Mutual Information in Frequency and Its Application to Measure Cross-Frequency Coupling in Epilepsy

Rakesh Malladi ⬤, Don H. Johnson ⬤, *Fellow, IEEE*, Giridhar P. Kalamangalam, Nitin Tandon, and Behnaam Aazhang, *Fellow, IEEE*

Abstract—We define a metric, mutual information in frequency (MI-in-frequency), to detect and quantify the statistical dependence between different frequency components in the data, referred to as cross-frequency coupling and apply it to electrophysiological recordings from the brain to infer cross-frequency coupling. The current metrics used to quantify the cross-frequency coupling in neuroscience cannot detect if two frequency components in non-Gaussian brain recordings are statistically independent or not. Our MI-in-frequency metric, based on Shannon's mutual information between the Cramér's representation of stochastic processes, overcomes this shortcoming and can detect statistical dependence in frequency between non-Gaussian signals. We then describe two data-driven estimators of MI-in-frequency: One based on kernel density estimation and the other based on the nearest neighbor algorithm and validate their performance on simulated data. We then use MI-in-frequency to estimate mutual information between two data streams that are dependent across time, without making any parametric model assumptions. Finally, we use the MI-in-frequency metric to investigate the cross-frequency coupling in seizure onset zone from electrocorticographic recordings during seizures. The inferred cross-frequency coupling characteristics are essential to optimize the spatial and spectral parameters of electrical stimulation based treatments of epilepsy.

Index Terms—Mutual information in frequency, dependent data, Cramér's spectral representation, cross-frequency coupling, epilepsy, seizure onset zone.

I. INTRODUCTION

E PILEPSY is a very common neurological disorder affecting nearly 1% of the world's population. Epilepsy is

R. Malladi is with the LinkedIn Corporation, Sunnyvale, CA 94085 USA (e-mail: malladirakesh@gmail.com).

D. H. Johnson and B. Aazhang are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA (e-mail: dhj@rice.edu; aaz@rice.edu).

G. P. Kalamangalam is with the Department of Neurology, University of Florida, Gainesville, FL 32611 USA (e-mail: Giridhar.Kalamangalam@neurology.ufl.edu).

N. Tandon is with the Department of Neurosurgery, University of Texas Health Center, Houston, TX 78712 USA (e-mail: Nitin.Tandon@uth.tmc.edu).

characterized by repeated, unprovoked seizures. Nearly a third of all epilepsy patients have medically refractory epilepsy (medication is not effective in these patients). For these patients, surgical resection of the seizure onset zone (SOZ) (the regions of the brain responsible for generating and sustaining seizure activity [3]) or electrical stimulation are possible treatment options. However, the efficacy of these treatments is variable and almost always never results in a cure [4], [5]. There is tremendous interest in leveraging the recent advances in electrical stimulation [6] and optogenetics [7] to develop spatiotemporally specific approaches to treat epilepsy. A crucial step in this endeavor is to develop an understanding of the coupling between neuronal oscillations in different frequency bands during seizures. This coupling or statistical dependence across frequency components between signals is referred to as cross-frequency coupling (CFC) [8], [9]. Our main objective is to learn the dynamics of cross-frequency coupling during seizures in epilepsy patients from the electrocorticographic (ECoG) data.

Elaborating the characteristics of epileptic seizures using cross-frequency coupling between ECoG data has been the focus of many papers. CFC has been used to predict the onset of seizure in [10] and detect epileptic seizures in [11]. CFC has also been used to localize the area for surgical resection in epilepsy patients [12]–[14]. Variations in CFC from preictal (before a seizure) to ictal (during a seizure) to postictal (after a seizures) in epilepsy patients have been analyzed in [15], [16]. In addition, the CFC in interictal stages is compared with that around seizures in [11], [17]–[19]. In this paper, we study CFC within and between various regions inside the seizure onset zone to determine the dominant frequencies involved in seizures and to learn the variations in coupling strength between various spatial regions inside SOZ. The results from this study are crucial to optimize the spectral and spatial parameters of next generation epilepsy treatments.

Cross-frequency coupling or dependence across frequencies in the data could be in a single recording or between recordings, not necessarily at the same frequency. Coherence can identify if two frequency components are statistically independent or not and quantify the dependence for linear, Gaussian processes [20]. There is no such equivalent metric for non-Gaussian signals. Since the time-series data recorded from the brain are neither linearly related nor Gaussian, neuroscientists typically use heuristic metrics that cannot identify if two frequency components are statistically independent or not and can only capture second-order dependencies. Some of the popular heuristics

estimate the phase-amplitude, amplitude-amplitude, phase-phase coupling between the low and high frequency components in the electrophysiological recordings from brain [8], [21]–[23]. In fact, a recent review article on CFC metrics suggests the use of cross-frequency "correlation" instead of "coupling" to describe these heuristic CFC metrics [22]. Furthermore, a list of confounds affecting the current CFC metrics is provided in [22]. A more comprehensive metric that detects statistical independence and thereby, capture both linear and nonlinear dependencies, would be invaluable in determining how neuronal oscillations at various frequencies are involved in the computation, communication, and learning in the brain. Here we propose a new methodology or metric to estimate the cross-frequency coupling (CFC) in neuroscience that overcomes the challenges of the existing approaches and as a proof-of-concept, we infer CFC characteristics of epileptic seizures using our metric.

Mutual information in frequency (MI-in-frequency), defined for linear Gaussian processes using coherence in [24], [25], can indeed be further developed into a general technique to estimate CFC. Inspired by prior work [26], we define MI-in-frequency between two frequencies in a signal (or two signals) as the Shannon's mutual information (MI) between the Cramér's spectral representations [27], [28] of the two signals at the corresponding frequencies. Cramér's spectral representation transforms a time-domain stochastic process into a stochastic process in the frequency domain, the samples of which can be estimated at each frequency from the time-domain data samples [29]. MI-in-frequency metric is equivalent to coherence measures for linear, Gaussian signals and can be thought of as "coherence" for non-Gaussian signals. The MI-in-frequency metric is one of the three mutual information based metrics used in [26] to analyze linear relationships between seismic data and [26] is not focussed on defining a single metric to capture the statistical dependence across frequency. We extend this approach to define a single metric, MI-in-frequency, to capture statistical dependencies across frequency for both linear and nonlinear data and use it measure CFC in the brain. We then describe two data-driven algorithms – one based on kernel density estimation (KDMIF) and the other based on nearest neighbor estimation (NNMIF) – to estimate MI-in-frequency without assuming any parametric model of the data. We considered these two approaches since they outperformed other approaches in estimating MI from i.i.d. data and there is no clear winner between them [30], [31]. We also demonstrate the superiority of MI-in-frequency over existing CFC metrics by comparing against modulation index [8], [21], a commonly used CFC metric, on simulated data.

In addition to estimating CFC between ECoG data, we use MI-in-frequency to develop a data-driven estimator for mutual information (MI). Note that MI estimation is a solved problem if the data samples are i.i.d. [32] or are sampled from linear, Gaussian processes [24], [25], [33], [34]. As mentioned earlier, real-world data is neither independent across time nor Gaussian and the underlying model is often unknown. Our data-driven MI estimation algorithm applies to dependent data, without making any parametric model assumptions. The key idea is to make the problem computationally tractable by focussing only on those frequencies in the two data streams that are statistically

dependent, which are identified by MI-in-frequency metric. Our MI estimator converges to the true value for Gaussian models and we validate its performance on nonlinear models.

Finally, we apply the MI-in-frequency estimators to infer the cross-frequency coupling in the seizure onset zone (SOZ), by analyzing electrocorticographic (ECoG) data from the SOZ of 9 patients with medial temporal lobe epilepsy in whom a total of 25 seizures were recorded. We investigate the dynamics of CFC in preictal, ictal and postictal periods within one SOZ electrode and between electrodes in different regions in the SOZ. We observe an increase in coupling in gamma and ripple high-frequency oscillations during seizures, with the largest increase within a SOZ electrode and a very small increase between electrodes in different regions inside SOZ. In addition, low-frequency coupling and linear interactions between SOZ electrodes also increase during the postictal state.

II. CRAMÉR'S SPECTRAL REPRESENTATION OF STOCHASTIC PROCESSES

Consider a stochastic processes $X(t)$, $t \in \mathbb{R}$. Let $S_X(\nu)$ for $\nu \in \mathbb{R}$ be the spectral distribution function of X and $s_X(\nu)$, its power spectral density, if it exists. Two basic spectral representations are associated with the stochastic process $X(t)$ - power spectral distribution and Cramér's representation [27], [28]. The Cramér's representation of $X(t)$ and its key properties are stated in the following theorem.

Theorem 1: (page 380 in [28]) Let $X(t)$ be a second order stationary, mean-square continuous and zero mean stochastic process. Then there exists a complex-valued, finite-variance, orthogonal increment process $\widetilde{X}(\nu)$ in the frequency domain $\nu \in \mathbb{R}$, such that

$$X(t) = \int_{-\infty}^{\infty} e^{j2\pi\nu t} d\widetilde{X}(\nu),$$

with $\mathbb{E}\left[d\widetilde{X}(\nu)\right] = 0$, and $\mathbb{E}\left[|d\widetilde{X}(\nu)|^2\right] = dS_X(\nu)$.

The process $\widetilde{X}(\nu) = \widetilde{X}_R(\nu) + j\widetilde{X}_I(\nu)$ satisfying the above theorem is the spectral process or the Cramér's representation of $X(t)$. $d\widetilde{X}(\nu)$ is the complex random variable representing the amplitude of oscillation in the interval from ν to $\nu + d\nu$ in $X(t)$. The integral in Theorem 1 is a Fourier-Stieltjes integral. Intuitively, Theorem 1 decomposes $X(t)$ into an orthogonal increment complex process in the frequency domain. Furthermore, if the $X(t)$ is real-valued, then $\widetilde{X}(-\nu) = \widetilde{X}^\star(\nu)$, $\mathbb{E}\left[d\widetilde{X}_R(\nu)d\widetilde{X}_I(\nu)\right] = 0$, and

$$\mathbb{E}\left[\left(d\widetilde{X}_R(\nu)\right)^2\right] = \mathbb{E}\left[\left(d\widetilde{X}_I(\nu)\right)^2\right] = \frac{1}{2}dS_X(\nu). \quad (1)$$

We have the following theorem for the special case of a real-valued Gaussian process $X(t)$.

Theorem 2: (page 385 in [28]) Let $X(t)$ be a real-valued stationary, mean-square continuous Gaussian process with zero mean and power spectral distribution function $S_X(\nu)$, $\nu \in \mathbb{R}$. Then the real and imaginary parts of its spectral process $\widetilde{X}_R(\nu)$ and $\widetilde{X}_I(\nu)$ are zero mean, mutually independent, identically distributed Gaussian processes satisfying (1).

Example: Consider the zero mean stationary Gaussian process $X(t) = A\cos(2\pi\nu_0 t + \Theta)$, where A is Rayleigh random variable with parameter σ_A that is independent of Θ, which is uniform in $[0, 2\pi]$. The increments of the spectral process of $X(t)$ are all zero, except at $\nu = \pm\nu_0$, where the increment is $\frac{A}{2}\exp(\pm j\Theta)$ [28]. This implies that the sample path of the real part of spectral process $\widetilde{X}(\nu)$ has two jumps of same magnitude and direction at frequencies $\pm\nu_0$, while that of the imaginary part has two jumps of same magnitude, but opposite directions at $\pm\nu_0$. The magnitude of the jump at ν_0 in the real and imaginary parts is $\frac{A}{2}\cos\Theta$ and $\frac{A}{2}\sin\Theta$ respectively, both of which are Gaussian random variables with mean zero and variance $\frac{1}{2}\sigma_A^2$. This spectral process is intuitive because we know $X(t)$ has all its energy only at frequencies $\pm\nu_0$ and the variance of the increments of the spectral process $d\widetilde{X}(\nu)$ is equal to the differential power spectral distribution of $X(t)$ which is nonzero only at $\pm\nu_0$. We therefore expect all sample paths of the random process $\widetilde{X}(\nu)$ with non-zero probability to be constant, except for jumps at $\pm\nu_0$.

Note that if the process is wide sense-stationary and Gaussian, then power spectral distribution would have all the information about the process and its relationship with Cramér's representation is given by Theorem 2. Otherwise, power spectral distribution only captures the second-order dependencies in the process. Since ECoG signals are not Gaussian, we use Cramér's representation to transform a time-domain stochastic process into a stochastic process in the frequency domain.

