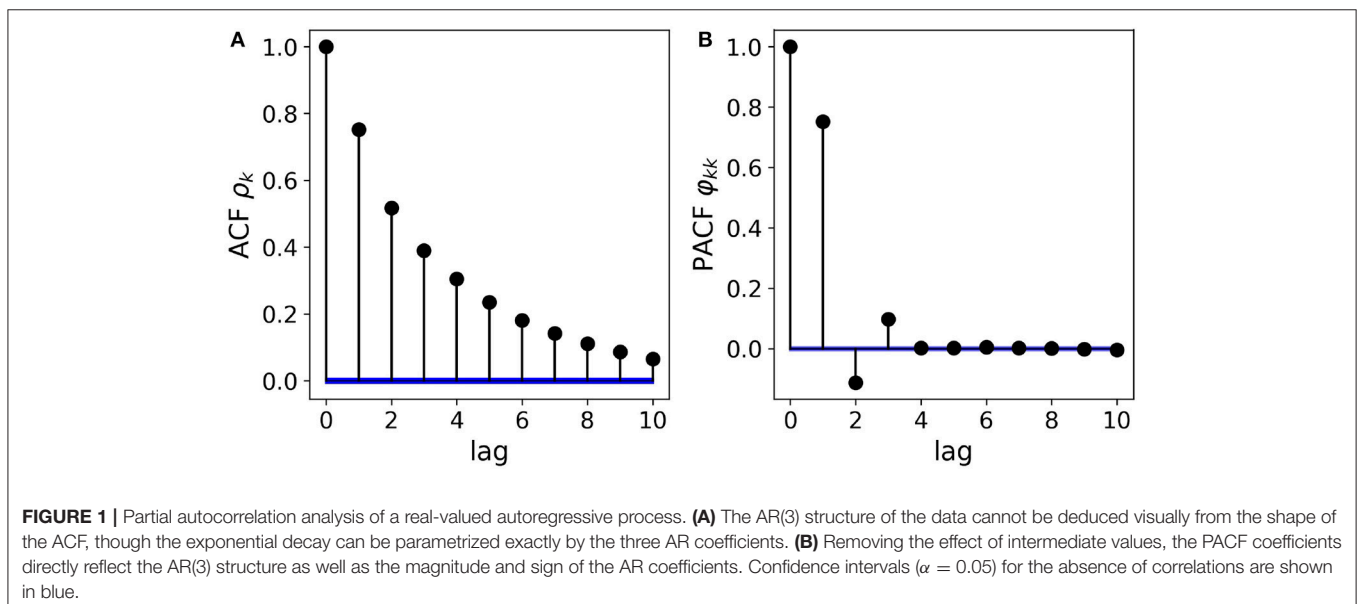
$$\mathbf{X}_n^{(k)} = (X_n, X_{n-1}, \dots, X_{n-k+1}). \quad (5)$$

Denoting a specific realization of the random variable  $X_i$  as  $x_i$ , the joint probability distribution of  $k$ -histories is given by

$$P(\mathbf{X}_n^{(k)}) = \Pr(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_{n-k+1} = x_{n-k+1}).$$

where  $x_i \in S$ , for all  $i = n - k + 1, \dots, n$ . In the following, the compact notation  $P(\mathbf{X}_n^{(k)})$  will be used.

The information content of a random variable  $X$  with possible values  $x_i \in S$  and associated probabilities  $P(X =$



$x_i) = p_i$  is measured by the **Shannon entropy**  $H(X) = -\sum_i p_i \log p_i$  (Kullback, 1959). The information content of the joint distribution representing the  $k$ -history  $\mathbf{X}_n^{(k)}$  is measured by the **joint entropy**, which is defined as (Kullback, 1959; Cover and Thomas, 2006):

$$H_X(n, k) := H(\mathbf{X}_n^{(k)}) \tag{6}$$

$$= -\sum_{\mathbf{x}_n^{(k)}} P(\mathbf{X}_n^{(k)}) \log P(\mathbf{X}_n^{(k)}) \tag{7}$$

where the sum runs over all possible values of  $X_n = x_n, \dots, X_{n-k+1} = x_{n-k+1}$ . The expression  $H_X(n, k)$  contains the time parameter  $n$ , such that the expression for  $H_X(n, k)$  can be used even in the case of non-stationary processes, whose statistical properties may depend on  $n$ . Under time-stationary conditions, the entropy is obtained by averaging over all time points and the resulting entropy will be abbreviated

$$H_k = \langle H_X(n, k) \rangle_n$$

where  $\langle \cdot \rangle_n$  denotes time averaging.

Adding information about the value of another random variable  $Y$  reduces the uncertainty about  $X$ , in case  $X$  and  $Y$  are statistically dependent. If  $X$  and  $Y$  are independent, the entropy of  $X$  does not change with the additional information about  $Y$ . To measure the influence of  $Y$  on  $X$ , **conditional entropy** is defined as  $H(X | Y) = H(X, Y) - H(Y)$ . In the following, two conditional entropy terms will be used.

The first term is a finite approximation to the entropy rate of the stochastic process  $X$ . The **entropy rate**  $h_X$  of a process quantifies the amount of surprise about the next symbol  $X_{n+1}$  emitted by the process, given knowledge about its past values  $\mathbf{X}_n^{(k)}$ . The theoretical or analytical value  $h_X$  is defined via an infinitely long history

$$h_X = \lim_{k \rightarrow \infty} H(X_{n+1} | \mathbf{X}_n^{(k)}).$$

When working with finite experimental data samples, the entropy rate has to be estimated from finite  $k$ -histories (Runge et al., 2012; Barnett and Seth, 2015; Faes et al., 2015; Xiong et al., 2017):

$$h_X(n, k) = H(X_{n+1} | \mathbf{X}_n^{(k)}). \tag{8}$$

Using the definition of conditional entropy,  $h_X(n, k)$  can be computed from joint entropies as:

$$h_X(n, k) = H(X_{n+1}, \mathbf{X}_n^{(k)}) - H(\mathbf{X}_n^{(k)}) \tag{9}$$

$$= H(\mathbf{X}_{n+1}^{(k+1)}) - H(\mathbf{X}_n^{(k)}). \tag{10}$$

Following the notation used for Shannon entropy, the time-stationary expression for the entropy rate will be denoted  $h_k = \langle h_X(n, k) \rangle_n$ .

The second conditional entropy term used is the **two-point conditional entropy**  $H(X_{n+k} | X_n)$ , that measures the amount of information about  $X_{n+k}$  contained in  $X_n$ .

Next, **mutual information** between two random variables is defined as  $I(X; Y) = H(X) - H(X | Y)$  and measures the information shared between both variables. Mutual information will be used to compute two quantities that are useful in characterizing symbol sequences.

First, **active information storage (AIS)** (Lizier et al., 2012) is complementary to the entropy rate. While the entropy rate measures how much information (or surprise) is contained in  $X_{n+1}$ , despite knowledge of its  $k$ -history  $\mathbf{X}_n^{(k)}$ , AIS measures the amount of common (or shared) information between  $X_{n+1}$  and its  $k$ -history. The active information storage term for a history of length  $k$  is defined as

$$a_X(n, k) = I(X_{n+1}; \mathbf{X}_n^{(k)}) \tag{11}$$

and the stationary expression is  $a_k = \langle a_X(n, k) \rangle_n$ .

For computational implementation, active information storage is decomposed into joint entropy terms:

$$\begin{aligned} I(X_{n+1}; \mathbf{X}_n^{(k)}) &= H(X_{n+1}) - H(X_{n+1} | \mathbf{X}_n^{(k)}) \\ &= H(X_{n+1}) + H(\mathbf{X}_n^{(k)}) - H(X_{n+1}, \mathbf{X}_n^{(k)}) \\ &= H(X_{n+1}) + H(\mathbf{X}_n^{(k)}) - H(\mathbf{X}_{n+1}^{(k+1)}). \end{aligned}$$

The second mutual information term used is  $I(X_{n+1}; X_k)$ , and yields an estimate of the statistical dependency between the random variables  $X_n$  and  $X_{n+k}$ . In a recent publication, we used the term **autoinformation function (AIF)** to denote the set of time-lagged mutual information terms computed for a number of time lags (von Wegner et al., 2017). The name AIF was derived from the formal analogy with the autocorrelation function (ACF) for metric time series. We defined the AIF coefficient at time lag  $k$  as:

$$\alpha_X(n, k) = I(X_{n+k}; X_n) \tag{12}$$

$$= H(X_{n+k}) - H(X_{n+k} | X_n) \tag{13}$$

$$= H(X_{n+k}) + H(X_n) - H(X_{n+k}, X_n) \tag{14}$$

and the stationary term is  $\alpha_k = \langle \alpha_X(n, k) \rangle_n$ . Rather than using linear correlation to measure the dependency between two time points, as the ACF does, the AIF employs mutual information between the random variables at time points  $n$  and  $n + k$ . The measure is symmetric, i.e.,  $I(X_n; X_{n+k}) = I(X_{n+k}; X_n)$  and therefore does not contain directional information. In analogy to the autocorrelation function, division of all coefficients by  $\alpha_X(n, 0)$  normalizes the AIF to  $\alpha_X(n, 0) = 1$ . The computational cost is independent of the time lag  $k$ , as all entropies are computed from one-dimensional ( $H(X_{n+k}), H(X_n)$ ) and two-dimensional ( $H(X_{n+k}, X_n)$ ) distributions.

