

RESEARCH ARTICLE

# Statistical Detection of EEG Synchrony Using Empirical Bayesian Inference

Archana K. Singh<sup>1\*</sup>, Hideki Asoh<sup>2</sup>, Yuji Takeda<sup>2</sup>, Steven Phillips<sup>2</sup>

**1** ATR Neural Information Analysis Laboratories, Kyoto 619-0288, Japan, **2** National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba 305-8568 Japan

\* [archana@mail.ecc.u-tokyo.ac.jp](mailto:archana@mail.ecc.u-tokyo.ac.jp); [sine.arc@gmail.com](mailto:sine.arc@gmail.com)



## Abstract

There is growing interest in understanding how the brain utilizes synchronized oscillatory activity to integrate information across functionally connected regions. Computing phase-locking values (PLV) between EEG signals is a popular method for quantifying such synchronizations and elucidating their role in cognitive tasks. However, high-dimensionality in PLV data incurs a serious multiple testing problem. Standard multiple testing methods in neuroimaging research (e.g., false discovery rate, FDR) suffer severe loss of power, because they fail to exploit complex dependence structure between hypotheses that vary in spectral, temporal and spatial dimension. Previously, we showed that a hierarchical FDR and optimal discovery procedures could be effectively applied for PLV analysis to provide better power than FDR. In this article, we revisit the multiple comparison problem from a new Empirical Bayes perspective and propose the application of the local FDR method (locFDR; Efron, 2001) for PLV synchrony analysis to compute FDR as a posterior probability that an observed statistic belongs to a null hypothesis. We demonstrate the application of Efron's Empirical Bayes approach for PLV synchrony analysis for the first time. We use simulations to validate the specificity and sensitivity of locFDR and a real EEG dataset from a visual search study for experimental validation. We also compare locFDR with hierarchical FDR and optimal discovery procedures in both simulation and experimental analyses. Our simulation results showed that the locFDR can effectively control false positives without compromising on the power of PLV synchrony inference. Our results from the application locFDR on experiment data detected more significant discoveries than our previously proposed methods whereas the standard FDR method failed to detect any significant discoveries.

## OPEN ACCESS

**Citation:** Singh AK, Asoh H, Takeda Y, Phillips S (2015) Statistical Detection of EEG Synchrony Using Empirical Bayesian Inference. PLoS ONE 10(3): e0121795. doi:10.1371/journal.pone.0121795

**Academic Editor:** Daniele Marinazzo, Universiteit Gent, BELGIUM

**Received:** November 26, 2014

**Accepted:** February 2, 2015

**Published:** March 30, 2015

**Copyright:** © 2015 Singh et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The PLV dataset and software code used for this analysis can be downloaded from <http://dx.doi.org/10.6084/m9.figshare.1309378>.

**Funding:** Funding provided by Japan Society for the Promotion of Science (<https://www.jsps.go.jp/english/>) grant KAKENHI (26280051) to SP and Ministry of Internal Affairs and Communications (<http://www.soumu.go.jp/english/>) for a Project "Novel and innovative R & D making use of brain structures" which supported AS for this research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Introduction

A new class of simultaneous inference problems has emerged with the recent development in EEG technology, which provides us with means of acquiring thousands of signals in the form of large-scale high-dimensional data. Establishing functional connectivity inference in topological time-frequency maps, which is increasingly the focus of neuroimaging research in recent

**Competing Interests:** The authors have declared that no competing interests exist.

years, epitomizes a classical case of such problems [1, 2, 3, 4, 5, 6, 7]. In particular, there is a greater appreciation for the complex role that synchrony plays in cognitive tasks, and how synchrony varies along spatial, temporal and spectral dimensions. For instance, when the brain engages in visual search, it locks frontal and parietal regions at certain time–frequency interval, and the topological, temporal, and spectral instances of this locking vary depending on whether the search employs top-down control of attention, as reported in previous studies on monkeys by [8] and on humans by [9].

An appropriate quantification of such topologically, spectrally, and temporally locked functional connectivity requires state-of-art mathematical and statistical methods. Mathematically, several options exist, and the most commonly used option is a phase locking value, which is computed from wavelet decomposed EEG signals as the absolute average phase difference over trials between two signals [10]. PLV measures are applicable to non-stationary signals, though the phase and amplitude components may not be independent for Gaussian signals [11]. The absolute PLVs are bounded between 0 (no phase locking) and 1 (maximum phase locking). In event related EEG studies, the research question is often concerned with the changes in synchronization in response to a stimulus, therefore absolute PLV values are normalized with respect to a pre-stimulus baseline. Normalized PLVs are no longer bound between 0 and 1 and provide instantaneous measures of changes in phase coupling between two signals at any desired frequency and time. In this article, all references to PLV should be regarded as normalized PLV throughout the text unless stated otherwise. Statistically, quantifying the significant synchrony from PLV is difficult, because their high-dimensional time–frequency measurements from multiple sensors incur a serious multiple comparison issue. A common practice to address the problem is to restrict the hypothesis space, e.g., by averaging the data from the pre-specified sensors, time and frequency so that there is only one (or a few) summary statistic(s) per subject per condition. Although this is a fair and valid practice, it requires the region of interest to be defined a priori, which is often not available especially for EEG synchrony research problems.

High dimensional data pose several challenges to the conventional approach for multiple comparison correction procedures. The main problem with this approach is that it treats large-scale testing as a simple extension of single hypothesis testing, which by design aims at rejecting the null hypothesis with high probability using the data only from the tested individual hypothesis. On the contrary and in practice, large scale testing is pursued with a different scientific motivation. A case in point is the PLV analysis, where the results are plotted in a functional connectivity map, showing the interesting discoveries concerning the hypothesis under investigation. For instance, we may be interested in exploring where, when, and at what frequencies do we observe phase locking in the brain in response to the task under investigation. Here, we expect each dimension of hypothesis space to have a distinct dependence structure depending on the expected joint distribution of discoveries (non-nulls) and non-discoveries (nulls). Therefore, the single testing approach, that requires simply pooling all the hypotheses from different dimensions together to determine a common cutoff without accounting for the joint distribution of the hypotheses, tends to be highly insensitive.

The problems of identifying statistically significant functional connectivity in high-dimensional space-time-frequency synchrony data have been reported by various authors in neuroimaging literature [12, 13, 14, 15, 16, 17, 18, 19]. However, only a few studies have suggested the solutions to this problem. In addition to frequentist inference, Bayesian inference for modeling effective connectivity e.g. a dynamic causal model is also available for time–frequency EEG data, which does not require a correction of multiple comparison [20]. However, it requires the sensors of interest to be defined a priori [21]. In the context of fMRI imaging of brain activity, the multiple comparison issue was addressed by random field theory that

statistically extracts and quantifies topological features of the SPM, e.g., number of regions, their spatial extent or peak height, and thus explicitly accounts for the dependence in neuroimaging data [16, 22, 23]. The use of permutation testing also improved the inference in fMRI studies to identify a cluster-level threshold [24]. For sparsely sampled EEG space-time-frequency data, both the topological and cluster-level inference require a priori knowledge about the region of interest in at least a single dimension. This restricts their applications to either space-time (when frequency band of interest is known a priori) or time-frequency (when sensor of interest is known a priori) analysis [15, 25, 26, 27].

Previously, we suggested two alternatives that avoid the difficulties with topological FDR just mentioned [17, 28]. One way is to organize hypotheses into a hierarchy of families. This approach is employed in the hierarchical false discovery rate (hFDR) procedure [29]. Another way is to modify the test statistic so that it explicitly shares the dependence information from across all the tests as suggested by [30] in his optimal discovery procedure (ODP). In this paper, we present a third alternative, which is motivated by Empirical Bayesian approach to control local false discovery rate [31]. We outline these approaches.

The hFDR method ameliorates this problem for testing multiple hypotheses that are naturally organized into a hierarchical structure [29]. Although we do not know in advance precisely which tests are dependent, domain knowledge, or the sorts of hypotheses we wish to entertain help determine a hypothesis-tree hierarchy. Suppose synchrony is known to localize within four frequency bands. Arranging 20 frequency-specific PLV tests into four families (one for each band) and treating each family as a single test explicitly assumes that each test within a family is dependent. Hence, our adjustment for effective significance of each test is reduced from 20 to 4, thus raising the sensitivity of our testing procedure. For each family-level test that exceeds threshold, the procedure is recursively applied. In practice, we found that hFDR was effective in detecting true effects without raising the level of false discoveries beyond the expected level with reference to high-dimensional PLV data [17]. The advantage of hFDR is that it allows for a hierarchical and multi-dimensional inference, e.g., when we seek to test the more detailed hypothesis of synchrony within a particular time band (within a particular frequency band). A potential difficulty with hFDR occurs when there is no natural or a priori reason for arranging multiple hypotheses into a particular hierarchy.

