

## Robustly Estimating the Flow Direction of Information in Complex Physical Systems

Guido Nolte,<sup>1</sup> Andreas Ziehe,<sup>2</sup> Vadim V. Nikulin,<sup>3</sup> Alois Schlögl,<sup>1</sup> Nicole Krämer,<sup>2</sup>  
Tom Brismar,<sup>4</sup> and Klaus-Robert Müller<sup>1,2</sup>

<sup>1</sup>Fraunhofer FIRST.IDA, Berlin, Germany

<sup>2</sup>Technical University of Berlin, Computer Science, Machine Learning Laboratory, Berlin, Germany

<sup>3</sup>Department of Neurology, Charite; Bernstein Center for Computational Neuroscience, Berlin, Germany

<sup>4</sup>Karolinska Institutet, Clinical Neurophysiology, Karolinska Hospital, Stockholm, Sweden

(Received 10 December 2007; published 10 June 2008)

We propose a new measure (phase-slope index) to estimate the direction of information flux in multivariate time series. This measure (a) is insensitive to mixtures of independent sources, (b) gives meaningful results even if the phase spectrum is not linear, and (c) properly weights contributions from different frequencies. These properties are shown in extended simulations and contrasted to Granger causality which yields highly significant false detections for mixtures of independent sources. An application to electroencephalography data (eyes-closed condition) reveals a clear front-to-back information flow.

DOI: 10.1103/PhysRevLett.100.234101

PACS numbers: 05.45.Tp, 05.45.Xt, 87.19.le

To understand interacting systems it is of fundamental importance to distinguish the driver from the recipient, and hence to be able to estimate the direction of information flow. The direction can be estimated with a temporal argument: the driver is earlier than the recipient implying that the driver contains information about the future of the recipient not contained in the past of the recipient while the reverse is not the case. This argument is the conceptual basis of Granger causality [1] which is probably the most prominent method to estimate the direction of causal influence in time series analysis. Granger causality was originally developed in econometry, but is applied to many different problems in physics, geosciences (cause of climate change), social sciences, and biology [2–5].

The difficulty in realistic measurements in complex systems is that asymmetries in detection power may as well arise due to other factors, specifically independent background activity having nontrivial spectral properties and eventually being measured in unknown superposition in the channels. In this case the interpretation of the asymmetry as a direction of information flow can lead to significant albeit false results [6]. The purpose of this Letter is to propose a novel estimate of flux direction which is highly robust against false estimates caused by confounding factors of very general nature.

More formally, we are interested in causal relations between a signal of interest consisting of two sources with time series  $x_i(t)$  for  $i = 1, 2$ . The measured data  $\mathbf{y}(t)$  are assumed to be a superposition of these sources and additive noise  $\eta(t)$  in the form

$$\mathbf{y}(t) = \mathbf{x}(t) + B\eta(t), \quad (1)$$

where  $\eta(t)$  is a set of  $M$  independent noise sources which are mixed into the measurement channels by an unknown  $2 \times M$  mixing matrix  $B$ .

The proposed method is based on the slope of the phase of cross-spectra between two time series with cross spectra defined for channels  $i$  and  $j$  as

$$S_{ij}(f) = \langle \hat{y}_i(f) \hat{y}_j^*(f) \rangle, \quad (2)$$

where  $\langle \cdot \rangle$  denotes expectation value.

The idea behind using phase slope is that interactions require some time, and if the speed at which different waves travel is similar, then the phase difference between sender and recipient increases with frequency and we expect a positive slope of the phase spectrum. This is most easily seen if we assume that the interaction is merely a delay by a time  $\tau$ , i.e.,  $y_2(t) = ay_1(t - \tau)$  with  $a$  being some constant. In the Fourier domain this relation reads  $\hat{y}_2(f) = a \exp(-i2\pi f\tau) \hat{y}_1(f)$ . For the cross spectrum between the two channels one has  $S_{12}(f) \sim \exp(i2\pi f\tau) \equiv \exp[i\Phi(f)]$ . The phase-spectrum  $\Phi(f) = 2\pi f\tau$  is linear and proportional to the time delay  $\tau$ . The slope of  $\Phi(f)$  can be estimated, and the causal direction is estimated to go from  $y_1$  to  $y_2$  ( $y_2$  to  $y_1$ ) if it is positive (negative).