Finally, the definition of partial autoinformation, the central concept of this work, is based on the concept of **conditional mutual information** which includes a third random variable

Z, on which the mutual information between X and Y is conditioned:

$$I(X; Y | Z) = H(X | Z) - H(X | Y, Z). \quad (15)$$

The information-theoretical match for the PACF should estimate the two-point dependency between  $X_n$  and  $X_{n+k}$ , while removing the influence of the intermediate variables  $\mathbf{X}_{n+k-1}^{(k-1)} = (X_{n+k-1}, \dots, X_{n+1})$ . This is achieved by computing the conditional mutual information (Equation 15) between  $X_n$  and  $X_{n+k}$ , given  $\mathbf{X}_{n+k-1}^{(k-1)}$ .

We therefore define the PAIF coefficient  $\pi_X(n, k)$  at time lag k as:

$$\pi_X(n, k) = I(X_{n+k}; X_n | X_{n+k-1} \dots X_{n+1}) \quad (16)$$

$$= I(X_{n+k}; X_n | \mathbf{X}_{n+k-1}^{(k-1)}). \quad (17)$$

Using the definition of conditional mutual information in terms of conditional entropies, and the expression of conditional entropy in terms of joint entropies, the computation of  $\pi_X(n, k)$  can be reduced to the estimation of joint entropies:

$$\pi_X(n, k) = H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k-1)}) - H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k-1)}, X_n) \quad (18)$$

$$= H(X_{n+k}, \mathbf{X}_{n+k-1}^{(k-1)}) - H(\mathbf{X}_{n+k-1}^{(k-1)}) - H(X_{n+k}, \mathbf{X}_{n+k-1}^{(k-1)}) + H(\mathbf{X}_{n+k-1}^{(k-1)}) \quad (19)$$

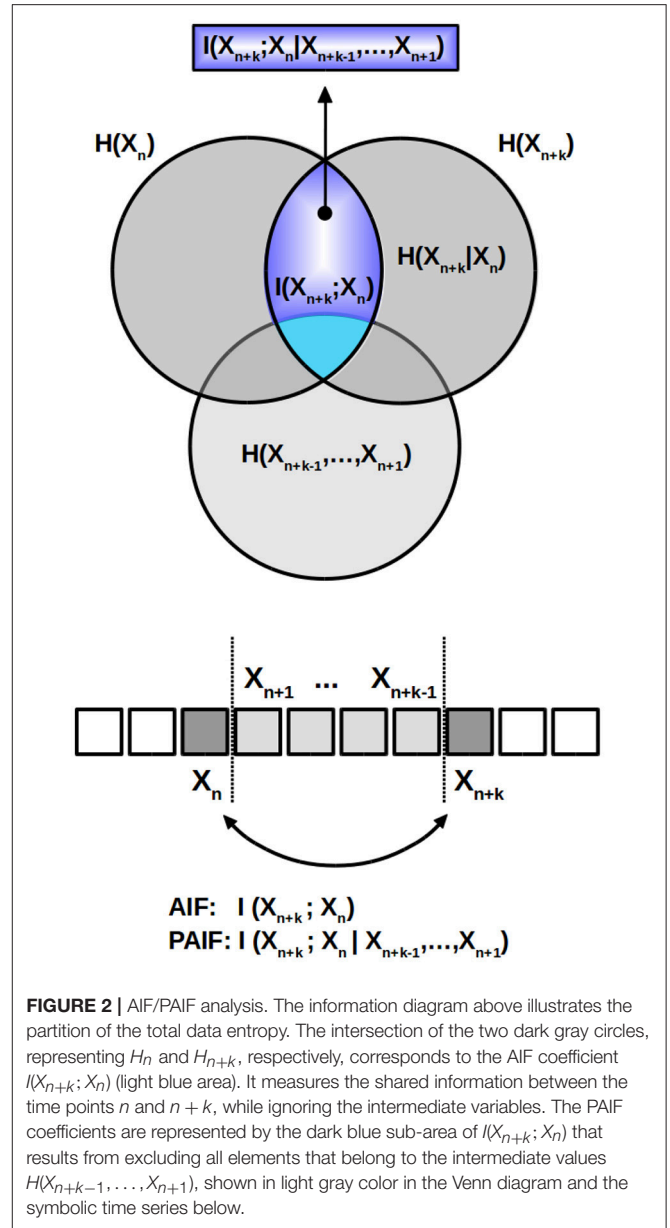
$$= H(\mathbf{X}_{n+k}^{(k)}) - H(\mathbf{X}_{n+k-1}^{(k-1)}) - H(\mathbf{X}_{n+k}^{(k)}) + H(\mathbf{X}_{n+k-1}^{(k-1)}). \quad (20)$$

The stationary expression is

$$\begin{aligned} \pi_k &= h_{k-1} - h_k \\ &= -H_{k-1} + 2H_k - H_{k+1}. \end{aligned}$$

For the first two coefficients  $\pi_X(n, 0)$  and  $\pi_X(n, 1)$ , there are no intermediate values  $\mathbf{X}_{n+k-1}^{(k-1)}$  to condition on. Analogous to the PACF algorithm, we set  $\pi_X(n, 0) = \alpha_X(n, 0)$  and  $\pi_X(n, 1) = \alpha_X(n, 1)$ . The computational load increases exponentially with history length k, as the discrete joint distribution  $P(\mathbf{X}_{n+k-1}^{(k-1)})$  over L labels has  $L^{k-1}$  elements.

The computation of the quantity of interest, the PAIF coefficients, is visualized in Figure 2. Above, the relationships between the quantities discussed here are shown as an information diagram, a special form of a Venn diagram. AIF coefficients are represented by the intersection of the two dark gray circles which represent  $H(X_{n+k})$  and  $H(X_n)$ , respectively. In the scheme below, where each element of the time series  $X_n$  is visualized as a square box, the AIF coefficients represent the shared information between  $X_{n+k}$  and  $X_n$ , without taking into account the effects of  $\mathbf{X}_{n+k-1}^{(k-1)}$  (light gray areas in the information diagram above and the symbolic sequence below). The PAIF corresponds to the part of  $I(X_{n+k}; X_n)$  that does not intersect with the lower circle, representing the intermediate values  $H(\mathbf{X}_{n+k-1}^{(k-1)})$ . The area that represents the PAIF is shown in dark blue in the information scheme.



**FIGURE 2 |** AIF/PAIF analysis. The information diagram above illustrates the partition of the total data entropy. The intersection of the two dark gray circles, representing  $H_n$  and  $H_{n+k}$ , respectively, corresponds to the AIF coefficient  $I(X_{n+k}; X_n)$  (light blue area). It measures the shared information between the time points  $n$  and  $n + k$ , while ignoring the intermediate variables. The PAIF coefficients are represented by the dark blue sub-area of  $I(X_{n+k}; X_n)$  that results from excluding all elements that belong to the intermediate values  $H(X_{n+k-1}, \dots, X_{n+1})$ , shown in light gray color in the Venn diagram and the symbolic time series below.

### 2.3. Markovianity Tests

A discrete Markov process  $(X_n)_{n \in \mathbb{Z}}$  of order  $M$  is defined via the property

$$P(X_{n+1} | \mathbf{X}_n^{(M+k)}) = P(X_{n+1} | \mathbf{X}_n^{(M)}) \quad (21)$$

for all positive integers  $k \geq 0$ . In words, the transition probabilities from  $X_n$  to the state  $X_{n+1}$  depend on the  $M$ -history of  $X_n$ , whereas inclusion of more values from the process' past, beyond  $X_{n-M+1}$ , does not convey further information about the transition probabilities.

