Symbolic Transfer Entropy

Matthäus Staniek[1,2,*] and Klaus Lehnertz[1,2,3,†]

[1]Department of Epileptology, Neurophysics Group, University of Bonn, Sigmund-Freud-Strasse 25, D-53105 Bonn, Germany
[2]Helmholtz-Institute for Radiation and Nuclear Physics, University of Bonn, Nussallee 14-16, 53115 Bonn, Germany
[3]Interdisciplinary Center for Complex Systems, University of Bonn, Römerstrasse 164, 53117 Bonn, Germany

(Received 21 January 2008; published 14 April 2008)

We propose to estimate transfer entropy using a technique of symbolization. We demonstrate numerically that symbolic transfer entropy is a robust and computationally fast method to quantify the dominating direction of information flow between time series from structurally identical and nonidentical coupled systems. Analyzing multiday, multichannel electroencephalographic recordings from 15 epilepsy patients our approach allowed us to reliably identify the hemisphere containing the epileptic focus without observing actual seizure activity.

DOI: 10.1103/PhysRevLett.100.158101 PACS numbers: 87.19.lm, 02.50.−r, 05.45.Tp

Understanding couplings between dynamical systems is a topic of general interest since synchronization and related complex interaction phenomena can be observed in many disciplines ranging from physics to the neurosciences [1,2]. The investigation of interactions addresses two major aspects, namely, the detection and quantification of strength and direction (or asymmetry) of couplings. In situations where the underlying equations of motion are not known, a detailed quantitative description can nevertheless be achieved by applying time series analysis techniques to experimentally acquired observables. Much work has been devoted to the problem of assessing directional couplings, and current approaches include estimates of phase dynamics [3–5], estimates based on the reconstruction of state spaces [6–8], methods that are based on recurrence properties of interacting systems [9], and information theoretic approaches [10] (see Ref. [11] for an overview). In principle, asymmetric dependences between coupled systems can be detected with measures that share some of the properties of mutual information [12] and take into account the dynamics of information transport. Transfer entropy [13], which is related to the concept of Granger causality [14], has been proposed to distinguish effectively driving and responding elements and to detect asymmetry in the interaction of subsystems. By appropriate conditioning of transition probabilities this quantity has been shown to be superior to the standard time delayed mutual information, which fails to distinguish information that is actually exchanged from shared information due to common history and input signals. Let $x_i = x(i)$ and $y_i = y(i)$, $i = 1, \ldots, N$, denote sequences of observations from systems X and Y. Transfer entropy incorporates time dependence by relating previous samples x_i and y_i to predict the next value x_{i+1}, and quantifies the deviation from the generalized Markov property, $p(x_{i+1}|x_i, y_i) = p(x_{i+1}|x_i)$, where p denotes the transition probability density. If there is no deviation from the generalized Markov property, Y has no influence on X. Transfer entropy, which is formulated as Kullback-Leibler entropy between $p(x_{i+1}|x_i, y_i)$ and $p(x_{i+1}|x_i)$, quantifies the incorrectness of this assumption, and is explicitly nonsymmetric under the exchange of x_i and y_i. In Refs. [15–17] various techniques have been proposed to estimate transfer entropy from observed data. Most techniques, however, make great demands on the data, require fine-tuning of parameters, and are highly sensitive to noise contributions, which limits the use of transfer entropy to field applications.