The idea here is now to define an average phase slope in such a way that (a) this quantity properly represents relative time delays of different signals and especially coincides with the classical definition for linear phase spectra, (b) it is insensitive to signals which do not interact regardless of spectral content and superpositions of these signals, and (c) it properly weights different frequency regions according to the statistical relevance. This quantity is termed “phase-slope index” (PSI) and is defined as

$$\tilde{\Psi}_{ij} = \Im \left( \sum_{f \in F} C_{ij}^*(f) C_{ij}(f + \delta f) \right), \quad (3)$$

where  $C_{ij}(f) = S_{ij}(f) / \sqrt{S_{ii}(f) S_{jj}(f)}$  is the complex coherence,  $S$  is the cross-spectral matrix,  $\delta f$  is the frequency resolution, and  $\Im(\cdot)$  denotes taking the imaginary part.  $F$  is

the set of frequencies over which the slope is summed. To see that the definition of  $\tilde{\Psi}_{ij}$  corresponds to a meaningful estimate of the average slope it is convenient to rewrite it as  $\tilde{\Psi}_{ij} = \sum_{f \in F} |C_{ij}(f)| |C_{ij}(f + \delta f)| \sin[\Phi(f + \delta f) - \Phi(f)]$ . For smooth phase spectra,  $\sin[\Phi(f + \delta f) - \Phi(f)] \approx \Phi(f + \delta f) - \Phi(f)$  and hence  $\tilde{\Psi}$  corresponds to a weighted average of the slope. We emphasize that since  $\tilde{\Psi}$  vanishes if the imaginary part of coherency vanishes it will be *insensitive* to mixtures of noninteracting sources [7,8].

Finally, it is convenient to normalize  $\tilde{\Psi}$  by an estimate of its standard deviation

$$\Psi = \tilde{\Psi} / \text{std}(\tilde{\Psi}), \quad (4)$$

with  $\text{std}(\tilde{\Psi})$  being estimated by the jackknife method. In the examples below we always show normalized measures of directionality, and we consider absolute values larger than 2 as significant.

We emphasize, that  $\Psi$  indicates the temporal order of two signals, which is then interpreted as a driver-responder relation. For bidirectional (or unknown) coupling a finding that, e.g.,  $A$  drives  $B$  does not imply that  $B$  has no impact on  $A$ . Rather, one cannot make a statement about the reverse direction.

Estimations of cross spectra is standard [7,9] but technical details may differ. Here, we first divide the whole data into epochs containing continuous data (4 s duration), then we divide each epoch further into segments of time  $T$ , here of 2 s duration corresponding to a frequency resolution of  $\delta f = 0.5$  Hz, multiply the data for each segment with a Hanning window (a cosine, raised by 1, with first minima at the edges of the segment), Fourier-transform the data, and estimate the cross spectra according to Eq. (2) as an average over all segments. The segments have 50% overlap within each epoch but not across epochs. To apply the jackknife method, for each pair of channels we calculate  $\tilde{\Psi}_k$  from data with the  $k$ th epoch removed for all  $k$ . The standard deviation of  $\tilde{\Psi}$  is finally estimated for  $K$  epochs as  $\sqrt{K}\sigma$  where  $\sigma$  is the standard deviation of the set of  $\tilde{\Psi}_k$  [10].

Our new method is compared to Granger causality using autoregressive (AR) models both for wide and narrow band analysis [11] with analogous normalization by the estimated standard deviation. To estimate the parameters of the model we here use the Levinson-Wiggins-Robinson [12] algorithm available in the open Biosig toolbox [13]. Granger causality is defined as the difference between the flux from channel 1 to 2 and the flux from channel 2 to 1 normalized to unit estimated standard deviation.

