
Inference of Whole Brain Electrophysiological Networks Through Multimodal Integration of Simultaneous Scalp and Intracranial EEG

Shihao Yang

Department of Systems Engineering
Stevens Institute of Technology
Hoboken, NJ 07030
syang57@stevens.edu

Feng Liu*

Department of Systems Engineering
Stevens Institute of Technology
Hoboken, NJ 07030
fliu22@stevens.edu

Abstract

Brain imaging research has transitioned over the past decades from identifying isolated regions of task-evoked activation to characterizing the spatiotemporal dynamics of large-scale brain networks. Electrophysiological signals are the direct manifestation of brain activity; thus, characterizing whole-brain electrophysiological networks (WBEN) can serve as a fundamental tool for neuroscience studies and clinical applications. In this work, we introduce a framework for integrating scalp EEG and intracranial EEG (iEEG) for WBEN estimation through a principled state-space modeling approach, where an Expectation-Maximization (EM) algorithm is designed to infer the state variables and brain connectivity simultaneously. We validated the proposed method on synthetic data, and the results revealed improved performance compared to traditional two-step methods using scalp EEG only, demonstrating the importance of including iEEG signals for WBEN estimation. For real data with simultaneous EEG and iEEG, we applied the developed framework to understand the information flows during encoding and maintenance phases of a working memory task. The information flows between subcortical and cortical regions are delineated, highlighting more significant information flows from cortical to subcortical regions during encoding than during maintenance. The results are consistent with previous research findings, but from a whole-brain perspective, which underscores the unique utility of the proposed framework.

1 Introduction

Brain networks represent the intricate and dynamic connectivity of neurons that facilitates communication across different brain regions. These networks are essential for supporting cognitive functions, from basic sensory processing to complex decision-making [1, 2]. Existing studies have suggested that accurately inferred brain connectivity patterns can help gain insights into the coordination and interactions between different brain regions [3, 4], reveal the brain network underpinnings of cognitive processes, and uncover the mechanisms and biomarkers of neuropsychiatric diseases [5, 6]. Over the past decades, functional Magnetic Resonance Imaging (fMRI) has been the most widely used brain imaging modality for functional brain network modeling and analysis [7]. By analyzing fMRI data, researchers can explore the connectomes of the human brain across different cognitive tasks [8, 9], consciousness states [10], degenerative diseases [11], and mental disorders [12]. However, fMRI has limitations, including non-portability and low temporal resolution, which restrict its use in applications that require characterization of instantaneous brain activity at sub-millisecond timescales [13].

*Corresponding author: Dr. Feng Liu, Email: fliu22@stevens.edu

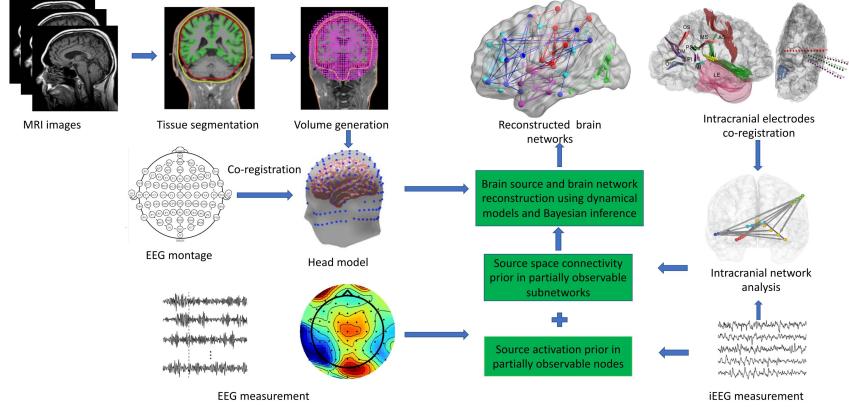


Figure 1: The overall pipeline of integration of EEG and iEEG for brain network reconstruction

Neuroimaging techniques with high temporal resolution, such as Magnetoencephalography (MEG) and Electroencephalography (EEG), can be used to measure electrical and magnetic brain signals, offering a direct measurement of brain activity rather than metabolic signals. To establish the electrophysiological connectome of the human brain, inferring whole-brain electrophysiological networks (WBEN) is important since it provides a direct network-level delineation of brain connectivity. Existing studies that use MEG/EEG to reconstruct brain electrophysiological networks have typically adopted a two-step procedure, with EEG/MEG source imaging (ESI; Step 1) followed by brain connectivity measures (Step 2), such as phase–amplitude coupling [14], coherence [15], phase synchronization [16], and Granger causality [17].

Existing ESI frameworks based on brain source localization suffer from low accuracy in estimating whole-brain activity across regions due to the ill-posedness of the inverse problem and are theoretically limited by the Restricted Isometry Property (RIP, the definition provided in Appendix A.1) [18]. While a number of recent works have sought to alleviate these limitations through refined spatial priors and filtering strategies [19–21], the fundamental challenges remain significant. To address the limitations of two-step approaches, Yang et al. proposed a one-step state-space model that jointly estimates source localization and dynamic connectivity by modeling ROI mean activities with time-varying autoregression [22]. Pirondini et al. developed computationally efficient algorithms combining spatial covariance estimation, linear state-space dynamics, and sparsity constraints, achieving improved source localization performance with significant reduction in computation time through steady-state Kalman filtering [23]. However, estimating whole-brain networks poses additional challenges for the ESI problem, as it further relaxes sparsity constraints [24]. More recently, Soleimani et al. demonstrated that Granger causal links can be directly inferred from MEG measurements without an intermediate source localization step, achieving superior performance with low false alarm rates through integrated parameter estimation and statistical analysis [25]. Additionally, Sanchez-Bornot et al. introduced multiple penalized state-space models with novel algorithms based on backpropagation, gradient descent, and alternating least squares, enabling simultaneous solution of source localization and functional connectivity problems for thousands of cortical sources using data-driven regularization [26].

In addition to non-invasive modalities, invasive neuroimaging technologies such as intracranial Electroencephalography (iEEG), including Electrocorticography (ECoG) and stereoelectroencephalography (sEEG), which place subdural electrodes on the brain surface (ECoG) or penetrating electrodes in subcortical regions (sEEG), can achieve more accurate connectivity mapping among different regions of interest with high temporal resolution [27, 28]. Recent studies have leveraged simultaneous scalp EEG and iEEG recordings to improve the reliability of electrophysiological source imaging. For example, Jiao et al. proposed an explainable deep learning framework (XDL-ESI) that unrolls optimization algorithms into neural networks and achieves accurate, interpretable source localization validated on simultaneous EEG-iEEG data [29]. Despite these advances, iEEG recordings remain limited in spatial coverage due to their invasive nature, making the brain only partially observable through iEEG measurement. As iEEG electrodes can only cover part of the brain, important neural activity or connectivity patterns might be undetected in other regions. For example, in the analysis of seizure onset zones in epilepsy, Shu et al. used MEG/EEG source imaging and identified interictal spikes that are missed by iEEG [30], highlighting the value of synergy between scalp EEG and inva

sive iEEG. The partial observability of iEEG and the challenges of localizing networked brain sources from EEG motivate the integration of scalp EEG and iEEG as complementary modalities. To address the above issues, we propose to integrate scalp EEG and iEEG to provide a accurate delineation of electrophysiological activation and connectivity at the whole-brain scale. Integrating iEEG and scalp EEG can yield a faithful reconstruction of WBEN, however, a principled multimodal integration modeling and inference framework has not been explored before. Early work leveraging state-space models solving the ESI and brain networks demonstrated significant potential [22, 25, 31]. In this paper, we propose a new inference framework based on state-space dynamical systems and Bayesian inference that leverages multimodal fusion of scalp EEG and iEEG for WBEN estimation. This work represents a new and unified computational paradigm that integrates scalp EEG and intracranial recordings for whole-brain network inference, treating both modalities as complementary observations of shared underlying neural dynamics within a rigorous Bayesian state-space paradigm. This framework enables comprehensive characterization of whole-brain electrophysiological networks when simultaneous recordings are available. The pipeline of the proposed approach is illustrated in Fig. 1.