In this Letter, we suggest to estimate transfer entropy by adopting a technique of symbolization, which has already been introduced with the concept of permutation entropy [18,19]. We follow Ref. [18] and define symbols by reordering the amplitude values of time series x_i and y_i. For a given, but otherwise arbitrary i, m amplitude values $X_i = \{x(i), x(i + l), \ldots, x(i + (m - 1)l)\}$ are arranged in an ascending order $\{x(i + (k_{i1} - 1)l) \leq x(i + (k_{i2} - 1)l) \leq \ldots \leq x(i + (k_{im} - 1)l)\}$, where l denotes the time delay, and m is the embedding dimension. In case of equal amplitude values the rearrangement is carried out according to the associated index k, i.e., for $x(i + (k_{i1} - 1)l) = x(i + (k_{i2} - 1)l)$ we write $x(i + (k_{i1} - 1)l) \leq x(i + (k_{i2} - 1)l)$ if $k_{i1} < k_{i2}$ thereby ensuring that every X_i is uniquely mapped onto one of the $m!$ possible permutations. A symbol is thus defined as $\hat{x}_i \equiv (k_{i1}, k_{i2}, \ldots, k_{im})$, and with the relative frequency of symbols we estimate joint and conditional probabilities of the sequence of permutation indices. Given symbol sequences $\{\hat{x}_i\}$ and $\{\hat{y}_i\}$ we define symbolic transfer entropy (STE) as

$$T^S_{Y,X} = \sum p(\hat{x}_{i+\delta}, \hat{x}_i, \hat{y}_i) \log \frac{p(\hat{x}_{i+\delta}|\hat{x}_i, \hat{y}_i)}{p(\hat{x}_{i+\delta}|\hat{x}_i)}, \qquad (1)$$

where the sum runs over all symbols and δ denotes a time step. The log is with base 2, thus $T^S_{Y,X}$ is given in bit. $T^S_{X,Y}$ is defined in complete analogy. The directionality index $T^S = (T^S_{X,Y} - T^S_{Y,X})$ quantifies the preferred direction of information flow and is expected to attain positive values for unidirectional couplings with X as the driver and nega-

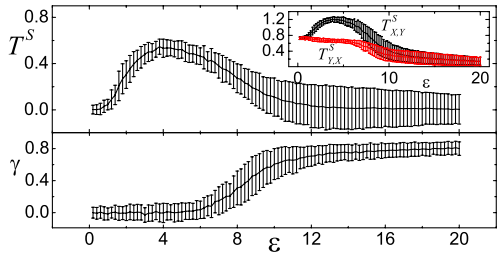

FIG. 1 (color online). Top: Directionality index T^S as a function of the coupling strength ϵ in diffusively coupled Lorenz oscillators. The inset shows the dependences of $T^S_{X,Y}$ (black) and $T^S_{Y,X}$ (red or gray) on ϵ. Bottom: Dependence of synchronization index γ [22] on ϵ ($\gamma \approx 0$ for independent systems and $\gamma = 1$ for fully synchronized systems). Error bars indicate standard deviations from 100 realizations of the systems using different initial conditions, and for each realization we randomly chose the control parameter $r \in [28 \pm 1]$.

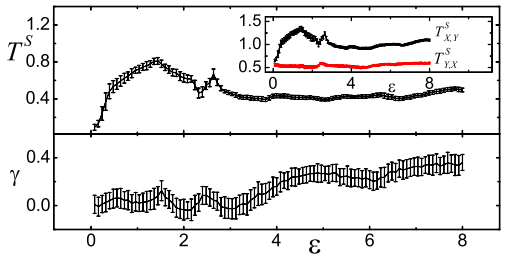

FIG. 2 (color online). Same as Fig. 1 but for a Lorenz system driven by a Rössler system. Error bars indicate standard deviations from 100 realizations of the systems, with initial conditions chosen randomly close to the attractors of both systems for each realization.

tive values for Y driving X. For symmetric bidirectional couplings we expect $T^S = 0$.

To demonstrate the use of STE, we study three examples: interactions between structurally identical and nonidentical nonlinear oscillators and interactions in long-lasting, multichannel electroencephalographic (EEG) recordings from epilepsy patients. Following Refs. [18,20] we chose the window $(m-1)l$ such as to cover about 1–2 basic periods of the systems when constructing symbol sequences. In addition, we here restrict ourselves to $\delta = 1$, presuming that a meaningful sampling rate is already chosen.