We illustrate typical results for two simple cases in Fig. 1. To study a more general class of signals we simulated data with structure

$$\mathbf{y}(t) = (1 - \gamma) \frac{\mathbf{x}(t)}{N_x} + \gamma \frac{B\boldsymbol{\eta}(t)}{N_\eta}. \quad (5)$$

Here, the signal  $\mathbf{x}(t)$  contains truly unidirectional informa-

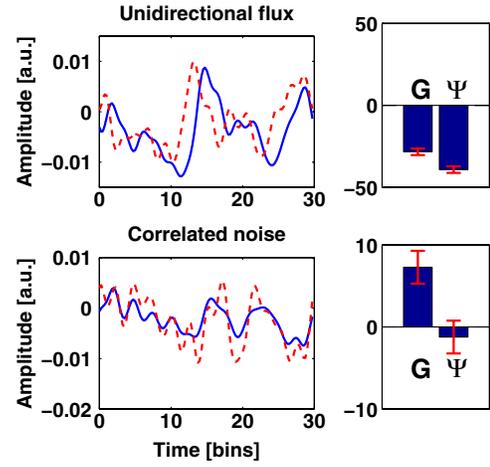


FIG. 1 (color online). Upper panels: Strong interaction from second (dashed line in left panels) to first (full line in left panels) signal estimated from 2000 data points generated with a simple AR model of order 1. Both methods detect this direction correctly. Lower panels: mixture of pink and white noise. In contrast to PSI, Granger causality erroneously still detects a significant direction. The error bars in the right panels indicate estimated 95% error margins corresponding to 2 standard deviations. Time series in the left panels were upsampled to 400 Hz.

tion flux and  $\boldsymbol{\eta}(t)$  is undirected noise. Both the signal part and the mixed noise part  $[B\boldsymbol{\eta}(t)]$  are normalized by the Frobenius norms of the respective data matrices ( $N_x$  and  $N_\eta$ ) and added with a parameter  $\gamma$  controlling for the relative strength.

The data were generated using AR models of order  $P = 5$  for two channels. An AR model is defined as

$$\mathbf{z}(t) = \sum_{p=1}^P A(p)\mathbf{z}(t-p) + \boldsymbol{\xi}(t), \quad (6)$$

where  $A(p)$  are the AR matrices up to order  $P$  and  $\boldsymbol{\xi}(t)$  is white Gaussian noise with covariance matrix  $\boldsymbol{\Sigma}$  chosen here to be the identity matrix. For computing Granger causality, the AR model was fitted with order  $P = 10$ .

All entries of AR matrices were selected randomly as independent Gaussian random numbers with  $A_{21}(p) = 0$  for the signal part  $\mathbf{x}(t)$ , corresponding to unidirectional flow from the second to first signal, and  $A_{12}(p) = A_{21}(p) = 0$  for the noise part  $\boldsymbol{\eta}(t)$ , corresponding to independent sources. Noise was mixed into channels with a random  $2 \times 2$  mixing matrix  $B$ . The time constant implicit in the AR model was assumed to be 10 ms, and we generated 60 000 data points for each system and channel. This corresponds to a Nyquist frequency of 50 Hz and to 10 min measurement. We analyzed systems for all  $\gamma$  in the range  $[0, 1]$  with step 0.1. For each  $\gamma$  we analyzed 1000 randomly selected stable systems with both methods and both for wide band (using all frequencies) and narrow band analysis. For the narrow band, we used a band of 5 Hz width, centered this band around the spectral peak of the (known) signal of interest

and analyzed only cases where the band includes at least 60% of the total power of the signal of interest.