2 Method

2.1 Basic problem definition

The linear discrete dynamic system of brain sources, as well as the linear model of EEG and iEEG observations, can be defined as:

$$\begin{aligned} x_t &= \sum_{k=1}^K \Phi_k x_{t-k} + n_t, \quad t = 1, \dots, T, \quad x_t \in \mathbf{R}^N, \\ y_t &= L x_t + w_t, \quad y_t \in \mathbf{R}^M, \\ z_t &= C x_t + e_t, \quad z_t \in \mathbf{R}^O, \end{aligned} \quad (1)$$

where N , M , and O are the number of the source regions, EEG electrodes, and iEEG electrodes, respectively. $\Phi_k \in \mathbf{R}^{N \times N}$ is the state transition matrix that delineates the impact of the source state at time $t - k$ to t . $n_t \in \mathbf{R}^N$, $n_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ is the noise in source state space which is assumed to be a multivariate Gaussian distribution with mean 0 and diagonal covariance matrix \mathbf{Q} . $L \in \mathbf{R}^{M \times N}$ is the lead field matrix. $w_t \in \mathbf{R}^M$, $w_t \sim \mathcal{N}(\mathbf{0}, \mathbf{P})$ is the measurement noise in EEG observation which is also assumed to be a multivariate Gaussian distribution with mean 0 and a covariance matrix \mathbf{P} that is assumed to be known by measuring on a realistic head model. And $C \in \mathbf{R}^{O \times N}$ is a full-row rank transformation matrix that selects the source signal where its region can be observed by iEEG directly. $e_t \in \mathbf{R}^O$, $e_t \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$ is the iEEG observational noise which is assumed to follow multivariate Gaussian distribution with mean 0 and a covariance matrix \mathbf{S} that is also can be measured in a similar manner as \mathbf{P} . Thus, according to the model definition, the parameters that need to be estimated are Φ and \mathbf{Q} . Define the unknown parameters as $\theta = \{\Phi, \mathbf{Q}\}$. Then, the log-likelihood can be written in the form:

$$\log p(y_{1:T}, z_{1:T} | \theta), \quad (2)$$

The log-likelihood of the model can be defined as

$$\log p(y_{1:T}, z_{1:T} | \theta), \quad \theta = \{\Phi, \mathbf{Q}\}. \quad (3)$$

Since the number of observations substantially exceeds the number of sources, the inverse estimation problem is highly ill-posed. To alleviate this phenomenon and simplify the problem, a regularization term was added based on the assumption that the connection from all other regions to a given region is sparse. In this case, the regularized maximum log-likelihood of parameters for the model can be defined as :

$$\theta^* = \arg \max_{\theta} \log p(y_{1:T}, z_{1:T} | \theta) - \lambda \sum_{n=1}^N \|\Phi_{n,:}\|_1, \quad (4)$$

where $\Phi_{n,:}$ is the n^{th} row of the state transition matrix. λ is a regularization weight for model estimation that can be decided manually according to the experience or by grid search [25].

2.2 An Expectation-Maximization estimation framework

In the domain of statistical methodology, the Expectation–Maximization (EM) algorithm serves as an iterative computational approach employed to determine the maximum likelihood, whether it pertains to local optima or the maximum of posterior estimations (MAP) of parameters in the context of a statistical model. These models, in particular, depend on latent variables hidden from direct observations, underscoring the importance of this sophisticated approach. The EM is characterized by iterative execution, with two pivotal steps: firstly, the Expectation (E) step, where a Q-function is calculated to encapsulate the expected value of the log-likelihood. Secondly, the Maximization (M) step, during which optimal parameters are derived to maximize the anticipated log-likelihood obtained during the E step. Consequently, these derived parameters play an important role in illustrating the distributional characteristics of the latent variables to prepare for the subsequent E step within the iterative loop. Since the data distribution in the source domain is unknown, the Expectation-Maximization framework takes the source state as the latent variable, which can be applied to find the optimal estimation of the log-likelihood and then obtain the approximated state in the source domain with the problem defined above.

We first illustrate the E-step by starting from Eq.(2) and using the facts that the observations of EEG and iEEG are conditional independent on the source state x . The log-likelihood can be rewritten and derived by introducing source state $x_{1:T}$ as

$$\begin{aligned} \log p(y_{1:T}, z_{1:T}, x_{1:T} | \theta) &= \log p(y_{1:T}, z_{1:T} | x_{1:T}, \theta) + \log p(x_{1:T} | \theta) \\ &= \log p(y_{1:T} | x_{1:T}, \theta) + \log p(z_{1:T} | x_{1:T}, \theta) + \log p(x_{1:T} | \theta). \end{aligned} \quad (5)$$

The first and second terms of the right-hand side of the second row can be easily obtained based on the Gaussian noise assumption while source x_t is given which are

$$\log p(y_{1:T} | x_{1:T}, \theta) = \sum_{t=1}^T \log p(y_t | x_t, \theta) = -\frac{T}{2} \log(2\pi|\mathbf{P}|) - \frac{1}{2} \sum_{t=1}^T \|y_t - Lx_t\|_{\mathbf{P}^{-1}}, \quad (6)$$

and

$$\log p(z_{1:T} | x_{1:T}, \theta) = \sum_{t=1}^T \log p(z_t | x_t, \theta) = -\frac{T}{2} \log(2\pi|\mathbf{S}|) - \frac{1}{2} \sum_{t=1}^T \|z_t - Cx_t\|_{\mathbf{S}^{-1}}, \quad (7)$$

where $|\cdot|$ is the matrix determinant, and $\|v\|_W = v^T W v$ is the quadratic form in the exponential term of multivariate Gaussian distribution. The last term can be obtained based on the linear dynamical model from (1), and utilize the presumption that \mathbf{Q} is a diagonal covariance matrix with the items $\{\sigma_i^2, i = 1, \dots, N\}$ on its diagonal, then we have

$$\log p(x_{1:T} | \theta) = \sum_{t=1}^T \log p(x_t | \theta) = -\frac{T}{2} \log(2\pi|\mathbf{Q}|) - \frac{1}{2} \sum_{i=1}^M \frac{1}{\sigma_i^2} \|\hat{\mathbf{x}}_i - \mathbf{X}\hat{\phi}_i\|_2^2, \quad (8)$$

where $\hat{\mathbf{x}}_i = [x_{i,K+1}, x_{i,K+2}, \dots, x_{i,T}]^T$, $\mathbf{X} = [[x_{1,K+1}, x_{1,K+2}, \dots, x_{1,T}]^T, \dots, [x_{1,1}, x_{1,2}, \dots, x_{1,T-K}]^T, [x_{N,K+1}, x_{N,K+2}, \dots, x_{N,T}]^T, \dots, [x_{N,1}, x_{N,2}, \dots, x_{N,T-K}]^T]$, and $\hat{\phi}_i = [[\Phi_k]_{i,j}]^T, k = 1, \dots, K, j = 1, \dots, N$. By substituting Eqs. (6)-(8) into Eq.(5), the formula can be reformulated, and then take the expectation to get the Q-function for EM as

$$\begin{aligned} \log p(y_{1:T}, z_{1:T}, x_{1:T} | \theta) &= -\frac{T}{2} \log(2\pi|\mathbf{P}|) - \frac{1}{2} \sum_{t=1}^T \|y_t - Lx_t\|_{\mathbf{P}^{-1}} \\ &\quad - \frac{T}{2} \log(2\pi|\mathbf{S}|) - \frac{1}{2} \sum_{t=1}^T \|z_t - Cx_t\|_{\mathbf{S}^{-1}} \\ &\quad - \frac{T}{2} \log(2\pi) - \frac{T}{2} \sum_{i=1}^N \log(\sigma_i^2) - \frac{1}{2} \sum_{i=1}^N \frac{1}{\sigma_i^2} \|\hat{\mathbf{x}}_i - \mathbf{X}\hat{\phi}_i\|_2^2 \\ &= \mathcal{C}_\theta - \frac{T}{2} \sum_{i=1}^N \log(\sigma_i^2) - \frac{1}{2} \sum_{i=1}^N \frac{1}{\sigma_i^2} \|\hat{\mathbf{x}}_i - \mathbf{X}\hat{\phi}_i\|_2^2, \end{aligned} \quad (9)$$

