

# Symbolic transfer entropy rate is equal to transfer entropy rate for bivariate finite-alphabet stationary ergodic Markov processes

Taichi Haruna<sup>1</sup> and Kohei Nakajima<sup>2,3</sup>

<sup>1</sup> Department of Earth & Planetary Sciences, Graduate School of Science, Kobe University, 1-1 Rokkodaicho, Nada, Kobe 657-8501, Japan, e-mail: [tharuna@penguin.kobe-u.ac.jp](mailto:tharuna@penguin.kobe-u.ac.jp)

<sup>2</sup> Department of Informatics, University of Zurich, Andreasstrasse 15, 8050 Zurich, Switzerland, e-mail: [nakajima@ifi.uzh.ch](mailto:nakajima@ifi.uzh.ch)

<sup>3</sup> Department of Mechanical and Process Engineering, ETH Zurich, Leonhardstrasse 27, 8092 Zurich, Switzerland

Received: date / Revised version: date

**Abstract.** *Transfer entropy* is a measure of the magnitude and the direction of information flow between jointly distributed stochastic processes. In recent years, its permutation analogues are considered in the literature to estimate the transfer entropy by counting the number of occurrences of orderings of values, not the values themselves. It has been suggested that the method of permutation is easy to implement, computationally low cost and robust to noise when applying to real world time series data. In this paper, we initiate a theoretical treatment of the corresponding rates. In particular, we consider the *transfer entropy rate* and its permutation analogue, the *symbolic transfer entropy rate*, and show that they are equal for any bivariate finite-alphabet stationary ergodic Markov process. This result is an illustration of the duality method introduced in [T. Haruna and K. Nakajima, Physica D 240, 1370 (2011)]. We also discuss the relationship among the transfer entropy rate, the time-delayed mutual information rate and their permutation analogues.

## 1 Introduction

Quantifying networks of information flows is critical to understand functions of complex systems such as living, social and technological systems. Schreiber [1] introduced the notion of *transfer entropy* to measure the magnitude

and the direction of information flow from one element to another element emitting stationary signals in a given system. It has been used to analyze information flows in real time series data from neuroscience [1–11], and many other fields [12–19].

The notion of *permutation entropy* introduced by Bandt and Pompe [20] has been proved that much of information contained in stationary time series can be captured by counting occurrences of orderings of values, not those of values themselves [21–26]. The method of permutation has been applied across many disciplines [27] and suggested that it is easy to implement, computationally low cost and robust to noise when applying to real world time series data. Among the previous works, one relevant theoretical result to this paper is that the entropy rate [28], which is one of the most fundamental quantities of stationary stochastic processes, is equal to the permutation entropy rate for any finite-alphabet stationary stochastic process [29, 30].

The *symbolic transfer entropy* [31] is a permutation analogue of the transfer entropy and has been used as an efficient and conceptually simple way of quantifying information flows in real time series data [31–34]. Another permutation analogue of the transfer entropy called *transfer entropy on rank vectors* has been introduced to improve the performance of the symbolic transfer entropy [35]. So far, most of the work on permutation analogues of the transfer entropy are in application side. This paper concerns the theoretical relationship among respective rates. In particular, we consider the rate of transfer entropy on rank vectors which we call *symbolic transfer entropy rate* and show that it is equal to the *transfer entropy rate* [36] for any bivariate finite-alphabet stationary ergodic Markov process. We also discuss the relationship

among the transfer entropy rate, the time-delayed mutual information rate and their permutation analogues.

Our approach is based on the duality between values and orderings introduced by the authors [37]. In [37], the excess entropy [38–45], which is an effective measure of complexity of stationary stochastic processes, and its permutation analogue is shown to be equal for any finite-alphabet stationary ergodic Markov process. In this paper, we extend this approach to the bivariate case and address the relationship between the transfer entropy rate and the symbolic transfer entropy rate.

This paper is organized as follows. In Section 2, we introduce the transfer entropy rate and the symbolic transfer entropy rate. We also discuss some combinatorial facts used in later sections. In Section 3, we give a proof of the equality between the transfer entropy rate and the symbolic transfer entropy rate which holds for bivariate finite-alphabet stationary ergodic Markov processes. In Section 4, we discuss the relationship among the transfer entropy rate, the time-delayed mutual information rate and their permutation analogues. Finally, in Section 5, we give concluding remarks.

## 2 Definitions and Preliminaries

Let  $A_n = \{1, 2, \dots, n\}$  be a finite alphabet consisting of natural numbers from 1 to  $n$ . In the following discussion,  $\mathbf{X} \equiv \{X_1, X_2, \dots\}$  and  $\mathbf{Y} \equiv \{Y_1, Y_2, \dots\}$  are jointly distributed finite-alphabet stationary stochastic processes, or equivalently,  $(\mathbf{X}, \mathbf{Y})$  is a bivariate finite-alphabet stationary stochastic process  $\{(X_1, Y_1), (X_2, Y_2), \dots\}$ , where

stochastic variables  $X_i$  and  $Y_j$  take their values in the alphabet  $A_n$  and  $A_m$ , respectively. We use the notation  $X_1^L \equiv (X_1, X_2, \dots, X_L)$  for simplicity. We write  $p(x_1^{L_1}, y_1^{L_2})$  for the joint probability of the occurrence of words  $x_1^{L_1} \equiv x_1 x_2 \dots x_{L_1} \in A_n^{L_1}$  and  $y_1^{L_2} \equiv y_1 y_2 \dots y_{L_2} \in A_m^{L_2}$  for  $L_1, L_2 \geq 1$ .

Originally, the notion of transfer entropy was introduced as a generalization of the entropy rate to bivariate processes [1]. Along this original motivation, here, we do not consider the *transfer entropy* but the *transfer entropy rate* [36] from  $\mathbf{Y}$  to  $\mathbf{X}$  which is defined by

$$t(\mathbf{X}|\mathbf{Y}) \equiv h(\mathbf{X}) - h(\mathbf{X}|\mathbf{Y}), \quad (1)$$

where  $h(\mathbf{X}) \equiv \lim_{L \rightarrow \infty} H(X_1^L)/L$  is the entropy rate of  $\mathbf{X}$ ,  $H(X_1^L) \equiv -\sum_{x_1^L \in A_n^L} p(x_1^L) \log_2 p(x_1^L)$  is the Shannon entropy of the occurrences of words of length  $L$  in  $\mathbf{X}$  and  $h(\mathbf{X}|\mathbf{Y})$  is the *conditional entropy rate* of  $\mathbf{X}$  given  $\mathbf{Y}$  defined by

$$h(\mathbf{X}|\mathbf{Y}) \equiv \lim_{L \rightarrow \infty} H(X_{L+1}|X_1^L, Y_1^L), \quad (2)$$

which always converges.  $t(\mathbf{X}|\mathbf{Y})$  has the properties that (i)  $0 \leq t(\mathbf{X}|\mathbf{Y}) \leq h(\mathbf{X})$  and (ii)  $t(\mathbf{X}|\mathbf{Y}) = 0$  if  $X_1^L$  is independent of  $Y_1^L$  for all  $L \geq 1$ .

