An Information-Theoretic Framework to Map the Spatio-Temporal Dynamics of the Scalp Electroencephalogram

Luca Faes[*], Daniele Marinazzo, Giandomenico Nollo, and Alberto Porta

Abstract— **We present the first application of the emerging framework of Information Dynamics to the characterization of the EEG activity. The framework provides entropy-based measures of information storage (self entropy, SE) and information transfer (joint transfer entropy (TE) and partial TE), which are applied here to detect complex dynamics of individual EEG sensors and causal interactions between different sensors. The measures are implemented according to a model-free and fully multivariate formulation of the framework, allowing the detection of nonlinear dynamics and direct links. Moreover, to deal with the issue of volume conduction, a compensation for instantaneous effects is introduced in the computation of joint and partial TE. The framework is applied to resting state EEG measured from healthy subjects in the eyes open (EO) and eyes closed (EC) conditions, evidencing condition-dependent patterns indicative of how information is distributed in the EEG sensor space. The SE was uniformly low during EO and significantly higher in the posterior areas during EC. The joint and partial TE were saturated by instantaneous effects, documenting how volume conduction blurs the detection of information flow in the EEG. However, the use of compensated TE measures led us to evidence meaningful patterns like the presence of local sinks of information flow and propagation motifs, and the emergence of prevalent front-to-back EEG propagation during EC. These findings support the feasibility of our information-theoretic approach to assess the spatio-temporal dynamics of the scalp EEG in different conditions.**

Index Terms— **causal connectivity, complex dynamics, EEG propagation, entropy estimation, multivariate time series analysis, transfer entropy, volume conduction.**

I. INTRODUCTION

THANKS to its noninvasiveness, portability and high temporal resolution, electroencephalography (EEG) is a well-established and widely used technique to investigate brain dynamics and brain interactions in humans. The study of brain dynamics based on the EEG is accomplished through the application of time series analysis methods to individual recordings, and aims at mapping the spatial distribution of the dynamical complexity of the brain activity measured at the level of the scalp. This approach is largely followed in EEG analysis to provide a quantitative description of the brain activity related to different physiological or pathological states [1-4]. Beyond the mapping of brain activity, the study of brain interactions is viewed as central for the understanding of the organized behavior of spatially distributed cortical regions. The estimation of brain interactions aims at describing the connectivity patterns which encode the direction and strength of the information flow among cortical areas. To this end, a big variety of estimation methods exists which model and quantify brain connectivity from the multichannel EEG [5]. Most of these methods are based on the concept of Granger causality implemented through the time- or frequency-domain representation of multivariate linear parametric models [6-9], and have been used to assess the patterns of EEG activity propagation in different conditions [10-15].

In spite of its wide application to the assessment of the brain function, meaningful EEG analysis still remains a challenging task because of the complex nature of the signals and of the neuroelectrical mechanisms underlying its generation. The dynamical nature of the EEG has been largely debated, with evidences suggesting that it may be regarded as a realization of a linear stochastic process [16] and that it may exhibit significant complex temporal fluctuations that reflect nonlinear processes [17]. Considering this, approaches able to detect both linear and nonlinear properties are recommended to fully characterize the intrinsic nature of the EEG [18]. The essentially nonlinear physiological basis of the EEG encourages also the development of connectivity estimators which depart from the linear modeling approach often followed for Granger-causal analysis [19]. Moreover, a big issue related to the estimation of directed interactions among EEG signals is the fact that the acquired data are a largely unknown superposition of the actual brain activities. This issue, commonly denoted as volume conduction, poses a serious challenge to EEG-based analyses of the information flow across different brain regions, because the underlying mixing of unmeasured

cortical sources favors the detection of false positive causality between the recorded signals [20-23]. While inverse source reconstruction approaches are often used to limit the issue of volume conduction, these approaches have to face an ill-posed inverse problem which imposes to make assumptions possibly inconsistent with the properties of the sources to be reconstructed. Moreover, since it has been shown that artifacts of volume conduction persist in the source activities reconstructed through inversion methods [24], the issue of compensating for these artifacts is still open.

The present study faces the study of brain dynamics and brain interactions assessed at the EEG sensor level within the emerging information-theoretic framework of information dynamics [25,26]. This framework provides entropy-based measures of *information storage* and *information transfer* that allow respectively the detection of complex dynamics in individual brain regions and of directed interactions between distant regions. Information storage refers to the presence of information in the past of a neural process that will serve to predict a fraction of the information contained in the future of the process [27]. Its application to multichannel EEG allows to map the spatial distribution of the complexity of brain dynamics: the lower the dynamical complexity at a given site, the higher the information stored at this site. Information transfer is a well-defined concept that implements the notion of Granger causality in a probabilistic framework through the definition of the so-called transfer entropy (TE) [28]. Contrary to bivariate analyses that investigate information transfer considering only two signals, we operate in a fully multivariate context through the computation of the overall information transferred to an assigned target signal from all other signals considered together, defined as joint TE, or the information transferred to the target from an individual source signal in the presence of the remaining signals, defined as partial TE (PTE). In the context of multichannel EEG, the PTE allows the detection of direct interactions between two brain sites, i.e., interactions not mediated by the activity at the other sites. The practical computation of information storage and information transfer is performed employing a recently devised model-free estimation approach that allows to deal with any type of linear and nonlinear dynamics [29]. This approach minimizes the bias in the estimation of information dynamics by exploiting a compensation scheme based on nearest-neighbor estimators and tackling the curse of dimensionality through a procedure for dimension reduction based on non-uniform embedding. Moreover, the problem of source mixing manifested in the EEG sensor space is faced introducing in TE analysis a compensation for instantaneous effects, i.e. effects resulting in simultaneous (non-delayed) information sharing between two or more processes [30]. The framework is applied to resting state EEG measured from healthy subjects during eyes open and eyes closed, in order to evaluate whether these two neurophysiological conditions are characterized by spatial patterns of information storage and transfer, and to assess the impact of volume conduction on the measures of information transfer.

The analysis framework presented in this study is implemented in the Matlab® ITS Toolbox, available for download at the link www.lucafaes.net/its.html.

II. METHODS

A. Subjects, Signals and Pre-Processing

The study comprises twenty-one young healthy subjects (11 females; age 22-39 yrs) with normal vision and reporting no history of neurological or mental diseases. EEG recordings were performed in an electrically and acoustically shielded, darkened room, where participants were comfortably lying down in a relaxed state. Signals were acquired with eyes closed (EC) for 40 s, and with eyes open (EO) for further 40 s.