$$\begin{aligned}
\mathcal{Q}(\theta; \theta^{(j)}) &= \mathcal{E}[\log p(y_{1:T}, z_{1:T}, x_{1:T} | \theta) | y_{1:T}, z_{1:T}, \theta^{(j)}] \\
&= \mathcal{C}_{\theta^{(j)}} - \frac{T}{2} \sum_{i=1}^N \log(\sigma_i^{(j)2}) - \frac{1}{2} \sum_{i=1}^N \frac{1}{\sigma_i^{(j)2}} \mathcal{E}[\|\hat{\mathbf{x}}_i^{(j)} - \underline{\mathbf{X}}^{(j)} \hat{\phi}_i^{(j)}\|_2^2] \\
&= \mathcal{C}_{\theta^{(j)}} - \frac{T}{2} \sum_{i=1}^N \log(\sigma_i^{(j)2}) - \frac{1}{2} \sum_{i=1}^N \frac{1}{\sigma_i^{(j)2}} \{ \mathcal{E}[\hat{\mathbf{x}}_i^{(j)T} \hat{\mathbf{x}}_i^{(j)} | y_{1:T}, z_{1:T}, \theta^{(j)}] \\
&\quad - 2\mathcal{E}[\hat{\mathbf{x}}_i^{(j)T} \underline{\mathbf{X}}^{(j)} | y_{1:T}, z_{1:T}, \theta^{(j)}] \hat{\phi}_i^{(j)} + \hat{\phi}_i^{(j)T} \mathcal{E}[\underline{\mathbf{X}}^{(j)T} \underline{\mathbf{X}}^{(j)} | y_{1:T}, z_{1:T}, \theta^{(j)}] \hat{\phi}_i^{(j)} \},
\end{aligned} \tag{10}$$

$$\begin{aligned}
\mathcal{C}_{\theta^{(j)}} &= -\frac{T}{2} \log(2\pi|\mathbf{P}|) - \frac{1}{2} \sum_{t=1}^T \|y_t - Lx_t^{(j)}\|_{\mathbf{P}^{-1}} \\
&\quad - \frac{T}{2} \log(2\pi|\mathbf{S}|) - \frac{1}{2} \sum_{t=1}^T \|z_t - Cx_t^{(j)}\|_{\mathbf{S}^{-1}} - \frac{T}{2} \log(2\pi),
\end{aligned} \tag{11}$$

where the bracket superscript represents the j^{th} iteration in EM, and $\mathcal{C}_{\theta^{(j)}}$ is a constant term when θ is given at j^{th} iteration. We can find that the $p(x_{1:T} | y_{1:T}, z_{1:T}, \theta)$ is also Gaussian due to the Gaussian property on x , y , and z given θ [32]. To find the Q-function, we can permute the equation and notice that the first and third expectation terms in the last row of the Eq.(10) consist of the second-order moment of the density $p(x_{1:T} | y_{1:T}, z_{1:T}, \theta)$ while the second expectation term can be expressed by the first-order moment of $p(x_{1:T} | y_{1:T}, z_{1:T}, \theta)$ whose mean as well as the covariance matrix can be estimated via Fixed Interval Smoothing (FIS), and the details of it will be listed in the next section. In the M-step, the optimal $\theta = \{\Phi, \mathbf{Q}\}$ that maximizes the Q-function defined above should be found. Since the number of observations is far less than that of the source regions, the inverse problem is ill-posed. The l_1 regularization on Φ is introduced based on the premise that the functional connection among a given region and others possesses sparse properties to reduce problem-solving difficulty. Thus, the equation of the maximization can be described with the form

$$\theta^{(j+1)} = \arg \max_{\theta} \mathcal{Q}(\theta; \theta^{(j)}) + \lambda \sum_{i=1}^N \|\Phi_{i,:|\theta}\|_1, \tag{12}$$

which can be efficiently addressed through the implementation of the Fast Adaptive Shrinkage/Thresholding Algorithm (FASTA). At this juncture, the EM)framework for source estimation has been established, providing a statistically rigorous foundation for the subsequent analytical procedures.

2.3 Source density estimation with FIS

Fixed interval smoothing is a statistical technique used in time series analysis and signal processing. It involves retrospectively estimating and improving the values of a time series over fixed time intervals, considering both past and future observations. This method is particularly useful for reducing noise or uncertainty in historical data and obtaining more accurate, smoothed estimates of the underlying trends or states within the time series. The principle of FIS consists of two parts: forward filtering and backward smoothing. The forward filter is executed to derive posterior estimates and covariances up to the given time t . Subsequently, the backward filter is applied to yield prior estimates and covariances, effectively extending the timeline backward to time t or providing a prior perspective in reverse chronology, in other words. Finally, the estimates and covariances derived from both forward and backward filtering at time t are integrated to produce the ultimate estimation of the state and covariance matrix. Recall the main problem in Eq. (1). Considering computing performance factors, the fusion observation can be described by combining y and z . The idea is that EEG and iEEG can be viewed as components of a unified electrophysiological system. Still, there is no correlation between the measuring noise of the two modalities, and the conditional density of merged observation is also a Gaussian. Thus, observation is redefined as

$$y_t^+ = L^+ x_t + w_t^+, \tag{13}$$

where

$$\begin{aligned}
L^+ &= \begin{pmatrix} L \\ C \end{pmatrix}, \quad L^+ \in \mathbf{R}^{(M+O) \times N}, \\
y_t^+ &= \begin{pmatrix} y_t \\ z_t \end{pmatrix}, \quad y_t^+ \in \mathbf{R}^{M+O}, \\
w_t^+ &= \begin{pmatrix} w_t \\ e_t \end{pmatrix}, \quad w_t^+ \sim \mathcal{N}(0, \mathbf{P}^+), \quad w_t^+ \in \mathbf{R}^{(M+O)}, \\
\mathbf{P}^+ &= \begin{pmatrix} \mathbf{P} & \mathbf{0}_{(M \times O)} \\ \mathbf{0}_{(O \times M)} & \mathbf{S} \end{pmatrix}, \quad \mathbf{P}^+ \in \mathbf{R}^{(M+O) \times (M+O)},
\end{aligned} \tag{14}$$

$$\begin{aligned}
\log p(y_{1:T}^+, x_{1:T} | \theta) &= \log p(y_{1:T}^+ | x_{1:T}, \theta) + \log p(x_{1:T} | \theta) \\
&= \mathcal{C}_\theta^+ - \frac{T}{2} \sum_{i=1}^N \log(\sigma_i^2) - \frac{1}{2} \sum_{i=1}^N \frac{1}{\sigma_i^2} \|\hat{\mathbf{x}}_i - \underline{\mathbf{X}}\hat{\phi}_i\|_2^2, \\
\mathcal{C}_\theta^+ &= -\frac{T}{2} \log(2\pi|\mathbf{P}^+|) - \frac{1}{2} \sum_{t=1}^T \|y_t^+ - L^+ x_t^{(j)}\|_{\mathbf{P}^{+1}}^2 - \frac{T}{2} \log(2\pi).
\end{aligned} \tag{15}$$

Then the log-likelihood problem can be transformed to Eq. (15), and the same transformation can be applied to Q-function. The estimation of the mean and the covariance matrix of $p(x_{1:T} | y_{1:T}, z_{1:T}, \theta)$ can be redefined as that of $p(x_{1:T} | y_{1:T}^+, \theta)$. And one can easily find that $p(x_{1:T} | y_{1:T}^+, \theta)$ is a Gaussian, since for two jointly Gaussians, the conditional distribution is also a Gaussian [32]. Then, the FIS can be applied to find the mean and covariance matrix for $p(x_{1:T} | y_{1:T}^+, \theta)$. The forward Kalman filtering can give the estimation on $p(x_t | y_{1:t}^+, \theta), \forall t$, while the backward Kalman smoothing can calculate the $p(x_t | y_{t:T}^+, \theta), \forall t$. The merging of two estimates can generate the final estimate on $p(x_t | y_{1:T}^+, \theta), \forall t$. Next, referring to the estimation framework [25], we start with Vector Auto Regressor (VAR) to generate the initial value for estimation. Since the source state at time t depends on the state of the former K time points, it is necessary to redefine the augmented source state as $\underline{\mathbf{x}}_t = [x_t; x_{t-1}; \dots; x_{t-K+1}], \underline{\mathbf{x}}_t \in \mathbf{R}^{(K+1)N}$ and the augmented dynamic model as (22) to transform VAR(K) problem into a VAR(1) one.