The fractions of significant false and significant correct detections as a function of  $\gamma$  are shown in Fig. 2. We observe that for increasing noise level the fraction of significant false detections for Granger causality comes close to 50% while PSI rarely makes significant detections at all. For PSI, the worst case observed is at  $\gamma = 0.8$  for the wide band with 6% significant false detections. This level can be reduced to about 3.5% if we increase the frequency resolution to 0.25 Hz. However, the price is some loss in statistical power and it is important to show that also the proposed method might fail, even if it is unlikely in the sense of the present simulation.

We observe similar significant correct detection rates for small and moderate noise levels. For high noise level Granger causality shows a much larger fraction of significant correct detections which, however, is meaningless given the large fraction of significant false detections.

We now apply the PSI to real electroencephalography (EEG) data in rest condition [14]. For this, 88 healthy subjects were recruited randomly by the aid of the Swedish population register. During the experiment, which lasted for 15 min, the subjects were instructed to relax and keep their eyes closed. Every minute the subjects were asked to open their eyes for 5 s. EEG was measured with a standard 10–20 system consisting of 19 channels. Data were analyzed using the linked mastoids reference. The protocol was approved by the Hospital Ethics Committee.

The most prominent feature of this measurement is the alpha peak at around 10 Hz. This rhythm is believed to represent a cortico-cortical or thalamo-cortical interaction. The direction of this interaction is an open question. While it is mostly believed that this rhythm originates in occipital (back part of the brain) or thalamic (which is deep) areas, and spreads to other (more frontal) areas this view has also been challenged [17,18].

For illustration we show results for PSI for one selected subject in Fig. 3. The power (upper right panel), averaged over the two occipital channels (O1 and O2), shows a very

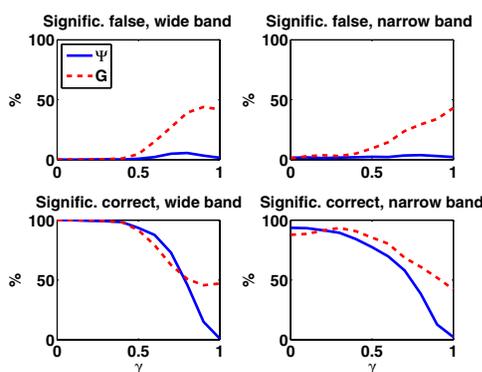


FIG. 2 (color online). Fraction of significant detections of Granger causality and PSI as a function of noise level  $\gamma$ .

strong peak at 9.5 Hz. PSI values were calculated for all channel pairs with frequency resolution 0.5 Hz using a frequency band of 5 Hz width centered around frequency  $f$ . In the lower right panel we show the net information flux at  $f = 9.5$  Hz defined for the  $i$ th channel by  $\Psi_{\text{net}}(i, f) = \sum_j \Psi_{ij}(f) / \text{std}[\sum_j \Psi_{ij}(f)]$ . We clearly observe that frontal channels are net drivers ( $\Psi_{\text{net}} > 0$ ) and occipital channels net recipients ( $\Psi_{\text{net}} < 0$ ).

To show preferred direction for all pairs of channel and for all frequencies we calculate the respective contribution to a given direction: for channels  $i$  and  $j$  with locations  $\mathbf{r}_i$  and  $\mathbf{r}_j$  in the two dimensional plane (as shown in the lower right panel), respectively, we calculate the normalized difference vector  $\delta\mathbf{r}_{ij} = (\mathbf{r}_j - \mathbf{r}_i) / |\mathbf{r}_j - \mathbf{r}_i|$  and project it onto  $\mathbf{u} = (-1, 0)^T$  for right-to-left direction and onto  $\mathbf{u} = (0, -1)^T$  for front-to-back direction. We finally calculate the contribution of  $\Psi_{ij}(f)$  to direction  $\mathbf{u}$  as  $\Psi_{ij}(f, \mathbf{u}) = \Psi_{ij}(f) \mathbf{u} \cdot \delta\mathbf{r}_{ij}$ .