General tests for the Markov property of low orders have been introduced in the 1950s and further tests for many

special cases are still being developed today. Early works used analytical expressions for the distribution of symbol counts, given a certain Markov structure, and developed likelihood ratio tests for the cases of known (Bartlett, 1951) and unknown (Hoel, 1954) transition probabilities. Further developments included  $\chi^2$  tests for hypotheses about the time-stationarity of transition probabilities, direct comparisons of different Markov orders (Anderson and Goodman, 1957; Goodman, 1958), as well as parameter estimation methods and tests for continuous time Markov processes (Billingsley, 1961). Using close relationships between  $\chi^2$  statistics and information-theoretical expressions, test statistics based on Kullback-Leibler metrics were summarized as a monograph and in a practice-oriented article containing many numerical examples by Kullback (Kullback, 1959; Kullback et al., 1962). Further approaches include the application of the Akaike information criterion to optimize the order estimate for a discrete Markov chain (Tong, 1975) as well as data compression oriented algorithms (Merhav et al., 1989) and extensive surrogate data tests (Pethel and Hahs, 2014). The results of the PAIF method developed here is compared to the Markov order test presented in van der Heyden et al. (1998). The latter test compares finite entropy rate estimates ( $h_X(n, k)$ ) of the data to be tested with surrogate statistics obtained from  $M$ -order Markov surrogates with the same transition probabilities  $P(X_{n+1} | \mathbf{X}_n^{(M)})$  as the data. The algorithm for the computation of the surrogates is given in detail in (van der Heyden et al., 1998) and is summarized in the following section 2.4. This test will be termed conditional entropy test. In this article, surrogate statistics for each test data set are computed from  $n = 100$  surrogate sequences for each Markov order  $M = 0, \dots, 5$ . The Markov order identified by the conditional entropy test is taken to be the value  $M$  for which all  $h_X(n, k)$  lie within the  $\alpha = 0.05$  confidence interval defined by the surrogates.

We recently published our Python implementation of the Markovianity tests of order 0-2 as well as symmetry and stationarity tests as given in Kullback et al. (1962), in article form (von Wegner and Laufs, 2018), and as open-source code. Although the code is part of an algorithm to process EEG microstate sequences, the tests can be exported and applied generically.

### 2.4. Markov Surrogate Data

A Markov process of order  $M$  is also defined via its transition probabilities  $P(X_{n+1} | \mathbf{X}_n^{(M)})$ , where the probability to go into state  $X_{n+1}$  is conditioned on the  $M$ -history  $\mathbf{X}_n^{(M)}$ . To synthesize a Markov process of order  $M$ , using the same transition probabilities as the underlying experimental time series ( $X_n$ ), the empirical  $M$ -order transition matrix is estimated first. To this end, all contiguous tuples of length  $(M + 1)$  taken from the time series, i.e., tuples of the form  $(X_{n-M+1}, \dots, X_n, X_{n+1})$  are considered. The maximum likelihood estimate for the transition probability  $P(X_{n+1} | \mathbf{X}_n^{(M)})$  based on this sample is given by

$$\hat{p}_{ML}(X_{n+1} | \mathbf{X}_n^{(M)}) = \frac{\#(X_{n+1}, \mathbf{X}_n^{(M)})}{\#(\mathbf{X}_n^{(M)})}$$

where  $\#(\cdot)$  denotes the number of times a specific outcome occurs in the empirical sequence  $(X_n)$  (Anderson and Goodman, 1957; van der Heyden et al., 1998). For instance,  $\#(\mathbf{X}_n^{(M)})$  is the number of realizations  $(X_n = x_n, \dots, X_{n-M+1} = x_{n-M+1})$ . While counting the tuples, the joint distribution of  $\mathbf{X}_n^{(M)}$  is recorded at the same time.

Following van der Heyden et al. (1998), the first  $M$  values of each surrogate Markov sequence are initialized with a sample from the joint distribution  $\mathbf{X}_n^{(M)}$ . From there, we have a  $M$ -history  $\mathbf{X}_n^{(M)}$  for every subsequent value  $X_{n+1}$ . The value of  $X_{n+1}$  is chosen according to the transition probabilities  $\hat{p}_{ML}(X_{n+1} | \mathbf{X}_n^{(M)})$  and the given  $M$ -history. Given a specific  $M$ -history  $\mathbf{X}_n^{(M)}$ , there are  $L$  transition probabilities  $q_0, \dots, q_{L-1}$ , where  $q_i = \hat{p}_{ML}(X_{n+1} = s_i | \mathbf{X}_n^{(M)})$ . The distribution of the state  $X_{n+1} = s_i$  is sampled correctly using a pseudo-random number  $r$ , uniformly distributed on the unit interval,  $r \sim \mathcal{U}_{[0,1]}$ , and the condition  $\sum_{l=0}^{i-1} q_l \leq r < \sum_{l=0}^i q_l$ .

We recently published a Python implementation for first-order Markov surrogates in the open-source package described in von Wegner and Laufs (2018), and have included the  $M$ -order Markov surrogates in the Github repository associated with this paper.

### 2.5. The Two-State Markov Process

The general concepts introduced above are easily applied to a two-state, first-order Markov process that can be written as

$$A \xrightleftharpoons[p]{q} B$$

with transition rates  $p$  and  $q$ . The self-transition rate for  $A \rightarrow A$  is  $1 - p$ , and the rate of  $B \rightarrow B$  is  $1 - q$ . The complete transition matrix  $T$  reads

$$T = \begin{pmatrix} 1 - p & p \\ q & 1 - q \end{pmatrix}$$

and has eigenvalues  $\lambda_0 = 1$  and  $\lambda_1 = 1 - (p + q)$ . The eigenvalue  $\lambda_0 = 1$  is assured by the Perron-Frobenius theorem as  $T$  is a stochastic matrix, i.e.,  $\sum_j T_{ij} = 1$  for all  $i$ . The normalized positive eigenvector to  $\lambda_0$  is the equilibrium or stationary distribution  $p_{st}$  of the process,

$$p_{st} = \left( \frac{q}{p + q}, \frac{p}{p + q} \right).$$

We set  $p_A = \frac{q}{p+q}$  and  $p_B = \frac{p}{p+q}$ . With the auxiliary functions  $\varphi, \psi : [0, 1] \rightarrow \mathbb{R}$  defined as  $\varphi(x) = -x \log x$  and  $\psi(x) = \varphi(x) + \varphi(1 - x)$ , the analytical quantities  $H_{pq}, h_{pq}$  and  $a_{pq}$  for the 2-state first-order Markov process acquire a very simple form.

The Shannon entropy of the 2-state Markov process is

$$\begin{aligned} H_{pq} &= -p_A \log p_A - p_B \log p_B \\ &= \varphi(p_A) + \varphi(1 - p_A) \\ &= \psi(p_A). \end{aligned}$$

Due to the Markov property, the entropy rate is  $h_{pq} = H(X_{n+1} | X_n)$  and evaluates to

$$h_{pq} = -p_A [(1 - p) \log(1 - p) + p \log p] - p_B [q \log q + (1 - q) \log(1 - q)] = p_A \psi(p) + p_B \psi(q).$$

The Markov property reduces the full expression for information storage  $I(X_{n+1}; \mathbf{X}_n^{(k)})$  to  $a_{pq} = I(X_{n+1}; X_n)$ :

$$a_{pq} = -p_A \log p_A - p_B \log p_B + p_A [(1 - p) \log(1 - p) + p \log p] + p_B [q \log q + (1 - q) \log(1 - q)] = \psi(p_A) - p_A \psi(p) - p_B \psi(q).$$

The total entropy is conserved between active information storage and the entropy rate:

$$H_{pq} = a_{pq} + h_{pq}. \tag{22}$$

To validate the proposed approach, these analytical results will later be compared with numerical results of a hidden Markov process classified as first-order Markovian by the PAIF method.

### 2.6. Higher-Order Markov Processes

To test the properties of the PAIF, two higher-order Markov processes with known properties are synthesized.

The first process is a third-order Markov process denoted  $MC_{1..M=3}$ , with transition probabilities that depend on  $\mathbf{X}_n^{(M=3)}$ . Given  $L$  states, there are  $L^M$  possible  $M$ -histories preceding  $X_{n+1}$ , such that  $P(X_{n+1} | \mathbf{X}_n^{(M)})$  in matrix form has shape  $(L^M, L)$ . The specific transition probabilities are random numbers fulfilling  $\sum_j P(X_{n+1} = s_j | \mathbf{X}_n^{(M)}) = 1$  for all  $M$ -histories  $\mathbf{X}_n^{(M)}$ . Sample paths are generated using the method described in section 2.4.