$$\underline{\mathbf{x}}_t = \underline{\Phi} \underline{\mathbf{x}}_{t-1} + \underline{\mathbf{n}}_t, \quad \forall t, \tag{16}$$

where

$$\begin{aligned}
\underline{\Phi} &= \begin{pmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{(K-1)} & \Phi_K \\ \mathbf{I}_{(M \times M)} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(M \times M)} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_{(M \times M)} & \mathbf{0} \end{pmatrix}, \quad \underline{\Phi} \in \mathbf{R}^{(K+1)N \times (K+1)N}, \\
\underline{\mathbf{n}}_t &= \begin{pmatrix} n_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \underline{\mathbf{n}}_t \in \mathbf{R}^{(K+1)N}, \quad \underline{\mathbf{n}}_t \sim \mathcal{N}(0, \underline{\mathbf{Q}}), \\
\underline{\mathbf{Q}} &= \begin{pmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},
\end{aligned} \tag{17}$$

are the augmented state transition matrix and disturbance in source, respectively. And $\underline{\mathbf{Q}}$ is the covariance matrix for the augmented disturbance, which is also diagonal, whose diagonal values are the same variance as \mathbf{Q} and 0 elsewhere. Then, the augmented version of merged observations is

$$y_t^+ = \underline{\mathbf{L}}^+ \underline{\mathbf{x}}_t + w_t^+, \quad \forall t, \tag{18}$$

where

$$\underline{\mathbf{L}}^+ = (\mathbf{L}^+ \quad \mathbf{0}), \quad \underline{\mathbf{L}}^+ \in \mathbf{R}^{M \times (K+1)N}. \tag{19}$$

In this way, the $p(\underline{\mathbf{x}}_t | y_{1:T}^+, \theta)$ can be calculated via FIS if the observations as well as θ are known. Next, according to [33] one can define

$$\begin{aligned}
\underline{\mathbf{x}}_{t_1|t_2} &= \mathcal{E}[\underline{\mathbf{x}}_{t_1}|y_{1:t_2}^+], \\
\boldsymbol{\Sigma}_{t_1|t_2} &= Cov(\underline{\mathbf{x}}_{t_1}|y_{1:t_2}^+), \\
\mathbf{B}_{t_1,t_2|T} &= Cov(\underline{\mathbf{x}}_{t_1}, \underline{\mathbf{x}}_{t_2}|y_{1:T}^+) = \begin{cases} \mathbf{B}_{t_2,t_1|T}^T, & \text{if } t_1 > t_2 \\ \boldsymbol{\Sigma}_{t_1|t_2}, & \text{if } t_1 = t_2 \\ \mathbf{K}_{t_1}^b \mathbf{B}_{t_1+1,t_2|T}^T, & \text{if } t_1 < t_2 \end{cases}, \\
\mathbf{K}_{t_1}^b &= \boldsymbol{\Sigma}_{t_1|t_1} \underline{\Phi}^T \boldsymbol{\Sigma}_{t_1+1|t_1}^{-1},
\end{aligned} \tag{20}$$

as the mean, covariance matrix as well as the cross-covariance matrix for any given t_1 and t_2 . Then, the model can be fitted with the FIS framework. We started with the initial value obtained via VAR in the forward filtering step. Then, for $t = 0, 1, 2, \dots, T-1$ can have

$$\begin{aligned}
\underline{\mathbf{x}}_{t+1|t} &= \underline{\Phi} \underline{\mathbf{x}}_{t|t}, \\
\boldsymbol{\Sigma}_{t+1|t} &= \underline{\Phi} \boldsymbol{\Sigma}_{t|t} \underline{\Phi}^T + \underline{\mathbf{Q}}, \\
\mathbf{K}_{t+1}^f &= \boldsymbol{\Sigma}_{t+1|t} \underline{\mathbf{L}}^{+T} (\underline{\mathbf{L}}^+ \boldsymbol{\Sigma}_{t+1|t} \underline{\mathbf{L}}^{+T} + \mathbf{P}^+)^{-1}, \\
\underline{\mathbf{x}}_{t+1|t+1} &= \underline{\mathbf{x}}_{t+1|t} + \mathbf{K}_{t+1}^f (y_{t+1}^+ - \underline{\mathbf{L}}^+ \underline{\mathbf{x}}_{t+1|t}), \\
\boldsymbol{\Sigma}_{t+1|t+1} &= \boldsymbol{\Sigma}_{t+1|t} - \mathbf{K}_{t+1}^f (\underline{\mathbf{L}}^+ \boldsymbol{\Sigma}_{t+1|t} \underline{\mathbf{L}}^{+T} + \mathbf{P}^+) \mathbf{K}_{t+1}^{fT}.
\end{aligned} \tag{21}$$

Taking the results from the filtering step, we can further do backward smoothing for $t = T-1, T-2, \dots, 0$ as

$$\begin{aligned}
\mathbf{K}_t^b &= \boldsymbol{\Sigma}_{t|t} \underline{\Phi}^T \boldsymbol{\Sigma}_{t+1|t}^{-1}, \\
\underline{\mathbf{x}}_{t|T} &= \underline{\mathbf{x}}_{t|t} + \mathbf{K}_t^b (\underline{\mathbf{x}}_{t+1|T} - \underline{\mathbf{x}}_{t+1|t}), \\
\boldsymbol{\Sigma}_{t|T} &= \boldsymbol{\Sigma}_{t|t} - \mathbf{K}_t^b (\boldsymbol{\Sigma}_{t+1|T} - \boldsymbol{\Sigma}_{t+1|t}) \mathbf{K}_t^{bT}.
\end{aligned} \tag{22}$$

Then, the cross-covariance matrix can be obtained according to (20), and finally, simply extract the first N rows of $\underline{\mathbf{x}}_{t|T}$ and the N^{th} order submatrix of the upper left corner of the matrix $\mathbf{B}_{t_1+1,t_2|T}$ for $\forall t_1, t_2 = 1, 2, \dots, T$, which are exactly the first- and second-order moment of $p(x_t|y_{1:T}^+, \theta), \forall t$, to finalize the E-step. The algorithmic pipeline is described in Appendix A.2.

3 Results

To validate the added value of the integration of scalp EEG and iEEG in estimating the WBEN, we first conducted experiments with simulated data, and then we tested the proposed inference framework on the Sternberg verbal working memory task to explore the connectivity maps between cortical and subcortical regions during the encoding and maintenance phases [34].

3.1 Numerical experiment on synthetic data

Realistic simulated data are generated and several baseline methods are used for comparison. Detailed experimental configurations are provided in the Appendix A.3.

Validation on the added value using simultaneous scalp EEG and iEEG: Firstly, the impact of iEEG coverages of the partially observable brain regions (state variables) on the estimation of brain connectivity was evaluated. In the experiment, the source space state variables are partial observable with a prescribed portion rendered by the iEEG electrodes. The coverage ratio was set to range from 0% to 50% with a stepsize of 10%. The SNR levels for EEG and iEEG observations were set to be -5dB and 30dB, respectively. The network complexity was set with the number of activations as 10 and in-degree as 2. In the full factorial design of experiments, 10 repetitive simulations were conducted.

We use *Acc* and *Sen* (definitions provided in Appendix A.3) to evaluate the network estimation performance. The impact of the percentages of observable variables is given in Fig. 2, accompanied with the numerical results given in Table 1. Not surprisingly, as the proportion of observable state space variables increases, both *Sen* and *Acc* score are increasing. It is worth noting that when the percentage of observable state space variables reaches 30%, there is a significant improvement in both *Acc* and *Sen*, as is shown in Fig. 3. The result shows that the performance of WBEN estimation can be significantly improved when appropriate amount of activated areas are observable from iEEG

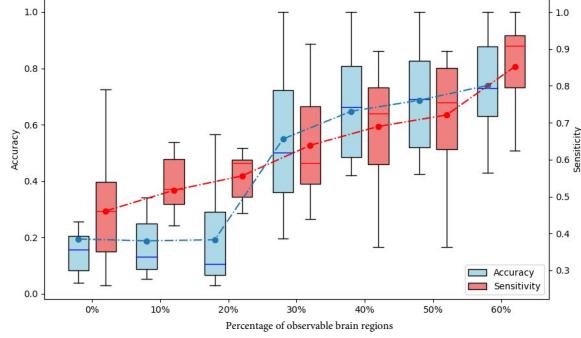


Figure 2: Evaluation of the *Sen* and *Acc* score varied with the percentage of observable brain regions on the WBEN estimation. The curve plots show the mean value of *Acc* and *Sen*.

electrodes, which highlights the value of using simultaneous recordings of scalp EEG and iEEG to characterize the whole brain network dynamics. An example with 30% partially observable brain nodes is shown in Fig. 4. When using two-step methods, due to inaccuracy caused by classical ESI approaches, the over-diffused source estimation results in highly dense brain networks. When using one-step approach with 0% of observable brain regions (without iEEG electrodes) [25], the performance is 0.461 for sensitivity and 0.189 for accuracy respectively, which a significant amount of false positive predictions.