In order to introduce the notion of *symbolic transfer entropy rate*, we define a total order on the alphabet  $A_n$  by the usual “less-than-or-equal-to” relationship. Let  $\mathcal{S}_L$  be the set of all permutations of length  $L \geq 1$ . We consider each permutation  $\pi$  of length  $L$  as a bijection on the set  $\{1, 2, \dots, L\}$ . Thus, each permutation  $\pi \in \mathcal{S}_L$  can be identified with the sequence  $\pi(1) \dots \pi(L)$ . The *permutation type*  $\pi \in \mathcal{S}_L$  of a given word  $x_1^L \in A_n^L$  is defined by

re-ordering  $x_1, \dots, x_L$  in ascending order, namely,  $x_1^L$  is of type  $\pi$  if we have  $x_{\pi(i)} \leq x_{\pi(i+1)}$  and  $\pi(i) < \pi(i+1)$  when  $x_{\pi(i)} = x_{\pi(i+1)}$  for  $i = 1, 2, \dots, L-1$ . For example,  $\pi(1)\pi(2)\pi(3)\pi(4)\pi(5) = 41352$  for  $x_1^5 = 24213 \in A_4^5$  because  $x_4 x_1 x_3 x_5 x_2 = 12234$ . The map  $\phi_n : A_n^L \rightarrow \mathcal{S}_L$  sends each word  $x_1^L$  to its unique permutation type  $\pi = \phi_n(x_1^L)$ .

We will use the notions of *rank sequences* and *rank variables* [29]. The *rank sequences* of length  $L$  are words  $r_1^L \in A_L^L$  satisfying  $1 \leq r_i \leq i$  for  $i = 1, \dots, L$ . The set of all rank sequences of length  $L$  is denoted by  $\mathcal{R}_L$ . It is clear that  $|\mathcal{R}_L| = L! = |\mathcal{S}_L|$ . Each word  $x_1^L \in A_n^L$  can be mapped to a rank sequence  $r_1^L$  by defining  $r_i \equiv \sum_{j=1}^i \delta(x_j \leq x_i)$  for  $i = 1, \dots, L$ , where  $\delta(P) = 1$  if the proposition  $P$  is true, otherwise  $\delta(P) = 0$ . We denote this map from  $A_n^L$  to  $\mathcal{R}_L$  by  $\varphi_n$ . It can be shown that the map  $\varphi_n : A_n^L \rightarrow \mathcal{R}_L$  is compatible with the map  $\phi_n : A_n^L \rightarrow \mathcal{S}_L$  in the following sense: there exists a bijection  $\iota : \mathcal{R}_L \rightarrow \mathcal{S}_L$  such that  $\iota \circ \varphi_n = \phi_n$  [37]. The *rank variables* associated with  $\mathbf{X}$  are defined by  $R_i \equiv \sum_{j=1}^i \delta(X_j \leq X_i)$  for  $i = 1, \dots, L$ . In general,  $\mathbf{R} \equiv \{R_1, R_2, \dots\}$  is a non-stationary stochastic process.

The *symbolic transfer entropy rate* from  $\mathbf{Y}$  to  $\mathbf{X}$  is defined by

$$t^*(\mathbf{X}|\mathbf{Y}) \equiv h^*(\mathbf{X}) - h^*(\mathbf{X}|\mathbf{Y}), \quad (3)$$

where  $h^*(\mathbf{X}) \equiv \lim_{L \rightarrow \infty} H^*(X_1^L)/L$  is the permutation entropy rate which is known to exist and is equal to  $h(\mathbf{X})$  [29],

$$H^*(X_1^L) \equiv - \sum_{\pi \in \mathcal{S}_L} p(\pi) \log_2 p(\pi)$$

is the Shannon entropy of the occurrences of permutations of length  $L$  in  $\mathbf{X}$ ,  $p(\pi) = \sum_{\phi_n(x_1^L)=\pi} p(x_1^L)$  and  $h^*(\mathbf{X}|\mathbf{Y})$  is given by

$$h^*(\mathbf{X}|\mathbf{Y}) \equiv \lim_{L \rightarrow \infty} (H^*(X_1^{L+1}, Y_1^L) - H^*(X_1^L, Y_1^L)) \quad (4)$$

if the limit in the right hand side exists. Here,  $H^*(X_1^{L_1}, Y_1^{L_2})$  is defined by

$$H^*(X_1^{L_1}, Y_1^{L_2}) \equiv - \sum_{\pi \in \mathcal{S}_{L_1}, \pi' \in \mathcal{S}_{L_2}} p(\pi, \pi') \log_2 p(\pi, \pi'),$$

where  $p(\pi, \pi') = \sum_{\phi_n(x_1^{L_1})=\pi, \phi_m(y_1^{L_2})=\pi'} p(x_1^{L_1}, y_1^{L_2})$ .

Let  $\mathbf{R}$  and  $\mathbf{S}$  be rank variables associated with  $\mathbf{X}$  and  $\mathbf{Y}$ , respectively. By the compatibility between  $\phi_k$  and  $\varphi_k$  for  $k = m, n$ , we have  $H(R_1^{L_1}, S_1^{L_2}) = H^*(X_1^{L_1}, Y_1^{L_2})$ . Thus,  $h^*(\mathbf{X}|\mathbf{Y})$  can be written as

$$h^*(\mathbf{X}|\mathbf{Y}) = \lim_{L \rightarrow \infty} H(R_{L+1}^L | R_1^L, S_1^L)$$

if  $h^*(\mathbf{X}|\mathbf{Y})$  exists.

Note that the above definition of the symbolic transfer entropy rate (3) is not the rate of the original symbolic transfer entropy introduced by Staniek and Lehnertz [31] but that of the transfer entropy on rank vectors [35] which is an improved version of it.

### 3 Main Result

In this section, we give a proof of the following theorem:

**Theorem 1** *For any bivariate finite-alphabet stationary ergodic Markov process  $(\mathbf{X}, \mathbf{Y})$ , we have the equality*

$$t(\mathbf{X}|\mathbf{Y}) = t^*(\mathbf{X}|\mathbf{Y}). \quad (5)$$

Before proceeding to the proof of Theorem 1, first we present some intermediate results used in the proof.