III. MUTUAL INFORMATION IN FREQUENCY

We first define MI between frequencies within a process and between two processes in continuous time. We then extend this definition to discrete-time stochastic processes. Consider $d\widetilde{X}(\nu_i)$ and $d\widetilde{Y}(\nu_j)$, the increments of spectral processes or the Cramér's representation of $X(t)$ and $Y(t)$ at frequencies ν_i and ν_j respectively. Let the joint probability density of the four dimensional random vector of the real and imaginary parts of $d\widetilde{X}(\nu_i)$ and $d\widetilde{Y}(\nu_j)$ be denoted by $P(d\widetilde{X}_R(\nu_i), d\widetilde{X}_I(\nu_i), d\widetilde{Y}_R(\nu_j), d\widetilde{Y}_I(\nu_j))$. The corresponding two-dimensional marginal densities are $P(d\widetilde{X}_R(\nu_i), d\widetilde{X}_I(\nu_i))$, $P(d\widetilde{Y}_R(\nu_j), d\widetilde{Y}_I(\nu_j))$. The MI-in-frequency between $X(t)$ at ν_i and $Y(t)$ at ν_j is defined as

$$\text{MI}_{XY}(\nu_i, \nu_j)$$
$$= \text{I}(\{d\widetilde{X}_R(\nu_i), d\widetilde{X}_I(\nu_i)\}; \{d\widetilde{Y}_R(\nu_j), d\widetilde{Y}_I(\nu_j)\}),$$
$$= \mathbb{E}\left\{\log\frac{P(d\widetilde{X}_R(\nu_i), d\widetilde{X}_I(\nu_i), d\widetilde{Y}_R(\nu_j), d\widetilde{Y}_I(\nu_j))}{P(d\widetilde{X}_R(\nu_i), d\widetilde{X}_I(\nu_i))P(d\widetilde{Y}_R(\nu_j), d\widetilde{Y}_I(\nu_j))}\right\},$$
$$\tag{2}$$

where $\text{I}(\{\cdot, \cdot\}; \{\cdot, \cdot\})$ is the standard mutual information between two pairs of two dimensional real-valued random vectors [34]. The MI between two different frequencies ν_i, ν_j in the same process $Y(t)$ is similarly defined as

$$\text{MI}_{YY}(\nu_i, \nu_j) = \text{I}(\{d\widetilde{Y}_R(\nu_i), d\widetilde{Y}_I(\nu_i)\}; \{d\widetilde{Y}_R(\nu_j), d\widetilde{Y}_I(\nu_j)\}). \tag{3}$$

The MI between the components of Y at frequencies $\nu_i = \nu_j = \nu$, $\text{MI}_{YY}(\nu, \nu)$, is ∞, a consequence of the fact that $[d\widetilde{Y}_R(\nu), d\widetilde{Y}_I(\nu)]$ is a continuous-valued random vector whose conditional differential entropy is not lower bounded. MI-in-frequency defined in (2), (3) is a non-negative number. If MI-in-frequency between two frequencies is zero, then they are independent and if not, MI-in-frequency is a measure of the statistical dependence between the two frequency components. MI-in-frequency between two processes is not symmetric in general, i.e., $\text{MI}_{XY}(\nu_i, \nu_j) \neq \text{MI}_{XY}(\nu_j, \nu_i)$. However, it is symmetric within a process, i.e., $\text{MI}_{YY}(\nu_i, \nu_j) = \text{MI}_{YY}(\nu_j, \nu_i)$.

Example: Continuing with our example in Section II, let $X(t) = A\cos(2\pi\nu_0 t + \Theta)$ and $Y(t) = X(t)^2$. Then $d\widetilde{Y}(\nu)$ is zero except at $\nu = 0$, where the spectral increment is $\frac{A^2}{2}$, and at $\nu = \pm 2\nu_0$, where the increment is $\frac{A^2}{4}\exp(\pm j2\Theta)$. As a result, the frequency components at $\pm\nu_0$ in X and at frequencies $\{0, \pm 2\nu_0\}$ in Y are statistically dependent and hence the MI-in-frequency obtained from (2) at these frequency pairs will be positive. In addition, the frequency components in Y at $\nu \in \{0, \pm 2\nu_0\}$ are dependent and hence the MI-in-frequency within Y at these frequencies will also be positive.

A. Gaussian Inputs to LTI Filters

Let's now consider the special case where $X(t)$, a Gaussian process with power spectral density $s_X(\nu)$ serves as the input to a linear, time-invariant (LTI) filter with transfer function $H_1(\nu)$ and $Y(t)$ is output observed in additive colored noise (white noise $W(t)$ passed through a LTI filter with transfer function $H_2(\nu)$). The processes $X(t)$ and $Y(t)$ are related by

$$y(t) = h_1(t) * x(t) + h_2(t) * w(t), \tag{4}$$

where $*$ denotes convolution operation, $x(t), y(t)$ and $w(t)$ are sample paths of $X(t)$, $Y(t)$ and $W(t)$ respectively. W is a Gaussian process with power spectral density $s_W(\nu)$ and independent of X. $h_1(t)$ and $h_2(t)$ are continuous-time impulse responses of LTI filters, whose transfer functions are $H_1(\nu)$ and $H_2(\nu)$ respectively. Let $d\widetilde{X}(\nu)$, $d\widetilde{W}(\nu)$ and $d\widetilde{Y}(\nu)$ be the spectral process increments of the Gaussian processes X, W and Y. We have from Theorem 2,

$$[d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu)] \sim \mathcal{N}\left(\mathbf{0}, \frac{1}{2}s_X(\nu)\mathbf{I}\right),$$

$$[d\widetilde{W}_R(\nu), d\widetilde{W}_I(\nu)] \sim \mathcal{N}\left(\mathbf{0}, \frac{1}{2}s_W(\nu)\mathbf{I}\right), \tag{5}$$

where $\mathcal{N}(\mu, \Sigma)$ represents Gaussian distribution with mean μ and covariance Σ, $\mathbf{0}$ is the two element zero vector and $\mathbf{I}$ is the 2×2 identity matrix. In addition, we can show for the model in (4) that

$$d\widetilde{Y}(\nu) = H_1(\nu)d\widetilde{X}(\nu) + H_2(\nu)d\widetilde{W}(\nu). \tag{6}$$

The proof of (6) is in the appendix. The MI-in-frequency defined in (2) is further simplified for the model in (4) using (5), (6) and stated in the following theorem.

Theorem 3: For the model given in (4), the MI between $X(t)$ at frequency ν_i and $Y(t)$ at frequency ν_j is zero, when $\nu_i \neq \nu_j$ and the MI between $X(t)$ and $Y(t)$ at frequency $\nu_i = \nu_j = \nu \neq 0$ is

$$\text{MI}_{XY}(\nu, \nu) = 2 \times \text{I}\left(\left\{d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu)\right\}; d\widetilde{Y}_R(\nu)\right)$$
$$= \log\left(1 + \frac{|H_1(\nu)|^2 s_X(\nu)}{|H_2(\nu)|^2 s_W(\nu)}\right). \quad (7)$$

The proof of the above theorem is in the appendix. Note that at $\nu = 0$, the MI-in-frequency between X and Y is equal to $\text{I}\left(\left\{d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu)\right\}; d\widetilde{Y}_R(\nu)\right)$, which is just half of the right hand side of (7). We intuitively expect different frequency components in the Gaussian input and its output from a linear system to be independent and Theorem 3 confirms that the proposed definition of MI-in-frequency agrees with this intuition. In addition, the MI between X and Y is ∞ when $|H_2(\nu)| = 0$, since the components of X and Y at such ν are linearly related. The MI between two different frequencies in $Y(t)$, generated from (4), is zero due to the linearity of the filters and Gaussian inputs. Furthermore, we can also show for the Gaussian processes X and Y related by (4) that MI-in-frequency is related to coherence $C_{XY}(\nu) \in [0, 1]$, by $\text{MI}_{XY}(\nu, \nu) = -\log(1 - C_{XY}(\nu))$. The proof is in the appendix. This result implies MI-in-frequency between Gaussian processes related by (4) can be estimated with the coherence. In addition, Theorem 3 also shows that MI-in-frequency between Gaussian processes related by (4) can be estimated by estimating the mutual information between $\left[d\widetilde{X}_R(\nu_i), d\widetilde{X}_I(\nu_i)\right]$ and $d\widetilde{Y}_R(\nu_j)$, a three dimensional estimate as opposed to a four dimensional estimate in general.

B. Discrete-Time Stochastic Processes

We now extend the definition of MI-in-frequency between continuous-time stochastic processes in (2), (3) to discrete-time stochastic processes. In practice, we only have access to data samples from a real-valued, discrete-time stochastic process, sampled at a given Nyquist sampling frequency F_s. Sampled signals have periodic spectra, with a period equalling F_s. In addition, components in the process with frequencies in the range $[F_s/2, F_s]$ correspond to negative frequencies [35]. Therefore, the actual frequency content in the signal is confined to $[0, F_s/2]$. We use normalized frequency $\lambda = \frac{\nu}{F_s} \in [0, 0.5]$ to describe the frequency axis in case of discrete-time stochastic processes, instead of ν which was used for continuous-time stochastic processes. The MI-in-frequency between discrete-time processes is therefore obtained by replacing ν_i, ν_j by the normalized frequencies $\lambda_1, \lambda_2 \in [0, 0.5]$ in (2), (3). Multivariate autoregressive models, commonly used to model electro-physiological signals recorded from brain [20], [36], are a special case of the discrete-time equivalent of (4). The analytic expression for MI at frequency λ for such discrete-time Gaussian processes is therefore similarly obtained by replacing the frequencies ν by λ in (7), which is also equal to $-\log(1 - C_{XY}(\lambda))$. This shows that for the special case of discrete-time Gaussian processes, MI-in-

frequency metric is equivalent to coherence and the definitions in [24], [25].

IV. DATA-DRIVEN ESTIMATION OF MI-IN-FREQUENCY

We describe two data-driven estimators–a kernel density based (KDMIF) and a nearest neighbor based (NNMIF) estimator to estimate MI-in-frequency, $\widehat{\text{MI}}_{XY}(\lambda_i, \lambda_j)$, between λ_i component of X and λ_j component of Y. The input to both these algorithms are the N samples of X and Y. The first step in both KDMIF and NNMIF estimators involves estimating the samples of spectral process increments $d\widetilde{X}(\lambda_i)$ and $d\widetilde{Y}(\lambda_j)$, of X at λ_i and of Y at λ_j respectively. In the second step, the KDMIF estimator uses the kernel density based MI estimator [32], [37], whereas NNMIF estimator uses the k-nearest neighbor based MI estimator [32], [38] to estimate MI from the samples of spectral process increments, $d\widetilde{X}(\lambda_i)$ and $d\widetilde{Y}(\lambda_j)$.

A. Kernel Density Based MI-in-Frequency (KDMIF) Estimator

1) Estimation of Samples of Spectral Process Increments: The first step of the algorithm is estimating the samples of spectral process increments of X and Y from N dependent data samples. We assume there is a finite memory in both these processes and choose a value for a parameter N_f, which is much larger than the length of dependence or memory in the data and determines the frequency resolution of our MI-in-frequency estimates. We assume data in different windows are independent of each other. Ideally, consecutive windows should be separated to ensure no dependence across windows and avoid the dependence across the window boundaries, but our simulation results demonstrate that not separating the windows doesn't affect performance significantly. N samples of X are split into N_s non-overlapping windows with $N_f = \frac{N}{N_s}$ data points in each window. Let us denote the samples in lth window of X and Y respectively by two N_f element one-dimensional vectors, $\mathbf{x}^l$ and $\mathbf{y}^l$, for $l = 1, 2, \ldots, N_s$.

Let us now focus on estimating samples of the random variable $d\widetilde{X}(\lambda_i)$. Let $\mathcal{F}\{\mathbf{x}^l\}(\alpha)$ denote the discrete-time Fourier transform (DTFT) of $\mathbf{x}^l$ at normalized frequency α. For $\lambda_i = \frac{i}{N_f} \in [0, 1]$ and $i \in [0, N_f - 1]$, let us define $d\widetilde{x}^l(\lambda_i)$ and integrated Fourier spectrum, $\widetilde{x}^l(\lambda_i)$, by

$$d\widetilde{x}^l(\lambda_i) = \mathcal{F}\{\mathbf{x}^l\}(\lambda_i) \text{ and } \widetilde{x}^l(\lambda_i) = \sum_{m=0}^{i} \mathcal{F}\{\mathbf{x}^l\}(\lambda_m). \quad (8)$$

It is stated in [29] that the random variable for which $\widetilde{x}^l(\lambda_i)$ is just one realization, tends to the spectral process of X at λ_i in mean of order γ, for any $\gamma > 0$, as the number of samples goes to infinity and assuming the underlying distribution is stationary and satisfies a mixing assumption. Also, $d\widetilde{x}^l(\lambda_i)$, which is the increment in $\widetilde{x}^l(\lambda_i)$ between λ_i and $\lambda_i + d\lambda$, is just the DTFT of the samples in window l. Calculating the DTFT with the FFT for each of the N_s windows separately yields an $N_f \times N_s$ matrix, whose i^{th} row, $\mathbf{d}\widetilde{\mathbf{x}}(\lambda_i) = \left[d\widetilde{x}^1(\lambda_i), d\widetilde{x}^2(\lambda_i), \ldots, d\widetilde{x}^{N_s}(\lambda_i)\right]$ is the complex-valued vector containing N_s samples of $d\widetilde{X}(\lambda_i)$,

the spectral process increments of X at $\lambda_i = \frac{i}{N_f}$. The lth element of $\mathbf{d\tilde{x}}(\lambda_i)$, $d\tilde{x}^l(\lambda_i) = d\tilde{x}_R^l(\lambda_i) + i d\tilde{x}_I^l(\lambda_i)$, is a particular realization of $d\tilde{X}(\lambda_i)$. A similar procedure is used to obtain the N_s samples of the spectral process increments of Y at $\lambda_j = \frac{j}{N_f}, j \in [0, N_f - 1]$ and the resulting samples are denoted by $\mathbf{d\tilde{y}}(\lambda_j) = [d\tilde{y}^1(\lambda_j), d\tilde{y}^2(\lambda_j), \ldots, d\tilde{y}^{N_s}(\lambda_j)]$.