The second process will be termed  $MC_{M=3}$  and is constructed in such a way that the  $X_n \rightarrow X_{n+1}$  transition only depends on  $X_{n-2}$ . Like the first process, this process can also be classified as third-order Markovian ( $M = 3$ ), with the particular property that the influence of  $X_n$  and  $X_{n-1}$  vanishes.

### 2.7. Hidden Markov Processes

A more general class of discrete processes is represented by probabilistic finite state machines (Crutchfield and Young, 1989), which implement hidden Markov models (HMMs). Hidden Markov models generate sequences of symbols defined over a set of observable states that correspond to our measurements. The observable symbols are emitted by a set of hidden states that follow a Markov process, usually of first order. Each hidden state emits the observable symbols according to its own probability distribution defined over the observable set. It is important to note that the sequence of emitted symbols does not necessarily follow a Markov law.

#### 2.7.1. Even Process

The even process is a non-Markov process with two hidden states ( $\{A, B\}$ ) and two observables ( $\{0, 1\}$ ). The process scheme is visualized in **Figure 4A**. The process can emit arbitrarily long sequences of zeros by repeated self-transitions of the hidden state  $A \rightarrow A$ . With probability  $p = 0.5$ , the state  $A$  can switch to  $B$  and

hereby emit a 1, which is followed by another 1 with probability  $p = 1$ . Thus, ones are always generated in pairs, i.e., in blocks of even length. The procedure generates dependencies that in theory reach into the infinite past and can therefore not be reduced to a Markov process.

#### 2.7.2. Golden Mean Process

Two different implementations of the Golden-mean process are used. First, a 2-state first-order Markovian implementation using two hidden states ( $\{A, B\}$ ) and two observable states ( $\{0, 1\}$ ) (Ara et al., 2016), and second, a fourth-order Markov implementation using seven hidden ( $\{A-G\}$ ) and two observable states ( $\{0, 1\}$ ) (Mahoney et al., 2016). The scheme of the 2-state process (**Figure 4B**) is structurally similar to the even process, but dynamically different. Ones are never emitted repeatedly, i.e., they are always preceded and followed by a zero, in contrast to the even process. The 7-state golden mean process is a so-called  $R, k$ -Markov process with Markov order  $R = 4$  and cryptic order  $k = 3$ , in our case (Mahoney et al., 2016).

### 2.8. Ising Model Data

The Ising model is a widely used discrete model from statistical physics (Hohenberg and Halperin, 1977). The model describes the ferromagnetic interaction of elementary spin variables, with two possible values  $\pm 1$ , as a function of temperature and the coupling coefficients between spins. The model can be realized with different geometries and in many cases, shows a phase transition at a critical temperature. We use a 2D square lattice geometry ( $L = 50$ ) and run the system for  $10^6$  time steps. Sample paths are generated by Monte Carlo simulation using a standard Gibbs sampling scheme (Bortz et al., 1975).

### 2.9. Simulated Ion Channel Data

The dynamics of a simple ion channel with one open and one closed state is modeled as a motion of a particle in the double-well potential  $V(x) = -\frac{a}{2}x^2 + \frac{b}{4}x^4$ , which shows two stable local minima at  $x_{1,2} = \pm\sqrt{\frac{a}{b}}$  and one unstable local maximum at  $x_0 = 0$  (Liebovitch and Czegledy, 1992; von Wegner et al., 2014). The system is excited by thermal noise, as implemented by iid Gaussian pseudo-random numbers  $\xi_n$ . The system is described by von Wegner et al. (2014)

$$X_{n+1} = X_n + (aX_n - bX_n^3)dt + \xi_n \tag{23}$$

and integrated with an Euler scheme and  $dt = 10^{-3}$ .

### 2.10. EEG Microstate Sequences

A resting state EEG data set from a 21 year old, healthy right-handed female during wakeful rest was selected and analyzed. The data set is part of a larger database for which we have reported the detailed pre-processing pipeline before (von Wegner et al., 2016, 2017). The 30 channel EEG raw data was sampled at 5 kHz using the standard 10 – 10 electrode configuration, band-pass filtered to the 1 – 30 Hz range using a zero-phase Butterworth filter with a slope of 24 dB/octave, down-sampled to 250 Hz and re-referenced to an average reference. Written informed consent was obtained from the subject and the study

was approved by the ethics committee of the Goethe University, Frankfurt, Germany. EEG microstates were identified using the first four principal components (PCA analysis) of the data set and the symbolic microstate sequence was obtained by competitively fitting the microstate maps back into the EEG data set as detailed in (von Wegner et al., 2016, 2017).

### 3. RESULTS

#### 3.1. Theoretical Results: Relations Between Measures

Using the time index  $n + k$  as a reference, the partial autoinformation coefficients  $\pi_X(n, k) = I(X_{n+k}; X_n | \mathbf{X}_{n+k-1}^{(k-1)})$  can be related to the entropy rate  $h_X(n + k - 1, k) = H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k)})$  and to active information storage  $a_X(n + k - 1, k) = I(X_{n+k}; \mathbf{X}_{n+k-1}^{(k)})$  as follows.

The entropy rate  $h_X(n + k - 1, k)$  can be written as the difference of two joint entropies of different lengths (Equation 9),  $h_X(n + k - 1, k) = H(\mathbf{X}_{n+k}^{(k+1)}) - H(\mathbf{X}_{n+k-1}^{(k)})$ .

Next, active information storage can be expressed as the difference of a joint entropy and the entropy rate:

$$\begin{aligned} a_X(n + k - 1, k) &= I(X_{n+k}; \mathbf{X}_{n+k-1}^{(k)}) \\ &= H(X_{n+k}) - H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k)}) \\ &= H(X_{n+k}) - h_X(n + k - 1, k). \end{aligned}$$

In the stationary case, we have

$$H_1 = a_k + h_k. \tag{24}$$

Similar to the case presented for the 2-state Markov process, it is observed that also in the general case, the entropy  $H(X_{n+k})$  is conserved, being the sum of active information storage and the entropy rate. In words, the information about the future state  $X_{n+k}$  is the sum of the actively stored information from time step  $n$  up to time step  $n + k - 1$ , and the entropy rate between time steps  $n + k - 1$  to  $n + k$ .

Finally, the PAIF coefficient  $\pi_X(n, k)$  can be written as the difference of entropy rates for different history lengths:

$$\begin{aligned} \pi_X(n, k) &= I(X_{n+k}; X_n | \mathbf{X}_{n+k-1}^{(k-1)}) \\ &= H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k-1)}) - H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k)}) \\ &= h_X(n + k - 1, k - 1) - h_X(n + k - 1, k). \end{aligned}$$

Alternatively,  $\pi_X(n, k)$  can also be decomposed into a difference of AIS terms for two different history lengths:

$$\begin{aligned} \pi_X(n, k) &= I(X_{n+k}; X_n | \mathbf{X}_{n+k-1}^{(k-1)}) \\ &= H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k-1)}) - H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k-1)}, X_n) \\ &= H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k-1)}) - H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k)}) \\ &= H(X_{n+k}) - H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k)}) - [H(X_{n+k}) \\ &\quad - H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k-1)})] \end{aligned}$$

$$\begin{aligned} &= I(X_{n+k}; \mathbf{X}_{n+k-1}^{(k)}) - I(X_{n+k}; \mathbf{X}_{n+k-1}^{(k-1)}) \\ &= a_X(n + k - 1, k) - a_X(n + k - 1, k - 1). \end{aligned}$$

Going from line 3 to line 4, we simply added and subtracted  $H(X_{n+k})$ . In words, the PAIF at time lag  $k$  is the difference between two AIF terms with history lengths  $k$  and  $k - 1$ , respectively.