EEG signals were recorded (Micromed Brain Quick System) from 19 channels with electrodes placed according to the 10–20 standard system. All electrodes were referred to a forehead common reference (Fpz) with Oz ground electrode. Electrooculographic (EOG) and ECG signals were also recorded. All signals were digitized with a sampling rate f_s=128 Hz and a precision of 16 bit.

EEG pre-processing consisted of: (i) digital band-pass filtering (0.3-40 Hz, FFT band-pass, zero phase-shift, unit gain filter) to remove baseline noise and extract information within the frequency bands of interest; (ii) if necessary, removal of artefacts from eyes blinks, eyes movements and cardiac activity by separation of these components from the brain activity through independent component analysis involving the 19 EEG signals and EOG and ECG signals; (iii) re-referencing of the EEG signals by subtracting from each channel the average of all other channels (common average referencing); (iv) selection of the most stationary 8 s windows, performed through an iterative test checking restricted weak stationarity of the signal through the approach described in [31].

B. Information-Theoretic Analysis

The time series measured from each subject are interpreted as realizations of an *M*-dimensional stochastic process that describes the EEG scalp dynamics (*M*=19 in this study). The analysis is performed considering one scalar sub-process as the target Y and collecting all other sub-processes in the (*M*-1)-dimensional driver process **X**. Denoting as Y_n the scalar random variable obtained sampling the process Y at the present time n, and as $Y_n^-=[Y_{n-1},Y_{n-2},\ldots]$ the vector variable describing the past of the process (the same notation holds for any scalar process $X \in \mathbf{X}$), the predictive information of Y measures the amount of information carried by Y_n that can be predicted from the past of the whole observed process $[\mathbf{X}_n^-,Y_n^-]$ [26,29,32]. The predictive information can be expressed as the sum of the information stored in Y and the information transferred to from **X** to Y, where the information storage and the information transfer are computed by the self entropy (SE) and the joint transfer entropy (TE), defined respectively as

$$S_Y = I(Y_n; Y_n^-) = H(Y_n) - H(Y_n \mid Y_n^-) \tag{1a}$$

$$T_{X \to Y} = I(Y_n; X_n^- \mid Y_n^-) = H(Y_n \mid Y_n^-) - H(Y_n \mid X_n^-, Y_n^-) \tag{1b}$$

The functionals $H(\cdot)$, $H(\cdot|\cdot)$, $I(\cdot;\cdot)$, and $I(\cdot;\cdot|\cdot)$ denote respectively entropy, conditional entropy, mutual information (MI) and conditional MI. Then, with reference to a specific driver $X \in \mathbf{X}$, the partial (conditioned) TE (PTE) is expressed as

$$\begin{aligned} T_{X \to Y \mid Z} &= I(Y_n; X_n^- \mid Y_n^-, Z_n^-) \\ &= H(Y_n \mid Y_n^-, Z_n^-) - H(Y_n \mid X_n^-, Y_n^-) \end{aligned} \tag{2}$$

where $\mathbf{Z} = \mathbf{X} \backslash X$ is the (M-2)-dimensional process containing all drivers except X.

Even though the existence of a significant PTE is usually taken as an indication of the presence of *lagged causality* from X to Y, the existence of an *instantaneous causality* might render useless this functional. Indeed, instantaneous effects might be erroneously taken as lagged and inflate TE and PTE because they are not conditioned out. Instantaneous causality between X and Y is present when X_n and Y_n are not independent given the past of the whole process $\{\mathbf{X}, Y\} = \{X, \mathbf{Z}, Y\}$ [33]; this condition is verified when nonzero values are taken by the conditional MI $I_{Y,X|\mathbf{Z}} = I(Y_n; X_n \mid X_n^-, Y_n^-, \mathbf{Z}_n^-)$. Defined as $\mathbf{X}_n^0$ the set of all and only the drivers $X \in \mathbf{X}$ for which instantaneous causality occurs between X and Y, more appropriate functionals for testing lagged causality are the compensated TE (cTE) from $\mathbf{X}$ to Y and compensated PTE (cPTE) from X to Y given $\mathbf{Z}$, defined respectively as [30]

$$T_{X \to Y}^0 = I(Y_n; X_n^- \mid Y_n^-, \mathbf{X}_n^0) \ , \tag{3a}$$

$$T_{X \to Y \mid Z}^0 = I(Y_n; X_n^- \mid Y_n^-, Z_n^-, \mathbf{X}_n^0) \ , \tag{3b}$$

The absence of instantaneous causality between X and Y can be verified after the compensation introduced by Eq. (3) observing that $I_{Y,X|\mathbf{Z}}^0 = I(Y_n; X_n \mid X_n^-, Y_n^-, \mathbf{Z}_n^-, \mathbf{X}_n^0) = 0$; while this condition is always true in theory, its verification in practical analysis serves as confirmation of the efficacy of the compensation for instantaneous effects.

C. Estimation of Information Dynamics

In this study, the measures of information dynamics are estimated following a procedure for non-uniform embedding which optimizes the lagged components to be included in the conditioning vectors according to a criterion for maximum relevance and minimum redundancy [29]. The approach is based on the progressive selection, from a set of *candidate components* Ω including the lagged variables that sample the past of the relevant processes up to a maximum lag L, of the variables which are the most informative about the target variable Y_n. For computing the SE, the initial set of candidate components is $\Omega^Y = \{Y_{n-\tau}, ..., Y_{n-L\tau}\}$, and the selection procedure

approximates the past Y_n^- with a sub-vector $Y_n^{d_Y}$ composed of the d_Y most relevant lagged variables of Y. For the computation of TE and PTE, the set of initial candidates is $\Omega^{XYZ} = \Omega^Y \cup \Omega^X \cup \Omega^Z$, where $\Omega^X = \{X_{n-1}, X_{n-1-\tau}, ..., X_{n-1-(L-1)\tau}\}$ and the same notation applies to each scalar component of $\mathbf{Z}$. Then, the past of the whole process, $[X_n^-, Y_n^-, \mathbf{Z}_n^-]$, is approximated with the d-dimensional vector $[X_n^{d_X}, Y_n^{d_Y}, \mathbf{Z}_n^{d_Z}]$ composed of lagged variables of X, Y and $\mathbf{Z}$. Each lagged variable is selected only if it contributes with significant information to the target, where statistical significance is assessed by means of a randomization test employing surrogate data. Details about the procedure for non-uniform embedding are in [29].