2) Estimating MI-in-Frequency: The MI-in-frequency estimate is now obtained from the N_s samples, $(d\tilde{x}_R^l(\lambda_i), d\tilde{x}_I^l(\lambda_i))$ and $(d\tilde{y}_R^l(\lambda_j), d\tilde{y}_I^l(\lambda_j))$, for $l = 1, 2, \ldots, N_s$, using a kernel density based plug-in nonparametric estimator [32]. The N_s data samples are split into N_{tr} training and N_{ts} test samples. The training data is used to estimate the four-dimensional joint probability density $P(d\tilde{X}_R(\lambda_i), d\tilde{X}_I(\lambda_i), d\tilde{Y}_R(\lambda_j), d\tilde{Y}_I(\lambda_j))$. The density is estimated using a kernel density estimator with Gaussian kernels, the optimal bandwidth matrix selected using smoothed cross-validation criterion [37] and implemented using "ks" package in R [39]. The joint density is marginalized to estimate the two-dimensional densities, $P(d\tilde{X}_R(\lambda_i), d\tilde{X}_I(\lambda_i))$ and $P(d\tilde{Y}_R(\lambda_j), d\tilde{Y}_I(\lambda_j))$, by recognizing that the bandwidth matrix for the two-dimensional marginal is the appropriate 2×2 sub-matrix from the 4×4 bandwidth matrix of the joint density. The estimates of the joint and the marginal densities at the N_{ts} test samples are plugged into the following equation (9) to estimate MI-in-frequency.

$$\widehat{MI}_{XY}(\lambda_i, \lambda_j)$$
$$= \frac{1}{N_{ts}} \sum_l \log \frac{\widehat{P}(d\tilde{x}_R^l(\lambda_i), d\tilde{x}_I^l(\lambda_i), d\tilde{y}_R^l(\lambda_j), d\tilde{y}_I^l(\lambda_j))}{\widehat{P}(d\tilde{x}_R^l(\lambda_i), d\tilde{x}_I^l(\lambda_i)) \widehat{P}(d\tilde{y}_R^l(\lambda_j), d\tilde{y}_I^l(\lambda_j))}. \tag{9}$$

B. Nearest Neighbor Based MI-in-Frequency (NNMIF) Estimator

1) Estimation of Samples of Spectral Process Increments: The first step in the nearest neighbor based MI-in-frequency estimator is exactly same as that of KDMIF estimator. Following the steps described in Section IV-A1, we estimate $d\tilde{x}^l(\lambda_i)$ and $d\tilde{y}^l(\lambda_j)$, for $l = 1, 2, \ldots, N_s$, the N_s samples of the spectral process increments of X at λ_i and Y at λ_j respectively.

2) Estimating MI-in-Frequency: $MI_{XY}(\lambda_i, \lambda_j)$ is now estimated from $d\tilde{x}^l(\lambda_i) \in \mathbb{R}^2$ and $d\tilde{y}^l(\lambda_j) \in \mathbb{R}^2$, for $l = 1, 2, \ldots, N_s$ using nearest neighbor based MI estimator [38]. We apply the first version of the algorithm in [38] to two-dimensional random variables $d\tilde{X}(\lambda_i)$ and $d\tilde{Y}(\lambda_j)$ to compute $\widehat{MI}_{XY}(\lambda_i, \lambda_j)$. Consider the joint four dimensional space $(d\tilde{X}(\lambda_i), d\tilde{Y}(\lambda_j)) \in \mathbb{R}^4$. The distance between two data points with indices $l_1, l_2 \in [1, N_s]$ is calculated using the infinity norm, according to $\max\{\|d\tilde{x}^{l_1}(\lambda_i) - d\tilde{x}^{l_2}(\lambda_i)\|, \|d\tilde{y}^{l_1}(\lambda_j) - d\tilde{y}^{l_2}(\lambda_j)\|\}$. Let ϵ_l denote the distance between the data sample $(d\tilde{x}^l(\lambda_i), d\tilde{y}^l(\lambda_j))$ and its Kth nearest neighbor, for $l = 1, 2, \ldots, N_s$. We used $K = 3$ in this paper [30]. Let n_x^l and n_y^l denote the number of samples of $d\tilde{X}(\lambda_i)$ and $d\tilde{Y}(\lambda_j)$ within an infinity norm ball of radius less than ϵ_l centered at $d\tilde{x}^l(\lambda_i)$ and $d\tilde{y}^i(\lambda_j)$ respectively. From [38], the MI-in-frequency between X and Y at normalized

---

Algorithm 1: Mutual Information Estimator.

Data: $(x[n], y[n])$, for $x[n], y[n] \in \mathbb{R}, n \in [0, N-1]$.
Result: $\hat{I}(X; Y)$
Algorithm:
  A) Estimate $\widehat{MI}_{XY}(\lambda_i, \lambda_j)$ at all possible pairs (λ_i, λ_j), using either the KDMIF or the NN-MIF estimator. Identify the sets Λ_x, Λ_y, such that for each $\lambda_{i_p} \in \Lambda_x$ there exists a $\lambda_{j_q} \in \Lambda_y$ such that $\widehat{MI}_{XY}(\lambda_{i_p}, \lambda_{j_q})$ is statistically significant and vice-versa. Let P, Q respectively denote the cardinality of Λ_x, Λ_y.
  B) Let $d\tilde{X}(\Lambda_x) = [d\tilde{X}(\lambda_{i_1}), \ldots, d\tilde{X}(\lambda_{i_P})] \in \mathbb{R}^{2P}$, $d\tilde{Y}(\Lambda_y) = [d\tilde{Y}(\lambda_{j_1}), \ldots, d\tilde{Y}(\lambda_{j_Q})] \in \mathbb{R}^{2Q}$. The mutual information between X and Y is given by
$$\hat{I}(X; Y) = \frac{1}{\max(P, Q)} \hat{I}(d\tilde{X}(\Lambda_x); d\tilde{Y}(\Lambda_y)),$$
  where the right hand side is estimated from N_s i.i.d. samples using any nonparametric MI estimator [32].

---

frequencies λ_i and λ_j is given by

$$\widehat{MI}_{XY}(\lambda_i, \lambda_j) = \psi(K) + \psi(N_s)$$
$$- \frac{1}{N_s} \sum_{l=1}^{N_s} (\psi(n_x^l + 1) + \psi(n_y^l + 1)), \tag{10}$$

where $\psi(\cdot)$ is the Digamma function.

C. Significance Testing

The statistical significance of the MI-in-frequency estimates obtained from both KDMIF and NNMIF estimators is now tested using the following procedure. We permute the samples in the vector $\mathbf{d\tilde{x}}(\lambda_i)$ randomly and estimate the MI-in-frequency between the permuted vector and the N_s samples of $d\tilde{Y}(\lambda_j)$. Instead of adding random phase or permuting the phase time series, which are typically used to test the statistical significance of phase-amplitude coupling metrics [40], we permute the samples of spectral process increments since our metric can detect coupling across phase and amplitude jointly. This process is repeated N_p times to obtain N_p permuted MI-in-frequency estimates, under the null hypothesis of independence. The permuted MI estimates will be almost zero since the permutations make the spectral processes almost independent. If the actual MI estimate, $\widehat{MI}_{XY}(\lambda_i, \lambda_j)$, is judged larger than all the permuted N_p estimates, then there is a statistically significant dependence between the processes at these frequencies.

V. MI BETWEEN DATA WITH TEMPORAL DEPENDENCIES

We now use MI-in-frequency to estimate mutual information between dependent data. The data-driven MI estimator, summarized in Algorithm 1, takes in N samples of X and Y as input and outputs the mutual information between X and Y,

$\hat{I}(X;Y)$, by estimating $\widehat{MI}_{XY}(\lambda_i, \lambda_j)$, where $\lambda_i = \frac{i}{N_f}, \lambda_j = \frac{j}{N_f}, \forall (i,j)$ such that $i,j \in [0, N_f - 1]$.

A. Identifying Coupled Frequencies

The first step in our MI estimator involves estimating the MI-in-frequency, $\widehat{MI}_{XY}(\lambda_i, \lambda_j)$, between $\lambda_i = \frac{i}{N_f}$ frequency component in X and $\lambda_j = \frac{j}{N_f}$ component in Y, for all (i,j) such that $i,j \in [0, N_f - 1]$ using either the KDMIF (Section IV-A) or the NNMIF (Section IV-B) algorithms. Statistical significance of the resulting estimates is assessed using the procedure described in Section IV-C. The resultant MI-in-frequency estimates across all frequency pairs can be graphically visualized by plotting the statistically significant MI-in-frequency estimates on a two-dimensional image grid, whose rows and columns correspond to frequencies of X and Y respectively. Let Λ_x and Λ_y respectively denote the set of frequency components of X and Y, such that for each $\lambda_{i_p} \in \Lambda_x$, there exists at least one $\lambda_{j_q} \in \Lambda_y$ for which $\widehat{MI}_{XY}(\lambda_{i_p}, \lambda_{j_q})$ is statistically significant and vice-versa.

B. Estimating Mutual Information

The final step in our algorithm estimates MI between the spectral process increments of X and Y at frequencies in Λ_x and Λ_y respectively. With P, Q denoting the cardinality of Λ_x, Λ_y respectively, let $d\widetilde{X}(\Lambda_x)$ and $d\widetilde{Y}(\Lambda_y)$ denote the $2P$ and $2Q$-dimensional random vector comprising the spectral process increments of X, Y at all frequencies in Λ_x and Λ_y respectively. We already computed N_s i.i.d. samples of these two random vectors to estimate MI-in-frequency estimates in the previous step of this algorithm. The desired MI estimate is computed from the mutual information between $d\widetilde{X}(\Lambda_x)$ and $d\widetilde{Y}(\Lambda_y)$, which is estimated using the k-nearest neighbor based estimator developed in [38], according to

$$\hat{I}(X;Y) = \frac{1}{\max(P,Q)} \hat{I}\left(d\widetilde{X}(\Lambda_x); d\widetilde{Y}(\Lambda_y)\right). \quad (11)$$

The MI estimator in (11) can be further simplified for discrete-time Gaussian processes. Without loss of generality, consider two Gaussian processes X and Y, related by

$$y[n] = h_1[n] * x[n] + h_2[n] * w[n], \quad (12)$$

where $h_1[n], h_2[n]$ are linear time-invariant (LTI) filters and W is white Gaussian noise independent of X. For the model in (12), which is the discrete-time equivalent of (4), the data-driven estimation in (11) can be further simplified to

$$\hat{I}(X;Y) = \frac{1}{N_f} \sum_{i=0}^{N_f/2} \widehat{MI}_{XY}(\lambda_i; \lambda_i), \text{ where } \lambda_i = \frac{i}{N_f}. \quad (13)$$

This result is obtained because linear models do not introduce cross-frequency dependencies and because negative frequencies do not carry any extra information. Furthermore, the relationship between the MI and the MI-in-frequency for two processes related by (12) is stated in the following theorem.

Theorem 4: Consider two discrete-time Gaussian stochastic processes X and Y related by (12). The mutual information between these processes, a scalar, is given by

$$I(X;Y) = \int_0^{0.5} MI_{XY}(\lambda, \lambda) d\lambda. \quad (14)$$

The proof of the above theorem is in the appendix. This theorem means that MI between two Gaussian processes over the entire time can be obtained by integrating the contribution from each frequency component. It is easy to see that the right hand side of (13) is just the Riemann sum of the integral on the right hand side of (14), which converges to the true value as N_f tends to infinity. This implies our MI estimator converges to the true value for discrete-time Gaussian processes.

Note that the MI estimation algorithm does not make any parametric assumptions on the underlying model between X and Y. The computation of MI via (11) can be greatly simplified by clustering the frequencies in Λ_x and Λ_y into groups such that there are no significant dependencies across groups and using the chain rule of mutual information. In addition, if we observe after the first step that significant MI-in-frequency estimates occur only at $(\lambda_i, \lambda_i), \forall i \in [0, N_f - 1]$, then the MI can be estimated using (13).

Finally, as we mentioned earlier, MI estimation between Gaussian processes is a solved problem in the sense that we can analytically compute it if the covariance of the Gaussian processes is known [3] and there are several estimators whose performance is thoroughly analyzed [4]. MI in frequency for Gaussian processes is analyzed by Brillinger [24]. In this paper, we extended Brillinger's work to define MI-in-frequency for any process. In the following section, we use simulated data to validate that the extensions we proposed to any process in this paper are still in agreement with the prior work on Gaussian processes and also work for non-Gaussian processes.

VI. PERFORMANCE EVALUATION ON SIMULATED DATA

The performance of the data-driven MI-in-frequency and mutual information estimators described in Sections IV and V respectively is validated on simulated data. The statistical significance of the estimates was assessed using the procedure described in Section IV-C. In addition, we compare the performance of the MI-in-frequency estimators against modulation index [8], [9], [22], a commonly used phase-amplitude coupling metric in neuroscience.