It makes sense to increase the significance of a test if there exist other tests with the similar results to maximize expected true positives for each fixed expected false positive. This is precisely the approach that is followed in ODP [30] and locFDR [31] procedure, neither of which requires hierarchically organized families of hypothesis. ODP maximizes the number of expected true positives for each fixed number of expected false positives. The procedure relies on a statistic specifically designed for multiple large-scale testing, which is computed as the ratio of the sum of the probability of a given test under the alternative distribution of each test to that of its null distribution. The relative significance of each test is also estimated with reference to the other tests. In this way, the relative significance of each observed statistic increases when multiple true positive results are likely to have similar values, e.g., when the tests are dependent. The estimation of ODP statistic is based on true null and alternate probability densities of the data. We developed an application of ODP and demonstrated its effectiveness in the context of high-dimensional EEG phase locking data [28]. The current ODP application is available only for the data that conforms to normal density, which holds reasonably well in the case of PLV data. However, for other applications of ODP, where distributional assumptions are unknown or different, it will take some substantial developments to arrive at a generally applicable set of methods.

The local FDR procedure offers an Empirical Bayesian perspective on the issue of multiple comparisons in high dimensional PLV data. The formulation of locFDR underlies an Empirical

Bayesian model, where priors of the parameters are common among all hypotheses, and are hierarchically estimated from the data. Local FDR is then estimated as the posterior probability that a hypothesis is null given its observed statistics [31]. The distribution of local FDR is modeled using a mixture density, and the inference is obtained by comparing the density of the null to the mixture distribution for any given statistic.

The local FDR procedure leverages high-dimensionality and dependence in the data to increase the sensitivity of the large-scale inference. Dependence is incorporated by sharing common information across the hypotheses for reducing both Type I (false discoveries) and Type II (missed discoveries) errors. Unlike hFDR, local FDR does not require any specific hierarchy among the hypotheses. Local FDR is more accurate than ODP, and has greater utility, because it can also be applied to other commonly used statistics in neuroimaging studies, such as t-statistic, z-score, and chi-square. Local FDR is also more robust to changes in symmetry of the statistic (e.g. z-score computed from the normalized PLV data) than ODP.

In this study, we focus on the application of locFDR and demonstrate its validity using simulations and experimental data in comparison with the other large-scale testing methods, ODP, hFDR, and the conventional FDR. The simulations reveal that these methods are more or less compatible in their detection sensitivity and FDR control. In experiment analysis, the locFDR option reveals the maximum number of significant pairs of synchrony at a conventional FDR cutoff of 5% in the expected region of interest. To the best of our knowledge, locFDR, hFDR, and ODP are the only currently available techniques to assess the significance of synchrony for large-scale testing that do not require a priori knowledge about the expected regions of interest (i.e., in time, frequency, or space). Therefore, a comparative review of these will serve as a useful reference for research investigations on brain connectivity using high-dimensional data.

## Materials and Methods

This section includes an overview of the concepts related to the phase locking value, frequentist and Bayesian formulations of FDR, and local FDR. For the purpose of comparing local FDR with other frequentist large-scale approaches, hFDR and ODP, they are also described briefly in the end.

### Phase Locking Value (PLV)

Phase-locking value (PLV [10]) is used as a measures of synchrony. PLVs are computed from stimulus-locked EEG data for each trial as measures of synchronization between brain regions. The phase  $\phi(t, f, n, e_i)$  at time  $t$ , frequency  $f$ , trial  $n$  and electrode  $e_i$  was computed by first convolving the data with a complex Morlet wavelet, defined as

$$w_{(t,f)} = (\sigma_t \sqrt{\pi})^{\frac{1}{2}} e^{-\frac{t^2}{2\sigma_t^2}} e^{2\pi i f t} \tag{1}$$

where  $\sigma_t = FTR/2\pi f$ . Following [10], FTR can be set as 7. The range of  $f$  can be specified as desired for the analysis. The absolute value of PLV for electrode pair  $(e_i, e_j)$  is computed as:

$$PLV_{(t,f,e_i,e_j)abs} = \frac{1}{N} \left| \sum_{h=1}^N e^{(\phi(t,f,n,e_i) - \phi(t,f,n,e_j))} \right| \tag{2}$$

where  $N$  is the number of trials. The absolute value of PLV varies between 0 (random phase difference, no phase locking) and 1 (constant phase difference, maximum phase locking). For event related studies, PLVs can be normalized with respect to a baseline, e.g. 200 ms pre-stimulus time period. The normalized values,  $PLV(t, f, e_i, e_j)norm$ , can be computed for each 1 ms

**Table 1. Variables associated with the number of true negatives (TN), false negatives (FN), false positives (FP) and true positives (TP), for multiple testing of  $m$  null hypotheses.**

	Declared Non-significant	Declared Significant	Total
True $H_0$	U (TN)	V (FP)	$m_0$
False $H_0$	T (FN)	S (TP)	$m - m_0$
Total	$m - R$	R	$m$

doi:10.1371/journal.pone.0121795.t001

time point as

$$PLV_{(t,f,e_i,e_j)norm} = (PLV_{(t,f,e_i,e_j)} - \mu_{base}) / \sigma_{base} \tag{3}$$

where  $\mu_{base}$  and  $\sigma_{base}$  are the mean and standard deviation of PLVs over the baseline period. PLVnorm represents normalized changes from the average baseline PLV at a given frequency and time and are no longer bound between 0 and 1. The positive PLVnorm values indicate increased synchronization and the negative values indicate decreased synchronization. PLVnorm values can be tested for significance using a parametric test, e.g., z-test and t-test. As PLV datasets involve simultaneous testing of many time-frequency windows across many electrode-pairs, their significance should be adjusted using multiple comparison procedures.

### False Discovery Rate (FDR)

FDR is defined as the expected proportion of falsely rejected hypotheses among the rejected ones, which is zero if there are no discoveries [32]. Let  $H_1, \dots, H_m$  denote the collection of null hypotheses corresponding to the p-values,  $p_1, \dots, p_m$ . The possible outcomes of  $m$  tests can be summarized as in Table 1:  $V$  and  $R$  represent the number of false positives and declared significant positives, respectively.

$$FDR = E\left(\frac{V}{R} | R > 0\right) Pr(R > 0) \tag{4}$$

### FDRBH Procedure for controlling False Discovery Rate

Benjamini and Hochberg [32] introduced a method for controlling FDR (henceforth referred to as FDRBH), which follows a *fixed error rate* approach, where FDR is fixed at a chosen level, and a rejection region is determined as follows. Let  $p_{(i)}$  and  $\alpha$  denote the ordered p-values and the pre-specified error rate, respectively. Then the rejection region for controlling FDR can be determined as  $\gamma = \max\{p_{(i)} : p_{(i)} \leq \alpha \frac{i}{m}\}$ , where all null hypotheses corresponding to  $p_{(i)} \leq \gamma$  are rejected. This method controls FDR at level  $\pi_0 \alpha$ , where  $\pi_0 = m_0/m$  is the proportion of true null hypotheses. Therefore, when all the null hypotheses are true ( $m_0 = m$ ), FDR is controlled at level  $\alpha$ , and when some of the null hypotheses are rejected ( $m_0 < m$ ), the procedure controls FDR at a level far below  $\alpha$ . The power of an FDR controlling procedure can be improved by substituting the estimated value of  $\pi_0$ .

### Positive False Discovery Rate (pFDR) and its estimation

Storey introduced an alternative measure and threshold for false discovery rate, *positive false discovery rate* (pFDR) and q-value [33]. pFDR is defined conditional on there being positive

findings, i.e., at least one discovery. Unlike FDRBH procedure, pFDR estimates false discovery rate given a fixed rejection threshold. This affords Storey’s procedure higher sensitivity over BH procedure, because it computes pFDR by estimating  $\pi_0$  over the specified rejection region. Assume that hypotheses  $H_1, \dots, H_m$  are tested using statistic  $z_1, \dots, z_m$  with corresponding p-values  $p_1, \dots, p_m$ . Let us denote a true null ( $H_i = 0$ ) as ‘null’. Then, pFDR can be defined for a given significance region  $Z$ :

$$pFDR(Z) = E\left(\frac{V(Z)}{R(Z)} \mid R(Z) > 0\right), \tag{5}$$

where  $V(Z) = \#\{\text{null} \mid z_i \in Z\}$  and  $R(Z) = \#\{z_i; z_i \in Z\}$  (see Table 1). The rejection region is determined from the observed p-values assuming that all null hypotheses are identical with an identical region  $\Gamma$  for all tests. Let  $\Gamma = [0, \gamma]$ , where  $\gamma \in [0, 1]$ , then we reject all the null hypotheses with p-values less than  $\gamma$ . Since  $p_i$  are uniformly distributed by assumption and  $\pi_0 m$  p-values are expected to be null, a conservative estimate of  $\pi_0$  can be given as follows:

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_i > \lambda\}}{(1 - \lambda)m} \tag{6}$$

where,  $\lambda$  is a tuning parameter,  $0 \leq \lambda \leq 1$ , and # indicates the number of times the condition within parentheses holds true.