The results can be summarized in a more compact form using the stationary expressions:

$$\pi_k = a_k - a_{k-1} \tag{25}$$

$$= h_{k-1} - h_k. \tag{26}$$

For stationary Markov processes, the joint Shannon entropy  $H_k$  exists and the  $k$ -order entropy rate estimates  $h_k$  converge in the limit of  $k \rightarrow \infty$  (Cover and Thomas, 2006). Using Equation 24, it follows that the AIS coefficients  $a_k$  also converge. Thus,  $\lim_{k \rightarrow \infty} a_k - a_{k-1} = 0$  and  $\lim_{k \rightarrow \infty} h_{k-1} - h_k = 0$ . Using Equation 25, we deduce that the PACF coefficients  $\pi_k$  also vanish in the large  $k$  limit:

$$\lim_{k \rightarrow \infty} \pi_k = 0.$$

##### 3.1.1. Markovianity

Using the Markov property defined in Equation 21, it is straightforward to prove that for a stationary Markov process of order  $M$ , the PAIF coefficients vanish ( $\pi_X(n, k) = 0$ ) for  $k > M$ :

$$\begin{aligned} \pi_X(n, k) &= I(X_{n+k}; X_n | \mathbf{X}_{n+k-1}^{(k-1)}) \\ &= H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k-1)}) - H(X_{n+k} | \mathbf{X}_{n+k-1}^{(k)}) \\ &= h_{k-1} - h_k \\ &= h_M - h_M \\ &= 0. \end{aligned}$$

Let the first- and second-order finite differences of an arbitrary discrete function  $f_k$  of integer parameter  $k$  be defined as  $\delta_k f_k = f_k - f_{k-1}$ , and  $\delta_k^2 f_k = f_{k+1} - 2f_k + f_{k-1}$ , then we get

$$\begin{aligned} \pi_k &= h_{k-1} - h_k \\ &= -\delta_k h_k \\ &= -\delta_k^2 H_k. \end{aligned}$$

Thus, the Markovianity test proposed in (van der Heyden et al., 1998) addresses a sequence of entropy rates  $h_k$ , for different history lengths  $k$ , which is the negative first-order difference of the sequence of Shannon entropies  $H_k$ . PAIF analysis uses the second-order difference of the sequence of Shannon entropies  $\pi_k = -\delta_k^2 H_k$ . The advantage of the PAIF analysis is the visual exploration of the coefficients, that are equal to zero for  $k > M$ , exactly like in visual PACF diagnostics for metric time series.



### 3.2. Higher-Order Markov Processes

The results for the two third-order Markov processes are shown in **Figures 3A–D**. The AIF and PAIF for the third-order Markov processes  $MC_{1\dots M=3}$  is shown in **Figures 3A,B**, respectively. The shape of the AIF does not reveal the third-order dependencies by visual inspection. The PAIF, however, clearly reflects the construction of the process, showing significant PAIF coefficients only up to time lag  $k = 3$ .

For the second process,  $MC_{M=3}$ , the entropy dynamics can already be estimated by visual inspection of the AIF, which shows a clear periodicity (**Figure 3C**). Significant PAIF coefficients only occur at time lags that are multiples of the Markov order,  $M = 3$ . The PAIF (**Figure 3D**) however, demonstrates the Markov structure of the process in a single significant coefficient  $\pi_3$ .

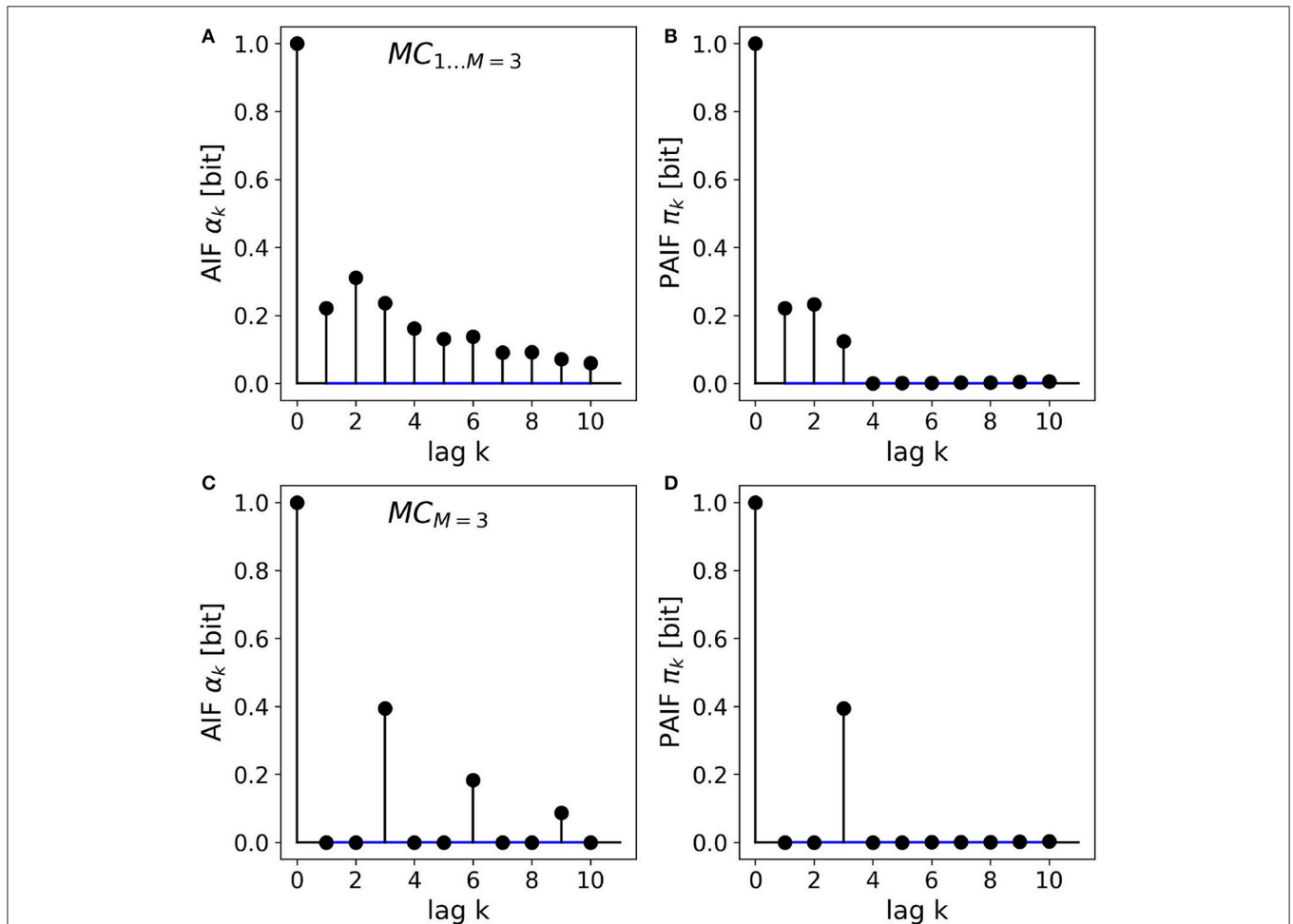
Kullback’s Markovianity tests of order 0-2 rejected the Markovian null hypotheses for both processes, as expected for Markov processes of order three, by construction. The conditional entropy test correctly identified the Markov order  $M = 3$  in both cases.

Confidence intervals ( $\alpha = 0.05$ ) constructed from uncorrelated surrogate time series are shown in blue. Due to their small magnitude, they visually appear as lines.