First, we study unidirectionally, diffusively coupled Lorenz oscillators (see Ref. [21] for details), where the diffusive coupling of strength ϵ is introduced in the x component of system Y. We integrated the equations of motion (fourth order Runge-Kutta algorithm with step size $\Delta s = 0.005$ and sampling interval $\Delta t = 0.03$), and with $m = 5$ and $l = 10$ we computed $T^S_{X,Y}$ and $T^S_{Y,X}$ using the x components of the systems as observables. Time series consisted of $N = 4096$ iterations after 10^4 transients. In Fig. 1 (upper part) we show the dependence of T^S on the coupling strength ϵ. At zero coupling, $T^S \approx 0$ but we observe $T^S_{X,Y} = T^S_{Y,X} \approx 0.75$, which would indicate an exchange of information between the uncoupled systems. This bias can be explained by both the limited number of data points as well as the narrow-band spectra of the systems [15] and vanishes for systems with broadband spectra and for $N \to \infty$ (data not shown here). Note that the bias of $T^S_{X,Y}$ and that of $T^S_{Y,X}$ results in a vanishing bias of the difference T^S. For increasing coupling strengths T^S increases until a maximum is reached at $\epsilon \approx 4$, which reflects the growing influence of the driver X on the dynamics of the responder Y (cf. the functional dependences of $T^S_{X,Y}$ and $T^S_{Y,X}$ on ϵ in the inset). T^S decreases for $\epsilon > 4$ and the preferred direction of information flow becomes less evident. In the lower part of Fig. 1 we show, for comparison, the synchronization index γ [22] as a function of the coupling strength, which indicates the establishment of synchronization for $\epsilon > 4$. Accordingly T^S decreases and the coupling direction can not be detected for independent ($\epsilon \approx 0$) and for fully synchronized systems ($\epsilon > 12$). Note, that T^S increases for $0 \leq \epsilon \leq 4$, while $\gamma \approx 0$.

Second, we study a Lorenz system driven by a Rössler system (see Refs. [8,23,24] for details; fourth order Runge-Kutta with $\Delta s = 0.005$ and $\Delta t = 0.03$). With $m = 5$ and $l = 10$ we estimated $T^S_{X,Y}$ and $T^S_{Y,X}$ using the y components of the systems as observables ($N = 4096$ data points after 10^4 transients). The dependence of the directionality index T^S on the coupling strength ϵ is shown in Fig. 2 (upper part). The inequality of $T^S_{X,Y}$ and $T^S_{Y,X}$ (cf. inset) entails slightly positive values for $\epsilon \approx 0$, which indicates a false detection of asymmetrical coupling. With an increasing coupling strength the driving character of the Rössler system is correctly reflected, and a best identification of the direction of information flow is achieved for $\epsilon \approx 1.5$. Despite the indication of only weak synchronization ($\gamma < 1$ even for stronger couplings), a detection of the direction of coupling is possible for a broad range of coupling strengths.

In Fig. 3 we present the influence of observational noise on the detectability of directional couplings. We added Gaussian white noise to the time series of the coupled

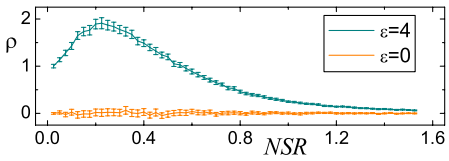

FIG. 3 (color online). Ratio ρ of directionality indices for noise-free and noise-contaminated time series of diffusively coupled Lorenz oscillators as a function of the noise-to-signal ratio NSR $= \sigma_n/\sigma_s$, where σ_n and σ_s denote the standard deviations of the noisy and the noise-free time series, respectively. Error bars indicate standard deviations from 100 realizations of the systems.