### Bayesian and global interpretation of pFDR

In a Bayesian sense, pFDR measures the probability that a significant test is a true null hypothesis  $pFDR(Z) = P(\text{null} \mid z_i \leq Z)$  or  $P(\text{null} \mid z_i \geq Z)$  for the left-tailed or right-tailed test. This is a natural Bayesian analogue to p-value,  $p\text{-value}(z_i) = P(z_i \leq Z \mid \text{null})$ , and is also known as q-value ( $z_i$ ). The significance of the observed statistic with respect to pFDR is computed as  $q\text{-value}(z_i) = \min(pFDR(Z))$ . Even though q-values are specific to each test, pFDR is essentially computed over the entire rejection tail, and hence it is also referred to as *global FDR*.

### Empirical Bayesian estimation of local False Discovery Rate (locFDR)

Efron et al. proposed an Empirical Bayesian concept for estimating false discovery rate using the mixture model [31]. In this method,  $\pi_0$  can be estimated using the area around the peak of the theoretical or empirical null distribution. They defined local false discovery rate (locFDR) as the probability of a hypothesis being in the null group given an observed value of the test statistic. Let us suppose that the hypothesis  $H_i$  is tested using a z-value,  $z_i$  resulting in the p-value,  $p_i$ . Assume that each outcome  $z_i$  belongs to either of the two possible classes, “not interesting” or “interesting” PLV effects, corresponding to whether  $z_i$  is generated according to the null or alternative hypothesis. Let  $p_0$  and  $p_1$  be the probability of  $z_i$  being a null or non-null with the densities,  $f_0(z)$  and  $f_1(z)$ , respectively.

$$\begin{aligned} p_0 &= P(\text{null}), & f_0(z) &= \text{density if } \textit{not interesting} \text{ effect} \\ p_1 &= P(\text{not - null}), & f_1(z) &= \text{density if } \textit{interesting} \text{ effect} \end{aligned}$$

Assuming both classes, we can define mixture density for z-values, and apply Bayes theorem to

obtain posteriori probabilities as follows.

$$\begin{aligned}
 f(z) &= p_0 f_0(z) + p_1 f_1(z), \\
 p_1(z) &= P(\text{not - null} | z_i = z) = 1 - p_0 \frac{f_0(z)}{f(z)}, \\
 p_0(z) &= P(\text{null} | z_i = z) = p_0 \frac{f_0(z)}{f(z)}.
 \end{aligned}
 \tag{7}$$

By definition, locFDR is  $p_0(z)$ , the posterior probability that a hypothesis given the observed value  $z$  is null. A full Bayesian analysis would require a priori estimation of  $p_0$ ,  $p_1$ ,  $f_0$  and  $f_1$ . In the proposed Empirical Bayesian model, locFDR is estimated empirically from the observed large-scale data assuming that most of the tests belong to null cases ( $\pi_0 \geq .9$ ). This is done by 1) estimating the  $f(z)$  from the observed statistics,  $z$ , which can be done by fitting a smooth curve  $\hat{f}(z)$  to the histogram of  $z$ , e.g., using Poisson fit, 2) defining the null density  $f_0(z)$  either theoretically or empirically, and 3) assuming  $p_0 = 1$  ensures a conservative control of false discoveries both in tail-area FDR and local false discovery rate [31, 32]. For better sensitivity (i.e., to avoid overestimation),  $\hat{\pi}_0$  from Eq 6 is substituted in the above model for computing locFDR.

For detecting significance of PLV, we assume  $f_0(z)$  to be standard normal density  $\phi(z) = \exp(-z^2/2)/\sqrt{2\pi}$  and  $p_0$  is estimated from the histogram data near  $z = 0$ . An R locFDR package is available for estimating local false discovery rate using empirically or theoretically defined null densities from <http://cran.r-project.org/web/packages/locfdr/index.html> [31].

### Optimal Discovery Procedure

The procedure is implemented by computing ODP statistic, which requires estimation of null and alternate density functions from the observed data [30]. Let  $m_0$  be the number of true null hypotheses. Let  $f_j$  and  $g_j$  represent the respective true null and alternative densities for the  $j^{th}$  hypothesis,  $j \in \{1, \dots, m\}$ , for the observed PLV data,  $x_i$ , from the  $i^{th}$  ( $i = 1, m$ ) electrode pair. The PLV data  $x_i$  is evaluated at the estimated probability density functions for all electrode pairs. Here, we assume a normal density function  $\phi(\cdot | \mu, \sigma^2)$  for PLV data.

$$\hat{S}_{ODP}(x_i) = \frac{\sum_{j=m_0+1}^m g_j(x_i | H_1)}{\sum_{j=1}^{m_0} f_j(x_i | H_0)}
 \tag{8}$$

where  $(\mu_{j1}, \sigma_{j1}^2)$  and  $(0, \sigma_{j0}^2)$  are the mean and variance under alternative and null distributions, respectively. The intuition behind ODP is that the relative significance of each observed  $x_i$  increases when multiple true positive results are likely to have similar values. It makes sense to increase the significance of a test if there exist other tests with the similar results to maximize expected true positives for each fixed expected false positive. For example, if  $x_i$  corresponds to a true alternative null hypothesis, then its density,  $\phi(x_i | \mu_{i1}, \sigma_{i1}^2)$  will make a substantial contribution to  $S_{ODP}(x_i)$ . Furthermore, if there are other true alternatives with  $\mu_{i1} \approx \mu_{j1}$ , then the likelihood of  $x_i$  under  $\phi(x_i | \mu_{j1}, \sigma_{j1}^2)$  will also make a substantial contribution to  $S_{ODP}(x_j)$ . This procedure is optimal in the sense that for any pre-specified false positive rate, the ODP will have maximum true positive rate, see for detailed proof [34].

## Hierarchical FDR

The detailed procedure for hierarchical FDR (hFDR) is described in [29], and in the context of PLV analysis in [17]. The hFDR procedure is implemented by organizing the hypotheses into a family-subfamily tree hierarchy, where each (sub)family is associated with a single hypothesis. For instance, the tests for PLV data are grouped into  $M$  frequency and  $N$  time families based on the frequency and time band associated with the test. The  $M$  frequency families constitute the first level of the hierarchy. In this case, there are  $N$  time subfamilies at the second level for each frequency family, and within each time subfamily are the test statistics, one for each electrode pair, at the third (lowest) level. The data associated with each test within a family is summarized (i.e., averaged) to become the data for that family and its associated single hypothesis. The testing begins at the first level by applying a single-sample t-test and FDRBH control to test the  $M$  hypotheses in frequency family. If any hypothesis is rejected, then testing continues by testing the corresponding time subfamily at the second level. This process continues by recursively checking each child hypothesis of a parent that was rejected, and terminating upon not rejecting any children.

The FDR bound on a hypothesis tree is defined recursively as the sum of the expected proportion of the number of false discoveries to total discoveries for each family. An approximate bound is estimated as  $bound = q \delta \frac{N_d + N_f}{N_d + 1}$ , where  $N_d$  is the number of observed discoveries,  $N_f$  is the number of families tested, and  $\delta$ , a multiplicative constant, is set to 1 [29]. This bound varies in an interval  $[q, 2q]$ , where  $q$  is the expected FDR level. When the number of discoveries far exceeds the number of family tested, the hierarchical FDR bound converges to  $q$ .

## Simulation

The purpose of simulations was to assess the specificity and sensitivity of locFDR for multidimensional EEG data. We assigned true PLV effects to particular frequency-time bands. Hence, we associated PLV differences (between conditions) with 2 frequency bands, 12 time bands, 10 participants, and 25 electrode pairs (i.e., all pairwise combinations of 5 frontal and 5 parietal electrodes) constituting a  $2 \times 12 \times 10 \times 25$  array. For frequency-time windows containing significant effects, 10 out of 25 pairs were defined as truly significant. The proportion of electrode pairs with true synchronies,  $\pi_1$ , varied from 8% to 48% over a range of time-frequency windows.

To simulate dependence in the signals of electrode pairs across the time windows for specific frequency bands, their PLVs were generated from a multivariate normal distribution with the parameters  $\mu$  and  $\Sigma = \sigma^2 R$ . The mean vector,  $\mu$  was assigned positive PLV effects to represent the true positives pertaining to alternative hypotheses, and zero PLV effects for null hypotheses representing true negatives. The covariance matrix,  $\Sigma = \sigma^2 R$  was constructed by assigning the variance,  $\sigma^2$  and correlation matrix,  $R$ , which were extracted from a real PLV dataset.