Results for all channel pairs and for all frequencies are shown for right-left information flow (upper left panel) and for front-back information flow (lower left panel). We do not observe any preferred direction in the right-left flow. In contrast, the information flow in the front-back direction shows a clear positive plateau at the alpha frequency (indicated with the letter “B”) meaning that typically the frontal channels are estimated as the drivers. We also observe a positive and negative peak (indicated with letters “A” and “C”) at frequencies around 7 and 12 Hz, respectively. Note that these peaks differ by the width of the frequency band. They are clearly artifacts due to inadequate settings of the band. Specifically, the alpha rhythm has a preferred phase (for given channel pair) which must be distinguished from the slope of the phase. E.g., PSI at around 7 Hz estimates the slope on the range from 4.5 to 9.5 Hz. The right edge of this interval just includes the

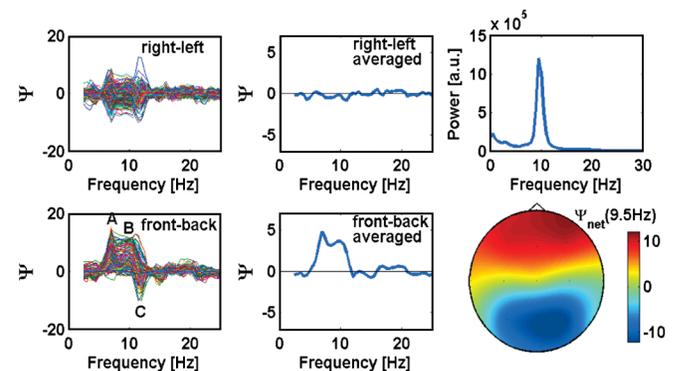


FIG. 3 (color). Upper right: signal power as a function of frequency averaged over the two occipital channels O1 and O2 showing a clear alpha peak at  $f = 9.5$  Hz. Lower right: net information flux at  $f = 9.5$  Hz. Left panels: PSI for all channel pairs and all frequencies projected on right-to-left and front-to-back direction, respectively. Middle panels: Average of PSI as a function of frequencies over all channel pairs.

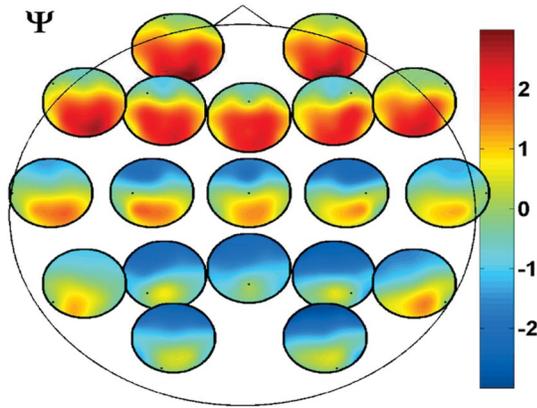


FIG. 4 (color). Phase-slope index for all pairs of channels averaged over all subjects each at the peak of the alpha rhythm. The  $i$ th small circle is located at the  $i$ th electrode position and is a contour plot of the  $i$ th row of the matrix with elements  $\Psi_{ij}$ . The red color in frontal circles indicates that the frontal electrodes are estimated as the drivers.

rising part of the phase but not the descending part leading to positive values for PSI if the phase is positive.

We found a similar structure in about 60% of the subjects. An average over all subjects of  $\Psi$  between all channels is shown in Fig. 4. We also found a substantial intersubject variability, both with regard to PSI and actual phase at the alpha peak. The origin is unclear and goes beyond the scope of this Letter. Granger causality did not yield any consistent spatial pattern.

Recent neuroimaging studies have challenged a simple view on a rest condition by showing a presence of default states in the cortex, which display complex patterns of neuronal activation [19,20]. We here show that not only specific areas are coactivated during rest state, but they also demonstrate at a gross level a preferential “default” mode of information flow in the cortex. Importantly, the drivers of such flow are mostly situated in the frontal areas, from where many top-down attentional influences are thought to be originated [21]. This suggests that the maintenance of vigilance is a process displaying a coordination of neuronal activity with well-defined drivers and recipients of information flow.