The estimation of the information storage was performed first approximating Y_n^- with the finite-dimensional vector $Y_n^{d_Y} \subseteq \Omega^Y$, and then expressing the SE as $S_Y = H(Y_n) - H(Y_n, Y_n^{d_Y}) + H(Y_n^{d_Y})$ and computing the three entropy terms by nearest neighbor entropy estimation. With this approach, the highest-dimensional entropy term $H(Y_n, Y_n^{d_Y})$ is estimated through neighbor search in the (d_Y+1)-dimensional space, while the lower-dimensional terms $H(Y_n)$ and $H(Y_n^{d_Y})$ are estimated through range searches in the spaces of dimension 1 and d_Y. This results in the estimate [29]:

$$\hat{S}_Y = \psi(N) + \psi(k) - \left\langle \psi(N_1 + 1) + \psi(N_{d_Y} + 1) \right\rangle \tag{4}$$

where ψ is the digamma function, N is the number of available realizations, and N_1 and N_{d_Y} are the number of points whose distance from Y_n and $Y_n^{d_Y}$ is strictly less than the distance from $[Y_n, Y_n^{d_Y}]$ to its k-th neighbor. ($\langle \cdot \rangle$ denotes average and the maximum distance is used here).

To estimate the information transfer, first we enlarged the vector $Y_n^{d_Y}$ with the components $X_n^{d_X} \subseteq \Omega^X$ and $\mathbf{Z}_n^{d_Z} \subseteq \Omega^\mathbf{Z}$ derived by non-uniform embedding to get the full embedding vector $V_n = [X_n^{d_X}, Y_n^{d_Y}, \mathbf{Z}_n^{d_Z}]$. Then, expressing the TE as $T_{\mathbf{X} \to Y} = H(Y_n, Y_n^{d_Y}) - H(Y_n^{d_Y}) - H(Y_n, V_n) + H(V_n)$, and expressing the PTE as $T_{X \to Y \mid \mathbf{Z}} = H(Y_n, Y_n^{d_Y}, \mathbf{Z}_n^{d_Z}) - H(Y_n^{d_Y}, \mathbf{Z}_n^{d_Z}) - H(Y_n, V_n) + H(V_n)$, the term $H(Y_n, V_n)$ is estimated through neighbor search in the $(d_X + d_Y + d_Z + 1)$-dimensional space, while the other terms are estimated through range searches in the subspaces of lower dimension. This results in the estimates:

$$\hat{T}_{\mathbf{X} \to Y} = \psi(k) + \left\langle \psi(N_{d_Y} + 1) - \psi(N_{d_{Y1}} + 1) - \psi(N_d + 1) \right\rangle$$
$$\hat{T}_{X \to Y \mid Z} = \psi(k) + \left\langle \psi(N_{d_{YZ}} + 1) - \psi(N_{d_{YZ1}} + 1) - \psi(N_d + 1) \right\rangle \tag{5}$$

where N_{d_Y}, $N_{d_{Y1}}$, $N_{d_{YZ}}$, $N_{d_{YZ1}}$, and N_d are the number of points whose distance from $Y_n^{d_Y}$, $[Y_n, Y_n^{d_Y}]$, $[Y_n^{d_Y}, \mathbf{Z}_n^{d_Z}]$, $[Y_n, Y_n^{d_Y}, \mathbf{Z}_n^{d_Z}]$, and V_n, respectively, is strictly less than the distance from $[Y_n, V_n]$
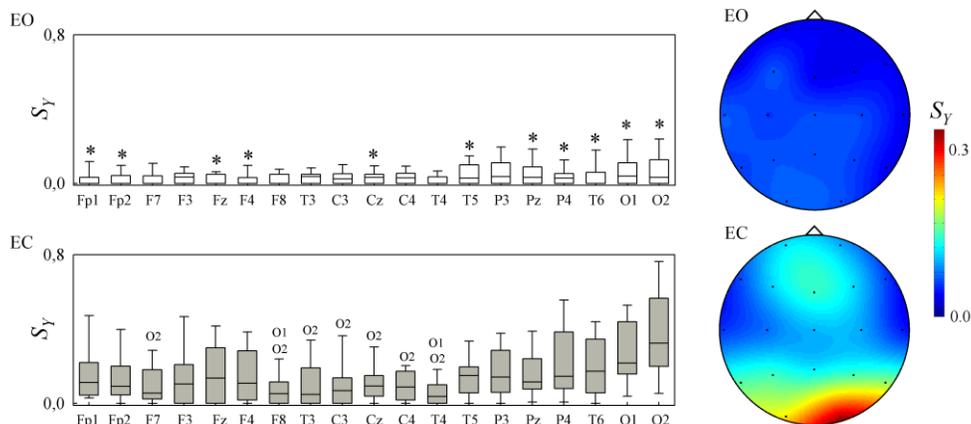
Fig. 1. Information storage measured by the self entropy (SE) S_Y computed for each electrode during eyes open (EO, top) and eyes closed (EC, bottom), and represented as box plot across subjects (left), or as color image (median across subjects, right). Asterisks indicate significant difference between EO and EC (Friedman + Sign Rank); for any given electrode, electrode labels above boxes indicate significant difference between the two electrodes (Kruskal Wallis + post-hoc).

to its k-th neighbor.

The presence of instantaneous causality between X and Y was tested approximating the past of the whole process with the estimated full embedding vector V_n and checking the statistical significance of the conditional MI $I_{Y,X|\mathbf{Z}}=I(Y_n ; X_n| V_n)$. This was achieved exploiting the same randomization test used in the non-uniform embedding scheme, applied to the conditional MI estimated through the nearest neighbor method. If instantaneous causality was detected between X_n and Y_n, $\mathbf{X}_n^0$ was incremented with X_n. The effectiveness of the compensation for instantaneous effects was tested checking the statistical significance of the conditional MI $I_{Y,X|\mathbf{Z}}^0=I(Y_n ; X_n| V_n ; \mathbf{X}_n^0)$. Then, the cTE and cPTE were estimated following the same procedure described above, but incrementing the sets of initial candidates Ω^X and $\Omega^{\mathbf{Z}}$ with the relevant zero-lag components taken from $\mathbf{X}_n^0$ before the execution of non-uniform embedding.

The statistical test for lagged causality used to decide whether the PTE and cPTE computed according to (2) and (3b) are significantly different from zero was performed implicitly through the application of the procedure for non uniform embedding: if the resulting embedding vector did not contain any lagged component from the driver X (i.e., if $X_n^{d_X}=[\cdot]$), the estimated PTE or cPTE was exactly zero and deemed as non-significant; on the contrary, if at least one lagged variable was selected from X by the randomization test (i.e., if $X_n^{d_X}\neq[\cdot]$), the PTE or cPTE was strictly positive and thus deemed as statistically significant [29]. The same test can be applied to test the statistical significance of the joint TE or cTE of (1b, 3b) by considering the presence of lagged components from all drivers $\mathbf{X}$.