We were also interested in comparing locFDR, ODP and hFDR methods on grounds of the false positives and false negatives that they incur. A direct comparison of these methods is difficult as each works on a distinct operating principle, using a different measure of false positives. Nevertheless, a comparative evaluation of their detection power and actual FDR incurred with synthetic data, where configurations of true discoveries is known, would be useful for exploring any specific conditions that may warrant the use of one method over the other. For ODP, the tuning parameter  $\lambda$  was automatically chosen using a bootstrap distribution with 100 resamples following [35]. For a detailed algorithm, refer to Appendix A in Singh et al. [28]. For hFDR analysis, we constructed a 3-level hierarchical FDR tree so as to cast the hypotheses belonging to the frequency dimension at the first level, those belonging to the time dimension at



the second level, and to each electrode pair at the third level. For FDRBH, all 600 (=  $2 \times 12 \times 25$ ) hypotheses were pooled in a single family. A 5% FDR threshold was used in locFDR, ODP, hFDR, and FDRBH. The simulations were performed in R. The numbers of detected discoveries, false positives, false negatives were obtained by averaging over 100 runs.

## Experimental data analysis

The experimental data was acquired from a previously published visual search study [9]. The purpose of the experiment was to test the hypothesis that top-down driven control of visual attention in humans is accompanied by frontal-parietal synchrony in the lower gamma-band. Top-down signals were induced using distractors that share a feature (e.g., color, or orientation) with the target, yielding a steep search slope (search time increasing with display set size)—inefficient search. Bottom-up signals were induced using distractors with no feature in common with the target, yielding a flat slope (search time independent of set size)—efficient search. Participants showed significantly greater synchrony between frontal and parietal electrodes in the lower gamma-band during inefficient than efficient search. These results supported the results from a monkey study of Buschman and Miller, who used spectral coherence to examine the synchrony between frontal and parietal local field potentials [8].

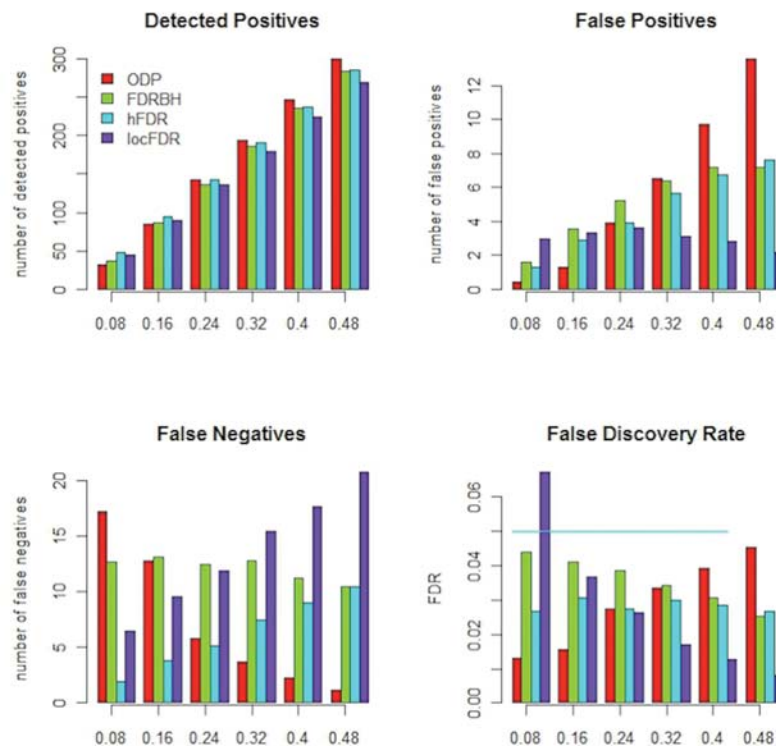
Following [9], the frequency,  $f$  ranged from 10 Hz to 58 Hz at intervals of 2 Hz for the calculation of absolute PLV (Eq 2), and the baseline period was from 200 ms to 0 ms prior to stimulus onset for the calculation of normalized PLV (Eq 3). Our region of interest was confined to 25 electrode pairs, i.e., five frontal electrodes (F7, F3, Fz, F4, F8) by five posterior electrodes (T5, P3, Pz, P4, T6), located according to the International 10–20 system. Frequency was partitioned into lower (22–34 Hz) and upper (36–48 Hz) gamma bands, corresponding to the original studies [8, 9, 17]. Time was partitioned into twelve 50 ms windows for the first 600 ms after stimulus (search display) onset. We applied locFDR, FDRBH, hFDR, and ODP using 5% and 10% thresholds. locFDR was performed using R locFDR package. The frequency-time maps of the resulting synchrony were plotted using matlab. The theoretical null was specified as  $N(0, 1)$  to estimate the  $p_0$ .

The PLV dataset and matlab code used for our analysis can be downloaded from <http://dx.doi.org/10.6084/m9.figshare.1309378>.

## Results

### Simulation

The results of simulation analysis are summarized in Fig 1. As the proportion of true discovery,  $\pi_1 = 1 - \pi_0$ , increases, locFDR exhibits a linear increase in the number of false negatives, a linear decrease in false discovery rate, whereas the number of false positives remains more or less constant. The local FDR was maintained well within the specified level of 5% except at extremely small proportion of true discovery; at  $\pi_1 = 0.08$ , the locFDR was approximately 7%. ODP shows a linear increase in the number of false positives, a linear decrease in the number of false negatives, and a linear increase in FDR as the proportion of true discoveries increases. FDR was maintained within the specified level of 5% in all cases. For hFDR, there was a linear increase in both the number of false positives and false negatives, but FDR remained constant and below 5% in all the cases. While the detection power shows the same trend across the methods, the number of detected discoveries differs slightly depending on the configuration of true discoveries.



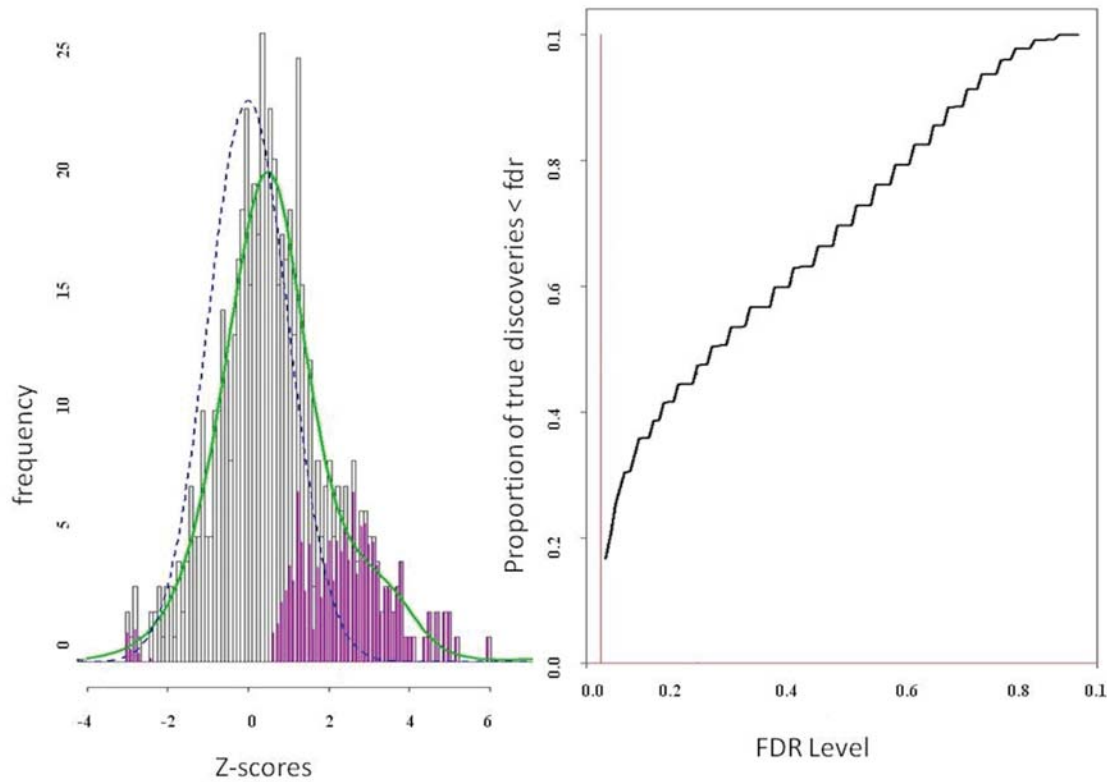
**Fig 1. Comparison between FDRBH, hFDR, ODP, and locFDR.** Each T-F window with assumed true synchrony effects has 10 pairs with true synchrony pairs; the x-axis represents the proportion of true positives; the graph panels represent: the total number of detected positives (top-left), the total number of false positives (top-right), the total number of false negatives (bottom-left), and false discovery rate (bottom-right).