Lorenz oscillators (see above) and observed $\rho = T_n^S/T^S$ for increasing noise-to-signal ratios (NSR). T_n^S denotes the directionality index for the noisy time series. At zero coupling, ρ fluctuates around zero for the range of NSR investigated here. We obtained similar results for the coupled Rössler and Lorenz systems. The resonancelike behavior of ρ at stronger couplings ($\epsilon = 4$ cf. Fig. 1) allows one to detect the direction of coupling more easily in the presence of noise. This phenomenon can be explained again by the narrow-band spectra of the systems, in which case the sparse distribution of symbols yields a less precise entropy estimate. When adding noise, the distribution of the symbols is broadened, which advances the estimation of entropies. For systems with broadband spectra (e.g., coupled Hénon systems) we observed ρ to decrease monotonically with increasing NSR. Nevertheless, the preferred direction of information flow can be detected even for high noise levels NSR > 1, which renders STE attractive for the analysis of field data.

With our third example we provide evidence for the high relevance of our approach to improve understanding of the complicated spatiotemporal dynamics of the human epileptic brain. Successful brain surgery (i.e., complete seizure control and minimum surgery-induced neurological deficits) in patients with medically intractable epilepsy strongly depends on an accurate definition of the brain volume that has to be resected. This requires the identification of the underlying epileptogenic network and a detailed understanding of the functional organization of cortical areas in which it is embedded, in order to spare nonaffected cortical regions. We here address the important and yet unsolved issue of whether interactions between the seizure generating area of the brain (epileptic focus) and remote areas can be identified, particularly during the seizure-free interval. Understanding the directionality of possible interactions within the epileptogenic network may help to improve the evaluation of epilepsy patients candidate for resective therapies and to gain deeper insights into mechanisms that lead to the generation of epileptic seizures. We retrospectively analyzed EEG data recorded from 15 patients undergoing presurgical evaluation for drug-resistant temporal lobe epilepsy. Since the localization of the epileptic focus could not be accomplished by means of noninvasive EEG recordings, intracranial electrodes [with $N_c = 20$ contacts; cf. Fig. 4(a)] were chronically implanted for the purpose of identifying the focal seizure origin. All patients achieved complete seizure control after surgery so the epileptic focus can be assumed to be contained within the resected area. EEG signals were referenced to a common average, sampled over a longer period (mean duration: 82.77 h; range: 6–226 h) at 200 Hz using a 16 bit analog-to-digital converter, and filtered within a frequency band of 0.5 to 85 Hz. All patients had signed informed consent that their clinical data might be used and