A. Comparing the KDMIF and NNMIF Estimators

Consider two stochastic processes X and Y, where X is a white Gaussian process with standard deviation σ_x and Y is obtained by

$$y[n] = h[n] * x[n] + w[n], \quad (15)$$

where W is a white Gaussian process with standard deviation σ_w that is independent of X and $h[n]$ is a linear time-invariant filter. We compared the performance of the kernel density based and nearest neighbor based estimators by benchmarking the estimates against the true value of MI-in-frequency and the mutual
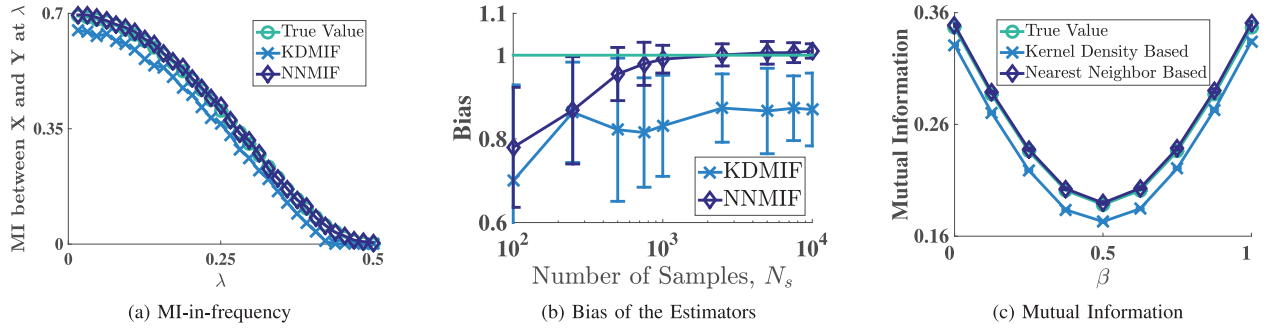
Fig. 1.   Comparing the performance of the kernel density based and nearest neighbor based estimators, KDMIF and NNMIF respectively, on simulated generated from (15) using a two-tap lowpass filter. (a) MI-in-frequency estimates obtained from KDMIF and NNMIF estimators along with the true value of MI-in-frequency are plotted against the normalized frequency λ for $\beta = 0.5$. (b) Plots the bias (mean of the ratio of the estimate and the true value in the filter passband) against the number of data samples used for estimation for $\beta = 0.5$. (c) Plots the MI estimate between X and Y obtained from kernel density and nearest neighbor algorithms along with the true value of MI for $\beta \in [0, 1]$.
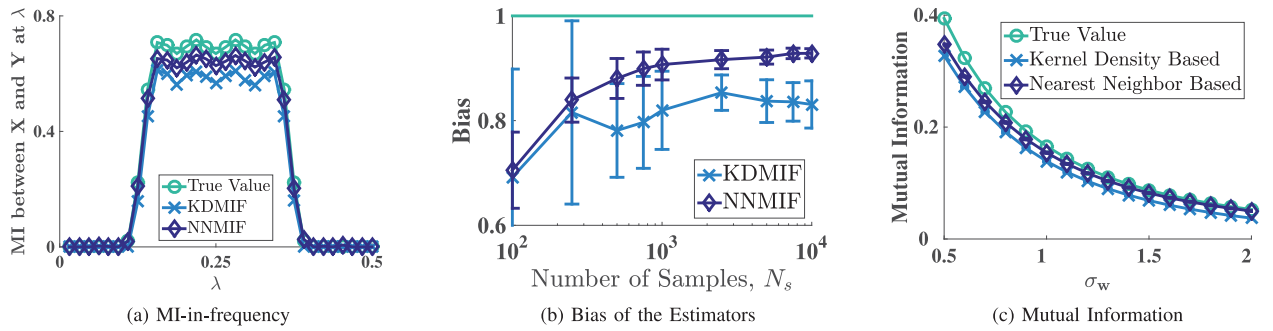

Fig. 2.   Comparing the performance of the kernel density based and nearest neighbor based estimators, KDMIF and NNMIF respectively, on simulated generated from (15) using a 33-tap bandpass filter with passband in $[0.15, 0.35]$ normalized frequency. (a) MI-in-frequency estimates obtained from KDMIF and NNMIF estimators along with the true value of MI-in-frequency are plotted against the normalized frequency λ for $\sigma_w = 1$. (b) Plots the bias (mean of the ratio of the estimate and the true value in the filter passband) against the number of data samples used for estimation for $\sigma_w = 1$. (c) Plots the MI estimate between X and Y from kernel density and nearest neighbor algorithms along with the true value of MI for different values of $\sigma_w \in [0.5, 2]$.

information between X and Y for the model in (15). We used two different filers: a two-tap low pass filter, $h[n] = [\beta, 1 - \beta]$, for $\beta \in [0, 1]$ and a 33-tap bandpass filter with passband in $[0.15, 0.35]$ normalized frequency range. We observed that modulation index, a popular CFC metric, was unable to correctly detect and quantify the strength of cross-frequency coupling for both these models.

1) Lowpass Filter: The samples of X and Y are generated from (15) with $\sigma_x = \sigma_w = 1$ and a lowpass filter with unit-impulse response $[\beta, 1 - \beta]$, for various values of $\beta \in [0, 1]$. The true value of MI-in-frequency at normalized frequency $\lambda \in [0, 0.5]$ is obtained substituting the parameters of this model in (7) and is plotted in Fig. 1(a) for $\beta = 0.5$. In addition, the MI-in-frequency estimated by the KDMIF and NNMIF algorithms from $N = 64 \times 10^4$ data samples, with $N_f = 64, N_s = 10^4$ is also plotted in Fig. 1(a). It is seen that the estimates from both algorithms follow the true value closely, without the knowledge of the underlying model. In addition, we evaluate the bias and the rate of convergence of both these algorithms as a function of N_s, with $N_f = 64$ in Fig. 1(b). The bias is defined as the average value of the ratio of MI-in-frequency estimate and its true value in the passband of the lowpass filter. We observe that the NNMIF algorithm converges faster and has lower bias than the KDMIF algorithm. We now use both these algorithms to estimate the mutual information between X and Y for $\beta \in [0, 1]$.

The analytical expression for the true value of MI[1] for this model is derived in [41]. It is evident from Fig. 1(c) that the MI estimates obtained from the nearest neighbor based estimator is closer to the true value than those from the kernel density based estimator.

2) Bandpass Filter: The samples of X are generated from a standard white Gaussian random process with $\sigma_x = 1$ and those of Y are generated from (15) using a 33-tap finite-impulse-response bandpass filter with passband in $[0.15, 0.35]$ normalized frequency range for different values of noise standard deviation, $\sigma_w \in [0.5, 2]$. We used the kernel density and the nearest neighbor based algorithms to estimate the MI-in-frequency and the mutual information between X and Y. The true value of MI-in-frequency is obtained from (7) and of mutual information is numerically calculated using power spectral density (chapter 10 in [34]). It is clear from Fig. 2(b) that the nearest neighbor based algorithm converges to the true value faster than the kernel density based algorithm. The nearest neighbor based algorithm also provides more accurate estimates of both MI-in-frequency and mutual information between X and Y, as evident from Fig. 2(a), (c) respectively. In addition, nearest neighbor based MI-in-frequency algorithm runs faster than kernel density based

---

[1]Note that for this particular model, mutual information is equal to the directed information from X to Y and the analytical expression is given in (18) in [41].
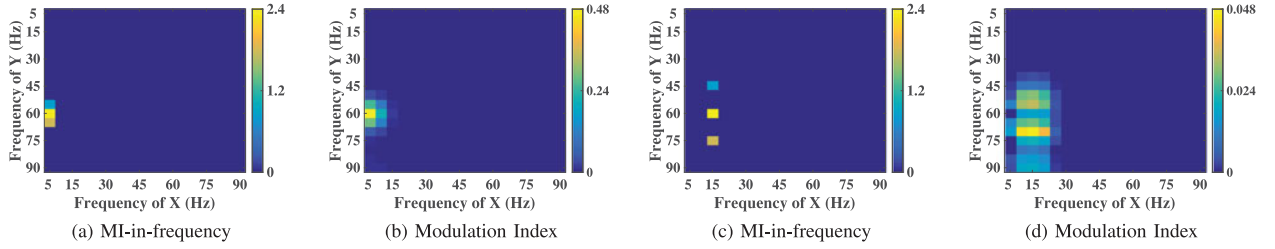
Fig. 3. Comparing the performance of MI-in-frequency against modulation index in detecting cross-frequency coupling in data generated from (16). (a) and (b) MI-in-frequency estimates obtained from nearest neighbor algorithm and modulation index are plotted respectively, when $f_l = 5$ Hz and $f_h = 60$ Hz in (16). Panels (c) and (d) respectively plot the MI-in-frequency estimates and modulation index estimates, when $f_l = 15$ Hz and $f_h = 60$ Hz in (16).

algorithm. We, therefore, conclude that the nearest neighbor based MI-in-frequency algorithm outperforms kernel density based algorithms and only depict the results obtained from nearest neighbor based algorithm in the remainder of the paper.

B. Comparison With Modulation Index

We now compare the effectiveness of MI-in-frequency against modulation index in detecting cross-frequency coupling, using the simulated model commonly used to validate CFC metrics [21], [22], [42]. Modulation index quantifies the relationship between the phase and amplitude envelopes extracted by the Hilbert transform [8]. Consider two random cosine waves, $s_l[n]$ and $s_h[n]$, at frequencies f_l and f_h respectively. Let f_s denote the sampling frequency. The samples of time-series X and Y are generated from the following model:

$$s_l[n] = A \cos\left(2\pi \frac{f_l}{f_s} n + \theta\right), \; s_h[n] = A \cos\left(2\pi \frac{f_h}{f_s} n + \theta\right)$$

$$x[n] = s_l[n] + w_1[n], y[n] = (1 + s_l[n]) s_h[n] + w_2[n],$$
(16)

where A is a Rayleigh random variable with parameter 1 and θ is a uniformly distributed random variable between 0 and 2π that is independent of A. $w_1[n], w_2[n]$ are samples of i.i.d white Gaussian noise process with standard deviation 1. We generated samples from this model with $f_l = 5$ Hz, $f_h = 60$ Hz and $f_s = 200$ Hz. MI-in-frequency between X and Y is estimated using the nearest neighbor based algorithm from $N = 40 \times 10^4$ samples with $N_s = 10^4$ and plotted in Fig. 3(a). Modulation index between X and Y estimated by using the Matlab toolbox [21], with the amplitude envelope estimated by the Hilbert transform and is plotted in Fig. 3(b). It is clear that both MI-in-frequency and modulation index successfully detect the cross-frequency coupling between 5 Hz component of X and $\{55, 60, 65\}$ Hz components of Y for these parameter values. We then generated X and Y from (16) with $f_l = 15$ Hz and all other parameter values unchanged. Fig. 3(c) plots the MI-in-frequency estimates obtained via NNMIF algorithm and as expected, we detect the CFC between 15 Hz component of X and $\{45, 60, 75\}$ Hz components of Y. However, modulation index, depicted in Fig. 3(d), was not able to correctly detect the CFC between X and Y for these parameter values. In addition, the strength of the modulation index decreased from around 0.5 when $f_l = 5$ Hz in Fig. 3(b) to 0.05 when $f_l = 15$ Hz in Fig. 3(d). This is because metrics like modulation index can only

detect the CFC correctly with good frequency resolution only when one of the frequencies involved is very small compared to the other frequency. Otherwise, the bandwidth of the filter used to extract the phase and the amplitude envelope should be larger, which will reduce the frequency resolution in the estimated CFC (note the smearing in Fig. 3(d), when compared to Fig. 3(b)) [22], [42]. In addition, we tested modulation index on data generated from (15) and (17) and found that modulation index is unable to detect the cross-frequency coupling for these relationships. This is not surprising since the modulation index like metrics are tuned to detect CFC when the underlying coupling is of the form in (16), whereas the MI-in-frequency defined in this paper overcomes this shortcoming, as evident from its performance on various simulated models.

C. Nonlinear Models

We now consider square nonlinearity, where the random processes X and Y are related by

$$y[n] = x[n]^2 + w[n],$$
(17)

where $w[n]$ is white Gaussian noise with standard deviation σ_w. Modulation index was not able to detect and quantify the cross-frequency coupling for this model. We estimated the MI-in-frequency between frequency components within Y, $\widehat{\text{MI}}_{YY}(\lambda_i, \lambda_j)$, between the frequency components of X and Y, $\widehat{\text{MI}}_{XY}(\lambda_i, \lambda_j)$, and the mutual information between X and Y, $\hat{I}(X;Y)$, from $N = 32 \times 10^4$ samples of X and Y with $N_s = 10^4$, for different values of noise standard deviation, $\sigma_w \in [0, 10]$. Computing the true value of MI-in-frequency and mutual information is nontrivial because of the nonlinearity. The performance of the algorithms is assessed by checking if they detect the cross-frequency coupling at expected frequency pairs and by checking if the mutual information estimates decrease with increasing noise power as expected. We considered two different models for the stochastic process X, such that its samples are dependent across time.