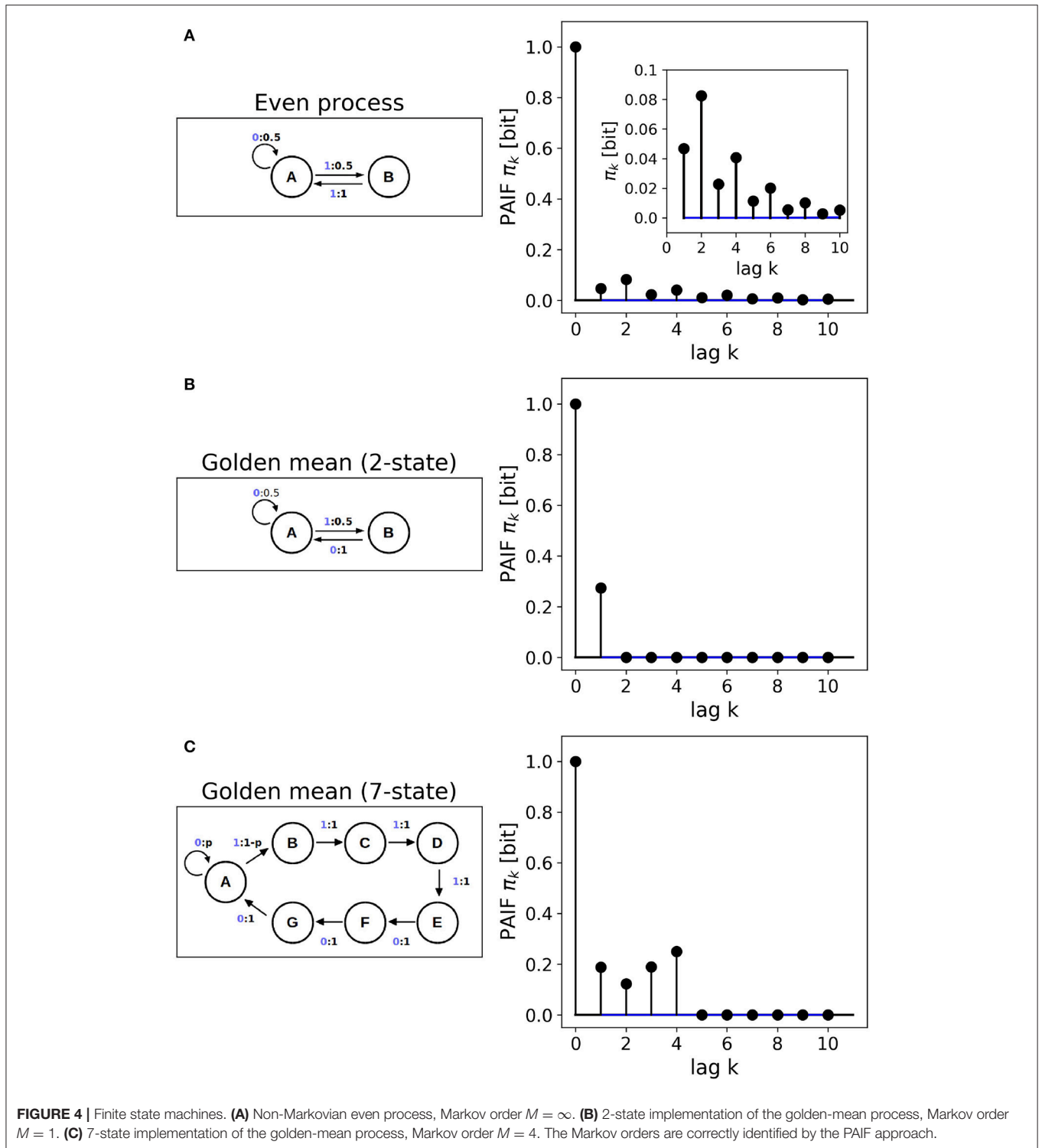
### 3.3. Hidden Markov Models

**Figure 4** shows the results obtained from HMM data. First, the non-Markovian even process is analyzed. To the right of the HMM scheme, **Figure 4A** shows the PAIF of a single sample path of length  $n = 10^6$ . The inset shows that for all tested time lags the PAIF coefficients lie above the iid confidence interval (blue lines). Thus, PAIF analysis suggests that we are observing a non-Markov process with extended memory effects.

**Figure 4B** shows the PAIF of the 2-state golden mean process. The PAIF has two significant coefficients  $\pi_0, \pi_1$  and decays to zero for all other time lags. The PAIF thus classifies the process correctly as a first-order Markov process, despite the hidden Markov implementation. Due to the Markov property, the process can also be represented by a transition matrix and



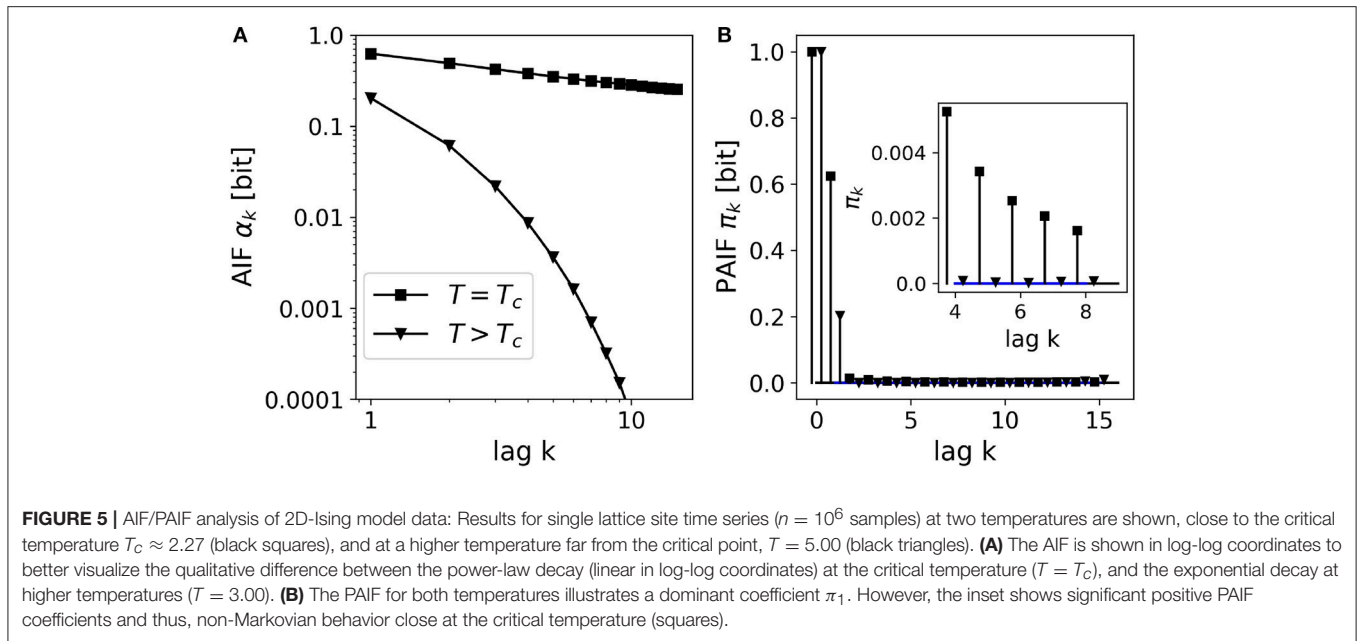
**FIGURE 3 |** AIF/PAIF analysis of two third-order Markov process samples. **(A)** The AIF of the  $MC_{1\dots M=3}$  process slowly decays toward zero and does not reveal the Markov order of the process. **(B)** The PAIF of  $MC_{1\dots M=3}$  shows a cutoff after  $k = 3$  coefficients, in accordance with the nominal Markov order. **(C)** The AIF of the  $MC_{M=3}$  process has period 3 and thus hints at the memory structure of the process. **(D)** The PAIF of the  $MC_{M=3}$  process clearly identifies the Markov order of the process by a distinct peak at the time lag corresponding to the correct model order  $M = k = 3$ .



an equilibrium distribution. The associated transition matrix  $T$  is

$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}$$

with stationary distribution  $p_{st} = [\frac{2}{3}, \frac{1}{3}]$ . Using these quantities, the theoretical results from Section 2.5 can be applied. Using finite histories ( $k = 2 \dots 10$ ), entropy conservation (Equation 24) is fulfilled with a maximum error of  $7.25 \times 10^{-4}$ , where the error was calculated as  $\frac{H_1 - h_k - a_k}{H_1}$ . Based on this analysis, the Shannon



entropy of a single symbol is  $H_1 = 0.919$  bit, and consists of an entropy rate of  $h_X = 0.669$  bit and active information storage of  $a_X = 0.253$  bit.

The 7-state HMM of the golden mean process is analyzed in **Figure 4C**. The fourth-order Markov structure of the implementation is clearly reflected by the PAIF that shows four positive coefficients. The PAIF captures the correct Markov order although the model contains seven hidden states and emits two observable symbols.

Kullback’s Markovianity tests of order 0–2 correctly classified the 2-state golden mean process as first-order Markovian. The p-values for orders 0-2 were  $p_0 = 0.000$ ,  $p_1 = 0.697$ , and  $p = 0.990$ , respectively. For the non-Markovian even process and the fourth-order Markovian 7-state golden mean process, low-order (0-2) Markovianity was correctly rejected. The conditional entropy test correctly identified the Markov properties of all three processes, i.e., found first- and fourth-order properties for the 2-state and 7-state golden mean processes, respectively, and an order  $M > 5$  for the even process.