We introduce the map  $\mu : \mathcal{S}_L \rightarrow \mathbb{N}^L$ , where  $\mathbb{N} = \{1, 2, \dots\}$  is the set of all natural numbers ordered by usual “less-than-or-equal-to” relationship, by the following procedure: first, given a permutation  $\pi \in \mathcal{S}_L$ , we decompose the sequence  $\pi(1) \cdots \pi(L)$  into *maximal ascending subsequences*. A subsequence  $i_j \cdots i_{j+k}$  of a sequence  $i_1 \cdots i_L$  is called a *maximal ascending subsequence* if it is ascending, namely,  $i_j \leq i_{j+1} \leq \cdots \leq i_{j+k}$ , and neither  $i_{j-1}i_j \cdots i_{j+k}$  nor  $i_j i_{j+1} \cdots i_{j+k+1}$  is ascending. Second, if  $\pi(1) \cdots \pi(i_1)$ ,  $\pi(i_1+1) \cdots \pi(i_2)$ ,  $\dots$ ,  $\pi(i_{k-1}+1) \cdots \pi(L)$  is the decomposition of  $\pi(1) \cdots \pi(L)$  into maximal ascending subsequences, then we define the word  $x_1^L \in \mathbb{N}^L$  by  $x_{\pi(1)} = \cdots = x_{\pi(i_1)} = 1, x_{\pi(i_1+1)} = \cdots = x_{\pi(i_2)} = 2, \dots, x_{\pi(i_{k-1}+1)} = \cdots = x_{\pi(L)} = k$ . Finally, we define  $\mu(\pi) = x_1^L$ . For example, the decomposition of  $25341 \in \mathcal{S}_5$  into maximal ascending subsequences is  $25, 34, 1$ . We obtain  $\mu(\pi) = x_1 x_2 x_3 x_4 x_5 = 31221$  by putting  $x_2 x_5 x_3 x_4 x_1 = 11223$ . By construction, we have  $\phi_n \circ \mu(\pi) = \pi$  when  $\mu(\pi) \in A_n^L$  for any  $\pi \in \mathcal{S}_L$ .

The map  $\mu$  can be seen as the dual to the map  $\phi_n$  (or  $\varphi_n$ ) in the following sense:

**Theorem 2 (Theorem 9 in [37])** *Let us put*

$$B_{n,L} \equiv \{x_1^L \in A_n^L | \exists \pi \in \mathcal{S}_L \text{ such that } \phi_n^{-1}(\pi) = \{x_1^L\}\},$$

$$C_{n,L} \equiv \{\pi \in \mathcal{S}_L | |\phi_n^{-1}(\pi)| = 1\},$$

where  $\phi_n^{-1}(\pi) \equiv \{x_1^L \in A_n^L | \phi_n(x_1^L) = \pi\}$  is the inverse image of  $\pi \in \mathcal{S}_L$  by the map  $\phi_n$ . Then,

(i)  $\phi_n$  restricted on  $B_{n,L}$  is a map into  $C_{n,L}$ ,  $\mu$  restricted on  $C_{n,L}$  is a map into  $B_{n,L}$ , and they form a pair of mutually inverse maps.

(ii)  $x_1^L \in B_{n,L}$  if and only if

for all  $1 \leq i \leq n-1$  there exist  $1 \leq j < k \leq L$

such that  $x_j = i+1$  and  $x_k = i$ . (6)

The proof of Theorem 2 can be found in [37].

Since  $h(\mathbf{X}) = h^*(\mathbf{X})$  holds for any finite-alphabet stationary process, proving (5) is equivalent to showing that the equality

$$\lim_{L \rightarrow \infty} H(R_{L+1}|R_1^L, S_1^L) = \lim_{L \rightarrow \infty} H(X_{L+1}|X_1^L, Y_1^L) \quad (7)$$

holds for any bivariate finite-alphabet stationary ergodic Markov process  $(\mathbf{X}, \mathbf{Y})$ . For simplicity, we assume that each  $(x, y) \in A_n \times A_m$  appears with a positive probability  $p(x, y) > 0$ . The essentially same proof can be applied to the general case.

**Lemma 1** *For any  $\epsilon > 0$  if we take  $L$  sufficiently large, then*

$$\sum_{\substack{x_1^L \text{ satisfies } (*), \\ y_1^L \text{ satisfies } (**)}} p(x_1^L, y_1^L) > 1 - \epsilon, \quad (8)$$

where  $(*)$  is the condition that for any  $x \in A_n$  there exist  $1 \leq i \leq \lfloor L/2 \rfloor < j \leq L$  such that  $x = x_i = x_j$  and  $(**)$  is the condition that for any  $y \in A_m$  there exist  $1 \leq i' \leq \lfloor L/2 \rfloor < j' \leq L$  such that  $y = y_{i'} = y_{j'}$ .

*Proof.* The ergodicity of  $(\mathbf{X}, \mathbf{Y})$  implies that the relative frequency of any word  $(x_1^k, y_1^k)$  converges in probability to  $p(x_1^k, y_1^k)$ . In particular, if  $F_{(x,y)}^N$  is the stochastic

variable defined by the number of indexes  $1 \leq i \leq N$  such that  $(X_i, Y_i) = (x, y)$  for  $(x, y) \in A_n \times A_m$ , then we have for any  $\epsilon > 0$  and  $\delta > 0$  there exists  $N_{(x,y),\epsilon,\delta}$  such that if  $N > N_{(x,y),\epsilon,\delta}$  then

$$\Pr\{|F_{(x,y)}^N/N - p(x, y)| < \delta\} > 1 - \epsilon.$$

Now, fix any  $\epsilon > 0$ . Choose  $\delta$  so that

$$0 < \delta < \min_{(x,y) \in A_n \times A_m} \{p(x, y)\}$$

and put  $N_0 \equiv \max_{(x,y) \in A_n \times A_m} \{N_{(x,y),\epsilon/(2nm),\delta}\}$ . Let  $S_{(x,y)}^N$  be the set of words  $(x_1^N, y_1^N)$  such that there exists  $1 \leq i \leq N$  that satisfies  $x_i = x$  and  $y_i = y$ , and  $S_N$  the set of words  $(x_1^N, y_1^N)$  such that for any  $(x, y) \in A_n \times A_m$  there exists  $1 \leq i \leq N$  that satisfies  $x_i = x$  and  $y_i = y$ .

If  $N > N_0$ , then we have for any  $(x, y) \in A_n \times A_m$

$$\begin{aligned} \Pr(S_{(x,y)}^N) &= \sum_{(x_1^N, y_1^N) \in S_{(x,y)}^N} p(x_1^N, y_1^N) \\ &= \Pr\{F_{(x,y)}^N > 0\} \\ &\geq \Pr\{|F_{(x,y)}^N/N - p(x, y)| < \delta\} \\ &> 1 - \epsilon/(2nm), \end{aligned}$$

where the inequality in the third line holds follows because we have  $p(x, y) > \delta$  by the choice of  $\delta$ .