All analyses performed in this study are based on the ITS Matlab® Toolbox (www.lucafaes.net/its.html) and can be probed launching the script EEG_EyesClosed.mat.

D. Data Analysis

Each analyzed data set consisted of M=19 time series of N=1024 samples. The computation of information dynamics was performed on the normalized time series obtained subtracting the mean from each EEG time series, and dividing the result by the standard deviation. The set of candidate components for non-uniform embedding was determined including L=5 past components for each series, and spacing each component with a lag τ determined as the decorrelation time of each individual time series, i.e. the lag at which the autocorrelation drops below a threshold equal to $1/e$ [34]. The procedure for non-uniform embedding was run using random shift surrogates to test for the statistical significance of the information brought to the target by the lagged variable selected at each step, and setting to k=10 the number of neighbors used for all entropy estimations [29].

Statistical analysis of the information storage (SE) and information transfer (joint TE) computed for the M=19 scalp regions in the two experimental conditions was performed using nonparametric statistics. Specifically, the Friedman's test was used to check for significant differences of SE, TE or cTE between the two conditions (EO vs. EC) after adjusting for possible spatial electrode distribution effects; when the test returned statistically significant differences, post-hoc analysis was performed for each electrode location using the Wilcoxon signed rank test. The Kruskal-Wallis analysis of variance was employed to test for regional differences of SE, TE or cTE in an assigned condition (EO or EC); when the test returned statistically significant differences, post-hoc analysis was performed between each pair of electrodes. In all statistical tests, compensation for multiple comparisons was performed applying a Bonferroni correction.

III. RESULTS

The duration of the window of past samples spanned in our analysis by the procedure for non-uniform embedding, $L\tau/f_s$, was equal to 192±70 ms during EO and 152±50 ms during EC

(mean±SD over all subjects and channels). During these analyses, the maximum length of the time-lagged causality effects detected by the procedure was 106±22 ms (EO) and 109±22 ms (EC) for the computation of TE/PTE, and 88±63 ms (EO) and 95±48 ms (EC) for the computation of cTE/cPTE.

Fig. 1 depicts the spatial mapping of the predictable EEG dynamics in the EO and EC conditions, expressed as the information storage computed through the SE. During EO, the amount of information stored in the EEG was relatively low and did not vary significantly across sensors. The storage increased significantly during EC, documenting higher regularity (i.e., lower complexity) of the brain dynamics. The increase was statistically significant in the frontal regions (Fp1, Fp2, Fz, F4) and even more marked in the posterior scalp regions (T5, Pz, P4, T6, O1, O2). The two occipital electrodes (O1 and particularly O2) stored also a significantly higher amount of information than many other electrodes (F7, F8, T3, C3, Cz, C4, T4).

Fig. 2 depicts the spatial distribution of the overall information transferred towards any scalp region, quantified by the joint TE directed to the target EEG sensor from all other sensors. The information transfer was remarkable in both EO and EC conditions, and resulted higher than the information storage in almost all scalp regions. The distribution of the information transfer was rather uniform: while some statistically significant regional differences were observed, especially in the EO condition, these differences were not substantial and did not evidence variations in the information flow along the back-to-front directions. No significant differences were detected between EO and EC.

The barplots of Fig. 3 report the results of the analysis of lagged causality and instantaneous causality performed counting, for each target electrode, the number of statistically significant PTE computed with the electrode considered as target (black) or considered as driver (white); moreover the percentages of significant conditional MI involving the target electrode is also reported (gray). The same information is reported on the grayscale images on the right, showing also the pairs of electrodes connected by a significant PTE (arrows in midplots) or exhibiting significant instantaneous correlation (lines in right smaller plots) in at least 8 out of 21 subjects. In both the EO and EC conditions, the percentages of incoming and outgoing directed links were comparable with each other and across regions, and were always substantially lower than the percentage of significant instantaneous links. The analysis of causal and instantaneous links found recurrently in at least one third of the subjects revealed that Granger causal relations were consistently found between pairs of adjacent electrodes, and instantaneous relations were ubiquitously found between almost any pair of electrodes. These patterns were observed almost identically during EO and EC.
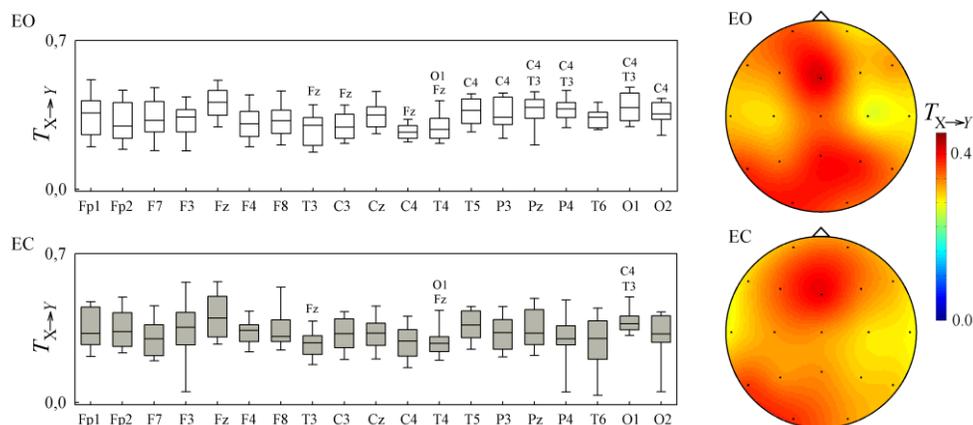
Fig. 2. Information transfer measured by the joint transfer entropy (TE) $T_{X \to Y}$ computed for each electrode during eyes open (EO, top) and eyes closed (EC, bottom), and represented as box plot across subjects (left), or as color image (median across subjects, right). For any given electrode, electrode labels above the box plot indicate significant difference between the two electrodes (Kruskal Wallis + post-hoc).
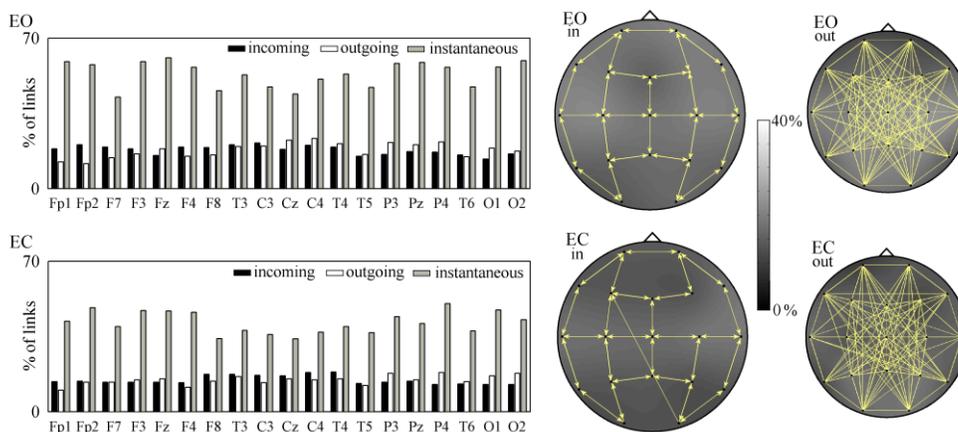