### 3.4. The Ising Model

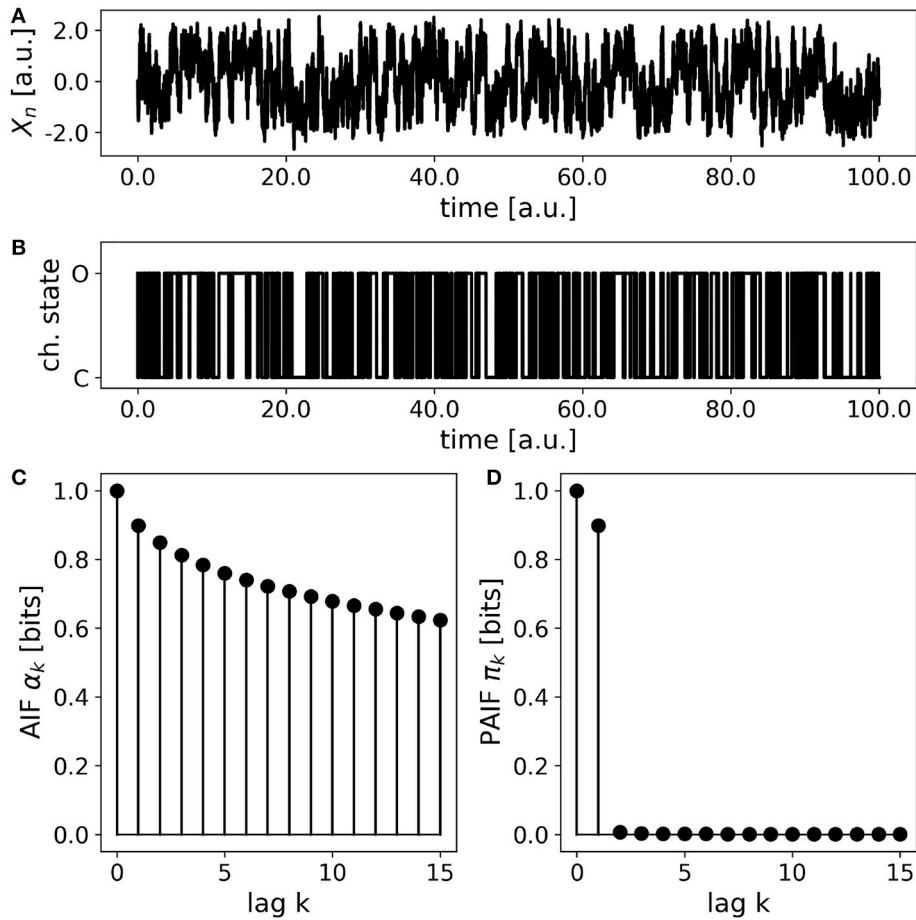
We simulated an Ising model on a 2D lattice ( $50 \times 50$  elements) at two temperatures, (i) around the critical temperature  $T_c = \frac{2}{1+\sqrt{2}} \approx 2.27$ , and (ii) at a higher temperature  $T = 5.00$ . From statistical physics, it is known that the system’s autocorrelation function shows a slow, power-law decay at the critical point, and an exponential decay far from the critical point where dynamics are dominated by thermal fluctuations. AIF/PAIF analysis was performed on time series of  $10^6$  samples of a randomly selected lattice site. The results are shown in **Figure 5**.

In contrast to the other figures in this manuscript, the AIF in **Figure 5A** is shown in log-log coordinates, to better visualize the difference between exponential and power-law behavior. The AIF at the critical point  $T_c$  shows an almost

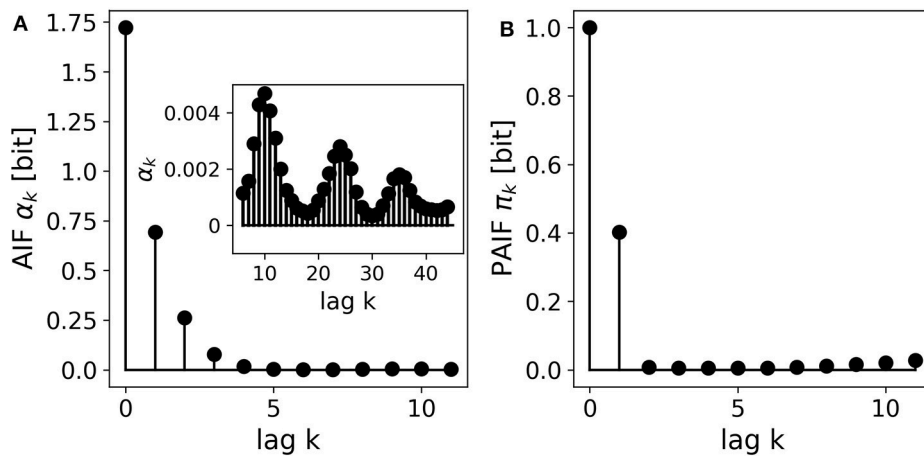
linear behavior in log-log coordinates (black squares), indicating very slow relaxation dynamics, as expected. For the higher temperature, far from the critical point ( $T = 5.0$ , black triangles), however, we observe a quickly decaying autoinformation trace, in accordance with results from classical time series analysis. **Figure 5B** shows the PAIF in linear coordinates, as in all other figures. It is observed that in both cases,  $T = T_c, T = 5.0$ , the PAIF profiles seem to be similar. We find two positive PAIF coefficients  $\pi_0, \pi_1$ , and significantly smaller PAIF coefficients for larger time lags. The inset, however, shows that at the critical temperature (squares), the PAIF coefficients lie above the confidence interval, demonstrating non-Markovian, long-range memory effects where the system undergoes a phase transition.

### 3.5. Simulated Ion-Channel Data

Simplified ion channel dynamics were generated by integration of Equation 23, representing the motion of a particle in a bistable potential, for instance an ion channel with two metastable states corresponding to the open (O) and close (C) state, respectively. To obtain a symbolic time series of O- and C-states, the continuous variable  $X_n$  is thresholded at a value of zero. Thereby, all positive values  $X_n > 0$  are assigned to the open state (O), and all negative values ( $X_n < 0$ ) are mapped to the close state. The AIF/PAIF analysis of the thresholded signal simulating electrophysiological ion channel data is shown in **Figure 6**. We observe a slowly decaying AIF (**Figure 6A**) without any information about the Markov order of the signal. The PAIF profile shows large coefficients  $\pi_0$  and  $\pi_1$ , followed by vanishing PAIF coefficients for  $k > 1$ . Though Markovian dynamics are expected for the continuous dynamics, it is not obvious that the Markov property could be detected after thresholding.



**FIGURE 6 |** Simulated ion channel data. **(A)** A continuous stochastic process  $X_n$  is obtained from a simulation of a double-well potential. A bistable behavior resembling ion channel recordings is observed. **(B)** Thresholding the continuous variable  $X_n$  into an open state (O,  $X_n > 0$ ) and a close state (C,  $X_n < 0$ ) yields a symbolic, binary process. **(C)** The AIF of the binary process shows a slow decay without revealing the Markov order of the process. **(D)** The PAIF suggests first-order Markov dynamics by vanishing PAIF coefficients  $\pi_k$  for  $k > 1$ .



**FIGURE 7 |** A 4-state resting state EEG microstate sequence. **(A)** The AIF shows a monotonous decay for smaller time lags  $k < 10$ . The inset shows the AIF for larger time lags ( $k_{max} = 50$ ) and reveals periodicities that could not be predicted at shorter time scales. **(B)** The PAIF indicates a mainly first-order Markovian structure but does not allow the computation of time lags as large as the AIF due to the exponentially growing size of the associated distributions.

### 3.6. EEG Microstate Sequences

The EEG microstate sequence shows a more complex behavior than the other presented examples. **Figure 7A** shows the AIF of the 4-state sequence, which seems to decay monotonously. The inset shows further information for longer time lags up to  $k = 40$ . We observe several periodic peaks at these time lags, an effect that we have discussed in detail in a recent publication (von Wegner et al., 2017). The PAIF in **Figure 7B** shows a dominant coefficient  $\pi_1$ . This finding suggests a mainly first-order Markov mechanism, and does not hint at the periodic behavior found in the AIF. Moreover, we observe increasing PAIF coefficients for larger time lags ( $k > 8$ ). This effect is caused by finite joint entropy estimates, which suffer from an insufficient sample size, given the history length  $k$ , and the size of the state space, here  $L = 4$ .

## 4. DISCUSSION

In the present article, an information-theoretical approach for the early diagnostic steps in symbolic time series analysis is established. In close analogy to classical time series analysis of continuous-valued random variables, a combined approach using two different measures that estimate the dependency between two time points is used. While autoinformation measures the statistical dependence between  $X_n$  and  $X_{n+k}$  directly, partial autoinformation removes the influence of the segment between both time points. The names AIF and PAIF were chosen to represent the close connection with the ACF/PACF approach. We have recently used the AIF to characterize stochastic processes and experimental EEG data (von Wegner et al., 2017, 2018), and the underlying functional can be found under the name of time-lagged mutual information in the literature. Partial autoinformation, however, is not found in the literature, to the best of the author's knowledge. Close connections to the entropy rate and the active information storage of the process, two well-studied information-theoretical quantities (Cover and Thomas, 2006; Lizier et al., 2012), are found and detailed. In particular, the newly introduced PAIF can be expressed either as the difference of two entropy rates with history lengths  $k - 1$  and  $k$ , respectively, or as the difference of two active information storage terms with different history lengths. These relationships also assure that the PAIF coefficients approach zero in the large  $k$  limit.