Then, having that

$$S_N \equiv \bigcap_{(x,y) \in A_n \times A_m} S_{(x,y)}^N,$$

it follows that

$$\Pr(S_N) > 1 - nm \times \epsilon/(2nm) = 1 - \epsilon/2.$$

Now, take  $L$  so that  $\lfloor L/2 \rfloor > N_0$ . Let  $U$  be the set of words  $(x_1^L, y_1^L)$  such that  $(x_1^{\lfloor L/2 \rfloor}, y_1^{\lfloor L/2 \rfloor}) \in S_{\lfloor L/2 \rfloor}$  and  $V$

the set of words  $(x_1^L, y_1^L)$  such that  $(x_{\lfloor L/2 \rfloor + 1}^L, y_{\lfloor L/2 \rfloor + 1}^L) \in S_{L - \lfloor L/2 \rfloor}$ . Then, we have

$$\Pr(U) \geq \Pr(S_{\lfloor L/2 \rfloor}) > 1 - \epsilon/2$$

and

$$\Pr(V) \geq \Pr(S_{L - \lfloor L/2 \rfloor}) > 1 - \epsilon/2.$$

Consequently, we obtain

$$\sum_{\substack{x_1^L \text{ satisfies } (*), \\ y_1^L \text{ satisfies } (**)}} p(x_1^L, y_1^L) = \Pr(U \cap V) > 1 - \epsilon.$$

□

We put

$$D_{n,m,L} \equiv \{(x_1^L, y_1^L) | x_1^L \text{ satisfies } (*) \text{ and } y_1^L \text{ satisfies } (**)\}$$

and

$$E_{n,m,L} \equiv \{(r_1^L, s_1^L) | \exists (x_1^L, y_1^L) \in D_{n,m,L} \text{ such that}$$

$$\varphi_n(x_1^L) = r_1^L, \varphi_m(y_1^L) = s_1^L\}.$$

Then, we have  $x_1^L \in B_{n,L}$  and  $y_1^L \in B_{m,L}$  for any  $(x_1^L, y_1^L) \in D_{n,m,L}$ . Indeed, if  $(x_1^L, y_1^L) \in D_{n,m,L}$ , then  $x_1^L$  and  $y_1^L$  satisfy  $(*)$  and  $(**)$ , respectively. For any  $1 \leq i \leq n-1$ , there exists  $1 \leq j \leq \lfloor L/2 \rfloor$  such that  $x_j = i+1$  and there exists  $\lfloor L/2 \rfloor < k \leq L$  such that  $x_k = i$  by  $(*)$ . Hence,  $x_1^L$  satisfies (6). By Theorem 2 (ii), we have  $x_1^L \in B_{n,L}$ . By the same way, we have  $y_1^L \in B_{m,L}$ .

Thus, the map

$$(x_1^L, y_1^L) \mapsto (\varphi_n(x_1^L), \varphi_m(y_1^L))$$

is a bijection from  $D_{n,m,L}$  to  $E_{n,m,L}$  due to the duality between  $\phi_k$  and  $\mu$  for  $k = m, n$ . Indeed, it is onto because

$E_{n,m,L}$  is the image of the map  $\varphi_n \times \varphi_m : A_n^L \times A_m^L \rightarrow \mathcal{R}_L \times \mathcal{R}_L$  restricted on  $D_{n,m,L}$ . It is also injective. For if  $(\varphi_n(x_1^L), \varphi_m(y_1^L)) = (\varphi_n(\bar{x}_1^L), \varphi_m(\bar{y}_1^L))$ , then  $\phi_n(x_1^L) = \iota \circ \varphi_n(x_1^L) = \iota \circ \varphi_n(\bar{x}_1^L) = \phi_n(\bar{x}_1^L)$  and similarly  $\phi_m(y_1^L) = \phi_m(\bar{y}_1^L)$ . By Theorem 2 (i),  $\phi_n$  and  $\phi_m$  are bijections from  $B_{n,L}$  to  $C_{n,L}$  and from  $B_{m,L}$  to  $C_{m,L}$ , respectively. Since  $x_1^L, \bar{x}_1^L \in B_{n,L}$  and  $y_1^L, \bar{y}_1^L \in B_{m,L}$ , it holds that  $x_1^L = \bar{x}_1^L$  and  $y_1^L = \bar{y}_1^L$ .

In particular, we have

$$p(x_1^L, y_1^L) = p(r_1^L, s_1^L)$$

and

$$p(r_{L+1}^L | r_1^L, s_1^L) = p(r_{L+1}^L | x_1^L, y_1^L)$$

for any  $(x_1^L, y_1^L) \in D_{n,m,L}$ , where  $r_1^L = \varphi_n(x_1^L)$  and  $s_1^L = \varphi_m(y_1^L)$ .

*Proof of Theorem 1.* Given any  $\epsilon > 0$ , let us take  $L$  large enough so that the inequality (8) holds. We shall evaluate each term in the right hand side of (9) (see below). First, the second term in (9) is bounded by  $\epsilon \log_2 n$  which can be arbitrary small. This is because

$$\sum_{(x_1^L, y_1^L) \notin D_{n,m,L}} p(x_1^L, y_1^L) \leq \epsilon$$

by Lemma 1 and the sum over  $x_{L+1}$  is at most  $\log_2 n$ .

Second, to show the third term also converges to 0 as  $L \rightarrow \infty$ , we use the Markov property: if  $(\mathbf{X}, \mathbf{Y})$  is ergodic Markov, then we can show that

$$\begin{aligned} \sum_{(r_1^L, s_1^L) \notin E_{n,m,L}} p(r_1^L, s_1^L) &= \sum_{(x_1^L, y_1^L) \notin D_{n,m,L}} p(x_1^L, y_1^L) \\ &< C\lambda^L \end{aligned}$$

$$\begin{aligned}
& H(X_{L+1}|X_1^L, Y_1^L) - H(R_{L+1}|R_1^L, S_1^L) \\
&= - \sum_{(x_1^L, y_1^L) \in D_{n,m,L}} p(x_1^L, y_1^L) \left( \sum_{x_{L+1}} p(x_{L+1}|x_1^L, y_1^L) \log_2 p(x_{L+1}|x_1^L, y_1^L) - \sum_{r_{L+1}} p(r_{L+1}|x_1^L, y_1^L) \log_2 p(r_{L+1}|x_1^L, y_1^L) \right) \\
&- \sum_{(x_1^L, y_1^L) \notin D_{n,m,L}} p(x_1^L, y_1^L) \sum_{x_{L+1}} p(x_{L+1}|x_1^L, y_1^L) \log_2 p(x_{L+1}|x_1^L, y_1^L) \\
&+ \sum_{(r_1^L, s_1^L) \notin E_{n,m,L}} p(r_1^L, s_1^L) \sum_{r_{L+1}} p(r_{L+1}|r_1^L, s_1^L) \log_2 p(r_{L+1}|r_1^L, s_1^L)
\end{aligned} \tag{9}$$

for some  $C > 0$  and  $0 \leq \lambda < 1$ . Indeed, we have

$$\sum_{(x_1^L, y_1^L) \notin D_{n,m,L}} p(x_1^L, y_1^L) \leq \sum_{\substack{x_1^L \text{ does not} \\ \text{satisfy } (*)}} p(x_1^L) + \sum_{\substack{y_1^L \text{ does not} \\ \text{satisfy } (**)}} p(y_1^L).$$