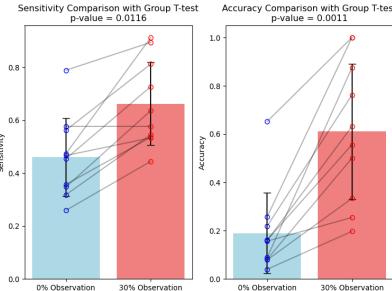


Figure 3: The significance of difference on accuracy and sensitivity between 0% and 30% source observation using group t-test.

Table 1: Evaluation of the percentage of observable brain regions on the WBEN estimation.

Metrics	0%	10%	20%	30%	40%	50%	60%
Sen	0.461 ± 0.155	0.517 ± 0.042	0.556 ± 0.070	0.639 ± 0.120	0.690 ± 0.147	0.721 ± 0.153	0.852 ± 0.131
Acc	0.189 ± 0.176	0.179 ± 0.161	0.180 ± 0.174	0.545 ± 0.224	0.657 ± 0.194	0.687 ± 0.184	0.735 ± 0.201

Impact of SNR on WBEN estimation: The scalp EEG SNR is a key factor for WBEN estimation. Since the SNR in the iEEG signal is known to be significant greater than that of the EEG signal [35], we mainly evaluate the impact of scalp EEG SNR on the WBEN estimation. In the experiment, the SNR of iEEG signal was set to 30dB with different levels of SNR for the scalp EEG signal. It is unknown how the performance of WBEN estimation will be improved when integrating the iEEG into the WBEN inference framework with a high level of noise presence in scalp EEG recordings. The SNR for the EEG signal was set from -10dB to 10dB [35], meanwhile, the percentage of observable brain regions from iEEG was set to be 30%, and the network complexity configuration is the same as in Sec.3.1. Table 2 shows the statistical result for the experiment. The pairwise comparison of 0% and 30% observable state variables is illustrated in Fig. 3. The group level t-test shows the significant difference when making 30% of state variables (brain regions) directly observable from iEEG electrodes, with a pronounced improvement in *Acc* score by reducing the false positive rates.

From Table 2, the classical two-step methods do not achieve satisfactory results even under high SNR conditions. In contrast, when employing the one-step state space modeling approach, performance remains comparable with either 0% or 30% iEEG observations given high SNR scalp EEG, highlighting the advantages of the one-step approach with state-space modeling paradigm for WBEN estimation. As SNR decreases with increased noise, *Acc* and *Sen* metrics deteriorate

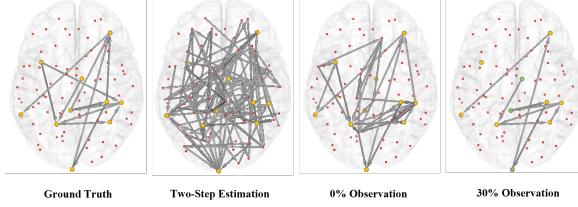


Figure 4: Visualization for network estimation in different methods, where the activated patch centers are highlighted with yellow color, and the iEEG observed patch centered are highlighted with green color. For two-step methods, the eLORETA that has the best performance is selected for comparison.

across all methods. However, performance using the one-step state space dynamic model degrades significantly, producing numerous incorrectly predicted connections at lower SNR values, whereas integrating iEEG measurements yields a relatively stable performance curve. This finding further confirms the value of incorporating iEEG for more accurate WBEN estimation, particularly when scalp EEG channels are contaminated with high noise levels.

Further analysis on the false positive rate, the different causal analysis baselines, and the impact of brain network complexity on the WBEN estimation are given in the Appendix A.6, A.5 and A.7, respectively.

3.2 Cortical-subcortical network analysis of working memory task

In this section, we analyzed the networks between cortical and subcortical brain regions during the Working Memory (WM) tasks. WM is commonly associated with learning, understanding, executive functioning, information processing, intelligence, and problem-solving in humans and various animals from infancy to old age [39]. Maintaining content in WM requires communication between an extensive network of brain regions [40]. In this section, the proposed method was applied to estimate and analyze the brain connectivity between cortical and subcortical regions at different WM phases, which are the encoding phase and maintenance phase, in a verbal WM task conducted by patients with epilepsy where simultaneous scalp EEG and iEEG were recorded. The data description and preprocessing are detailed in Appendix A.4.

The experiment used a second-order vector auto-regressor model for the estimation framework stated in section 2.2. The regional connectivity of brain activity during the encoding and maintenance phases was estimated separately for each human participant’s tasks. Lastly, the final estimation was obtained by first taking the sum over the estimated state transition matrices for each task and then averaging all task-leveled state transition matrices according to different phases. To ensure the stability of the result, only the connections from region i to j that contains signal power no less than 10% of the signal power from region j to itself were kept, i.e., $\Phi(i, j) > 0.1 * \Phi(j, j)$. Moreover,

Table 2: Evaluation of performance with different levels of scalp EEG SNR.

Method	Metrics	SNR=5	SNR=0	SNR=-5	SNR=-10
MNE	Sen	0.328 ± 0.131	0.253 ± 0.149	0.162 ± 0.120	0.170 ± 0.192
	Acc	0.013 ± 0.006	0.011 ± 0.005	0.008 ± 0.004	0.006 ± 0.003
dSPM	Sen	0.199 ± 0.144	0.203 ± 0.144	0.152 ± 0.134	0.149 ± 0.159
	Acc	0.010 ± 0.008	0.011 ± 0.007	0.007 ± 0.003	0.006 ± 0.003
sLORETA	Sen	0.255 ± 0.160	0.269 ± 0.159	0.165 ± 0.123	0.176 ± 0.189
	Acc	0.011 ± 0.006	0.011 ± 0.007	0.008 ± 0.004	0.006 ± 0.004
eLORETA	Sen	0.442 ± 0.101	0.412 ± 0.119	0.383 ± 0.135	0.363 ± 0.143
	Acc	0.013 ± 0.006	0.012 ± 0.006	0.010 ± 0.005	0.010 ± 0.005
ALCMV [36]	Sen	0.427 ± 0.134	0.390 ± 0.145	0.412 ± 0.161	0.384 ± 0.139
	Acc	0.002 ± 0.001	0.002 ± 0.001	0.002 ± 0.001	0.002 ± 0.001
ASTAR [37]	Sen	0.113 ± 0.011	0.107 ± 0.013	0.125 ± 0.022	0.095 ± 0.013
	Acc	0.003 ± 0.001	0.002 ± 0.001	0.003 ± 0.001	0.003 ± 0.002
VSSI-ARD [38]	Sen	0.448 ± 0.117	0.421 ± 0.128	0.361 ± 0.124	0.338 ± 0.148
	Acc	0.003 ± 0.002	0.003 ± 0.001	0.003 ± 0.003	0.003 ± 0.001
EEG with 0% of iEEG obs.	Sen	0.707 ± 0.194	0.592 ± 0.169	0.461 ± 0.155	0.402 ± 0.132
	Acc	0.429 ± 0.292	0.253 ± 0.245	0.189 ± 0.176	0.034 ± 0.023
EEG with 30% of iEEG obs.	Sen	0.780 ± 0.136	0.716 ± 0.115	0.639 ± 0.120	0.491 ± 0.103
	Acc	0.404 ± 0.270	0.501 ± 0.269	0.545 ± 0.224	0.595 ± 0.167