doi:10.1371/journal.pone.0121795.g001

## Experiment

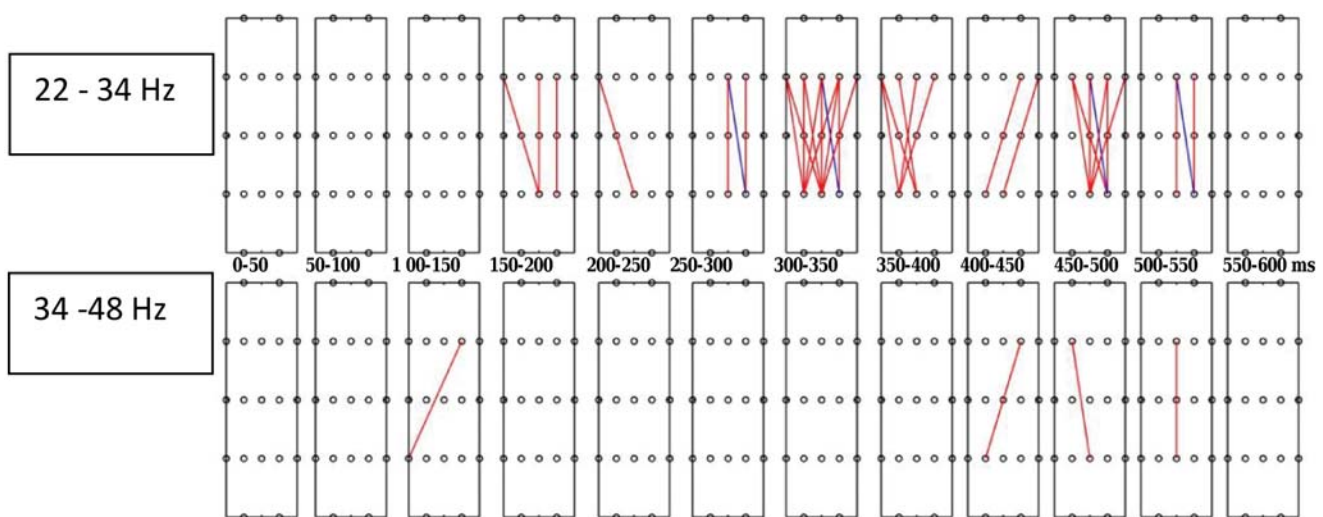
For the experimental data the null density, mixture density, and power (as a function of FDR) were estimated by locFDR (Fig 2). The locFDR method estimated 10% of the tests as belonging to the non-null group, that is, the proportion of pairs with true discovery ( $\pi_1 = 1 - \pi_0$ ), was estimated as 0.1. For comparison of locFDR results with those from hFDR, ODP and FDRBH, we first set the significant threshold at the conventional 5% level. The locFDR method detected 32 significant pairs (Fig 3). Most of the pairs are concentrated in the lower gamma band (22–34 Hz) over 300–500 ms post-stimulus time interval for significantly greater synchrony in the inefficient than efficient condition (red lines). The hFDR option detected 26 discoveries over 300–500 ms post-stimulus time interval in the lower gamma band (22–34 Hz) and none in upper gamma band (36–48 Hz) (see Fig 2 in [17]). ODP and FDRBH failed to detect any significant discoveries.

Next, we raised the significance threshold to 10% FDR level and obtained the results from all the four options. The results showed converging evidences of synchrony from locFDR, hFDR and ODP options predominantly in 22–34 Hz frequency band and 300–550 ms post stimulus time interval for the inefficient than efficient condition (Fig 4). The FDRBH option showed only 10 pairs with significantly greater synchrony in the lower gamma band for inefficient than efficient condition (results for FDRBH are not shown).



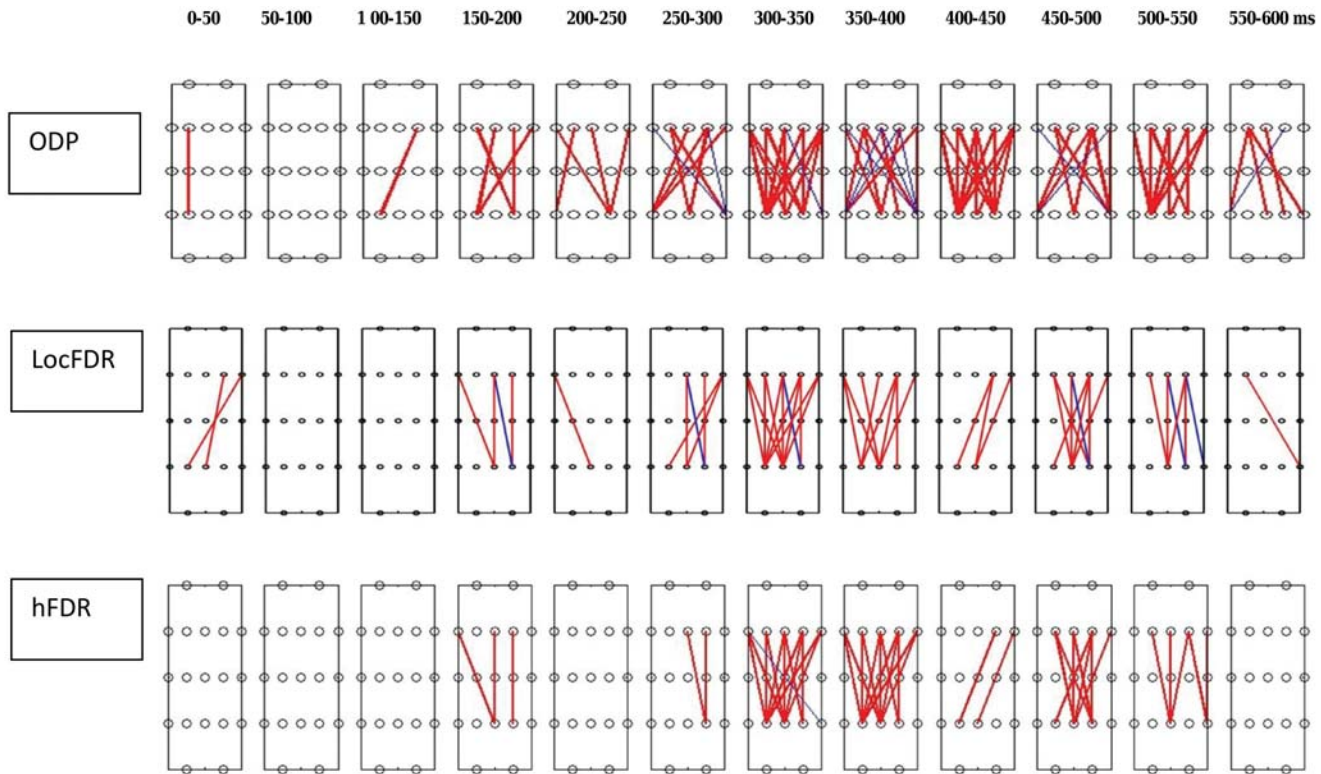
**Fig 2. Empirical estimates of null and mixture density are shown on the left panel.** The blue dashed line and green solid lines represents the null and mixture density, respectively. The pink bars indicate non-null counts. Power estimation as a function of estimated local FDR is shown on the right panel.

doi:10.1371/journal.pone.0121795.g002



**Fig 3. The synchrony map for experimental data from locFDR application.** The top and bottom rows correspond to lower (22–34 Hz) and higher (36–48 Hz) gamma bands, and twelve 50 ms time windows spanning 600 ms post-stimulus duration within the band. The map shows 5 (frontal) × 5 (parietal) electrode pairs showing significantly greater phase-locking for the inefficient than efficient search conditions (red lines) for 5.

doi:10.1371/journal.pone.0121795.g003



**Fig 4. Synchrony maps from hFDR, ODP, and locFDR applications.** Each map shows 5 (frontal) × 5 (parietal) electrode pairs corresponding to the lower (22–34 Hz) gamma band and twelve 50 ms time windows spanning 600 ms post-stimulus duration. The electrode pairs with significantly greater phase-locking for the inefficient than efficient search condition are indicated in red lines. The significant threshold is set at 10.

doi:10.1371/journal.pone.0121795.g004

## Discussion

This article describes the application of a novel local false discovery rate procedure [31] as a Bayesian alternative to the problem of detecting statistically significant synchrony in large-scale testing of PLVs. The last couple of decades have witnessed a gradual paradigm shift in the way we interpret the significance from large-scale hypothesis testing, which has become pervasive in all neurophysiological applications. This shift began with the paper by [32], who showed that controlling false discovery rate instead of family-wise error rate would lead to higher detection power without compromising on Type I error control. Since then, there have been continuing efforts to improve the sensitivity of FDR controlling method. Several new approaches have been adopted for large-scale testing of high-dimensional neuroimaging data, e.g., including cluster-wise false discovery rate procedure [36], topological false discovery rate control [23], hierarchical false discovery rate control [17], optimal false discovery rate control [28] etc. among others.