To conclude, we presented a new method to estimate the direction of causal relations from time series’ based on the phase slope of the cross-spectra. We here defined an average of the phase slope such that this average is insensitive to arbitrary mixtures of independent sources. We verified this for random linear systems also showing that Granger causality is highly sensitive to mixtures of independent noise sources. Additionally, we showed that in situations with combined unidirectional flow and undirected noise

our method correctly distinguished the two phenomena. Application on real EEG data underlined the versatility of our method as a universal tool for estimating causal flow in noisy physical systems.

We acknowledge partial funding from DFG, BMBF, BBCN Berlin and EU.

- 
- [1] C. Granger, *Econometrica* **37**, 424 (1969).
  - [2] R. Kaufmann and D. Stern, *Nature (London)* **388**, 39 (1997).
  - [3] D. Marinazzo, M. Pellicoro, and S. Stramaglia, *Phys. Rev. E* **73**, 066216 (2006).
  - [4] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. Bressler, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9849 (2004).
  - [5] J. Sato, E. Amaro, D. Takahashi, M. Felix, M. Brammer, and P. Morettin, *NeuroImage* **31**, 187 (2006).
  - [6] Z. Albo, G.D. Prisco, Y. Chen, G. Rangarajan, W. Truccolo, J. Feng, R. Vertes, and M. Ding, *Biol. Cybern.* **90**, 318 (2004).
  - [7] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, *Clin. Neurophysiol.* **115**, 2292 (2004).
  - [8] G. Nolte, F. Meinecke, A. Ziehe, and K. Müller, *Phys. Rev. E* **73**, 051913 (2006).
  - [9] P. Nunez, R. Srinivasan, A. Westdorf, R. Wijesinghe, D. Tucker, R. Silberstein, and P. Cadusch, *Electroencephalogr. Clin. Neurophysiol.* **103**, 499 (1997).
  - [10] MATLAB code at <http://ml.cs.tu-berlin.de/causality/>.
  - [11] M. Ding, Y. Chen, and S. Bressler, *Handbook of Time Series Analysis* (Wiley, New York, 2006), pp. 437–459.
  - [12] S. Marple, *Digital Spectral Analysis with Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1987).
  - [13] A. Schlögl, BIOSIG—an open source software library for biomedical signal processing, <http://BIOSIG.SF.NET>.
  - [14] P. Nunez, R. Silberstein, Z. Shi, M. Carpenter, R. Srinivasan, D. Tucker, S. Doran, P. Cadusch, and R. Wijesinghe, *Clin. Neurophysiol.* **110**, 469 (1999).
  - [15] F. L. da Silva, *Electroencephalogr. Clin. Neurophysiol.* **79**, 81 (1991).
  - [16] M. Schreckenberger, C. Lange-Asschenfeldt, M. Lochmann, K. Mann, T. Siessmeier, H. Buchholz, P. Bartenstein, and G. Gründer, *NeuroImage* **22**, 637 (2004).
  - [17] J. Ito, A. Nikolaev, and C. van Leeuwen, *Biol. Cybern.* **92**, 54 (2005).
  - [18] N. Oishi, T. Mima, K. Ishii, K. Bushara, T. Hiraoka, Y. Ueki, H. Fukuyama, and M. Hallett, *NeuroImage* **36**, 1301 (2007).
  - [19] H. Laufs, K. Krakow, P. Sterzer, E. Eger, A. Beyerle, A. Salek-Haddadi, and A. Kleinschmidt, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11053 (2003).
  - [20] D. Mantini, M. Perucci, C.D. Gratta, G. Romani, and M. Corbetta, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13170 (2007).
  - [21] C. Gilbert and M. Sigman, *Neuron* **54**, 677 (2007).