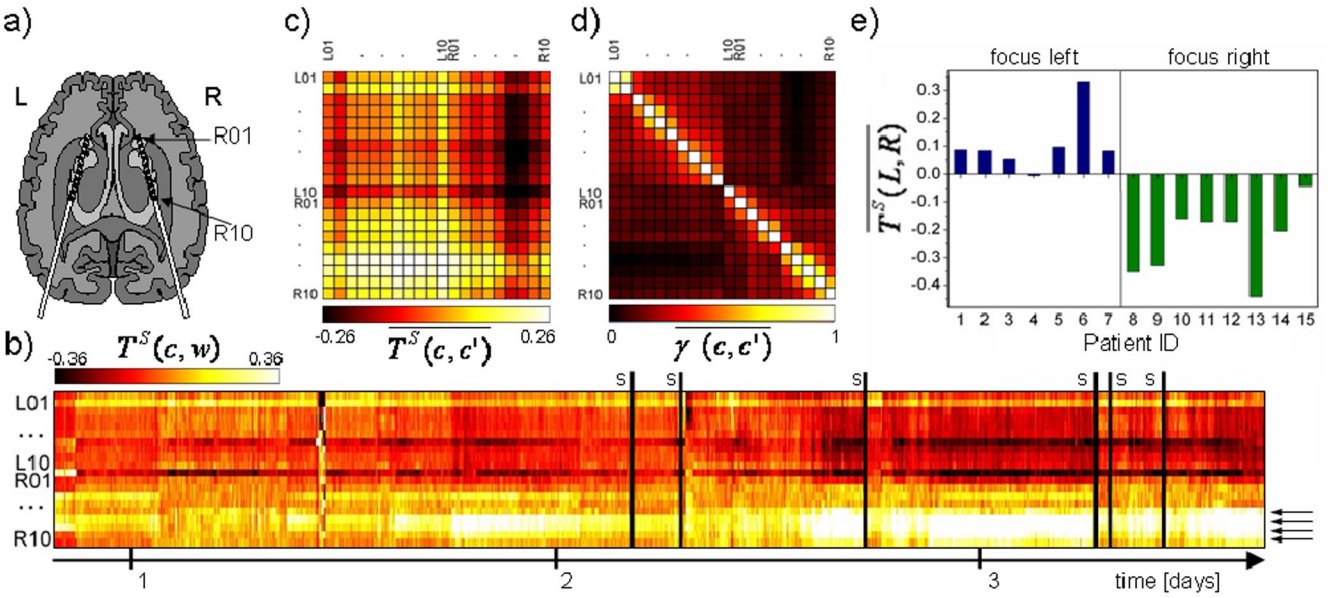

FIG. 4 (color online). (a) Electrode implantation scheme for intrahippocampal depth electrodes. (b) Exemplary temporal evolution of the preferred direction of information flow $T^S(c, w)$. EEG was recorded over 62.8 h. S denotes electrographic seizure onset, tics on time axis denote midnight, and arrows indicate the epileptic focus as identified with established presurgical evaluation techniques. (c) Exemplary directionality matrix. Color coded entries $\overline{T^S(c, c')}$ represent the temporal average of $T^S(c, c', w)$ for each pair (c, c') of electrode contacts and attain positive values if a brain structure sampled with a channel corresponding to row index c drives another brain structure sampled with a channel corresponding to the column index c'. EEG recording from the seizure-free interval only (total duration: 44.1 h). (d) Same as (c) but for synchronization indices $\gamma(c, c', w)$. (e) The preferred direction of information flow from the left to the right brain hemisphere $\overline{T^S(L, R)}$ correctly indexes the focal brain hemisphere in all but one of the patients.

published for research purposes, and the study protocol had previously been approved by the local ethics committee. Using a sliding window approach, we performed a time resolved estimation of T^S and γ from nonoverlapping EEG segments of 20.48 s duration [25], and embedding parameters were set to $m = 5$ and $l = 3$. For each window w and electrode contact c we define the preferred direction of information flow to the remaining $N_c - 1$ contacts as $T^S(c, w) = (N_c - 1)^{-1} \sum_{c' \neq c} [T^S_{c,c'}(w) - T^S_{c',c}(w)]$. In Fig. 4(b) we show, as an example, the temporal evolution of $T^S(c, w)$ for an EEG recording lasting about 2.5 days from a patient with a right-sided epileptic focus. Despite fluctuations, particularly the brain region sampled with contacts $R06$–$R09$ appears as a strong driving region. Periods of wakefulness or sleep had only a minor influence on the temporal evolution of the preferred direction of information flow. Given the little variance over time we averaged, for each channel combination (c, c'), the values of $T^S(c, c', w) = T^S_{c,c'}(w) - T^S_{c',c}(w)$ and $\gamma(c, c', w)$ from all windows covering the seizure-free interval (we discarded values related to seizures and activities four hours prior to and 30 min after a seizure [26]), which provides a compressed view of direction $[\overline{T^S(c, c')}$; Fig. 4(c)] and strength of interactions $[\overline{\gamma(c, c')}$; Fig. 4(d)] between all sampled brain structures. For this patient, the brain region underlying contacts $R06$–$R09$ appears to drive other regions even from the opposite brain hemisphere, despite the low level of interhemispheric synchronization [cf. Fig. 4(d)]. Interestingly, this driving structure corresponds to the epileptic focus as identified with established presurgical evaluation techniques. In all but one of the patients, the direction of information flow between the left and the right brain hemisphere $\overline{T^S(L, R)}$ [i.e., the mean of all $\overline{T^S(c, c')}$ with $c \in \{L01, \ldots, L10\}$ and $c' \in \{R01, \ldots, R10\}$] indicated the driving influence of the hemisphere containing the epileptic focus, despite the crude spatial and temporal averaging and without observing actual seizure activity [cf. Fig. 4(e)].