1) Random Cosine With Squared Nonlinearity: The samples of X are generated from a random cosine wave,

$$x[n] = A \cos(2\pi \lambda_0 n + \theta),$$
(18)

where A is a Rayleigh random variable with parameter 1, θ is a uniform random variable between 0 and 2π that is independent of A and $\lambda_0 = \frac{4}{32}$. It is easy to see that frequency components of X are statistically independent and this is confirmed by the
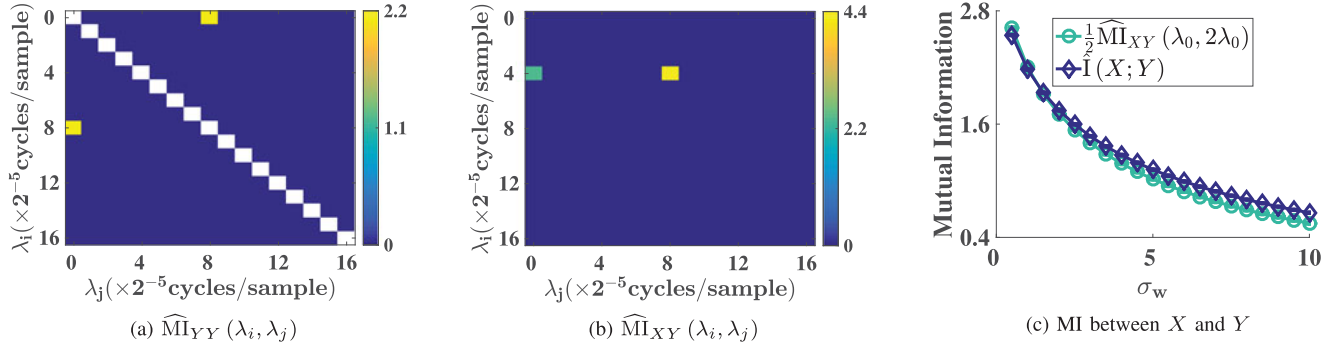
(a) $\widehat{\mathrm{MI}}_{YY}(\lambda_i, \lambda_j)$

(b) $\widehat{\mathrm{MI}}_{XY}(\lambda_i, \lambda_j)$

(c) MI between X and Y

Fig. 4. (a) MI-in-frequency estimates from the nearest neighbor based algorithm between the frequency components within the random processes Y, obtained from the single cosine data-generation model, (18) with $\sigma_w = 1$. Note that the MI-in-frequency estimates along the principal diagonal are not plotted, since they are equal to ∞. (b) MI-in-frequency estimates between random processes X and Y related by the single cosine data-generation model with $\sigma_w = 1$. It is clear that MI-in-frequency estimator correctly identifies the pairwise frequency dependencies. (c) MI-in-frequency between X at λ_0 and Y at $2\lambda_0$, $\widehat{\mathrm{MI}}_{XY}(\lambda_0, 2\lambda_0)$, obtained from (10) along with the MI estimate between X and Y, $\hat{\mathrm{I}}(X;Y)$, obtained from Algorithm 1 for various values of the noise standard deviation, σ_w.

(a) $\widehat{\mathrm{MI}}_{YY}(\lambda_i, \lambda_j)$

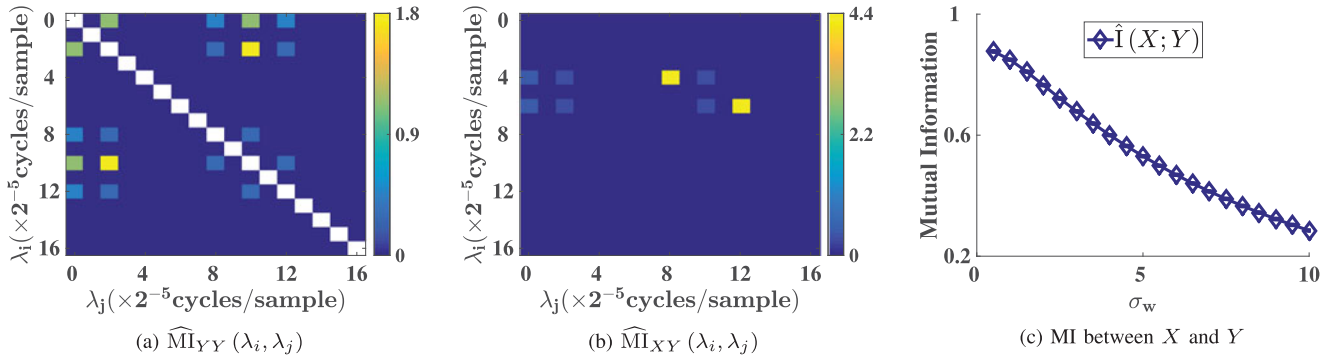(b) $\widehat{\mathrm{MI}}_{XY}(\lambda_i, \lambda_j)$

(c) MI between X and Y

Fig. 5. (a) MI-in-frequency estimates from the nearest neighbor based algorithm between the frequency components within the random processes Y, obtained from the two cosine data-generation model, (19). The MI-in-frequency estimates are not plotted along the diagonal, since they are equal to ∞. (b) MI-in-frequency estimates between random processes X and Y related by the two cosine data-generation model. It is clear that MI-in-frequency estimator correctly identifies the pairwise frequency dependencies between X and Y. (c) $\hat{\mathrm{I}}(X;Y)$, the MI estimate between X and Y obtained from Algorithm 1 for various values of the noise standard deviation, σ_w.

NNMIF estimator. However, because of the square nonlinearity in (17), the DC component of Y and the $2\lambda_0$ component of Y will be statistically dependent and this is confirmed by Fig. 4(a), which plots the MI-in-frequency between components of Y generated with $\sigma_w = 1$ using the NNMIF algorithm. The common information between these two processes will be present between λ_0 component of X and the $\{0, 2\lambda_0\}$ components of Y. This cross-frequency dependence is confirmed by Fig. 4(b), which plots the estimates of MI-in-frequency between X and Y obtained by the NNMIF algorithm from (10): we observe that significant dependencies occur only at $(\lambda_0, 0)$ and $(\lambda_0, 2\lambda_0)$ frequency pairs. As a result, $P = 1, Q = 2$. The MI estimate from Algorithm 1, $\hat{\mathrm{I}}(X;Y) = \frac{1}{2}\hat{\mathrm{I}}\left(d\widetilde{X}(\lambda_0); \{d\widetilde{Y}(0), d\widetilde{Y}(2\lambda_0)\}\right)$ is plotted in Fig. 4(c). The MI estimate decreases with increasing σ_w as expected. In addition, we note for this model that the DC component of Y does not contain any extra information about X, given the $2\lambda_0$ component of Y. Therefore, we expect $\frac{1}{2}\hat{\mathrm{I}}(d\widetilde{X}(\lambda_0); \{d\widetilde{Y}(0), d\widetilde{Y}(2\lambda_0)\}) = \frac{1}{2}\widehat{\mathrm{MI}}_{XY}(\lambda_0; 2\lambda_0)$, a result verified in Fig. 4(c), since the two curves are very close.

2) Two Random Cosines With Squared Nonlinearity: The samples of random process X are generated according to

$$x[n] = A_1 \cos(2\pi\lambda_1 n + \theta_1) + A_2 \cos(2\pi\lambda_2 n + \theta_2), \quad (19)$$

where A_1, A_2 are independent Rayleigh random variables with parameter 1, θ_1, θ_2 are independent uniformly distributed random variables between 0 and 2π that are independent of A_1, A_2, and $\lambda_1 = \frac{4}{32}, \lambda_2 = \frac{6}{32}$. As before, the frequency components of X are statistically independent. However, after some basic algebra, it is easy to see that the all possible pairs of frequency components of Y in $\{0, \lambda_2 - \lambda_1, 2\lambda_1, \lambda_2 + \lambda_1, 2\lambda_2\}$ are statistically dependent, except for $(2\lambda_1, 2\lambda_2)$ frequency pair, and we expect to see statistically significant MI-in-frequency estimates between these frequency components. This is confirmed by Fig. 5(a), which plots the MI-in-frequency estimates within Y, generated with $\sigma_w = 1$ and obtained by the NNMIF algorithm. The pairwise frequency dependencies between X and Y occur at $(\lambda_1, 0)$, $(\lambda_1, \lambda_2 - \lambda_1)$, $(\lambda_1, 2\lambda_1)$, $(\lambda_1, \lambda_2 + \lambda_1)$, $(\lambda_2, 0)$, $(\lambda_2, \lambda_2 - \lambda_1)$, $(\lambda_2, \lambda_2 + \lambda_1)$ and $(\lambda_2, 2\lambda_2)$. Fig. 5(b) plots the estimates of pairwise MI-in-frequency between X

TABLE I
CLINICAL DETAILS OF THE PATIENTS ANALYZED

Patient	Number of Seizures Analyzed	Age/Sex	Seizure Onset Zone	Outcome of Surgery
P1	3	22/M	RAH 1-2, RPH 4, RAMY 2-3	Class IA
P2	3	61/M	LAH 2-4, LPH 2	Class IIIA
P3	2	29/M	PD 4, 5 LF 28, LP 4	Class IA
P4	3	21/F	MST 1, TP 1, HD 1	Class IA
P5	3	24/M	LPH 5, 6, LPSM 8, LMH 5, RMH 4, 5, RPSM 7	Class IB
P6	3	35/M	AH 3-5, PH 4 AMY 2, 3	Class IA
P7	3	26/M	AH 1, 2, 5, PH 5 TOP 3, 4	Class IIB
P8	3	41/M	LAH 5, LAMY 3	N/A
P9	2	18/F	RAH 3-5, LPH 6, RPH 5-7	Class IB

The full forms of the electrodes in seizure onset zone column in Table I: RAH - right anterior hippocampus, RPH - right posterior hippocampus, RAMY - right amygdala, PD - posterior hippocampal depth, MST - mid-subtemporal lobe, TP - temporopolar, HD - hippocampal depth and AST - anterior sub-temporal lobe, LMH - left mid hippocampus, AH - anterior hippocampus, PH - posterior hippocampus, AMY - amygdala, TOP - temporo-occipito-parietal. The outcomes are in Engel epilepsy surgery outcome scale [43], [44]: "Class IA - completely seizure free since surgery, class IB - non disabling simple partial seizures only since surgery, class IIB - rare disabling seizures since surgery ("almost seizure-free"), class IIIA - worthwhile seizure reduction, class IV - no worthwhile improvement".

and Y generated with $\sigma_w = 1$ and obtained by the data-driven NNMIF algorithm using (10). The algorithm correctly identifies all the dependent frequency pairs and $P = 2, Q = 5$. We then apply the algorithm described in Section V and plot the estimates the MI for different values of noise standard deviation σ_w in Fig. 5(c). Again, the MI decreases with increasing noise power, as expected. These different models validate the superiority of MI-in-frequency over other existing metrics to detect cross-frequency coupling and also demonstrate the performance and accuracy of the data-driven MI-in-frequency and MI estimators.

VII. CFC IN SEIZURE ONSET ZONE

Epilepsy is a common neurological disorder characterized by repeated, unprovoked seizures. The seizure onset zone (SOZ) comprises regions of the brain that are responsible for generating and sustaining seizures [3]. Surgical resection of the seizure onset zone is the prescribed treatment for a large portion of medically refractory epilepsy patients with focal epilepsy. However, surgical resection risks damage to critical functional zones that are frequently adjacent or even overlapping with the seizure focus, depending on location of the focus [45]. An ideal solution might be a closed-loop neuromodulation strategy that stimulates the epileptic [41], [46] and other networks [47] at the optimal frequency with spatial and temporal specificity [6], [7]. In this paper, we focus on learning more about the characteristic frequencies and the spatial specificity of epileptic networks. Specifically, we investigate cross-frequency coupling between various regions in the seizure onset zone during the evolution of

seizures and identify the frequencies with strong coupling. We estimate the cross-frequency coupling (CFC) from ECoG data recorded from the SOZ electrodes using our nearest neighbor based MI-in-frequency estimator. We infer the characteristics of CFC within and between various regions inside the seizure onset zone.

We analyzed ECoG data, sampled at $F_s = 1$ kHz, from a total of 25 seizures recorded from nine patients with medial temporal lobe epilepsy. Clinical details of the patients, along with the seizure onset zone identified from ECoG data [41], are summarized in Table I. The seizure start and end time were marked by the neurologist. We analyzed ECoG recordings from SOZ electrodes during preictal (window spanning up to 3 minutes immediately before the seizure starts), ictal (during seizures) and postictal (window spanning up to 3 minutes immediately after the seizure ends) periods. We only focussed on the oscillations in alpha (7.5–12.5 Hz), beta (12.5–30 Hz), gamma (30–80 Hz) and ripples (80–200 Hz), excluding 60 Hz line noise and its harmonics. The CFC oscillations are analyzed at spectral resolution of 10 Hz by choosing $N_f = 100$, and the exact frequencies considered are $\{10, 20, \dots, 200\}$ Hz, excluding $\{60, 120, 180\}$ Hz. The resulting 17×17 CFC matrix from each ECoG electrode and between all pairs of ECoG electrodes in the SOZ is estimated using nearest neighbor based estimator (Section IV-B) during preictal, ictal and postictal periods during all the twenty five seizures.