EEG phase locking analysis involves a complex dataset with a small proportion of significant synchrony effects in multiple dimensions of time, frequency, and electrode space, each with a distinct dependence structure. The inference in classic hypothesis testing theory is essentially interpreted by thresholding p-values that are obtained individually from tail-area of the null distribution under each individual hypothesis, and FDRBH is no exception. Inferences derived from classical hypothesis tend to be too conservative to reveal even the true significant discoveries when applied for statistical testing of PLV.

There are several limitations with classic single hypothesis testing. The main drawback is that it prevents the dependence structure among the expected true discoveries to be incorporated in the simultaneous evaluation of significance from multiple hypotheses testing. The second limitation is due to the tail-area based testing; it may lead to a flawed inference when the distribution of the statistic is asymmetrical or bumpy. Another important issue concerns the interpretation of the p-value itself. Although a p-value tells us whether or not an observed statistic is significant, it does not tell us whether it is generated by the null or alternative distribution. For example, the p-value of z-statistic,  $P(Z \geq 1.96|null)$  has a sensible interpretation, but  $P(Z = 1.96) = 0$  is uninformative.

In our previous articles, we adopted hierarchical FDR method and optimal discovery procedures to overcome some of the problems as we just mentioned [17, 28]. The hFDR testing clusters the multi-dimensional distribution of PLVs into a single tree of nested families. The families that are likely to contain true negatives are excluded hierarchically, reducing the dimensionality. The remaining significant clusters provide a multi-scale representation of the PLV signal implicitly accounting for underlying dependence. ODP testing relies on a test statistic that is designed for multiple testing by explicitly sharing the dependence information from across all the tests, and it optimizes FDR by estimating expected true and false positives. The significance of ODP is tested using  $pFDR = P(null|Z \geq z)$ , which is like the p-value with a reversed conditioning. It can be estimated specific to each observed statistic (q-value) from the points falling in the entire tail-area,  $Z \geq 1.96$ , and hence it is also called global FDR. pFDR is also viewed as the Bayesian posterior probability that a significant test belongs to a true null hypothesis [33, 37]. Bayesian interpretation of FDR serves as a connection between the frequentist and Bayesian theory.

Efron et al. maneuvered the Bayesian concept of the global FDR a bit more intricately to overcome most of the aforementioned problems [31, 37, 38]. It estimates local false discovery rate by approximating a fully hierarchical Bayesian model, in which the priors are common for all tests and estimated from the observed data assuming empirical or theoretical null distribution and thus the dependence structure is incorporated. The local FDR is computed as the posterior probability that a hypothesis is null given an observed statistic,  $P(null|Z = z)$ . The large-scale testing leverages this local FDR inference for any observed statistic, which helps in judging the significance of the observed statistic on its own without referring to the tail area, e.g.,  $Z \geq 1.6$ . This approach allows us to frame better questions and confront the significance issue in a direct manner rather than by relying on the indirect reasoning that lies at the core of all p-valued methods.

The locFDR procedure was developed for microarray data analysis involving thousands of genes tested for multiple strain effects that result in millions of simultaneous tests. As high-dimensional PLV data acquired from EEG shares similar issues of large-scale testing, where the distribution of synchrony measures depends on time, frequency, and spatial location, we sought to investigate whether Efron's locFDR method can serve as an effective tool for testing the significance of PLV inference. The procedure works well as we have seen both in the experimental data and simulation data examples. The locFDR inference with experimental data was superior than FDRBH, hFDR and ODP while conventionally controlling FDR at 5% cutoff; it detected more significant pairs than hFDR, whereas FDRBH and ODP did not detect any pair. At a more lenient cutoff of 10%, the Empirical Bayesian inference converges with the results from frequentist hFDR and ODP methods in the 22–34 Hz frequency band and 300–550 ms post stimulus time interval for the inefficient than efficient condition (Fig 4). The results from FDRBH at 10% cutoff showed only 10 pairs. The practice of setting Type I error threshold at 5% dates back to classic single testing, specifically using family-wise error rate (FWER) controlling procedures. For FDR, there are no such standards, though most existing FDR controlling

methods impose the need to fix an acceptable FDR level before any data are seen. The locFDR quantity itself is a measure of significance of Type I error control, and should be reported for each interesting signal concerning the experimental hypothesis. Researchers may use their discretion to select a threshold, which is appropriate for their research domain, e.g., for microarray inference, 10% or 20% thresholds are acceptable [37]. The FDR curve that shows power of the test as a function of FDR (Fig 2) can serve as a reference for a reasonable choice.

In the simulation analysis, locFDR detected more discoveries than ODP when the proportion of discoveries was small, at  $\pi_1 = .08$  and  $\pi_1 = .16$ . This situation typically describes our research problem with a high dimensional PLV data, where very few pairs are expected to contain the interesting or significant effects. Recall here that  $\pi_1$  was estimated to be 0.1 for the experiment data. Although we did not examine the robustness of each method against PLV strength directly (by varying magnitude of PLV for designated true positives across different simulations), we effectively evaluate robustness by varying the number (or ratio) of true positives, etc. (see Fig 1), which provides an indication of robustness. The reasoning is that if the strength of PLV is decreased, then (naturally) we expect fewer (uncorrected) detections, on which the correction methods are based, since a detection is obtained by the underlying ( $z$ ) statistic. Hence, the robustness (of the correction method) can be simply evaluated by varying the number (ratio) of true positives, etc., as shown in the figure.

As locFDR is computed empirically from the null and mixture densities, it accommodates both asymmetric and symmetric distributions equally effectively. In contrast, the performance of ODP depends on the homogeneity of the dependence structure among hypotheses, and thus it is expected to perform better with asymmetric data. Asymmetry is implied when most PLV effects are in a particular direction, i.e. they are either mostly positive or mostly negative, which indeed seems to be the case with our experimental data that showed more significant PLV in inefficient-efficient contrast than in efficient-inefficient contrast (Fig 4). The simulated data was also asymmetric (with greater proportion of positive PLV effects than negative PLV effects) as it was generated using the parameters obtained from a real PLV dataset. For a comparison, we ran the simulation with an induced symmetry by assuming both positive and negative PLV effects in similar proportions. We observed that symmetry reduced the number of detections by ODP considerably when the proportion of true discoveries was smaller than 25%. Symmetry did not affect the performance of locFDR and hFDR (results not shown here).

In the experiment analysis, hFDR, ODP, and locFDR offered almost similar qualitative results. While the inference obtained from all the three methods are comparable, each method has its own advantages. hFDR approach is more amenable for datasets that are naturally hierarchical and is particularly powerful when the synchrony effects are concentrated in a few families of hypotheses. However, the PLV effects may not be bound to a given dimension (or a few families). In such cases, ODP and locFDR may be more powerful, both of which explicitly account for the inherent dependence structure by sharing the common information from all the tests. They both assign a direct significance measure of FDR to each of the test, eliminating the need to determine a cutoff. Their disadvantage over hFDR is that they work well only with large number of hypotheses (at least several hundred). The local FDR is more accurate than ODP for estimating FDR. The significance in ODP is determined using global FDR from the entire tail area, which gives an aggregated estimate. The locFDR procedure estimates the local FDR, which is determined for each test using the specific point in the tail, and it is more accurate. ODP requires computing a new statistic, which restricts its application. In comparison, locFDR is more widely applicable, e.g., to normal, chi-square, t-static, and F distributions, which are commonly used in our research domain [38, 39]. If the distribution is uncertain, the null can be defined empirically from the data. In this article, we have assumed the normal distribution, which holds reasonably well in the case of normalized PLVs [28, 40], and hence it

can be tested with z-test or t-test. In the case of un-normalized absolute PLVs, the Gaussian distribution is unlikely to hold as they are bounded between 0 and 1. According to a recent study [11], von Mises and circularly symmetric Gaussian distributions are adequate for absolute PLV data and can be tested by Chi-square statistic for goodness of fit of these distributions.