Fig. 3. Percentage of statistically significant causal links detected by the partial TE (PTE) ($T_{X \to Y|Z} > 0$) entering and leaving each electrode during eyes open (EO, top) and eyes closed (EC, bottom), and represented as bar plots (left) or as grayscale image (mid, right). In left plots, gray bars indicate the percentage of significant instantaneous causal relations ($I_{Y,X|Z} > 0$) detected for each electrode. In mid images, yellow arrows depict the causal links detected as statistically significant in at least one third of the subjects. In right images, yellow lines depict the instantaneous links detected as statistically significant in at least one third of subjects.

Fig. 4 reports the distribution of the information transfer computed as in Fig. 2 but using the joint cTE in place of the joint TE. The compensation of instantaneous causality led to substantially lower amounts of information transfer (see Fig. 4 vs. Fig. 2), which became generally lower than the information storage (see Fig. 4 vs. Fig. 1). Moreover, the cTE evidenced regional propagation patterns peculiar of each condition that were not visible using the TE. During EO, the frontal lateral electrodes (F8 and particularly F7) received the largest information flow, and a high amount of information was transferred also towards the right occipital and temporal electrodes;

the parietal areas (P3, P4) were those receiving the lower information flow. During EC, the information flowing to F7 and F8 was still high, but remarkable flows of information were directed also towards Cz and, even more strongly, towards the occipital electrodes (especially O2) and the right temporal area. The information flow directed to T6, O1 and O2 was significantly higher than during EO, so that these areas received significantly higher information than that received by some frontal regions (e.g., F3,Fz,F4,F8). The information transfer was still low in the parietal areas P3 and P4.
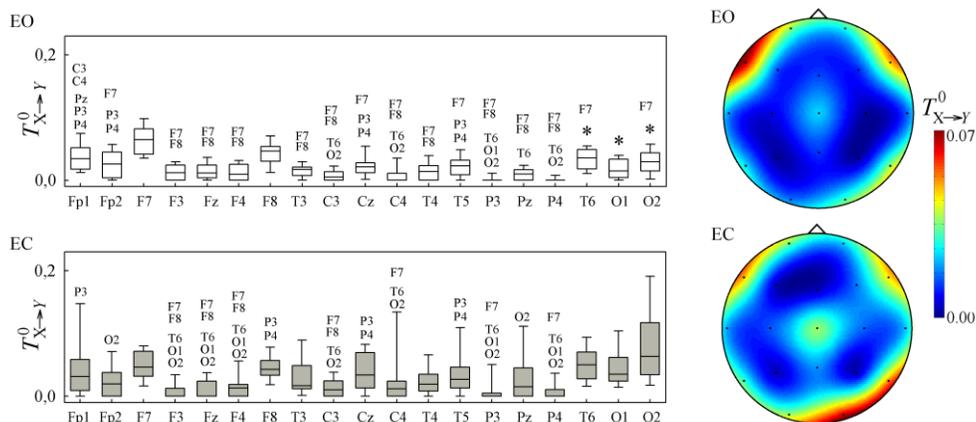
Fig. 4. Information transfer measured by the compensated joint TE (cTE) $T^0_{X \to Y/\mathbf{Z}}$ computed for each electrode during eyes open (EO, top) and eyes closed (EC, bottom), and represented as box plot across subjects (left), or as color image (median across subjects, right). Asterisks indicate significant difference between EO and EC (Friedman + Sign Rank); for any given electrode, electrode labels above the box plot indicate significant difference between the two electrodes (Kruskal Wallis + post-hoc).
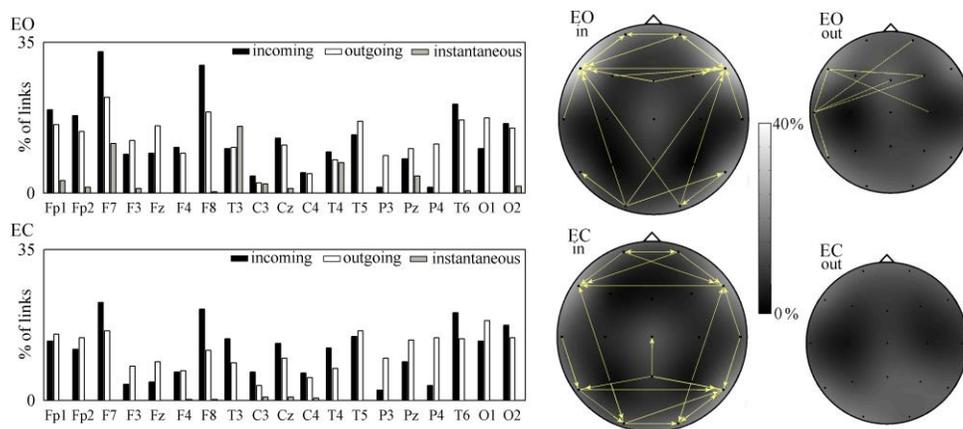

Fig. 5. Percentage of statistically significant causal links detected by the compensated partial TE (cPTE) ($T^0_{X \to Y/\mathbf{Z}} > 0$) entering and leaving each electrode during eyes open (EO, top) and eyes closed (EC, bottom), and represented as bar plots (left) or as grayscale image (mid, right). In the left plots, gray bars indicate the percentage of significant instantaneous causal relations ($I^0_{Y,X|\mathbf{Z}} > 0$) detected for each electrode. In mid images, yellow arrows depict the causal links detected as statistically significant in at least one third of the subjects. In right images, yellow lines depict the instantaneous links detected as statistically significant in at least one third of the subjects.