We then grouped the ECoG electrodes into distinct anatomical regions based on their label and analyzed the average CFC within a SOZ electrode, between two electrodes in the same anatomical region and between electrodes in different anatomi-
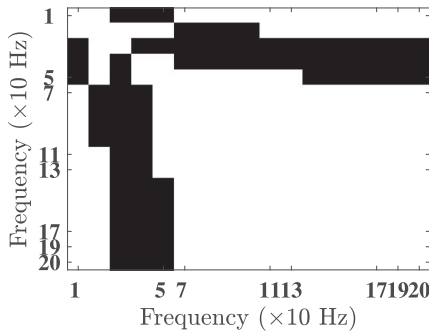
Fig. 6.   Binary mask plotting the frequency pairs with statistically significant differences across all the hypotheses tested after applying false discovery rate correction. White and black colored regions represent frequency pairs with and without statistically significant variation respectively.

cal regions. For instance, consider patient P1. ECoG electrodes in the SOZ of patient P1 are grouped into three different anatomical regions–RAH, RPH, and RAMY (Table I). We estimated 5 CFC matrices, one per SOZ electrode, to infer the average CFC within an electrode in SOZ in this patient. We estimated 20 CFC matrices between all pairs of electrodes in the SOZ. Of these, 4 CFC matrices (2 to learn the CFC between the 2 SOZ electrodes in RAH and 2 to learn the CFC between the 2 SOZ electrodes in RAMY regions) are grouped to learn the average CFC between electrodes in the same anatomical region in SOZ. The remaining 16 CFC matrices are grouped to learn the CFC between different regions inside the SOZ. The estimated CFC matrices are grouped into these three spatial categories for all the nine patients during preictal, ictal and postictal periods. We only presented the results for CFC within a SOZ electrode and between electrodes in different SOZ regions during preictal, ictal and postictal periods.

We used the permutation procedure outlined in Section IV-C to estimate the CFC under the null hypothesis and assess the significance of the estimated CFC values across the six conditions considered (CFC during preictal, ictal, postictal periods within a SOZ electrode and between electrodes in different SOZ regions) using Wilcoxon signed-rank test [48]. We also used the Wilcoxon signed-rank test to identify the frequency pairs with significant variation in CFC between preictal and ictal periods and between ictal and postictal periods, both within a SOZ electrode and between electrodes in different SOZ regions (four hypotheses in total). In addition, we used the Mann-Whitney U-test [48] to identify frequency pairs with significant changes in CFC within a SOZ electrode and between electrodes in different SOZ regions across preictal, ictal and postictal periods (three hypotheses in total). We estimated 3621 p-values in total ($13 \times 17 \times 16 + 5 \times 17$) and applied false discovery rate correction at a significance level of 0.01 to account for multiple comparisons [49]. The frequency pairs with significant statistical variation across all the hypotheses considered are depicted using a binary mask in Fig. 6, in which black and white colored regions respectively represent frequency pairs without statistically significant variation and with statistically significant variation. Lack of statistical significance at the black regions in

Fig. 6 could be because of insufficient data or could be due to a neuronal transition mechanism as the brain moves from preictal to ictal to postictal state. It is important to note that if we tested only a subset of the thirteen hypotheses, then some of the frequency pairs in black colored regions in Fig. 6 could become statistically significant.

The median CFC within an electrode in SOZ during preictal, ictal and postictal periods grouped across all twenty five seizures in nine patients analyzed is plotted in Fig. 7. In Fig. 7(a), median CFC in the preictal period is plotted, while the difference between median CFC in the ictal and preictal period, and between postictal and ictal period is plotted in Fig. 7(b) and (c) respectively. We need to multiply the binary mask in Fig. 6 with the plots in Fig. 7 to obtain frequency pairs with significant statistical variation. The (i, j)th element in the matrix in Fig. 7(a) is the median MI-in-frequency between the $10i$ and $10j$ Hz frequency components during preictal period across all SOZ electrodes in the twenty five seizures analyzed. The principal diagonal in the three CFC matrices is not plotted since MI-in-frequency between same frequencies in a signal is infinity. It is clear from this figure that ripple frequencies are heavily synchronized during preictal stage within an electrode in SOZ. The synchronization between all frequency pairs, particularly in gamma and ripples, seemed to increase during the seizure when compared to just before the seizure. And finally, the synchronization between high-frequency bands decreased, and low frequencies become more synchronized amongst themselves and with high-frequencies in the postictal period compared to the ictal period within an electrode in SOZ.

The median CFC between electrodes in different SOZ regions grouped across all twenty five seizures in nine patients analyzed is plotted in Fig. 8. We need to multiply the binary mask in Fig. 6 with the plots in Fig. 8 to obtain frequency pairs with significant statistical variation. The median CFC during the preictal period is plotted in Fig. 8(a). It is clear from the principal diagonal that neighboring regions in SOZ have weak linear interactions (possibly due to their spatial proximity) just before a seizure starts. From Figs. 7(a) and 8(a), it is clear that the CFC strength is much lower between regions when compared to within an electrode. From Fig. 8(b), we observe a small increase in CFC between regions as the brain transitions to seizure state. However, the increase is much smaller between regions when compared to the increase observed in Fig. 7(b), which suggests that different SOZ regions potentially drive the rest of the brain into a seizure state independently, which implies any nonsurgical treatment should target these different regions simultaneously to disrupt the epileptic network. As the brain transitions to postictal state, we observe a sharp increase in linear coupling between electrodes in different SOZ regions, which suggests that postictal periods, unlike ictal periods, are characterized by an increase in linear interactions. These results highlight the role of gamma and ripple high-frequency oscillations (HFOs) during seizures and the dynamic reorganization of synchronization between neuronal oscillations inside the seizure onset zone during the course of a seizure. These results also suggest that multiple regions inside the seizure onset zone might have to be
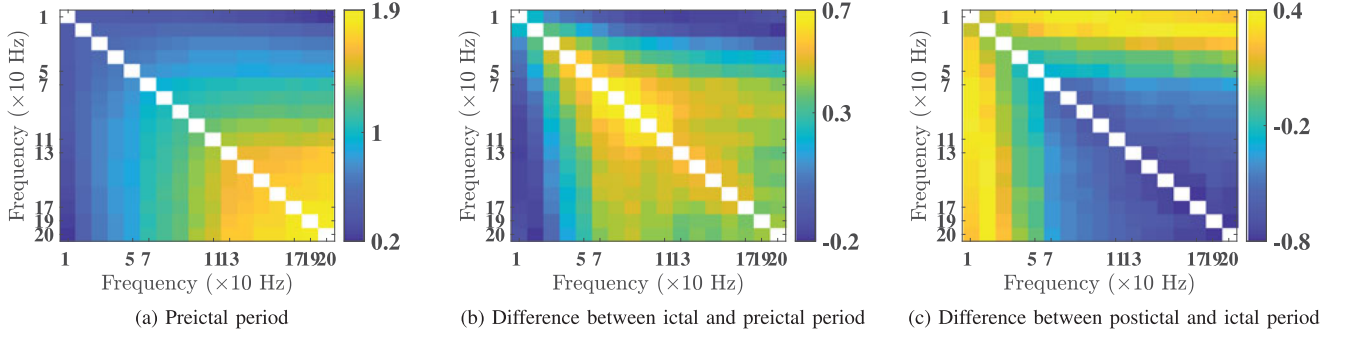
(a) Preictal period   (b) Difference between ictal and preictal period   (c) Difference between postictal and ictal period

Fig. 7. Cross-frequency coupling within an electrode inside the seizure onset zone. (a) MI-in-frequency estimates over the frequencies $\{10, 20, \ldots, 200\}$ Hz excluding $\{60, 120, 180\}$ Hz are obtained from each electrode in SOZ during preictal period and the median of the resulting CFC estimates from all the SOZ electrodes in the twenty five seizures from the nine temporal lobe epilepsy patients analyzed is plotted. (b) MI-in-frequency estimates are obtained from each electrode in SOZ in the ictal period and the difference between the median CFC estimate in ictal and preictal period is plotted. Similarly, (c) plots the difference in the median CFC between postictal and ictal periods.

(a) Preictal period   (b) Difference between ictal and preictal period   (c) Difference between postictal and ictal period
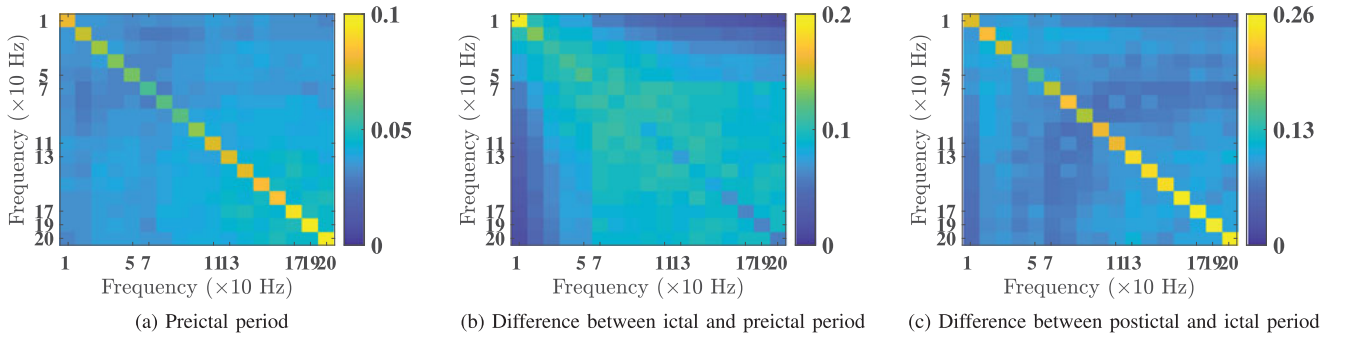
Fig. 8. Cross-frequency coupling between electrodes in different regions inside the seizure onset zone. (a) MI-in-frequency estimates over the frequencies $\{10, 20, \ldots, 200\}$ Hz excluding $\{60, 120, 180\}$ Hz are obtained between electrodes in different SOZ regions during the preictal period and the median of the resulting CFC estimates from the twenty five seizures in the nine temporal lobe epilepsy patients analyzed is plotted. (b) MI-in-frequency estimates are obtained between electrodes in different SOZ regions from the ictal period and the difference between the median CFC estimate from the ictal and preictal period is plotted. Similarly, (c) plots the difference in the median CFC between postictal and ictal periods.

targeted simultaneously using neuromodulation techniques to control seizure activity.

VIII. DISCUSSION AND CONCLUSION

Detecting and quantifying relationships between multiple data streams recorded from a physical system is of interest in many science and engineering disciplines. However, since the underlying model is often unknown and nonlinear, detecting and quantifying the relationships in data is very challenging in most real-world applications. Brownian distance covariance [50], maximal information coefficient [51] are some of the recent works that attempt to overcome this challenge in the most general case. Furthermore, in neuroscience, we are also interested in decomposing the relationships in frequency domain and estimating cross-frequency coupling (CFC) from electrophysiological recordings. Motivated to understand nonlinear frequency coupling in electrophysiological recordings from the brain and inspired by [26], we defined MI-in-frequency between stochastic processes that are not necessarily Gaussian and estimated it using data-driven estimators. We found that the nearest neighbor based MI-in-frequency estimator outperforms the kernel-based MI-in-frequency estimator. MI-in-frequency can be thought of as "coherence" for non-Gaussian signals. At a first glance, CFC

could be estimated by first filtering the data into appropriate frequency bands and then applying the techniques in [21], [50], [51]. However, [22] summarizes all the caveats and confounds in estimating CFC using this approach. In contrast, the MI-in-frequency metric estimates CFC without explicitly band-pass filtering the data into appropriate frequency bands.

We then compared the performance of MI-in-frequency against modulation index [8], [21], a popular CFC metric used to measure phase-amplitude coupling that involves bandpass filtering, on simulated data and observed that MI-in-frequency outperforms the existing metrics used to estimate CFC. The main advantages of the MI-in-frequency approach over existing methods to estimate CFC are that it detects statistical independence, detects dependencies across phase and amplitude jointly, applies to linear and nonlinear dependencies, and is not dependent on parameters like the filter bandwidth. Our approach will need more data when compared with coherence since MI-in-frequency detects both linear and nonlinear dependencies in frequency. From the simulation results on linear models, we need about 10^3 samples to be within 10% of the true value. For the ECoG data sampled at 1 kHz and a desired spectral resolution of 10 Hz, this implies the total number of data samples is of the order of 100 seconds or a couple of minutes, which is roughly the size of preictal, ictal and postictal windows used in

Section VII. In summary, we developed a metric to detect statistical independence in frequency which outperforms existing CFC metrics and for the first time, utilized frequency domain to estimate mutual information over time between dependent data.

The MI-in-frequency metric can be further extended along several directions and some of them are outlined here. We can move to wavelet based analysis to improve the fixed time-frequency resolution of our Fourier-based approach in future work. The assumption of data stationarity in observation window (also assumed by most CFC metrics) can be potentially relaxed by utilizing time-frequency distributions and developing heuristics to measure the dependencies across frequency. However, the inherent trade-off involved is that we are not guaranteed to detect statistical independence. It is also possible to define and estimate conditional MI-in-frequency to eliminate indirect coupling estimated between two signals because of a third signal which is coupled to both.

We then apply the MI-in-frequency estimators to infer the coupling between neuronal oscillations before, during and after seizures in the seizure onset zone. Spatially, we used the electrode labels to identify the different regions in the SOZ. This is just one possible way to analyze the spatial variation in CFC. Some of the other possible options include using the distance between electrodes or using the underlying neuronal cell types to split the electrodes into different regions in SOZ. Our MI-in-frequency metric provides a framework that can be utilized to learn the CFC characteristics for any desired spatial grouping. In addition, the frequency resolution of our estimated CFC was constant and wavelet transform, instead of Fourier transform, can be utilized to provide greater resolution at lower frequencies.