Recent research in brain dynamics has established a number of methods for quantifying functional connectivity in EEG studies (see [1, 3] for a review of some of these methods). However, no standard method exists to justify the statistical significance of the connectivity inference from these methods, specifically when there is no a priori knowledge about the expected region of synchrony. Therefore, our article, which illustrates locFDR method for computing the false discovery rate associated with each PLV test, would benefit functional connectivity research.

While the Bayesian inference for EEG synchrony is available in the dynamic causal model of SPM [21], it eschews the frequentist multiple comparison issue. Friston et al. suggest that reporting of the posterior probability of activation is sufficient for Bayesian inference and to avoid the frequentist multiple comparison issue [20]. However, if we want to control wrong labeling of voxels (or pairs of EEG signals in our problem) across tests, the multiple comparison correction is still necessary on Bayesian posterior probabilities [41]. The proposed Empirical Bayes method brings together both the frequentist and Bayesian approaches, in the sense that we can report and control the estimated local FDR at each test, which is defined as the posterior probability that a test belongs to null group, and controlling it. The locFDR approach is readily applicable to the other available methods for measuring synchrony, e.g., mutual information, generalized synchronization [42], single-trial phase locking [43], structural synchrony [13], empirical mode detection PLV [44], phase resetting [45, 46], a method based on Cohens class of time-frequency distributions [47], and a recently published graph partitioning method for modeling brain connectivity [48].

## Acknowledgments

We thank Shigeyuki Oba for sharing his views on Local FDR and ODP methods.

## Author Contributions

Analyzed the data: AKS YT SP. Contributed reagents/materials/analysis tools: AKS HA. Wrote the paper: AKS SP YT HA.

## References

1. David O, Cosmelli D, Friston KJ. Evaluation of different measures of functional connectivity using a neural mass model. *NeuroImage*. 2004; 21(2):659–673. doi: [10.1016/j.neuroimage.2003.10.006](https://doi.org/10.1016/j.neuroimage.2003.10.006) PMID: [14980568](https://pubmed.ncbi.nlm.nih.gov/14980568/)
2. Schevon CA, Cappell J, Emerson R, Isler J, Grieve P, Goodman R, et al. Cortical abnormalities in epilepsy revealed by local EEG synchrony. *NeuroImage*. 2007; 35(1):140–148. doi: [10.1016/j.neuroimage.2006.11.009](https://doi.org/10.1016/j.neuroimage.2006.11.009) PMID: [17224281](https://pubmed.ncbi.nlm.nih.gov/17224281/)
3. Dauwels J, Vialatte F, Musha T, Cichocki A. A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. *NeuroImage*. 2010; 49(1):668–693. doi: [10.1016/j.neuroimage.2009.06.056](https://doi.org/10.1016/j.neuroimage.2009.06.056) PMID: [19573607](https://pubmed.ncbi.nlm.nih.gov/19573607/)
4. Chialvo DR. Emergent complex neural dynamics. *Nature Physics*. 2010; 6:744–750. doi: [10.1038/nphys1803](https://doi.org/10.1038/nphys1803)
5. Ponten SC, Daffertshofer A, Hillebrand A, Stam CJ. The relationship between structural and functional connectivity: Graph theoretical analysis of an EEG neural mass model. *NeuroImage*. 2010; 52(3):985–994. doi: [10.1016/j.neuroimage.2009.10.049](https://doi.org/10.1016/j.neuroimage.2009.10.049) PMID: [19853665](https://pubmed.ncbi.nlm.nih.gov/19853665/)

6. Deshpande G, Santhanam P, Hu X. Instantaneous and causal connectivity in resting state brain networks derived from functional MRI data. *NeuroImage*. 2011; 54(2):1043–1052. doi: [10.1016/j.neuroimage.2010.09.024](https://doi.org/10.1016/j.neuroimage.2010.09.024) PMID: [20850549](https://pubmed.ncbi.nlm.nih.gov/20850549/)
7. Haufe S, Nikulin VV, Müller KR, Nolte G. A critical assessment of connectivity measures for EEG data: A simulation study. *NeuroImage*. 2013; 64(0):120–133. doi: [10.1016/j.neuroimage.2012.09.036](https://doi.org/10.1016/j.neuroimage.2012.09.036) PMID: [23006806](https://pubmed.ncbi.nlm.nih.gov/23006806/)
8. Buschman TJ, Miller EK. Top-down versus bottom-up control of attention in the prtal and posterior parietal cortices. *Science*. 2007; 315(5820):1860–1862. doi: [10.1126/science.1138071](https://doi.org/10.1126/science.1138071) PMID: [17395832](https://pubmed.ncbi.nlm.nih.gov/17395832/)
9. Phillips S, Takeda Y. Greater frontal-parietal synchrony at low gamma-band frequencies for inefficient than efficient visual search in human EEG. *International Journal of Psychophysiology*. 2009; 75(1):39–43. doi: [10.1016/j.ijpsycho.2009.11.001](https://doi.org/10.1016/j.ijpsycho.2009.11.001) PMID: [19903501](https://pubmed.ncbi.nlm.nih.gov/19903501/)
10. Lachaux J, Rodrigue E, Martnerie J, Varela FJ. Measuring phase synchrony in brain signals. *Human brain mapping*. 1999; 8(4):194–208. doi: [10.1002/\(SICI\)1097-0193\(1999\)8:4%3C194::AID-HBM4%3E3.0.CO;2-C](https://doi.org/10.1002/(SICI)1097-0193(1999)8:4%3C194::AID-HBM4%3E3.0.CO;2-C) PMID: [10619414](https://pubmed.ncbi.nlm.nih.gov/10619414/)
11. Aydore S, Pantazis D, Leahy RM. A note on the phase locking value and its properties. *NeuroImage*. 2013; 74(0):231–244. doi: [10.1016/j.neuroimage.2013.02.008](https://doi.org/10.1016/j.neuroimage.2013.02.008) PMID: [23435210](https://pubmed.ncbi.nlm.nih.gov/23435210/)
12. Bhattacharya J, Petsche H. Phase synchrony analysis of EEG during music perception reveals changes in functional connectivity due to musical expertise. *Signal Processing*. 2005; 85(11):2161–2177. doi: [10.1016/j.sigpro.2005.07.007](https://doi.org/10.1016/j.sigpro.2005.07.007)
13. Fingelkurts A, Fingelkurts A, Krause C, Kaplan A, Borisov S, Sams M. Structural (operational) synchrony of EEG alpha activity during an auditory memory task. *NeuroImage*. 2003; 20(1):529–542. doi: [10.1016/S1053-8119\(03\)00305-7](https://doi.org/10.1016/S1053-8119(03)00305-7) PMID: [14527613](https://pubmed.ncbi.nlm.nih.gov/14527613/)
14. Kitajo K, Doesburg M, Yamanaka K, Nozaki, Ward LM, Yamamoto Y. Effects of Parkinson's disease on brain-wave phase synchronisation and cross-modulation. *EPL (Europhysics Letters)*. 2010; 89(4):48001. doi: [10.1209/0295-5075/89/48001](https://doi.org/10.1209/0295-5075/89/48001)
15. Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*. 2007; 164(1):177–190. doi: [10.1016/j.jneumeth.2007.03.024](https://doi.org/10.1016/j.jneumeth.2007.03.024) PMID: [17517438](https://pubmed.ncbi.nlm.nih.gov/17517438/)
16. Chumbley JR, Friston KJ. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *NeuroImage*. 2009; 44(1):62–70. doi: [10.1016/j.neuroimage.2008.05.021](https://doi.org/10.1016/j.neuroimage.2008.05.021) PMID: [18603449](https://pubmed.ncbi.nlm.nih.gov/18603449/)
17. Singh AK, Phillips S. Hierarchical control of false discovery rate for phase locking measures of EEG synchrony. *NeuroImage*. 2010; 50(1):40–47. doi: [10.1016/j.neuroimage.2009.12.030](https://doi.org/10.1016/j.neuroimage.2009.12.030) PMID: [20006711](https://pubmed.ncbi.nlm.nih.gov/20006711/)
18. Krieg, J, Fonseca, ATD, Martínez-Montes, E, Marquis, P, Liégeois-Chauvel, C, Bénar, CG. A comparison of methods for assessing alpha phase resetting in electrophysiology, with application to intracerebral EEG in visual areas. *NeuroImage*. 2011; In Press, Corrected Proof:–.
19. Daly I, Sweeney-Reed C, Nasuto S. Testing for significance of phase synchronisation dynamics in the EEG. *Journal of Computational Neuroscience*. 2013; 34(3):411–432. doi: [10.1007/s10827-012-0428-2](https://doi.org/10.1007/s10827-012-0428-2) PMID: [23104010](https://pubmed.ncbi.nlm.nih.gov/23104010/)
20. Friston KJ, Glaser DE, Henson RNA, Kiebel S, Phillips C, Ashburner J. Classical and Bayesian inference in neuroimaging: applications. *NeuroImage*. 2002;p. 484–512.
21. Litvak V, Mattout J, Kiebel SJ, Phillips C, Henson RNA, Kilner J, et al. EEG and MEG data analysis in SPM8. *Comput Intell Neurosci*. 2011; 2011(852961):32.
22. Worsley KJ, Evans AC, Marrett S, Neelin P. A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*. 1992; 12(6):900–918. doi: [10.1038/jcbfm.1992.127](https://doi.org/10.1038/jcbfm.1992.127) PMID: [1400644](https://pubmed.ncbi.nlm.nih.gov/1400644/)
23. Chumbley J, Worsley KJ, Flandin G, Friston KJ. Topological FDR for neuroimaging. *NeuroImage*. 2010; 49(4):3057–3064. doi: [10.1016/j.neuroimage.2009.10.090](https://doi.org/10.1016/j.neuroimage.2009.10.090) PMID: [19944173](https://pubmed.ncbi.nlm.nih.gov/19944173/)
24. Bullmore ET, Suckling J, Overmeyer S, Rabe-Hesketh S, Taylor E, Brammer MJ. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE transactions on medical imaging*. 1999; 18(1):32–42. doi: [10.1109/42.750253](https://doi.org/10.1109/42.750253) PMID: [10193695](https://pubmed.ncbi.nlm.nih.gov/10193695/)
25. Kilner J, Kiebel SJ, Friston KJ. Applications of random field theory to electrophysiology. *Neuro-science Letters*. 2005; 374:174–178. doi: [10.1016/j.neulet.2004.10.052](https://doi.org/10.1016/j.neulet.2004.10.052)
26. Kilner J, Friston KJ. Topological Inference for EEG and MEG. *The Annals of Applied Statistics*. 2010; 4(3):1272–1290. doi: [10.1214/10-AOAS337](https://doi.org/10.1214/10-AOAS337)
27. Kiebel SJ, Friston KJ. Statistical Parametric Mapping for Event-Related Potentials I: Generic Considerations. *NeuroImage*. 2004; 22(2):492–502. doi: [10.1016/j.neuroimage.2004.02.013](https://doi.org/10.1016/j.neuroimage.2004.02.013) PMID: [15193578](https://pubmed.ncbi.nlm.nih.gov/15193578/)