Since

$$\begin{aligned}
\sum_{\substack{x_1^L \text{ does not} \\ \text{satisfy } (*)}} p(x_1^L) &\leq \sum_{x \in A_n} \left( \sum_{\substack{x_i \neq x, \\ 1 \leq i \leq N}} p(x_1^N) + \sum_{\substack{x_i \neq x, \\ N < i \leq L}} p(x_{N+1}^L) \right) \\
&\leq 2 \sum_{x \in A_n} \sum_{\substack{x_i \neq x, \\ 1 \leq i \leq N}} p(x_1^N)
\end{aligned}$$

and similarly

$$\sum_{\substack{y_1^L \text{ does not} \\ \text{satisfy } (**)}} p(y_1^L) \leq 2 \sum_{y \in A_m} \sum_{\substack{y_i \neq y, \\ 1 \leq i \leq N}} p(y_1^N),$$

where  $N = \lfloor L/2 \rfloor$ , it is sufficient to show that the probabilities

$$\beta_{x,\mathbf{X},L} \equiv \sum_{\substack{x_i \neq x, \\ 1 \leq i \leq N}} p(x_1^N)$$

for all  $x \in A_n$  and

$$\beta_{y,\mathbf{Y},L} \equiv \sum_{\substack{y_i \neq y, \\ 1 \leq i \leq N}} p(y_1^N)$$

for all  $y \in A_m$  converge to 0 exponentially fast.

Let  $P$  be the transition matrix for the Markov process  $(\mathbf{X}, \mathbf{Y})$ . We denote its  $(x, y)(x', y')$ -th element by  $p_{(x,y)(x',y')}$

which indicates the transition probability from state  $(x, y)$

to  $(x', y')$ . We denote the stationary distribution associated with  $(\mathbf{X}, \mathbf{Y})$  by  $\mathbf{p} = (p_{(x,y)})_{(x,y) \in A_n \times A_m}$  which uniquely

exists because of the ergodicity of the process. The probability of the occurrence of a word  $(x_1^L, y_1^L)$  is given by

$p(x_1^L, y_1^L) = p_{(x_1, y_1)} p_{(x_1, y_1)(x_2, y_2)} \cdots p_{(x_{L-1}, y_{L-1})(x_L, y_L)}$ . For any  $x \in A_n$ , we define the matrix  $P_x$  whose  $(x', y')(x'', y'')$ -th element is defined by

$$(P_x)_{(x',y')(x'',y'')} = \begin{cases} 0 & \text{if } x' = x \\ p_{(x',y')(x'',y'')} & \text{otherwise.} \end{cases}$$

Then, we can write

$$\beta_{x,\mathbf{X},L} = \langle (P_x)^{N-1} \mathbf{u}_x, \mathbf{p} \rangle,$$

where the vector  $\mathbf{u}_x = (u_{(x',y')})$  is defined by  $u_{(x',y')} =$

0 if  $x' = x$  and otherwise  $u_{(x',y')} = 1$  and  $\langle \cdots \rangle$  is the

usual inner product in the  $n \times m$ -dimensional Euclidean

space. Since  $P_x$  is a non-negative matrix, we can apply

the Perron-Frobenius theorem to it. We can show that its

Perron-Frobenius eigenvalue (the non-negative eigenvalue

whose absolute value is the largest among the eigenvalues)

$\lambda_x$  is strictly less than 1 by the same manner as in the

proof of Lemma 13 in [37]. We can also show that for any

$\delta > 0$  there exists  $C_{\delta,x} > 0$  such that for any  $k \geq 1$

$$\|(P_x)^k \mathbf{u}_x\| \leq C_{\delta,x} (\lambda_x + \delta)^k \|\mathbf{u}_x\|,$$

where  $\|\cdot\|$  is the Euclidean norm. The proof for this fact is found in, for example, Section 1.2 of [46]. Hence, if we choose  $\delta > 0$  sufficiently small so that  $\lambda_x + \delta < 1$  and put  $\gamma_x \equiv (\lambda_x + \delta)^{1/2}$  and  $C_x = C_{\delta,x} (\lambda_x + \delta)^{-2} \|\mathbf{u}_x\| \|\mathbf{p}\|$ , then we have  $\beta_{x,\mathbf{X},L} \leq C_x \gamma_x^L$ . By the same manner, we can obtain the similar bound for  $\beta_{y,\mathbf{Y},L}$  for all  $y \in A_m$ .

Since the sum over  $r_{L+1}$  is at most  $\log_2(L+1)$ , the absolute value of the third term is bounded by the quantity  $C \lambda^L \log_2(L+1)$  which goes to 0 as  $L \rightarrow \infty$ . Note that there is a  $O(\log L)$  diverging term coming from the sum over  $r_{L+1}$ . The assumed ergodic Markov property is used to overcome this divergence by showing the quantity

$$\sum_{(r_1^L, s_1^L) \notin E_{n,m,L}} p(r_1^L, s_1^L)$$

converges to 0 exponentially fast.

Finally, the first term is shown to be 0 by the same discussion as in the proof of Lemma 1 in [29] : if  $(x_1^L, y_1^L) \in D_{n,m,L}$ , then each symbol  $x \in A_n$  appears at least once in the word  $x_1^L$  (indeed, it appears at least twice). If  $a_x$  is the number of  $1 \leq x \leq n$  occurring in the word  $x_1^L$ , then  $a_x > 0$  for all  $1 \leq x \leq n$ . Hence, given  $(x_1^L, y_1^L) \in D_{n,m,L}$ ,  $x_{L+1} = x$  if and only if  $r_{L+1} = 1 + \sum_{x'=1}^x a_{x'}$ . Indeed, we have

$$r_{L+1} = \sum_{i=1}^{L+1} \delta(x_i \leq x_{L+1}) = 1 + \sum_{x'=1}^{x_{L+1}} a_{x'}.$$

Hence, if  $x_{L+1} = x$ , then we have  $r_{L+1} = 1 + \sum_{x'=1}^x a_{x'}$ .

For the converse, if  $r_{L+1} = 1 + \sum_{x'=1}^x a_{x'}$ , then we have  $\sum_{x'=1}^{x_{L+1}} a_{x'} = \sum_{x'=1}^x a_{x'}$ . Since  $a_{x'} > 0$  for all  $1 \leq x' \leq n$ , this happens only when  $x_{L+1} = x$ .

Thus, given  $(x_1^L, y_1^L) \in D_{n,m,L}$ , the probability distribution

$$p(r_{L+1}|x_1^L, y_1^L)$$

is just a re-indexing of  $p(x_{L+1}|x_1^L, y_1^L)$ , which implies that the first term is exactly equal to 0. This completes the proof of the theorem.  $\square$

From the proof, we can also see that  $t^*(\mathbf{X}|\mathbf{Y}) \leq t(\mathbf{X}|\mathbf{Y})$  holds for any bivariate finite-alphabet stationary ergodic process  $(\mathbf{X}, \mathbf{Y})$  if  $h^*(\mathbf{X}|\mathbf{Y})$  exists for the process.