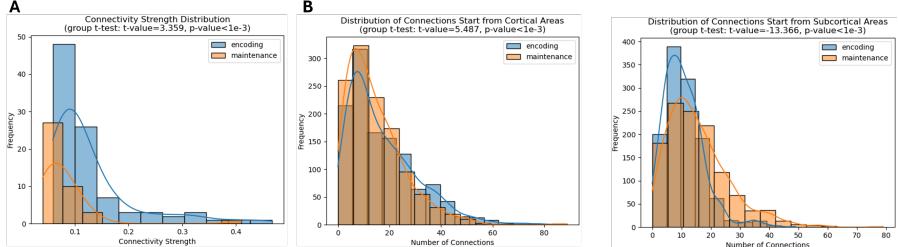


Figure 5: Cortical-subcortical connectivity estimation analyses across encoding and maintenance phases in verbal working memory. **A** depicts the distribution of absolute connectivity strength between cortical and subcortical regions during both phases based on averaged results, with significant differences confirmed by group t-test ($t = 3.359$, $p < 0.001$). **B** illustrates the distribution of connection frequencies across phases derived from event-level analyses: left panel represents outgoing information flows from cortical regions ($t = 5.487$, $p < 0.001$), while right panel shows outgoing information flows from subcortical regions ($t = -13.336$, $p < 0.001$).

the HO atlas regions in the cortical area were further merged into lobe-level granularity, while caudate, putamen, and pallidum were aggregated as basal ganglia to further pursue robust macro-scale results. The estimated dynamic networks (Appendix A.9) revealed significant phase-dependent differences. During encoding, predominant information flow occurred from cortical to subcortical regions, specifically from frontal, temporal (including auditory processing areas), and parietal lobes to thalamus and basal ganglia. Meanwhile, subcortical-originating connections primarily targeted other subcortical structures, with directed pathways from basal ganglia and hippocampus to thalamus, and from amygdala to basal ganglia and thalamus. Conversely, maintenance phase exhibited a reversed directionality pattern, with prominent connections from thalamus and basal ganglia to frontal and parietal lobes, and from hippocampus to temporal regions (including auditory processing areas). Statistical analyses in Fig. 5 confirmed significant differences in both directional connectivity and distribution patterns between subcortical and cortical regions across phases. We further analyzed the connectivity strength and the direction of information flow between cortical and subcortical regions during the encoding and maintenance phases. As shown in Fig. 5, the overall connectivity strength during the encoding phase is stronger than that of the maintenance phase. Besides, the amount of connection that flows start from cortical regions is generally greater during the encoding phase than the maintenance phase, whereas brain connections during the maintenance phase are more active in the information flow that is directed from subcortical regions to elsewhere.

4 Conclusion

We proposed the first unified computational framework for integrating simultaneously recorded scalp EEG and iEEG data to estimate the whole-brain electrophysiological networks. This framework enables the delineation of neurophysiological networks at a whole brain scale and with millisecond-level temporal resolution. Results validate the complementary value of both modalities, demonstrating that strategic multimodal scalp EEG and iEEG deployment can significantly improve network reconstruction accuracy even with high volume of measurement noise. Numerical experiments confirm the robustness under broad experiment configurations, while application to Sternberg verbal working memory task yielded insightful findings consistent with previous studies [40–47], particularly the characterization of cortical-subcortical information flows during encoding and maintenance phase [40], which is the first analysis of this type at the whole brain scale. The proposed framework can serve as a fundamental computational tool for electrophysiological brain network analysis with applications to clinical studies and neuroscience research.

Acknowledgment

Research reported in this publication was supported by the National Institute of Neurological Disorders and Stroke (NINDS) of the National Institutes of Health (NIH), United States under Award Number R21NS135482 (PI: Liu). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- [1] B. He, A. Sohrabpour *et al.*, “Electrophysiological source imaging: A noninvasive window to brain dynamics,” *Annual Review of Biomedical Engineering*, 2018.
- [2] D. S. Bassett and M. S. Gazzaniga, “Understanding complexity in the human brain,” *Trends in Cognitive Sciences*, vol. 15, no. 5, pp. 200–209, 2011.
- [3] T. D. Wager and E. E. Smith, “Neuroimaging studies of working memory,” *Cognitive, Affective, & Behavioral Neuroscience*, vol. 3, no. 4, pp. 255–274, 2003.
- [4] M. Rubinov and O. Sporns, “Complex network measures of brain connectivity: uses and interpretations,” *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, 2010.
- [5] K. Supek, V. Menon, D. Rubin, M. Musen, and M. D. Greicius, “Network analysis of intrinsic functional brain connectivity in alzheimer’s disease,” *PLoS Computational Biology*, vol. 4, no. 6, p. e1000100, 2008.
- [6] M. P. van den Heuvel and O. Sporns, “A cross-disorder connectome landscape of brain dysconnectivity,” *Nature Reviews Neuroscience*, vol. 20, no. 7, pp. 435–446, 2019.
- [7] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, “Neurophysiological investigation of the basis of the fMRI signal,” *Nature*, vol. 412, no. 6843, pp. 150–157, 2001.
- [8] B. Biswal, F. Zerrin-Yetkin, V. M. Haughton, and J. S. Hyde, “Functional connectivity in the motor cortex of resting human brain using echo-planar MRI,” *Magnetic Resonance in Medicine*, vol. 34, no. 4, pp. 537–541, 1995.
- [9] S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, A. R. Laird *et al.*, “Correspondence of the brain’s functional architecture during activation and rest,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 31, pp. 13 040–13 045, 2009.
- [10] T. Uehara, T. Yamasaki, T. Okamoto, T. Koike, S. Kan, S. Miyauchi, J.-i. Kira, and S. Tobimatsu, “Efficiency of a “small-world” brain network depends on consciousness level: a resting-state fMRI study,” *Cerebral Cortex*, vol. 24, no. 6, pp. 1529–1539, 2014.
- [11] E. L. Dennis and P. M. Thompson, “Functional brain connectivity using fmri in aging and alzheimer’s disease,” *Neuropsychology Review*, vol. 24, pp. 49–62, 2014.
- [12] Q.-Z. Wu, D.-M. Li, W.-H. Kuang, T.-J. Zhang, S. Lui, X.-Q. Huang, R. C. Chan, G. J. Kemp, and Q.-Y. Gong, “Abnormal regional spontaneous neural activity in treatment-refractory depression revealed by resting-state fMRI,” *Human Brain Mapping*, vol. 32, no. 8, pp. 1290–1299, 2011.
- [13] S.-G. Kim, W. Richter, and K. Uğurbil, “Limitations of temporal resolution in functional MRI,” *Magnetic Resonance in Medicine*, vol. 37, no. 4, pp. 631–636, 1997.
- [14] A. B. Tort, R. Komorowski, H. Eichenbaum, and N. Kopell, “Measuring phase-amplitude coupling between neuronal oscillations of different frequencies,” *Journal of Neurophysiology*, vol. 104, no. 2, pp. 1195–1210, 2010.
- [15] R. Srinivasan, W. R. Winter, J. Ding, and P. L. Nunez, “EEG and MEG coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics,” *Journal of Neuroscience Methods*, vol. 166, no. 1, pp. 41–52, 2007.
- [16] J. Fell and N. Axmacher, “The role of phase synchronization in memory processes,” *Nature Reviews Neuroscience*, vol. 12, no. 2, pp. 105–118, 2011.
- [17] F. D. V. Fallani, L. Astolfi, F. Cincotti, D. Mattia, A. Tocci, S. Salinari, M. Marciani, H. Witte, A. Colosimo, and F. Babiloni, “Brain network analysis from high-resolution EEG recordings by the application of theoretical graph indexes,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 5, pp. 442–452, 2008.