28. Singh AK, Asoh H, Phillips S. Optimal detection of functional connectivity from high-dimensional EEG synchrony data. *NeuroImage*. 2011; 58(1):148–156. doi: [10.1016/j.neuroimage.2011.05.082](https://doi.org/10.1016/j.neuroimage.2011.05.082) PMID: [21704709](https://pubmed.ncbi.nlm.nih.gov/21704709/)
29. Yekutieli D. Hierarchical False Discovery Rate Controlling Methodology. *Journal of the American Statistical Association*. 2008; 103(481):309–316. doi: [10.1198/016214507000001373](https://doi.org/10.1198/016214507000001373)
30. Storey JD. The optimal discovery procedure: a new approach to simultaneous significance testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2007; 69(3):347–368. doi: [10.1111/j.1467-9868.2007.005592.x](https://doi.org/10.1111/j.1467-9868.2007.005592.x)
31. Efron B, Tibshirani R, Storey JD, Tusher V. Empirical Bayes Analysis of a Microarray Experiment. *Journal of the American Statistical Association*. 2001; 96(456):1151–1160. doi: [10.1198/016214501753382129](https://doi.org/10.1198/016214501753382129)
32. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B*. 1995; 57(1):289–300.
33. Storey JD. The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics*. 2003; 31:2013–2035. doi: [10.1214/aos/1074290335](https://doi.org/10.1214/aos/1074290335)
34. Storey JD, Dai JY, Leek JT. The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. *Biostatistics*. 2007; 8(2):414–432. doi: [10.1093/biostatistics/kxl019](https://doi.org/10.1093/biostatistics/kxl019) PMID: [16928955](https://pubmed.ncbi.nlm.nih.gov/16928955/)
35. Storey JD. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002; 64(3):479–498. doi: [10.1111/1467-9868.00346](https://doi.org/10.1111/1467-9868.00346)
36. Heller R, Stanley D, Yekutieli D, Rubin N, Benjamini Y. Cluster-based analysis of fMRI data. *NeuroImage*. 2006; 33(2):599–608. doi: [10.1016/j.neuroimage.2006.04.233](https://doi.org/10.1016/j.neuroimage.2006.04.233) PMID: [16952467](https://pubmed.ncbi.nlm.nih.gov/16952467/)
37. Efron B, Tibshirani R. Empirical bayes methods and false discovery rates for microarrays. *Genetic Epidemiology*. 2002; 23(1):70–76. doi: [10.1002/gepi.1124](https://doi.org/10.1002/gepi.1124) PMID: [12112249](https://pubmed.ncbi.nlm.nih.gov/12112249/)
38. Efron B. Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. *Journal of the American Statistical Association*. 2004; 99(465):96–104. doi: [10.1198/016214504000000089](https://doi.org/10.1198/016214504000000089)
39. Schwartzman A, Dougherty RF, Lee J, Ghahremani D, Taylor JE. Empirical null and false discovery rate analysis in neuroimaging. *NeuroImage*. 2009; 44(1):71–82. doi: [10.1016/j.neuroimage.2008.04.182](https://doi.org/10.1016/j.neuroimage.2008.04.182) PMID: [18547821](https://pubmed.ncbi.nlm.nih.gov/18547821/)
40. Doesburg SM, Roggeveen AB, Kitajo K, Ward LM. Large-Scale Gamma-Band Phase Synchronization and Selective Attention. *Cerebral Cortex*. 2008; 18(2):386–396. doi: [10.1093/cercor/bhm073](https://doi.org/10.1093/cercor/bhm073) PMID: [17556771](https://pubmed.ncbi.nlm.nih.gov/17556771/)
41. Woolrich MW. Bayesian inference in fMRI. *NeuroImage*. 2012; 62(2):801–810. doi: [10.1016/j.neuroimage.2011.10.047](https://doi.org/10.1016/j.neuroimage.2011.10.047) PMID: [22063092](https://pubmed.ncbi.nlm.nih.gov/22063092/)
42. Quiñero Quiroga R, Kraskov A, Kreuz T, Grassberger P. Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Phys Rev E*. 2002; 65(4):041903. doi: [10.1103/PhysRevE.65.041903](https://doi.org/10.1103/PhysRevE.65.041903)
43. Lachaux JP, Rodriguez E, Quyen MLV, Lutz A, Martinerie J, Varela FJ. Studying single-trials of phase synchronous activity in the brain. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*. 2000; 10(10):2429–2439. doi: [10.1016/S0218-1274\(00\)00156-0](https://doi.org/10.1016/S0218-1274(00)00156-0)
44. Sweeney-Reed C, Nasuto S. A novel approach to the detection of synchronisation in EEG based on empirical mode decomposition. *Journal of Computational Neuroscience*. 2007; 23:79–111. doi: [10.1007/s10827-007-0020-3](https://doi.org/10.1007/s10827-007-0020-3) PMID: [17273939](https://pubmed.ncbi.nlm.nih.gov/17273939/)
45. Makinen V, Tiitinen H, May P. Auditory event-related responses are generated independently of ongoing brain activity. *NeuroImage*. 2005; 24(4):961–968. doi: [10.1016/j.neuroimage.2004.10.020](https://doi.org/10.1016/j.neuroimage.2004.10.020) PMID: [15670673](https://pubmed.ncbi.nlm.nih.gov/15670673/)
46. Thatcher RW, North DM, Biver CJ. Intelligence and EEG phase reset: A two compartmental model of phase shift and lock. *NeuroImage*. 2008; 42(4):1639–1653. doi: [10.1016/j.neuroimage.2008.06.009](https://doi.org/10.1016/j.neuroimage.2008.06.009) PMID: [18620065](https://pubmed.ncbi.nlm.nih.gov/18620065/)
47. Aviyente S, Bernat EM, Evans WS, Sponheim SR. A phase synchrony measure for quantifying dynamic functional integration in the brain. *Human Brain Mapping*. 2011; 32(1):80–93. doi: [10.1002/hbm.21000](https://doi.org/10.1002/hbm.21000) PMID: [20336687](https://pubmed.ncbi.nlm.nih.gov/20336687/)
48. Chang YT, Pantazis D, Leahy RM. To cut or not to cut? Assessing the modular structure of brain networks. *NeuroImage*. 2014; 91(0):99–108. doi: [10.1016/j.neuroimage.2014.01.010](https://doi.org/10.1016/j.neuroimage.2014.01.010) PMID: [24440531](https://pubmed.ncbi.nlm.nih.gov/24440531/)