## 4 On the relationship with the time-delayed mutual information rate

Apart from permutation, it is natural to ask whether the equality for the conditional entropy rate

$$\lim_{L \rightarrow \infty} H(X_{L+1}|X_1^L, Y_1^L) = \lim_{L \rightarrow \infty} \frac{1}{L} H(X_1^{L+1}|Y_1^L) \quad (10)$$

holds or not, which is parallel to the equality for the entropy rate  $\lim_{L \rightarrow \infty} H(X_{L+1}|X_1^L) = \lim_{L \rightarrow \infty} \frac{1}{L} H(X_1^{L+1})$  which holds for any finite-alphabet stationary stochastic process  $\mathbf{X}$  [28]. In this section, we will see that this question has an intimate relationship with the relationship between the transfer entropy rate and the *time-delayed mutual information rate*.

In general, (10) does not hold. For example, if  $\mathbf{X} = \mathbf{Y}$ , then we have  $\lim_{L \rightarrow \infty} H(X_{L+1}|X_1^L, Y_1^L) = h(\mathbf{X})$ , while  $\lim_{L \rightarrow \infty} \frac{1}{L} H(X_1^{L+1}|Y_1^L) = 0$ . However, note that the in-



equality

$$\lim_{L \rightarrow \infty} H(X_{L+1}|X_1^L, Y_1^L) \geq \lim_{L \rightarrow \infty} \frac{1}{L} H(X_1^{L+1}|Y_1^L) \quad (11)$$

holds for any bivariate finite-alphabet stationary stochastic process  $(\mathbf{X}, \mathbf{Y})$ . Indeed, we have

$$\begin{aligned} & \lim_{L \rightarrow \infty} H(X_{L+1}|X_1^L, Y_1^L) \\ &= \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{i=1}^{L+1} H(X_i|X_1^{i-1}, Y_1^{i-1}) \\ &\geq \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{i=1}^{L+1} H(X_i|X_1^{i-1}, Y_1^L) \\ &= \lim_{L \rightarrow \infty} \frac{1}{L} H(X_1^{L+1}|Y_1^L), \end{aligned}$$

where the first equality is due to the Cesàro mean theorem (if  $\lim_{L \rightarrow \infty} b_L = b$  then  $\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{i=1}^L b_i = b$ ) and the last equality follows from the chain rule for the Shannon entropy. In the following, we give a sufficient condition for (10).

**Proposition 1** *If there exists  $N > 0$  such that if  $i > N$  then  $X_i$  is independent of  $Y_i^{i+j}$  given  $X_1^{i-1}$  and  $Y_1^{i-1}$  for any  $j \geq 0$ , that is,*

$$\begin{aligned} & \Pr(X_i = x_i, Y_i^{i+j} = y_i^{i+j} | X_1^{i-1} = x_1^{i-1}, Y_1^{i-1} = y_1^{i-1}) \\ &= \Pr(X_i = x_i | X_1^{i-1} = x_1^{i-1}, Y_1^{i-1} = y_1^{i-1}) \\ &\times \Pr(Y_i^{i+j} = y_i^{i+j} | X_1^{i-1} = x_1^{i-1}, Y_1^{i-1} = y_1^{i-1}) \end{aligned}$$

for any  $j \geq 0$ ,  $x_k \in A_n$  ( $1 \leq k \leq i$ ) and  $y_l \in A_m$  ( $1 \leq l \leq i+j$ ), then (10) holds, namely, we have the equality

$$\lim_{L \rightarrow \infty} H(X_{L+1}|X_1^L, Y_1^L) = \lim_{L \rightarrow \infty} \frac{1}{L} H(X_1^{L+1}|Y_1^L).$$

*Proof.* Let us put  $a_{i,L} \equiv H(X_{i+1}|X_1^i, Y_1^L)$ . If we fix the index  $i$ , then  $a_{i,L}$  is a decreasing sequence of  $L$ . By the

chain rule for the Shannon entropy, we have

$$H(X_1^{L+1}|Y_1^L) = \sum_{i=0}^L H(X_{i+1}|X_1^i, Y_1^L) = \sum_{i=0}^L a_{i,L}.$$

However, by the assumption, we have  $a_{i,i} = a_{i,i+1} = a_{i,i+2} = \dots$  for  $i > N$ . Hence, we have

$$H(X_1^{L+1}|Y_1^L) = \sum_{i=0}^N a_{i,L} + \sum_{i=N+1}^L a_{i,i}.$$

Since the former sum is finite, by the Cesàro mean theorem, we obtain

$$\begin{aligned} \lim_{L \rightarrow \infty} \frac{1}{L} H(X_1^{L+1}|Y_1^L) &= \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{i=N+1}^L a_{i,i} \\ &= \lim_{L \rightarrow \infty} a_{L,L} \\ &= \lim_{L \rightarrow \infty} H(X_{L+1}|X_1^L, Y_1^L). \end{aligned}$$

□

Note that if the assumption holds, then it holds for  $N = 1$  by stationarity. If  $(\mathbf{X}, \mathbf{Y})$  is a stationary Markov process, then we can show by direct calculation that the assumption of Proposition 1 is equivalent to the following simpler condition by using the Markov property:

$$p(x_2, y_2 | x_1, y_1) = p(x_2 | x_1, y_1) p(y_2 | x_1, y_1) \quad (12)$$

for any  $x_1, x_2 \in A_n$  and  $y_1, y_2 \in A_m$ .

If (10) holds, then we obtain

$$t(\mathbf{X}|\mathbf{Y}) = \lim_{L \rightarrow \infty} \frac{1}{L} I(X_1^{L+1}; Y_1^L), \quad (13)$$

where  $I(\mathbf{A}; \mathbf{B})$  is the mutual information between stochastic variables  $\mathbf{A}$  and  $\mathbf{B}$ . We call the quantity at the right hand side of (13) *time-delayed mutual information rate* and denote it by  $i_{+1}(\mathbf{X}; \mathbf{Y})$ . Note that we have

$$t(\mathbf{X}|\mathbf{Y}) \leq i_{+1}(\mathbf{X}; \mathbf{Y})$$

for any bivariate finite-alphabet stationary stochastic process  $(\mathbf{X}, \mathbf{Y})$  by the inequality (11).

For any bivariate finite-alphabet stationary stochastic process  $(\mathbf{X}, \mathbf{Y})$ , it is straightforward to see that

$$i_{+1}(\mathbf{X}; \mathbf{Y}) = h(\mathbf{X}) + h(\mathbf{Y}) - h(\mathbf{X}, \mathbf{Y})$$

holds. Hence, we have

$$i_{+1}(\mathbf{X}; \mathbf{Y}) = i_{+1}(\mathbf{Y}; \mathbf{X}) = i(\mathbf{X}; \mathbf{Y}) = i(\mathbf{Y}; \mathbf{X}).$$

Here,

$$i(\mathbf{X}; \mathbf{Y}) \equiv \lim_{L \rightarrow \infty} \frac{1}{L} I(X_1^L; Y_1^L)$$

is the *mutual information rate* between  $\mathbf{X}$  and  $\mathbf{Y}$ . Thus, when we consider the *rate* for mutual information between two jointly distributed finite-alphabet stationary stochastic processes  $\mathbf{X}$  and  $\mathbf{Y}$ , which is defined in the limit  $L \rightarrow \infty$ , time delay has no significance in contrast to the time-delayed mutual information which has been used in time series embedding [47] and detection of nonlinear interdependence between two time series at different time points [48, 49].