To conclude, with symbolic transfer entropy we introduced a convenient, robust, and computationally fast [27] method that allows one to quantify the preferred direction of information flow between time series from observed data. These features render STE a promising tool for the analysis of field data. With our numerical investigations and the analysis of time series of brain electrical activity we have shown that asymmetric dependences between structurally identical and nonidentical coupled but not yet fully synchronized systems can be reliably detected.

We thank Hannes Osterhage and Jens Prusseit for useful comments. This work was supported by No. SFB-TR3 (subproject A2) of the Deutsche Forschungsgemeinschaft.

*mattstan@gmx.de
†klaus.lehnertz@ukb.uni-bonn.de

[1] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge University Press, Cambridge, U.K., 2001).
[2] E. Pereda, R. Quian Quiroga, and J. Bhattacharya, Prog. Neurobiol. **77**, 1 (2005).
[3] M. G. Rosenblum and A. S. Pikovsky, Phys. Rev. E **64**, 045202(R) (2001).
[4] D. A. Smirnov and B. P. Bezruchko, Phys. Rev. E **68**, 046209 (2003).
[5] D. A. Smirnov, M. B. Bodrov, J. L. P. Velazquez, R. A. Wennberg, and B. P. Bezruchko, Chaos **15**, 024102 (2005).
[6] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, Phys. Rev. E **54**, 6708 (1996).
[7] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, Physica (Amsterdam) **134D**, 419 (1999).
[8] R. Quian Quiroga, J. Arnhold, and P. Grassberger, Phys. Rev. E **61**, 5142 (2000).
[9] M. C. Romano, M. Thiel, J. Kurths, and C. Grebogi, Phys. Rev. E **76**, 036211 (2007).
[10] S. Frenzel and B. Pompe, Phys. Rev. Lett. **99**, 204101 (2007).
[11] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, Phys. Rep. **441**, 1 (2007).
[12] C. E. Shannon and W. Weaver, *The Mathematical Theory of Information* (University Press, Urbana, IL, 1949).
[13] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[14] C. Granger, Econometrica **37**, 424 (1969).
[15] A. Kaiser and T. Schreiber, Physica (Amsterdam) **166D**, 43 (2002).
[16] P. F. Verdes, Phys. Rev. E **72**, 026222 (2005).
[17] M. Lungarella, A. Pitti, and Y. Kuniyoshi, Phys. Rev. E **76**, 056117 (2007).
[18] C. Bandt and B. Pompe, Phys. Rev. Lett. **88**, 174102 (2002).
[19] C. Bandt, G. Keller, and B. Pompe, Nonlinearity **15**, 1595 (2002).
[20] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, U.K., 2003).
[21] I. Belykh, V. Belykh, and M. Hasler, Chaos **16**, 015102 (2006).
[22] Z. Liu, Europhys. Lett. **68**, 19 (2004).
[23] M. Le Van Quyen, J. Martinerie, C. Adam, and F. Varela, Physica (Amsterdam) **127D**, 250 (1999).
[24] M. Paluš and M. Vejmelka, Phys. Rev. E **75**, 056211 (2007).
[25] S. Blanco, H. Garcia, R. Quian Quiroga, L. Romanelli, and O. Rosso, IEEE Eng. Med. Biol. Mag. **14**, 395 (1995).
[26] F. Mormann, T. Kreuz, C. Rieke, R. G. Andrzejak, A. Kraskov, P. David, C. E. Elger, and K. Lehnertz, Clin. Neurophysiol. **116**, 569 (2005).
[27] Measuring information flow between two time series (4096 data points) can be performed in less than 0.1 s on a PC with a CPU clock rate of 2.4 GHz.