Fig. 5 is structured like Fig. 3, reporting the analysis of lagged and instantaneous causal links assessed after compensation of instantaneous effects. As a result of the compensation, the number of significant instantaneous causal relations decreased dramatically, reducing to some residual links confined to the connections involving the electrodes F7 and T3 during EO, and becoming completely absent during EC. This more parsimonious representation of the patterns of lagged causality allowed to distinguish the two considered conditions in terms of significant PTE values. First, the percentage of incoming and outgoing connections became unbalanced for some electrodes, e.g., revealing a tendency of the fronto-lateral areas (electrodes F7 and F8) to receive more information than that sent out to the other areas. During EO, the PTE patterns representing the most consistent causal relations were those directed to F7 and F8 from the frontal electrodes, from the temporal electrodes T3 and T4, and from the occipital electrodes O1 and O2. This suggests the existence of a

frontal propagation circuit, fed also by lateral and occipital activity. During EC, the frontal circuit is still present (even if weaker) but also a posterior circuit emerged which was fed by front-to-back propagation (T5←Pz→T6; T3→T5→O1, T4→T6→O2; P3→O1, P4→O2).

The emergence during EC of EEG propagation directed towards the posterior and occipital regions was revealed also counting the total number of front-to-back and back-to-front significant links detected by cPTE, which was 133 vs. 79 (sum over all subjects). The same counting performed during EO returns 95 vs. 89, suggesting that the prevalence of front-to-back propagation is limited to the EC condition. It is worth noting that this behavior was evident only after compensating for instantaneous effects, since the total number of front-to-back and back-to-front significant PTE links was 261 vs. 243 during EO and 220 vs. 205 during EC.

IV. DISCUSSION

The present study reports the first exhaustive application of the emerging framework of information dynamics [25,26] to the characterization of EEG activity. Through the development of a fully multivariate representation and model-free estimation of information dynamics, we assessed the processing of information inside the network of EEG sensors by quantifying the amounts of information stored at each sensor and transferred across different sensors. While information storage and information transfer are constituent elements of the predictive information of a network of interacting dynamic processes which are related to each other [32], previous attempts to characterize EEG dynamics were limited to the investigation of only one of these aspects. The information storage was never directly computed for EEG dynamics, as more traditional complexity measures are commonly preferred to assess the complexity of the EEG sensor activity [1-3]. As to the information transfer, a thorough model-free computation of Granger causal measures like the TE has not been performed before on EEG data. Previous recent TE investigations on electro-magnetic brain activities were limited to bivariate analyses restricted to pairs of signals [28,35], mostly because the curse of dimensionality limits severely the reliable model-free estimation of entropy and MI measures for high-dimensional variables. In this study, multivariate model-free analysis of the information transfer was made possible by the estimation framework that we recently proposed [29], which allows to compute reliable estimates of the overall (direct and indirect) causal information flow arriving at any target electrode from all the other electrodes (joint TE), as well as of the structure of the direct interactions estimated in a truly multivariate context that allows ruling out indirect interactions (PTE). Remarkably, the adopted estimators return values of SE, joint TE and PTE which are strictly positive only when associated with significant information storage or transfer. This property led us to assess the statistical significance of the estimated PTE and assess the corresponding percentages of significant causal links.

The framework of information dynamics provides entropy measures that characterize the overall dynamics of the analyzed multivariate time series, rather than concentrating on specific oscillations as it is commonly done in frequency domain analyses. Nevertheless, being sensitive to the amplitude variations in the observed time series, information measures tend to capture the dynamics of the predominant oscillations within the time series. In the context of our EEG analysis, the predominant oscillations are likely those in the alpha band, especially in the EC condition [40,42]. This is suggested in our results by the detection of maximal embedding delays in the range of 90-100 ms, which are compatible with the period of the alpha waves, and by the patterns of information storage that seem to reflect alpha EEG activity as seen in Fig. 1 and discussed in the following. The analysis of the information stored in the EEG dynamics (Fig. 1) revealed characteristic spatial patterns and differences between the analyzed conditions that can be related to known neurophysiological behaviors. It is worth recalling that the information storage assessed through the SE reflects the regularity of the dynamics intended as a quantity complementary to that measured by well-known complexity measures such as Approximate Entropy [36], Sample Entropy [37] or nonlinear predictability [38]. According to this interpretation, the relatively low amounts of information storage measured during EO reflect the richness of the EEG dynamics in this condition, with a high dynamical complexity resulting from the contemporaneous presence of oscillatory activity in several frequency bands. On the contrary, during EC the EEG exhibited a significantly higher information storage in several scalp regions. This lower complexity during EC is very likely related to the emergence of alpha activity in the EEG, which is characterized by dominant regular oscillations at ~10 Hz [39]. This interpretation is confirmed by the fact that the highest amounts of information were stored in the posterior and occipital regions where the alpha activity is known to be more visible [40].

While the information storage can be assessed in a relatively straightforward way through the quantification of the self-predictability of the EEG, the assessment of the remaining component of predictive information, i.e. the information transfer, is complicated by computational and physical issues. The computational problem amounts to dealing with multiple time series, each sampled at multiple lags in the past, for the computation of joint TE and PTE. In this study, this problem was faced through the exploitation of non-uniform embedding and the accurate entropy estimation based on nearest neighbors [29]. The physical problem consists in the known fact that the measured scalp EEG potentials do not reveal exclusively genuine brain activity from localized cortical regions beneath the acquiring electrode, but are rather a mixing of the activity from multiple non-localized cortical regions which are conveyed to the acquiring electrode through volume conduction. Since source mixing is instantaneous and Granger-causal measures look for time-lagged influences, effects of volume conduction have been often implicitly discarded in the analysis of directed connectivity performed in the EEG sensor space [10-12,14,15]. However, theoretical and empirical model-based analyses have demonstrated that instantaneous effects arising among EEG time series as a result of volume conduction have an adverse impact on the estimation of time-lagged causality, and should thus be avoided [21,23,24,41]. Simulation studies have shown that this issue holds also for model-free TE estimators [28,30]. The problem is confirmed also by the results of the present study, showing that TE and PTE have high modulus, do not vary substantially across regions and between conditions, and detect a fully connected network of scalp connectivity where each electrode is bidirectionally linked with the adjacent electrodes (Figs. 2,3). Therefore, also in agreement with previous reports [22,24,28], we advise for the utilization of analysis methods able to deal with effects such as those inherent to volume conduction.