We observed that the high-frequency synchronization within an ECoG electrode in SOZ increases during seizures and decreases immediately after the seizure, which is accompanied by an increase in low-frequency coupling. However, the coupling between different anatomical regions in SOZ does not increase noticeably during seizures and is also followed by a large increase in linear interactions immediately after a seizure. These observations suggest that seizure activity is characterized by nonlinear interactions and is potentially due to the independent efforts by various regions within SOZ, which implies that all these regions are potential spatial targets for electrical stimulation. Furthermore, we did a preliminary investigation to learn if there are the differences in CFC between interictal periods and seizure periods. Fig. 9 plots the CFC within an ECoG electrode and between ECoG electrodes in different regions in SOZ during interictal period in two patients (P1 and P2). Comparing Fig. 9 with Figs. 7(a) and 8(a), it looks like the CFC within a SOZ electrodes at higher frequencies slightly increases, while CFC between electrodes in different regions across the diagonal (or equivalently, linear interactions) slightly decreases as the brain transitions from interictal to preictal periods. We plan to extend this analysis to a larger patient cohort. Building a real-time seizure prediction system utilizing the variations in CFC between interictal and seizure periods is the focus of our current [52] and future work. In addition, the CFC characteristics were patient-specific and we presented the median CFC across
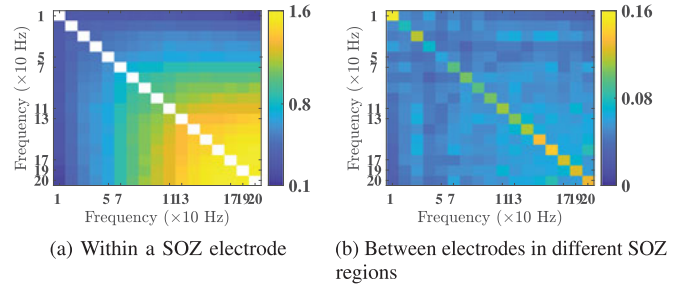

(a) Within a SOZ electrode   (b) Between electrodes in different SOZ regions

Fig. 9. Cross-frequency coupling during interictal periods. (a) MI-in-frequency estimates over the frequencies $\{10, 20, \ldots, 200\}$ Hz excluding $\{60, 120, 180\}$ Hz are obtained from each electrode in SOZ during interictal period and the median of the resulting CFC estimates from all the SOZ electrodes in patients P1 and P2 is plotted. (b) MI-in-frequency estimates over the frequencies $\{10, 20, \ldots, 200\}$ Hz excluding $\{60, 120, 180\}$ Hz are obtained between electrodes in different SOZ regions during the interictal period and the median of the resulting CFC estimates in patients P1 and P2 is plotted.

all the patients considered. Going forward, the MI-in-frequency metric should be applied to infer the CFC between channels in SOZ and outside SOZ to learn how SOZ drives the rest of the brain into a seizure state in each epilepsy patient. The results from such an analysis will improve our understanding of the CFC mechanisms underlying seizure activity and will serve as the first step towards the development of a patient-specific, closed-loop, non-surgical treatment for epilepsy.

APPENDIX

a) Proof of (6)

We have from (4),

$$y(t) = \int_{-\infty}^{\infty} h_1(t-\tau)x(\tau)d\tau + \int_{-\infty}^{\infty} h_2(t-\tau)w(\tau)d\tau \qquad (20)$$

$$\Rightarrow \int_{-\infty}^{\infty} e^{j2\pi\nu t} d\widetilde{Y}(\nu) = \int_{-\infty}^{\infty} h_1(t-\tau) \int_{-\infty}^{\infty} e^{j2\pi\nu\tau} d\widetilde{X}(\nu) d\tau$$

$$+ \int_{-\infty}^{\infty} h_2(t-\tau) \int_{-\infty}^{\infty} e^{j2\pi\nu\tau} d\widetilde{W}(\nu) d\tau \text{ (from Theorem 1)}$$

$$= \int_{-\infty}^{\infty} e^{j2\pi\nu t} \int_{-\infty}^{\infty} h_1(t-\tau) e^{-j2\pi\nu(t-\tau)} d\tau d\widetilde{X}(\nu)$$

$$+ \int_{-\infty}^{\infty} e^{j2\pi\nu t} \int_{-\infty}^{\infty} h_2(t-\tau) e^{-j2\pi\nu(t-\tau)} d\tau d\widetilde{W}(\nu) \qquad (21)$$

$$= \int_{-\infty}^{\infty} e^{j2\pi\nu t} \left\{ H_1(\nu) d\widetilde{X}(\nu) + H_2(\nu) d\widetilde{W}(\nu) \right\}.$$

$$\Rightarrow d\widetilde{Y}(\nu) = H_1(\nu) d\widetilde{X}(\nu) + H_2(\nu) d\widetilde{W}(\nu). \qquad (22)$$

b) Proof of Theorem 3

We will first prove that $\mathrm{MI}_{XY}(\nu_1, \nu_2)$ is zero, when X and Y are related by (4) and $\nu_1 \neq \nu_2$. Since the processes $X(t)$ and $W(t)$ are independent, their spectral processes are also independent. In addition, we also know from Theorem 2 that the spectral increments of the Gaussian process $X(t)$ are independent. It is clear from (6) that given $H_1(\nu)$ and

$H_2(\nu)$, $\left[d\widetilde{Y}_R(\nu_2), d\widetilde{Y}_I(\nu_2)\right]$ is completely determined by the two-dimensional random vectors $\left[d\widetilde{X}_R(\nu_2), d\widetilde{X}_I(\nu_2)\right]$ and $\left[d\widetilde{W}_R(\nu_2), d\widetilde{W}_I(\nu_2)\right]$, both of which are independent of the two-dimensional random vector $\left[d\widetilde{X}_R(\nu_1), d\widetilde{X}_I(\nu_1)\right]$ when $\nu_1 \neq \nu_2$. This implies the MI between $\left[d\widetilde{Y}_R(\nu_2), d\widetilde{Y}_I(\nu_2)\right]$ and $\left[d\widetilde{X}_R(\nu_1), d\widetilde{X}_I(\nu_1)\right]$, which is defined as $\mathrm{MI}_{XY}(\nu_1, \nu_2)$, is zero.

We will now derive the analytical expression for $\mathrm{MI}_{XY}(\nu, \nu)$, for $\nu \neq 0$. Let $H_1(\nu) = H_{1R}(\nu) + jH_{1I}(\nu)$ and $H_2(\nu) = H_{2R}(\nu) + jH_{2I}(\nu)$. We can see from (5), (6) that

$$\left[d\widetilde{Y}_R(\nu), d\widetilde{Y}_I(\nu)\right] \sim \mathcal{N}\left(\mathbf{0}, \left(\frac{1}{2}s_X(\nu)|H_1(\nu)|^2 + \frac{1}{2}s_W(\nu)|H_2(\nu)|^2\right)\mathbf{I}\right), \quad (23)$$

where $\mathcal{N}$ represents Gaussian distribution, $\mathbf{0}$ is a two element zero vector and $\mathbf{I}$ is the 2×2 identity matrix. In addition,

$$\left[d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu), d\widetilde{Y}_R(\nu), d\widetilde{Y}_I(\nu)\right] \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right), \quad (24)$$

where $\Sigma_{11} = \frac{1}{2}s_X(\nu)\mathbf{I}$, $\Sigma_{22} = \frac{1}{2}\sigma_{\widetilde{Y}}^2(\nu)\mathbf{I}$, $\sigma_{\widetilde{Y}}^2(\nu) = (s_X(\nu)|H_1(\nu)|^2 + s_W(\nu)|H_2(\nu)|^2)$, $\mathbf{I}$ is the 2×2 identity matrix and $\mathbf{0}$ is a four element zero vector. In addition,

$$\Sigma_{12} = \Sigma_{21}^{\mathrm{T}} = \begin{bmatrix} \frac{1}{2}s_X(\nu)H_{1R}(\nu) & \frac{1}{2}s_X(\nu)H_{1I}(\nu) \\ -\frac{1}{2}s_X(\nu)H_{1I}(\nu) & \frac{1}{2}s_X(\nu)H_{1R}(\nu) \end{bmatrix}.$$

Now, the MI between X and Y at frequency ν is given by

$$\mathrm{MI}_{XY}(\nu, \nu) = \mathrm{I}\left(\left\{d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu)\right\}; \left\{d\widetilde{Y}_R(\nu), d\widetilde{Y}_I(\nu)\right\}\right)$$
$$= \mathrm{I}\left(\left\{d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu)\right\}; d\widetilde{Y}_R(\nu)\right)$$
$$+ \mathrm{I}\left(\left\{d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu)\right\}; d\widetilde{Y}_I(\nu)|d\widetilde{Y}_R(\nu)\right) \quad (25)$$
$$= \mathrm{I}\left(\left\{d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu)\right\}; d\widetilde{Y}_R(\nu)\right)$$
$$+ \mathrm{I}\left(\left\{d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu)\right\}; d\widetilde{Y}_I(\nu)\right), \quad (26)$$

where (25) follows from the chain rule of mutual information [34] and (26) follows because the real and imaginary parts of the spectral process of a Gaussian process are independent from Theorem 2. In addition, $\left[d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu), d\widetilde{Y}_R(\nu)\right]$ is a Gaussian distributed random vector with zero mean and covariance matrix Σ', which is easily obtained from (24). Since the mutual information between components of a Gaussian random vector depends only on the determinants of the joint distribution's covariance matrices and that of marginals [34], we can easily show that

$$\mathrm{I}\left(\left\{d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu)\right\}; d\widetilde{Y}_R(\nu)\right) = \frac{1}{2}\log\frac{|\Sigma_{11}|\left(\frac{1}{2}\sigma_{\widetilde{Y}}^2\right)}{|\Sigma'|}$$
$$= \frac{1}{2}\log\left(1 + \frac{|H_1(\nu)|^2 s_X(\nu)}{|H_2(\nu)|^2 s_W(\nu)}\right). \quad (27)$$

Similarly, we can also show that

$$\mathrm{I}\left(\left\{d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu)\right\}; d\widetilde{Y}_I(\nu)\right) =$$
$$\frac{1}{2}\log\left(1 + \frac{|H_1(\nu)|^2 s_X(\nu)}{|H_2(\nu)|^2 s_W(\nu)}\right). \quad (28)$$

From (26), (27) and (28), we have

$$\mathrm{MI}_{XY}(\nu, \nu) = 2 \times \mathrm{I}\left(\left\{d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu)\right\}; d\widetilde{Y}_R(\nu)\right)$$
$$= \log\left(1 + \frac{|H_1(\nu)|^2 s_X(\nu)}{|H_2(\nu)|^2 s_W(\nu)}\right). \quad (29)$$

At $\nu = 0$, MI-in-frequency between X and Y is equal to $\mathrm{I}\left(\left\{d\widetilde{X}_R(\nu), d\widetilde{X}_I(\nu)\right\}; d\widetilde{Y}_R(\nu)\right)$, since the imaginary part of Y is zero.

c) Relationship Between MI in Frequency and Coherence

The coherence $C_{XY}(\nu) \in [0, 1]$ between two processes X and Y related by (4) is given by

$$C_{XY}(\nu) = \frac{|s_{XY}(\nu)|^2}{s_X(\nu)s_Y(\nu)}$$
$$= \frac{|H_1(\nu)|^2 s_X(\nu)}{s_X(\nu)|H_1(\nu)|^2 + s_W(\nu)|H_2(\nu)|^2}.$$
$$\Rightarrow -\log(1 - C_{XY}(\nu)) = \log\left(1 + \frac{|H_1(\nu)|^2 s_X(\nu)}{|H_2(\nu)|^2 s_W(\nu)}\right)$$
$$= \mathrm{MI}_{XY}(\nu, \nu). \quad (30)$$

d) Proof of Theorem 4

Now we consider two discrete-time Gaussian stochastic processes $X[n]$ and $Y[n]$ that are related by

$$y[n] = h_1[n] * x[n] + h_2[n] * w[n], \quad (31)$$

where $h_1[n]$ and $h_2[n]$ are the impulse responses of two discrete-time linear, time-invariant filters. (31) is the discrete-time equivalent of (4). It was shown in chapter 10 in [33] that mutual information between the discrete-time Gaussian stochastic processes $X[n]$ and $Y[n]$ is related to coherence according to

$$\mathrm{I}(X; Y) = -\int_0^{0.5} \log(1 - C_{XY}(\lambda))\, d\lambda. \quad (32)$$

From (30) and (32), we have

$$\mathrm{I}(X; Y) = \int_0^{0.5} \mathrm{MI}_{XY}(\lambda, \lambda)\, d\lambda. \quad (33)$$

References

[1] R. Malladi, D. H. Johnson, G. Kalamangalam, N. Tandon, and B. Aazhang, "Measuring cross-frequency coupling using mutual information and its application to epilepsy," in *Cosyne Abstracts*, Salt Lake City, UT, USA, 2017.

[2] R. Malladi, D. H. Johnson, G. Kalamangalam, N. Tandon, and B. Aazhang, "Data-driven estimation of mutual information using frequency domain and its application to epilepsy," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2017.

[3] H. O. Lüders, I. Najm, D. Nair, P. Widdess-Walsh, and W. Bingman, "The epileptogenic zone: General principles," *Epileptic Disorders*, vol. 8, no. 2, pp. 1–9, 2006.

[4] F. Rosenow and H. Lüders, "Presurgical evaluation of epilepsy," *Brain*, vol. 124, no. 9, pp. 1683–1700, 2001.