The result on the relationship between  $t(\mathbf{X}|\mathbf{Y})$  and  $i_{+1}(\mathbf{X}; \mathbf{Y})$  when  $(\mathbf{X}, \mathbf{Y})$  is a stationary Markov process can be summarized as follows:

**Proposition 2** *Let  $(\mathbf{X}, \mathbf{Y})$  be a bivariate finite-alphabet stationary ergodic Markov process over the alphabet  $A_n \times A_m$ . If the condition*

$$p(x_2, y_2 | x_1, y_1) = p(x_2 | x_1, y_1) p(y_2 | x_1, y_1)$$

*holds for any  $x_1, x_2 \in A_n$  and  $y_1, y_2 \in A_m$ , then we have the equality*

$$t(\mathbf{X}|\mathbf{Y}) = i_{+1}(\mathbf{X}; \mathbf{Y}),$$

*which is equivalent to (13).*

Another interesting case is when both  $(\mathbf{X}, \mathbf{Y})$  and  $\mathbf{Y}$  are Markov. In this case, a necessary and sufficient condition for (13) can be derived easily:

**Proposition 3** *If both  $(\mathbf{X}, \mathbf{Y})$  and  $\mathbf{Y}$  are finite-alphabet stationary ergodic Markov processes, then we have*

$$t(\mathbf{X}|\mathbf{Y}) = i_{+1}(\mathbf{X}; \mathbf{Y})$$

*if and only if the condition*

$$p(x_2, y_2 | x_1, y_1) = p(x_2 | x_1, y_1) p(y_2 | y_1) \quad (14)$$

*holds for any  $x_1, x_2 \in A_n$  and  $y_1, y_2 \in A_m$ .*

*Proof.* Proving the claim is equivalent to showing

$$h(\mathbf{X}|\mathbf{Y}) = h(\mathbf{X}, \mathbf{Y}) - h(\mathbf{Y}).$$

By using the Markov property, we obtain

$$\begin{aligned} h(\mathbf{X}|\mathbf{Y}) - h(\mathbf{X}, \mathbf{Y}) + h(\mathbf{Y}) \\ = H(X_1, X_2 | Y_1) - H(X_1, X_2 | Y_1, Y_2) \geq 0. \end{aligned}$$

In the last inequality, we have the equality if and only if  $Y_2$  is independent of  $X_1^2$  given  $Y_1$ , that is,

$$p(x_1, x_2, y_2 | y_1) = p(x_1, x_2 | y_1) p(y_2 | y_1) \quad (15)$$

for any  $x_1, x_2 \in A_n$  and  $y_1, y_2 \in A_m$ , which is equivalent to the condition in the proposition.  $\square$

Let us introduce the *symbolic time-delayed mutual information rate* by

$$i_{+1}^*(\mathbf{X}; \mathbf{Y}) \equiv \lim_{L \rightarrow \infty} \frac{1}{L} I^*(X_1^{L+1}; Y_1^L), \quad (16)$$

where

$$I^*(X_1^{L+1}; Y_1^L) \equiv H^*(X_1^{L+1}) + H^*(Y_1^L) - H^*(X_1^{L+1}, Y_1^L)$$

and discuss the relationship with the transfer entropy rate, the symbolic transfer entropy rate and the (time-delayed) mutual information rate.  $i_{+1}^*(\mathbf{X}; \mathbf{Y})$  exists for any bivariate finite-alphabet stationary stochastic process as we will see below.

Similar properties with the time-delayed mutual information rate hold for the symbolic time-delayed mutual information rate: first, we note that  $t^*(\mathbf{X}|\mathbf{Y}) \leq i_{+1}^*(\mathbf{X}; \mathbf{Y})$  holds for any bivariate finite-alphabet stationary stochastic process  $(\mathbf{X}, \mathbf{Y})$  such that  $t^*(\mathbf{X}|\mathbf{Y})$  exists because the similar inequality with (11) holds for permutation analogues of corresponding quantities. Second, the symbolic time-delayed mutual information rate also admits the following expression:

$$i_{+1}^*(\mathbf{X}; \mathbf{Y}) = h^*(\mathbf{X}) + h^*(\mathbf{Y}) - h^*(\mathbf{X}, \mathbf{Y}), \quad (17)$$

where  $h^*(\mathbf{X}, \mathbf{Y}) = \lim_{L \rightarrow \infty} H(R_1^L, S_1^L)/L$ . Thus, if we introduce the *symbolic mutual information rate* between  $\mathbf{X}$  and  $\mathbf{Y}$  by

$$i^*(\mathbf{X}; \mathbf{Y}) = \lim_{L \rightarrow \infty} \frac{1}{L} I^*(X_1^L; Y_1^L),$$

then we have

$$i_{+1}^*(\mathbf{X}; \mathbf{Y}) = i_{+1}^*(\mathbf{Y}; \mathbf{X}) = i^*(\mathbf{X}; \mathbf{Y}) = i^*(\mathbf{Y}; \mathbf{X}).$$

Since the symbolic time-delayed mutual information rate is a sum of permutation entropy rates, we have

$$i_{+1}(\mathbf{X}; \mathbf{Y}) = i_{+1}^*(\mathbf{X}; \mathbf{Y})$$

for any bivariate finite-alphabet stationary stochastic process  $(\mathbf{X}, \mathbf{Y})$ .

By combining Theorem 1, Proposition 2, Proposition 3 and (4), we obtain the following proposition:

**Proposition 4** *Let  $(\mathbf{X}, \mathbf{Y})$  be a bivariate finite-alphabet stationary ergodic Markov process over the alphabet  $A_n \times A_m$ . Then,*

(i) *if*

$$p(x_2, y_2 | x_1, y_1) = p(x_2 | x_1, y_1) p(y_2 | x_1, y_1)$$

*holds for any  $x_1, x_2 \in A_n$  and  $y_1, y_2 \in A_m$ , then we have*

$$t^*(\mathbf{X}|\mathbf{Y}) = t(\mathbf{X}|\mathbf{Y}) = i_{+1}(\mathbf{X}; \mathbf{Y}) = i_{+1}^*(\mathbf{X}; \mathbf{Y}).$$

(ii) *If  $\mathbf{Y}$  is a Markov process and*

$$p(x_2, y_2 | x_1, y_1) = p(x_2 | x_1, y_1) p(y_2 | y_1)$$

*holds for any  $x_1, x_2 \in A_n$  and  $y_1, y_2 \in A_m$ , then we have*

$$t^*(\mathbf{X}|\mathbf{Y}) = t(\mathbf{X}|\mathbf{Y}) = i_{+1}(\mathbf{X}; \mathbf{Y}) = i_{+1}^*(\mathbf{X}; \mathbf{Y}).$$

## 5 Concluding Remarks

In this paper, we proved that the equality between the transfer entropy rate and the symbolic transfer entropy rate holds for any bivariate finite-alphabet stationary ergodic Markov process. The assumption is strong, however, it is the first theoretical result on permutation analogues of the transfer entropy rate. We also discussed the relationship between these quantities and the time-delayed mutual information rate and its permutation analogue.