Our results suggest also that reasonably interpretable pat-

terns of directed connectivity may be retrieved in the EEG sensor space provided that a correction for source mixing is applied. In this study, we exploited the idea of compensating for non-physiological instantaneous effects by conditioning on the present of the driver processes that share significant information with the present of the target process while performing TE computation [30]. Adopting this compensation, we were able to elicit significant regional differences in the causal information flow directed to specific EEG sensors (Fig. 4), as well as characteristic patterns of time lagged causality assessed by the PTE (Fig. 5). Our main findings were: the existence of regional sinks of information flow, located mostly in the fronto-lateral regions, and also in the occipital regions during EC; the formation of propagation motifs involving mostly the frontal sensors during EO, and also the temporal and occipital sensors during EC; and the emergence of a prevalent front-to-back propagation of the EEG activity during EC. The presence of significant amounts of information directed to the frontal regions, also originating from the occipital areas in the EO condition (Fig. 5), is consistent with the flow detected in healthy subjects during wakefulness using a frequency domain approach to causal inference [10]. Our result about the prevalence of front-to-back EEG propagation during EC may be counterintuitive, as it is mostly believed that the alpha rhythm, which is prevalent during EC, originates in the occipital regions and spreads towards the frontal areas of the brain [42]. However, this view has been challenged by an increasing number of studies finding that the propagation direction of the alpha waves is predominantly from anterior to posterior cortical regions [43-45], which are thus supported by our findings.

The correction for source mixing applied in this study removed all zero-lag correlations between EEG activities during EC, while some residual instantaneous effects involving the left fronto-temporal regions were observed during EO (Fig. 5). The reasons for this incomplete compensation may be computational and theoretical. Computationally, the test for instantaneous causality performed through the randomization procedure proposed in [29] may lack sensitivity when the corresponding conditional MI involves embedding vectors of relatively high dimension. Theoretically, a rigorous treatment of instantaneous effects should set instantaneous causality between two processes X and Y when their current variables X_n and Y_n are not independent conditionally on *any* combination of past and present variables of the observed multivariate process [33,46]; while in this report we simplified the test rejecting independence between X_n and Y_n conditioned to $\mathbf{X}_n^-$ and Y_n^- only, further analyses should assess the benefit and computational tractability of the complete test which looks at the dependencies conditioning to any subset taken from $\mathbf{Z}_n$. Moreover, approaches attempting to assign a direction to instantaneous links can be considered to assess the causal structure of the network with finer detail [47]. Finally, future studies should also compare the approach for compensation of instantaneous causality used here with alternative ways proposed in

the literature to deal with the detrimental effects of volume conduction. These alternatives comprise approaches in the EEG sensor space, such as corrections involving time inversion tests combined with Granger causality estimation [24] and time-shift tests combined with model-free estimation of TE [28], or the use of measures deemed as insensitive to volume conduction [20,44], as well as approaches that follow the different perspective of performing inverse source reconstruction prior to the estimation of Granger-causal measures [13,21,23].

V. CONCLUSIONS

This work documents the appropriateness of assessing the dynamics of information of the scalp EEG network through the computation of the information stored at each electrode and transferred across electrodes. Our analysis emphasizes the importance of performing a compensation for instantaneous effects in the analysis of TE, since we showed that volume conduction effects dominate the transfer of information across EEG sensors blurring the detection of meaningful patterns of information flow. This inclusive integrated approach to the analysis of information dynamics, though not allowing to draw definitive conclusions about the processing of information of the deep electrocortical sources that generate the observed activity, led us to retrieve condition-dependent patterns descriptive of how information is distributed in the EEG sensor space. We showed that significant transfer of information may occur towards scalp regions which do not store high amounts of information (e.g., frontal areas during EO), but also towards regions with substantial stored information (e.g., occipital areas during EC). Together with the prevalence of front-to-back propagation of alpha activity detected in the EC condition, these findings encourage the exploitation of EEG information dynamics in neurocognitive or clinical investigations.

REFERENCES

[1] J. Bhattacharya, "Complexity analysis of spontaneous EEG," *Acta Neurobiol. Exp.*, vol. 60, no. 4, pp. 495-501, 2000.
[2] X. S. Zhang, R. J. Roy, and E. W. Jensen, "EEG complexity as a measure of depth of anesthesia for patients," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 12, pp. 1424-1433, 2001.
[3] S. Erla et al., "k-Nearest neighbour local linear prediction of scalp EEG activity during intermittent photic stimulation," *Med. Eng. Phys.*, vol. 33, no. 4, pp. 504-512, 2011.
[4] R. Okazaki et al., "Changes in EEG complexity with electroconvulsive therapy in a patient with autism spectrum disorders: a multiscale entropy approach," *Frontiers Human Neurosci.*, vol. 9, 2015.
[5] V. Sakkalis, "Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG," *Comput. Biol. Med.*, vol. 41, no. 12, pp. 1110-1117, 2011.
[6] S. Q. Hu et al., "Causality Analysis of Neural Connectivity: Critical Examination of Existing Methods and Advances of New Methods," *IEEE Trans. Neural Networks*, vol. 22, no. 6, pp. 829-844, 2011.
[7] L. Faes, S. Erla, and G. Nollo, "Measuring connectivity in linear multivariate processes: definitions, interpretation, and practical analysis," *Comp. Math. Methods Med.*, vol. 2012, pp. 140513, 2012.
[8] L. Barnett and A. K. Seth, "The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference," *J. Neurosci. Methods*, vol. 223, pp. 50-68, 2014.

[9] A. Porta and L. Faes, "Wiener-Granger causality in network physiology with applications to cardiovascular control and neuroscience," *Proceedings of the IEEE*, vol. 104, no. 2, pp. 282-309, 2016.

[10] M. Kaminski, K. Blinowska, and W. Szclenberger, "Topographic analysis of coherence and propagation of EEG activity during sleep and wakefulness," *Electroencephalogr. Clin. Neurophysiol.*, vol. 102, no. 3, pp. 216-227, 1997.

[11] R. Kus, M. Kaminski, and K. J. Blinowska, "Determination of EEG activity propagation: pair-wise versus multichannel estimate," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 9, pp. 1501-1510, 2004.

[12] M. G. Philiastides and P. Sajda, "Causal influences in the human brain during face discrimination: a short-window directed transfer function approach," *IEEE Trans. Biomed. Eng*, vol. 53, no. 12 Pt 2, pp. 2602-2605, 2006.

[13] L. Astolfi et al., "Comparison of different cortical connectivity estimators for high-resolution EEG recordings," *Hum. Brain. Mapp.*, vol. 28, no. 2, pp. 143-157, 2007.

[14] K. Blinowska et al., "Transmission of Brain Activity During Cognitive Task," *Brain Topography*, vol. 23, no. 2, pp. 205-213, 2010.