[5] G. K. Bergey *et al.*, "Long-term treatment with responsive brain stimulation in adults with refractory partial seizures," *Neurology*, vol. 84, no. 8, pp. 810–817, 2015.

[6] S. Sunderam, B. Gluckman, D. Reato, and M. Bikson, "Toward rational design of electrical stimulation strategies for epilepsy control," *Epilepsy Behav.*, vol. 17, no. 1, pp. 6–22, 2010.

[7] E. Krook-Magnuson and I. Soltesz, "Beyond the hammer and the scalpel: Selective circuit control for the epilepsies," *Nature Neurosci.*, vol. 18, no. 3, pp. 331–338, 2015.

[8] R. T. Canolty *et al.*, "High gamma power is phase-locked to theta oscillations in human neocortex," *Science*, vol. 313, no. 5793, pp. 1626–1628, 2006.

[9] R. T. Canolty and R. T. Knight, "The functional role of cross-frequency coupling," *Trends Cogn. Sci.*, vol. 14, no. 11, pp. 506–515, 2010.

[10] C. Alvarado-Rojas *et al.*, "Slow modulations of high-frequency activity (40–140 Hz) discriminate preictal changes in human focal epilepsy," *Sci. Rep.*, vol. 4, 2014, Art. no. 4545.

[11] K. Edakawa *et al.*, "Detection of epileptic seizures using phase–amplitude coupling in intracranial electroencephalography," *Sci. Rep.*, vol. 6, 2016, Art. no. 25422.

[12] M. Guirgis, Y. Chinvarun, M. del Campo, P. L. Carlen, and B. L. Bardakjian, "Defining regions of interest using cross-frequency coupling in extratemporal lobe epilepsy patients," *J. Neural Eng.*, vol. 12, no. 2, 2015, Art. no. 026011.

[13] S. A. Weiss *et al.*, "Seizure localization using ictal phase-locked high gamma a retrospective surgical outcome study," *Neurology*, vol. 84, no. 23, pp. 2320–2328, 2015.

[14] S. Liu *et al.*, "Exploring the time–frequency content of high frequency oscillations for automated identification of seizure onset zone in epilepsy," *J. Neural Eng.*, vol. 13, no. 2, 2016, Art. no. 026026.

[15] H. Zhou, Y. Li, Y.-L. Hsin, and W. Liu, "Phase-amplitude coupling analysis for seizure evolvement using Hilbert Huang transform," in *Proc. IEEE 38th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2016, pp. 1022–1025.

[16] R. Zhang *et al.*, "Temporal-spatial characteristics of phase-amplitude coupling in electrocorticogram for human temporal lobe epilepsy," *Clin. Neurophysiol.*, vol. 128, no. 9, pp. 1707–1718, 2017.

[17] M. Amiri, B. Frauscher, and J. Gotman, "Phase-amplitude coupling is elevated in deep sleep and in the onset zone of focal epileptic seizures," *Frontiers Human Neurosci.*, vol. 10, p. 387, 2016.

[18] M. Cotic, Y. Chinvarun, M. del Campo, P. L. Carlen, and B. L. Bardakjian, "Spatial coherence profiles of ictal high-frequency oscillations correspond to those of interictal low-frequency oscillations in the ECoG of epileptic patients," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 76–85, Jan. 2016.

[19] B. Frauscher, N. von Ellenrieder, F. Dubeau, and J. Gotman, "Different seizure-onset patterns in mesiotemporal lobe epilepsy have a distinct interictal signature," *Clin. Neurophysiol.*, vol. 128, no. 7, pp. 1282–1289, 2017.

[20] L. Faes and G. Nollo, *Multivariate Frequency Domain Analysis of Causal Interactions in Physiological Time Series*. Rijeka, Croatia: InTech, 2011.

[21] A. C. Onslow, R. Bogacz, and M. W. Jones, "Quantifying phase–amplitude coupling in neuronal network oscillations," *Prog. Biophys. Mol. Biol.*, vol. 105, no. 1, pp. 49–57, 2011.

[22] J. Aru *et al.*, "Untangling cross-frequency coupling in neuroscience," *Current Opinion Neurobiol.*, vol. 31, pp. 51–61, 2015.

[23] R. Pascual-Marqui *et al.*, "The dual frequency RV-coupling coefficient: A novel measure for quantifying cross-frequency information transactions in the brain," arXiv:1603.05343, 2016.

[24] D. R. Brillinger, "Second-order moments and mutual information in the analysis of time series," *Recent Adv. Stat. Methods*, pp. 64–76, 2002.

[25] R. Salvador *et al.*, "A simple view of the brain through a frequency-specific functional connectivity measure," *Neuroimage*, vol. 39, no. 1, pp. 279–289, 2008.

[26] D. R. Brillinger and A. Guha, "Mutual information in the frequency domain," *J. Stat. Planning Inference*, vol. 137, no. 3, pp. 1076–1084, 2007.

[27] H. Cramér and M. Leadbetter, *Stationary and Related Stochastic Processes: Sample Function Properties and their Applications* (Wiley series in Probability and Mathematical Statistics: Tracts on Probability and Statistics). New York, NY, USA: Wiley, 1967.

[28] H. J. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences*. vol. 2. New York, NY, USA: Wiley, 1979.

[29] D. R. Brillinger, *Time Series: Data Analysis and Theory*. Philadelphia, PA, USA: Soc. Ind. Appl. Math., 2001.

[30] S. Khan *et al.*, "Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data," *Phys. Rev. E*, vol. 76, no. 2, 2007, Art. no. 026209.

[31] E. Schaffernicht, R. Kaltenhaeuser, S. S. Verma, and H.-M. Gross, "On estimating mutual information for feature selection," in *Proc. Int. Conf. Artif. Neural Netw.*, 2010, pp. 362–367.

[32] Q. Wang, S. R. Kulkarni, and S. Verdú, "Universal estimation of information measures for analog sources," *Found. Trends Commun. Info. Theory*, vol. 5, no. 3, pp. 265–353, 2009.

[33] M. Pinsker, *Information and Information Stability of Random Variables and Processes* (Holden-Day series in time series analysis). San Francisco, CA, USA: Holden-Day, 1964.

[34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2012.

[35] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*. vol. 2. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

[36] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Prog. Neurobiol.*, vol. 77, no. 1, pp. 1–37, 2005.

[37] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York, NY, USA: Wiley, 2015.

[38] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, no. 6, 2004, Art. no. 066138.

[39] T. Duong, "ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R," *J. Stat. Softw.*, vol. 21, no. 7, pp. 1–16, 2007.

[40] D. Dvorak and A. A. Fenton, "Toward a proper estimation of phase–amplitude coupling in neural oscillations," *J. Neurosci. Methods*, vol. 225, pp. 42–56, 2014.

[41] R. Malladi, G. Kalamangalam, N. Tandon, and B. Aazhang, "Identifying seizure onset zone from the causal connectivity inferred using directed information," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 7, pp. 1267–1283, Oct. 2016.

[42] J. I. Berman *et al.*, "Variable bandwidth filtering for improved sensitivity of cross-frequency coupling metrics," *Brain Connectivity*, vol. 2, no. 3, pp. 155–163, 2012.

[43] J. Engel, "Update on surgical treatment of the epilepsies summary of the second international palm desert conference on the surgical treatment of the epilepsies (1992)," *Neurology*, vol. 43, no. 8, pp. 1612–1612, 1993.

[44] C. Tonini *et al.*, "Predictors of epilepsy surgery outcome: A meta-analysis," *Epilepsy Res.*, vol. 62, no. 1, pp. 75–87, 2004.

[45] U. Gleissner, R. Sassen, M. Lendt, H. Clusmann, C. Elger, and C. Helmstaedter, "Pre-and postoperative verbal memory in pediatric patients with temporal lobe epilepsy," *Epilepsy Res.*, vol. 51, no. 3, pp. 287–296, 2002.

[46] S. Karunakaran *et al.*, "The interictal mesial temporal lobe epilepsy network," *Epilepsia*, vol. 59, no. 1, pp. 244-258, 2017.

[47] K. Kim, A. D. Ekstrom, and N. Tandon, "A network approach for modulating memory processes via direct and indirect brain stimulation: Toward a causal approach for the neural basis of memory," *Neurobiol. Learn. Memory*, vol. 134, pp. 162–177, 2016.

[48] G. W. Corder and D. I. Foreman, *Nonparametric Statistics: A Step-by-step Approach*. New York, NY, USA: Wiley, 2014.

[49] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc. Series B (Methodolog.)*, vol. 57, pp. 289–300, 1995.

[50] G. J. Székely and M. L. Rizzo, "Brownian distance covariance," *Ann. Appl. Stat.*, vol. 3, no. 4, pp. 1236–1265, 2009.

[51] D. N. Reshef *et al.*, "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, 2011.

[52] S. Hooper *et al.*, "On developing an FPGA based system for real time seizure prediction," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2017.

Rakesh Malladi received the B.Tech and M.Tech degrees in electrical engineering from the Indian Institute of Technology Madras, Chennai, India, in 2011 and the Ph.D. degree from the Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA, in 2017. During the Summer of 2010, he was with IBM Research, Bangalore, India, during the Summer of 2013, he was with Texas Instruments, Dallas, TX, USA, and during the Summer of 2015, he was a Research Intern with Cyberonics, Houston, TX, USA. Since 2017, he has been a Machine Learning Engineer with LinkedIn Corporation, Sunnyvale, CA, USA. His research interests include signal processing, machine learning, computational neuroscience, and information theory.
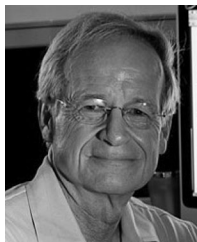
Don H. Johnson (F'90) received the S.B. and S.M. degrees in 1970, the E.E. degree in 1971, and the Ph.D. degree in 1974, all in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA. In 1974, he joined the M.I.T. Lincoln Laboratory, Lexington, MA, USA, as a Staff Member, to work on digital speech systems. In 1977, he joined the Faculty with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA, where he is currently the J.S. Abercrombie Professor Emeritus. His research interests include art forensic analysis from x-ray images and statistical signal processing. He has coauthored the graduate-level book *Array Signal Processing* and he holds nine patents. He is a Former President of the Signal Processing Society, and a recipient of the IEEE Millenium Medal and the Signal Processing Society's Meritorious Service Award.

Giridhar P. Kalamangalam received the MBBS degree from the Jawaharlal Institute, Pondicherry, India, in 1989, and the M.Sc. and D.Phil. degrees in applied mathematics from the University of Oxford, Oxford, U.K., in 1991 and 1995, respectively. Following postgraduate general medical training from Membership of the Royal Colleges of Physicians, U.K., he was a Neurologist from Glasgow, U.K., with subspecialty expertise in epilepsy with Cleveland Clinic, Cleveland, Oh, USA. He was an Assistant and then an Associate Professor and a Attending Epileptologist with the University of Texas Health Science Center, Houston, TX, USA, until 2017. Since September 2017, he has been the Wilder Family Endowed Professor and the Division Chief of epilepsy with the University of Florida College of Medicine, Gainesville, FL, USA. His research interests include physiological dynamics and the neuroimaging of epilepsy and cognitive neuroscience.

Nitin Tandon is a Professor of neurosurgery with the McGovern Medical School, University of Texas Health Science Center in Houston, TX, USA, and practices with the Memorial Hermann Hospital, Texas Medical Center, Houston, TX, USA, where he directs a world-class epilepsy surgery and brain mapping program. He has performed more than 3000 brain operations, with more than 1200 for brain tumors and 600 for epilepsy. He has pioneered innovative technologies in azimuth based stereo-electroencephalography, laser interstitial ablation, and other minimally invasive approaches to brain tumors and epilepsy. His research interests include fundamental insight into the processes of cognition and epilepsy using intracranial recordings and multimodal comparisons linking noninvasive imaging data and tractography datasets with intracranial recordings and cortical stimulation. This will pave the way for innovative rehabilitative and neuroprosthetic approaches to the remediation of aphasic and dyslexic disorders and intractable epilepsy. His research is funded by a BRAIN initiative U01, an R01, and the NSF grants.

Behnaam Aazhang (F'99) received the B.S. (Hons.), M.S., and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1981, 1983, and 1986, respectively. From 1981 to 1985, he was a Research Assistant with the Coordinated Science Laboratory, University of Illinois. In August 1985, he joined the Faculty of Rice University, Houston, TX, USA, where he is currently the J.S. Abercrombie Professor with the Department of Electrical and Computer Engineering and a Professor and the Director of Center on Neuroengineering with Multi-University Research Center, Houston, TX, USA. From 2006 to 2014, he held an Academy of Finland Distinguished Visiting Professorship appointment with the University of Oulu, Oulu, Finland. He received an Honorary Doctorate degree from the University of Oulu (the highest honor that the university can bestow), in 2017. He is a Fellow of the American Association for the Advancement of Science, an Distinguished Lecturer of the IEEE Communication Society. He is the recipient of the IEEE ComSoc CTTC Outstanding Service Award "For innovative leadership that elevated the success of the Communication Theory Workshop" in 2016 and Outstanding Technical Achievement Award "For consistent, fundamental contributions to multiuser communication theory for wireless networks" in 2017. He is a recipient of 2004 IEEE Communication Society's Stephen O. Rice best paper award for a paper with A. Sendonaris and E. Erkip. He has been listed in the Thomson-ISI Highly Cited Researchers and has been a Keynote and Plenary Speaker of several conferences.