Next natural question is how we can weaken the assumption for (5). At present, the authors are aware that the equality (5) can be at least extended to any finite-state finite-alphabet hidden Markov process whose state transition matrix is irreducible by a similar, but technically improved discussion as in the ergodic Markov case. The crucial point is how to overcome  $O(\log L)$  diverging term arising in the difference between the block Shannon entropy and its permutation analogue. Research results along this line will be presented elsewhere.

Our result in this paper is restricted. However, we illustrated how the duality between  $\phi_n$  and  $\mu$ , which is called the duality between values and orderings in [37], can be used in different setting considered in [37]. We hope that the duality method opens up a systematic study on the relationship between the information-theoretic quantities and their permutation analogues for finite-state stationary stochastic processes.

TH was supported by JST PRESTO program. The authors would like to thank D. Kugiumtzis for his useful comments and discussion. The authors also acknowledge anonymous referees for their helpful comments to improve the manuscript.

## References

1. T. Schreiber, *Phys. Rev. Lett.* **85**, 461 (2000)
2. M. Besserve, B. Schoelkopf, N. K. Logothetis, S. Panzeri, *J. Comput. Neurosci.* **29**, 547 (2010)
3. A. Buehlmann, G. Deco, *PLoS Comput. Biol.* **6**, e1000934 (2010)
4. M. Garofalo, T. Nieuw, P. Massobrio, S. Martinoia, *PLoS ONE* **4**, e6482 (2009)
5. C. J. Honey, R. Koetter, M. Breakspear, O. Sporns, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10240 (2007)
6. T. Katura, N. Tanaka, A. Obata, H. Sato, A. Maki, *Neuroimage* **31**, 1592 (2006)
7. N. Luedtke, N. K. Logothetis, S. Panzeri, *Magn. Reson. Imaging* **28**, 1113 (2010)
8. M. Lungarella, O. Sporns, *PLoS Comput. Biol.* **2**, e144 (2006)
9. S. A. Neymotin, K. M. Jacobs, A. A. Fenton, W. W. Lytton, *J. Comput. Neurosci.* **30**, 69 (2011)
10. S. Sabesan, L. B. Good, K. S. Tsakalis, A. Spanias, D. M. Treiman, L. D. Iasemidis, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **17**, 244 (2009)
11. V. A. Vakorin, O. A. Krakovska, A. R. McIntosh, *J. Neurosci. Methods* **184**, 152 (2009)
12. M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, N. F. Thornhill, *IEEE Trans. Control Syst. Technol.* **15**, 12 (2007)
13. H. Kamberaj, A. van der Vaart, *Biophys. J.* **97**, 1747 (2009)
14. R. Marschinski, H. Kantz, *Eur. Phys. J. B* **30**, 275 (2002)
15. M. Materassi, A. Wernik, E. Yordanova, *Nonlinear Processes Geophys.* **14**, 153 (2007)
16. L. J. Moniz, E. G. Cooch, S. P. Ellner, J. D. Nichols, J. M. Nichols, *Ecol. Modell.* **208**, 145 (2007)
17. J. M. Nichols, M. Seaver, S. T. Trickey, M. D. Todd, C. Olson, L. Overbey, *Phys. Rev. E* **72**, 046217 (2005)
18. J. Pahle, A. K. Green, C. J. Dixon, U. Kummer, *BMC Bioinf.* **9**, 139 (2008)

19. T. Q. Tung, T. Ryu, K. H. Lee, D. Lee, in *Proceedings of 20th IEEE International Symposium on Computer-Based Medical Systems*, edited by P. Kokol, V. Podgorelec, D. MiceticTurk, M. Zorman, M. Verlic (2007), p. 383
20. C. Bandt, B. Pompe, *Phys. Rev. Lett.* **88**, 174102 (2002)
21. C. Bandt, G. Keller, B. Pompe, *Nonlinearity* **15**, 1595 (2002)
22. C. Bandt, F. Shiha, *J. Time Ser. Anal.* **28**, 646 (2007)
23. J. M. Amigó, M. B. Kennel, *Physica D* **231**, 137 (2007)
24. M. Misiurewicz, *Nonlinearity* **16**, 971 (2003)
25. K. Keller, M. Sinn, *Physica D* **239**, 997 (2010)
26. K. Keller, A. M. Unakafov, V. A. Unakafova, *Physica D* **241**, 1477 (2012)
27. J. M. Amigó, *Permutation Complexity in Dynamical Systems*. (Springer-Verlag Berlin Heidelberg, 2010)
28. T. M. Cover, J. A. Thomas, *Elements of Information Theory*. (John Wiley & Sons, Inc, 1991)
29. J. M. Amigó, M. B. Kennel, L. Kocarev, *Physica D* **210**, 77 (2005)
30. J. M. Amigó, *Physica D* **241**, 789 (2012)
31. M. Staniek, K. Lehnertz, *Phys. Rev. Lett.* **100**, 158101 (2008)
32. A. M. Kowalski, M. T. Martin, A. Plastino, L. Zunino, *Phys. Lett. A* **374**, 1819 (2010)
33. M. Martini, T. A. Kranz, T. Wagner, K. Lehnertz, *Phys. Rev. E* **83**, 011919 (2011)
34. A. Papana, D. Kugiumtzis, P. G. Larsson, *Phys. Rev. E* **83**, 036207 (2011)
35. D. Kugiumtzis, *J. Nonlin. Sys. Appl.* **3**, 73 (2012)
36. P. O. Amblard, O. J. J. Michel, *J. Comput. Neurosci.* **30**, 7 (2011)
37. T. Haruna, K. Nakajima, *Physica D* **240**, 1370 (2011)
38. J. P. Crutchfield, N. H. Packard, *Physica D* **7**, 201 (1983)
39. J. P. Crutchfield, D. P. Feldman, *Chaos* **15**, 25 (2003)
40. D. P. Feldman, C. S. McTague, J. P. Crutchfield, *Chaos* **18**, 043106 (2008)
41. R. Shaw, *The Dripping Faucet as a Model Chaotic System*. (Aerial Press, Santa Cruz, California, 1984)
42. P. Grassberger, *Int. J. Theor. Phys.* **25**, 907 (1986)
43. W. Bialek, I. Nemenman, N. Tishby, *Neural Computation* **13**, 2409 (2001)
44. W. Li, *Complex Systems* **5**, 381 (1991)
45. D. V. Arnold, *Complex Systems* **10**, 143 (1996)
46. A. Katok, B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*. (Cambridge University Press, 1995)
47. A. M. Fraser, H. L. Swinney, *Phys. Rev. A* **33**, 1134 (1986)
48. K. Kaneko, *Physica D* **23**, 436 (1986)
49. J. A. Vastano, H. L. Swinney, *Phys. Rev. Lett.* **60**, 1773 (1988)