[15] F. Protopapa et al., "Granger causality analysis reveals distinct spatio-temporal connectivity patterns in motor and perceptual visuo-spatial working memory," *Frontiers Comp. Neurosci.*, vol. 8, pp. 00146, 2014.

[16] K. J. Blinowska and M. Malinowski, "Nonlinear and Linear Forecasting of the Eeg Time-Series," *Biol. Cybern.*, vol. 66, no. 2, pp. 159-165, 1991.

[17] N. Hazarika, A. C. Tsoi, and A. A. Sergejew, "Nonlinear considerations in EEG signal classification," *IEEE Trans. Signal Processing*, vol. 45, no. 4, pp. 829-836, 1997.

[18] C. J. Stam, "Nonlinear dynamical analysis of EEG and MEG: review of an emerging field," *Clin. Neurophysiol.*, vol. 116, no. 10, pp. 2266-2301, 2005.

[19] D. Marinazzo et al., "Nonlinear connectivity by Granger causality," *Neuroimage.*, vol. 58, pp. 330-338, 2011.

[20] G. Nolte et al., "Identifying true brain interaction from EEG data using the imaginary part of coherency," *Clin. Neurophysiol.*, vol. 115, no. 10, pp. 2292-2307, 2004.

[21] G. Gomez-Herrero, M. Atienza, K. Egiazarian, and J. L. Cantero, "Measuring directional coupling between EEG sources," *Neuroimage.*, vol. 43, no. 3, pp. 497-508, 2008.

[22] G. Nolte and K. R. Muller, "Localizing and estimating causal relations of interacting brain rhythms," *Frontiers in Human Neuroscience*, vol. 4, pp. 209, 2010.

[23] S. Haufe et al., "Modeling Sparse Connectivity Between Underlying Brain Sources for EEG/MEG," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 1954-1963, 2010.

[24] S. Haufe, V. V. Nikulin, K. R. Muller, and G. Nolte, "A critical assessment of connectivity measures for EEG data: A simulation study," *Neuroimage*, vol. 64 pp. 120-133, 2013.

[25] J. T. Lizier, *The local information dynamics of distributed computation in complex systems*, Berlin Heidelberg: Springer, 2013.

[26] L. Faes and A. Porta, "Conditional entropy-based evaluation of information dynamics in physiological systems," in *Directed Information Measures in* Neuroscience, R. Vicente, M. Wibral, and J. T. Lizier, Eds. Berlin: Springer-Verlag, 2014, pp. 61-86.

[27] M. Wibral et al., "Local active information storage as a tool to understand distributed neural information processing," *Front. Neuroinf.*, vol. 8, pp. 1, 2014.

[28] R. Vicente et al., "Transfer entropy-a model-free measure of effective connectivity for the neurosciences," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 45-67, 2011.

[29] L. Faes et al., "Estimating the decomposition of predictive information in multivariate systems," *Phys. Rev. E*, vol. 91, no. 3, pp. 032904, 2015.

[30] L. Faes, G. Nollo, and A. Porta, "Compensated transfer entropy as a tool for reliably estimating information transfer in physiological time series," *Entropy*, vol. 15, no. 1, pp. 198-219, 2013.

[31] V. Magagnin et al., "Non-stationarities significantly distort short-term spectral, symbolic and entropy heart rate variability indices," *Physiol Meas.*, vol. 32, no. 11, pp. 1775-1786, 2011.

[32] L. Faes, A. Porta, and G. Nollo, "Information Decomposition in Bivariate Systems: Theory and Application to Cardiorespiratory Dynamics," *Entropy*, vol. 17, no. 1, pp. 277-303, 2015.

[33] L. Faes, A. Porta, and G. Nollo, "Algorithms for the inference of causality in dynamic processes: application to cardiovascular and cerebrovascular variability," in *37th Int. Conf. of the IEEE Eng. Med. Biol. Soc. - EMBC 2015*, pp. 1789-1792, 2015.

[34] M. Small, *Applied nonlinear time series analysis: applications in physics, physiology and finance*, World Scientific, 2005.

[35] C. S. Huang et al., "Identifying changes in EEG information transfer during drowsy driving by transfer entropy," *Frontiers Hum. Neurosci.*, vol. 9 pp. 570, 2015.

[36] S. M. Pincus, "Approximated entropy (ApEn) as a complexity measure," *Chaos*, no. 5, pp. 110-117, 1995.

[37] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Am. J. Physiol. Heart Circ. Physiol.*, vol. 278, no. 6, pp. H2039-H2049, 2000.

[38] A. Porta et al., "Complexity and nonlinearity in short-term heart period variability: comparison of methods based on local nonlinear prediction," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 1, pp. 94-106, 2007.

[39] K. Kirschfeld, "The physical basis of alpha waves in the electroencephalogram and the origin of the "Berger effect"," *Biol. Cybern.*, vol. 92, no. 3, pp. 177-185, 2005.

[40] C. Babiloni et al., "Sources of cortical rhythms in adults during physiological aging: a multicentric EEG study," *Hum. Brain. Mapp.*, vol. 27, no. 2, pp. 162-172, 2006.

[41] L. Faes et al., "A framework for assessing frequency domain causality in physiological time series with instantaneous effects," *Phil. Trans. Roy. Soc. A*, vol. 371, no. 1997, pp. 20110618, 2013.

[42] P. L. Nunez, "Wavelike properties of the alpha rhythm," *IEEE Trans. Biomed. Eng.*, vol. 21 pp. 473-482, 1974.

[43] J. Ito, A. R. Nikolaev, and C. van Leeuwen, "Spatial and temporal structure of phase synchronization of spontaneous alpha EEG activity," *Biol. Cybern.*, vol. 92, no. 1, pp. 54-60, 2005.

[44] G. Nolte et al., and K. R. Muller, "Robustly estimating the flow direction of information in complex physical systems," *Phys. Rev. Lett.*, vol. 100, no. 23, pp. 234101, 2008.

[45] T. Inouye et al., "Potential Flow of Alpha-Activity in the Human Electroencephalogram," *Neurosci. Lett.*, vol. 187, no. 1, pp. 29-32, Feb.1995.

[46] D. Chicharro and S. Panzeri, "Algorithms of causal inference for the analysis of effective connectivity among brain regions," *Front. Neuroinf.*, vol. 8 pp. 64, 2014.

[47] A. Hyvarinen and S. M. Smith, "Pairwise likelihood ratios for estimation of non-Gaussian structural equation models," *J. Mach. Learn. Res.*, vol. 14 pp. 111–152, 2013.