- [18] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [19] F. Liu, L. Wang, Y. Lou, R.-C. Li, and P. L. Purdon, “Probabilistic structure learning for EEG/MEG source imaging with hierarchical graph priors,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 321–334, 2020.
- [20] S. Yang, M. Jiao, J. Xiang, N. Fotedar, H. Sun, and F. Liu, “Rejuvenating classical brain electrophysiology source localization methods with spatial graph fourier filters for source extents estimation,” *Brain Informatics*, vol. 11, no. 1, p. 8, 2024.
- [21] M. Kouti, K. Ansari-Asl, and E. Namjoo, “EEG dynamic source imaging using a regularized optimization with spatio-temporal constraints,” *Medical & Biological Engineering & Computing*, vol. 62, no. 10, pp. 3073–3088, 2024.
- [22] Y. Yang, E. Aminoff, M. Tarr, and R. E. Kass, “A state-space model of cross-region dynamic connectivity in MEG/EEG,” *Advances in Neural Information Processing Systems*, vol. 29, 2016.
- [23] E. Pirondini, B. Babadi, G. Obregon-Henao, C. Lamus, W. Q. Malik, M. S. Hämäläinen, and P. L. Purdon, “Computationally efficient algorithms for sparse, dynamic solutions to the EEG source localization problem,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 6, pp. 1359–1372, 2017.
- [24] E. J. Candes, “The restricted isometry property and its implications for compressed sensing,” *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.
- [25] B. Soleimani, P. Das, I. D. Karunathilake, S. E. Kuchinsky, J. Z. Simon, and B. Babadi, “NLGC: Network localized granger causality with application to meg directional functional connectivity analysis,” *NeuroImage*, vol. 260, p. 119496, 2022.
- [26] J. Sanchez-Bornot, R. C. Sotero, J. S. Kelso, Ö. Şimşek, and D. Coyle, “Solving large-scale MEG/EEG source localisation and functional connectivity problems simultaneously using state-space models,” *NeuroImage*, vol. 285, p. 120458, 2024.
- [27] D. Duncan, R. B. Duckrow, S. M. Pincus, I. Goncharova, L. J. Hirsch, D. D. Spencer, R. R. Coifman, and H. P. Zaveri, “Intracranial EEG evaluation of relationship within a resting state network,” *Clinical Neurophysiology*, vol. 124, no. 10, pp. 1943–1951, 2013.
- [28] S. P. Burns, S. Santaniello, R. B. Yaffe, C. C. Jouny, N. E. Crone, G. K. Bergey, W. S. Anderson, and S. V. Sarma, “Network dynamics of the brain and influence of the epileptic seizure onset zone,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 49, pp. E5321–E5330, 2014.
- [29] M. Jiao, X. Xian, B. Wang, Y. Zhang, S. Yang, S. Chen, H. Sun, and F. Liu, “XDL-ESI: Electrophysiological sources imaging via explainable deep learning framework with validation on simultaneous EEG and iEEG,” *NeuroImage*, vol. 299, p. 120802, 2024.
- [30] S. Shu, S. Luo, M. Cao, K. Xu, L. Qin, L. Zheng, J. Xu, X. Wang, and J.-H. Gao, “Informed MEG/EEG source imaging reveals the locations of interictal spikes missed by sEEG,” *NeuroImage*, vol. 254, p. 119132, 2022.
- [31] P. Manomaitsaowapak, A. Nartkulpat, and J. Songsiri, “Granger causality inference in EEG source connectivity analysis: a state-space approach,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 7, pp. 3146–3156, 2021.
- [32] B. D. Anderson and J. B. Moore, *Optimal filtering*. North Chelmsford, MA, USA: Courier Corporation, 2012.
- [33] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

[34] V. Dimakopoulos, L. Stieglitz, L. Imbach, and J. Sarnthein, “Dataset of intracranial EEG, scalp EEG, and beamforming sources from epilepsy patients performing a verbal working memory task,” 2023, version 1.0.1. [Online]. Available: <https://doi.org/10.18112/openneuro.ds004752.v1.0.1>

[35] T. Ball, M. Kern, I. Mutschler, A. Aertsen, and A. Schulze-Bonhage, “Signal quality of simultaneously recorded invasive and non-invasive EEG,” *NeuroImage*, vol. 46, no. 3, pp. 708–716, 2009.

[36] A. Yektaeian Vaziri and B. Makkiabadi, “Accelerated algorithms for source orientation detection and spatiotemporal LCMV beamforming in EEG source localization,” *Frontiers in Neuroscience*, vol. 18, p. 1505017, 2025.

[37] G. Wan, M. Jiao, X. Ju, Y. Zhang, H. Schweitzer, and F. Liu, “Electrophysiological brain source imaging via combinatorial search with provable optimality,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 10, 2023, pp. 12491–12499.

[38] K. Liu, Z. L. Yu, W. Wu, Z. Gu, and Y. Li, “Imaging brain extended sources from EEG/MEG based on variation sparsity using automatic relevance determination,” *Neurocomputing*, vol. 389, pp. 132–145, 2020.

[39] N. Cowan, “Working memory underpins cognitive development, learning, and education,” *Educational Psychology Review*, vol. 26, pp. 197–223, 2014.

[40] V. Dimakopoulos, P. Mégevand, L. H. Stieglitz, L. Imbach, and J. Sarnthein, “Information flows from hippocampus to auditory cortex during replay of verbal working memory items,” *Elife*, vol. 11, p. e78677, 2022.

[41] N. C. Müller, B. N. Konrad, N. Kohn, M. Muñoz-López, M. Czisch, G. Fernández, and M. Dresler, “Hippocampal–caudate nucleus interactions support exceptional memory performance,” *Brain Structure and Function*, vol. 223, pp. 1379–1389, 2018.

[42] A. B. Moore, Z. Li, C. E. Tyner, X. Hu, and B. Crosson, “Bilateral basal ganglia activity in verbal working memory,” *Brain and Language*, vol. 125, no. 3, pp. 316–323, 2013.

[43] C. Chang, S. Crottaz-Herbette, and V. Menon, “Temporal dynamics of basal ganglia response and connectivity during verbal working memory,” *NeuroImage*, vol. 34, no. 3, pp. 1253–1269, 2007.

[44] B. Crosson, H. Benefield, M. A. Cato, J. R. Sadek, A. B. Moore, C. E. Wierenga, K. Gopinath, D. Soltsik, R. M. Bauer, E. J. Auerbach *et al.*, “Left and right basal ganglia and frontal activity during language generation: contributions to lexical, semantic, and phonological processes,” *Journal of the International Neuropsychological Society*, vol. 9, no. 7, pp. 1061–1077, 2003.

[45] M. S. Fustiñana, T. Eichlisberger, T. Bouwmeester, Y. Bitterman, and A. Lüthi, “State-dependent encoding of exploratory behaviour in the amygdala,” *Nature*, vol. 592, no. 7853, pp. 267–271, 2021.

[46] J. Kamiński, S. Sullivan, J. M. Chung, I. B. Ross, A. N. Mamelak, and U. Rutishauser, “Persistently active neurons in human medial frontal and medial temporal lobe support working memory,” *Nature Neuroscience*, vol. 20, no. 4, pp. 590–601, 2017.

[47] J. Li, D. Cao, S. Yu, X. Xiao, L. Imbach, L. Stieglitz, J. Sarnthein, and T. Jiang, “Functional specialization and interaction in the amygdala–hippocampus circuit during working memory processing,” *Nature Communications*, vol. 14, no. 1, p. 2921, 2023.

[48] T. T. Cai, L. Wang, and G. Xu, “New bounds for restricted isometry constants,” *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4388–4394, 2010.

[49] B. Fischl, “Freesurfer,” *NeuroImage*, vol. 62, no. 2, pp. 774–781, 2012.

[50] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen, “MNE software for processing MEG and EEG data,” *NeuroImage*, vol. 86, pp. 446–460, 2014.