The ability of the PAIF to identify the order of a stationary Markov process is shown analytically by re-writing the PAIF in terms of conditional entropies. A short proof shows that the PAIF coefficients of a stationary Markov process of order  $M$  are zero ( $\pi_k = 0$ ) for  $k > M$ . The practical performance of the method is validated numerically, using test data with known Markov orders, and by comparison with the results of two other tests (Kullback et al., 1962; van der Heyden et al., 1998). All test examples used in this article are correctly classified by the PAIF approach, in the same way the PACF performs for continuous autoregressive processes. A close relationship between the PAIF

and the conditional entropy test (van der Heyden et al., 1998) is established by re-writing both in terms of joint entropies  $H_k$ . We found that while the conditional entropy test addresses the first-order discrete difference of  $H_k$  with respect to  $k$ , the PAIF actually tests the corresponding second-order discrete derivative. This completes the goal of establishing an information-theoretical tool analogous to classical PACF analysis.

Our experimental data examples also reveal some important limitations of the approach. The PAIF coefficients for the 4-state EEG microstate sequence ( $n = 153,225$  samples) increase for time lags above approximately  $k > 8$ . Comparison with Markov surrogate samples shows that this increase is due to the limited sample size, and is not a feature of the EEG data set (data not shown). The effect is easily understood by a simple numerical example. If for the same data set, we wanted to compute the PAIF coefficients for the same time lags as used in the AIF (**Figure 7**), joint probability distributions with  $L^k$  bins will occur. Thus, to extend the PAIF analysis of a  $L = 4$ -state process to  $k = 50$ , distributions with  $4^{50} > 10^{30}$  elements have to be estimated, clearly exceeding the length of the data sample numerous times. The example also shows that this is an intrinsic limitation of the approach, as it always occurs for information-theoretical quantities involving joint entropies, and is not specific to the PAIF introduced here.

Finally, the present article exclusively deals with discrete stochastic processes. Future investigations should include the corresponding quantities for continuous random variables, and Gaussian processes in particular. For example, it has been shown in the past that for Gaussian random variables, Granger causality is equivalent to transfer entropy (Barnett et al., 2009). By analogy, it can be conjectured that the PAIF and PACF approaches are likely to be related, if not equivalent, for Gaussian processes.

It will be interesting to see further applications of the presented approach to theoretical and experimental data and to investigate further theoretical connections to other quantities already in use.

## AUTHOR CONTRIBUTIONS

FvW designed the study, performed all presented analyses, and wrote the manuscript.

## FUNDING

This work was funded by the Bundesministerium für Bildung und Forschung (grant 01 EV 0703) and LOEWE Neuronale Koordination Forschungsschwerpunkt Frankfurt (NeFF) as well as by the LOEWE program CePTER.

## ACKNOWLEDGMENTS

The author thanks Helmut Laufs for permission to use the EEG data set used in this article. The author also thanks the Reviewers for valuable input and suggestions, in particular the work by van der Heyden et al.

## REFERENCES

- Anderson, T. W., and Goodman, L. A. (1957). Statistical inference about Markov chains. *Ann. Math. Stat.* 28, 89–110.
- Ara, P. M., James, R. G., and Crutchfield, J. P. (2016). Elusive present: hidden past and future dependency and why we build models. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* 93:022143. doi: 10.1103/PhysRevE.93.022143
- Barnett, L., Barrett, A. B., and Seth, A. K. (2009). Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* 103:238701. doi: 10.1103/PhysRevLett.103.238701
- Barnett, L., and Seth, A. K. (2015). Granger causality for state-space models. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* 91:040101. doi: 10.1103/PhysRevE.91.040101
- Bartlett, M. S. (1951). The frequency goodness of fit test for probability chains. *Math. Proc. Cambridge Philos. Soc.* 47, 86–95.
- Billingsley, P. (1961). Statistical methods in Markov chains. *Ann. Math. Stat.* 32, 12–40.
- Bortz, A. B., Kalos, M. H., and Lebowitz, J. L. (1975). A new algorithm for Monte Carlo simulation of Ising spin systems. *J. Comput. Phys.* 17, 10–18.
- Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day. Hoboken, NJ: Wiley.
- Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. New York, NY: Wiley-Interscience.
- Crutchfield, J. P. and Young, K. (1989). Inferring statistical complexity. *Phys. Rev. Lett.* 63, 105–108.
- Daw, C. S., Finney, C. E. A., and Tracy, E. R. (2003). A review of symbolic analysis of experimental data. *Rev. Sci. Instrum.* 74, 915–930. doi: 10.1063/1.1531823
- Faes, L., Porta, A., and Nollo, G. (2015). Information decomposition in bivariate systems: theory and application to cardiorespiratory dynamics. *Entropy* 17, 277–303. doi: 10.3390/e17010277
- Goodman, L. A. (1958). Exact probabilities and asymptotic relationships for some statistics from m-th order Markov chains. *Ann. Math. Stat.* 29, 476–490.
- Hägström, O. (2002). *Finite Markov Chains and Algorithmic Applications*. Cambridge, UK: Cambridge University Press.
- Hoel, P. G. (1954). A test for Markoff chains. *Biometrika* 41, 430–433.
- Hohenberg, P. C., and Halperin, B. I. (1977). Theory of dynamic critical phenomena. *Rev. Mod. Phys.* 49, 435–479.
- Kullback, S. (1959). *Information Theory and Statistics*. Mineola, NY: Dover Publications, Inc.
- Kullback, S., Kupperman, M., and Ku, H. H. (1962). Tests for contingency tables and Markov chains. *Technometrics* 4, 573–608.
- Liebavitch, L. S., and Czegledy, F. P. (1992). A model of ion channel kinetics based on deterministic, chaotic motion in a potential with two local minima. *Ann. Biomed. Eng.* 20, 517–531.
- Lizier, J. T., Prokopenko, M., and Zomaya, A. Y. (2012). Local measures of information storage in complex distributed computation. *Inform. Sci.* 208, 39–54. doi: 10.1016/j.ins.2012.04.016
- Mahoney, J. R., Aghamohammadi, C., and Crutchfield, J. P. (2016). Occam's quantum strop: synchronizing and compressing classical cryptic processes via a quantum channel. *Sci. Rep.* 6:20495. doi: 10.1038/srep20495
- Merhav, N., Gutman, M., and Ziv, J. (1989). On the estimation of the order of a markov chain and universal data compression. *IEEE Trans. Info Theor.* 35, 1014–1019.
- Mézard, M. and Montanari, A. (2009). *Information, Physics, and Computation*. Oxford University Press, Oxford, UK.
- Pethel, S. D., and Hahs, D. W. (2014). Exact significance test for Markov order. *Phys. D* 269, 42–47. doi: 10.1016/j.physd.2013.11.014
- Runge, J., Heitzig, J., Petoukhov, V., and Kurths, J. (2012). Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys. Rev. Lett.* 108:258701. doi: 10.1103/PhysRevLett.108.258701
- Tong, H. (1975). Determination of the order of a markov chain by akaike's information criterion. *J. Appl. Probabil.* 12, 488–497.
- van der Heyden, M. J., Diks, C. G. C., Hoekstra, B. P. T., and DeGoede, J. (1998). Testing the order of discrete markov chains using surrogate data. *Phys. D* 117, 299–313.
- von Wegner, F., and Laufs, H. (2018). Information-theoretical analysis of EEG microstate sequences in Python. *Front. Neuroinform.* 10:30. doi: 10.3389/fninf.2018.00030
- von Wegner, F., Laufs, H., and Tagliazucchi, E. (2018). Mutual information identifies spurious Hurst phenomena in resting state EEG and fMRI data. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* 97:022415. doi: 10.1103/PhysRevE.97.022415
- von Wegner, F., Tagliazucchi, E., Brodbeck, V., and Laufs, H. (2016). Analytical and empirical fluctuation functions of the EEG microstate random walk - short-range vs. long-range correlations. *Neuroimage* 141, 442–451. doi: 10.1016/j.neuroimage.2016.07.050
- von Wegner, F., Tagliazucchi, E., and Laufs, H. (2017). Information-theoretical analysis of resting state EEG microstate sequences - non-Markovity, non-stationarity and periodicities. *Neuroimage* 158, 99–111. doi: 10.1016/j.neuroimage.2017.06.062
- von Wegner, F., Wieder, N., and Fink, R. H. A. (2014). Microdomain calcium fluctuations as a colored noise process. *Front. Genet.* 5:376. doi: 10.3389/fgene.2014.00376
- Xiong, W., Faes, L., and Ivanov, P. C. (2017). Entropy measures, entropy estimators, and their performance in quantifying complex dynamics: effects of artifacts, nonstationarity, and long-range correlations. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* 95:062114. doi: 10.1103/PhysRevE.95.062114

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 von Wegner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.