- [51] S. Haufe and A. Ewald, “A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies,” *Brain Topography*, vol. 32, pp. 625–642, 2019.
- [52] G. A. Rousselet, “Does filtering preclude us from studying ERP time-courses?” *Frontiers in Psychology*, vol. 3, p. 131, 2012.
- [53] M. S. Hämäläinen and R. J. Ilmoniemi, “Interpreting magnetic fields of the brain: minimum norm estimates,” *Medical Biological Engineering and Computing*, vol. 32, no. 1, pp. 35–42, 1994.
- [54] R. D. Pascual-Marqui, “Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details,” *Methods and Findings in Experimental and Clinical Pharmacology*, vol. 24, no. Suppl D, pp. 5–12, 2002.
- [55] A. M. Dale, A. K. Liu, B. R. Fischl, R. L. Buckner, J. W. Belliveau, J. D. Lewine, and E. Halgren, “Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity,” *Neuron*, vol. 26, no. 1, pp. 55–67, 2000.
- [56] R. D. Pascual-Marqui, A. D. Pascual-Montano, D. Lehmann, K. Kochi, M. Esslen, L. Jancke, P. Anderer, B. Saletu, H. Tanaka, K. Hirata *et al.*, “Exact low resolution brain electromagnetic tomography (eLORETA),” *NeuroImage*, vol. 31, no. Suppl 1, p. S86, 2006.
- [57] A. M. Dale, B. Fischl, and M. I. Sereno, “Cortical surface-based analysis: I. segmentation and surface reconstruction,” *NeuroImage*, vol. 9, no. 2, pp. 179–194, 1999.
- [58] C. Gratton, T. O. Laumann, A. N. Nielsen, D. J. Greene, E. M. Gordon, A. W. Gilmore, S. M. Nelson, R. S. Coalson, A. Z. Snyder, B. L. Schlaggar *et al.*, “Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation,” *Neuron*, vol. 98, no. 2, pp. 439–452, 2018.
- [59] E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. T. Constable, “Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity,” *Nature Neuroscience*, vol. 18, no. 11, pp. 1664–1671, 2015.
- [60] B. Ibrahim, S. Suppiah, N. Ibrahim, M. Mohamad, H. A. Hassan, N. S. Nasser, and M. I. Saripan, “Diagnostic power of resting-state fMRI for detection of network connectivity in alzheimer’s disease and mild cognitive impairment: A systematic review,” *Human Brain Mapping*, vol. 42, no. 9, pp. 2941–2968, 2021.
- [61] W. Xie, R. T. Toll, and C. A. Nelson, “EEG functional connectivity analysis in the source space,” *Developmental Cognitive Neuroscience*, vol. 56, p. 101119, 2022.
- [62] J. M. Palva, S. H. Wang, S. Palva, A. Zhigalov, S. Monto, M. J. Brookes, J.-M. Schoffelen, and K. Jerbi, “Ghost interactions in MEG/EEG source space: A note of caution on inter-areal coupling measures,” *NeuroImage*, vol. 173, pp. 632–643, 2018.
- [63] S. H. Wang, M. Lobier, F. Siebenhühner, T. Puoliväli, S. Palva, and J. M. Palva, “Hyperedge bundling: A practical solution to spurious interactions in MEG/EEG source connectivity analyses,” *NeuroImage*, vol. 173, pp. 610–622, 2018.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: The abstract and introduction clearly articulate all claims, comprehensively detail our contributions, and explicitly address the key assumptions and limitations underlying this research.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: Limitation can be found in Appendix A.11.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory Assumptions and Proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: Our paper relies on the EM algorithm for the inference of dynamic system, so the assumptions and proof is not provided here, which can be found in the textbook Chapter 13.3 in *Pattern Recognition and Machine Learning* By Bishop 2006. .

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental Result Reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The method and numerical experiments are well documented to make sure this research is reproducible. The code supporting the finding of this research will be made public available. The data we used is publicly available.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: The data we used is public available. <https://doi.org/10.18112/openneuro.ds004752.v1.0.1>. The code for the model will be made public accessible upon acceptance. At the current stage, no code is made public available.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental Setting/Details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: Yes, experimental setting is presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment Statistical Significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Random trials were used. Statistical significance were tested.

Guidelines: .

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments Compute Resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [\[Yes\]](#)

Justification: The computational resources for this research is presented in Appendix A.3
Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code Of Ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [\[Yes\]](#)

Justification: We confirm that we followed the NeurIPS code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader Impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [\[Yes\]](#)

Justification: This tool can be used as an analytical tool to characterize the brain networks. It can help neurosurgeons make better decisions, however, it may mislead the decision making if this tool did not used properly.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: **[Yes]**

Justification: Safeguards are given in the Appendix A.12.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: **[Yes]**

Justification: Public data is used in this paper, and the original paper and link for the public dataset is presented in our paper.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New Assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: This paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and Research with Human Subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: the paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [Yes]

Justification: The authors' university IRB board approved its exemption.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

A Appendix

A.1 Definition of restricted isometry property

The definition of Restricted Isometry Property (RIP) is that for a lead field matrix $\Phi \in \mathbb{R}^{m \times n}$ and sparsity level s , the RIP is mathematically defined as:

$$(1 - \delta_s) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_s) \|x\|_2^2, \quad (23)$$

where $\delta_s \in (0, 1)$ is the Restricted Isometry Constant (RIC) and the inequality holds for all s -sparse vectors x .

The RIP can be equivalently expressed using operator norms:

$$\delta_s = \max_{|S| \leq s} \|\Phi_S^T \Phi_S - I\|_{2 \rightarrow 2}, \quad (24)$$

where Φ_S denotes the submatrix of Φ with columns indexed by set S , and I is the identity matrix.

In terms of eigenvalues, RIP ensures:

$$1 - \delta_s \leq \lambda_{\min}(\Phi_S^T \Phi_S) \leq \lambda_{\max}(\Phi_S^T \Phi_S) \leq 1 + \delta_s. \quad (25)$$

Cai et al. proved that as sparsity s increases, even optimal measurement matrices cannot keep δ_s sufficiently small, making recovery unreliable [48]. Since EEG leadfields often violate RIP at moderate–high sparsity, ESI accuracy degrades under multi-source conditions, motivating the use of strong priors or data-driven constraints.

A.2 Algorithmic framework and procedure

Algorithm 1 EM Framework for Parameter Estimation

Require: EEG measurement $\{y_t\}_{t=1}^T$, iEEG measurement $\{z_t\}_{t=1}^T$, leadfield matrix for EEG L and iEEG C , estimated noise covariance matrix of EEG \mathbf{P} and iEEG \mathbf{S} , other parameters for VAR model, regularization, and halting condition.

- 1: Integrate the two modalities to get an augmented measurement $\{y_t^+\}_{t=1}^T$, leadfield matrix L^+ , and noise covariance matrix \mathbf{P}^+ . Initialize parameter $\theta^{(0)}$ according to any conventional source localization method, e.g., MNE.
- 2: **while** the convergence or halting condition is not met **do**
- 3: **E-step:** Compute $\mathcal{Q}(\theta; \theta^{(j)})$ based on the conditional density $p(x_{1:T} | y_{1:T}^+, \theta^{(j)})$ estimated via FIS.
- 4: **M-step:** Maximize $\theta^{(j+1)} = \underset{\theta}{\operatorname{argmax}} \{ \mathcal{Q}(\theta; \theta^{(j)}) + \lambda \sum_{i=1}^N \|\Phi_{i,:}^{(j)}\|_1 \}$ using FASTA.
- 5: $j = j + 1$
- 6: **end while**

Ensure: Optimal parameter θ^*

A.3 Simulated experiment settings

Brain Forward Model: To generate synthetic EEG data, we used a realistic head model to compute the leadfield matrix based on the T1-MRI images from FreeSurfer [49]. The brain tissue segmentation and tissue surface generation were conducted using FreeSurfer. A 128-channel BioSemi EEG cap layout was used, and the EEG channels were co-registered with the scalp surface and further validated on the MNE-Python toolbox [50]. The source space was parcelled and resampled. Then a three-layer boundary element method (BEM) was built based on the reconstructed surfaces, resulting in a leadfield matrix denoted as L .

Realistic data generation: To generate a causal time series, the Berlin Brain Connectivity Benchmark (BBCB) [51] is used with randomly generated state transition matrices Φ_k with $k \in \{1, \dots, K\}$. To ensure the convergence of source signals, each eigenvalue in Φ is further validated to be less or equal to 1. Then we add an independent random Gaussian noise to each source signal at every time step. Lastly, an acausal third-order Butterworth filter with zero phase delay was applied with band-pass frequency being [0.1Hz, 40Hz] [52]. With the generated source signal, the observation time series can be derived by multiplying the leadfield matrix with the source signal and adding the channel-wise

correlated random Gaussian noise with a given signal-to-noise ratio (SNR) level. The iEEG noise is generated separately with a relatively higher SNR level [35].

Evaluation Metrics: In the simulated experiment, the ground truth of the connectivity network is defined based on the generated state transition matrix Φ . If the i, j -th element $\Phi_{(i,j)}$ is zero, there is no link from i to j , otherwise there is a link from i to j . In this experiment, K is set to be 1 so that the state at time t is only dependent on the state at time $t - 1$. The Sensitivity (Sen) and Accuracy (Acc) metrics for connectivity estimation are defined to evaluate the performance given as:

$$Sen = \frac{N_c}{N_{tot}}, \quad Acc = \frac{N_c}{N_c + N_w}, \quad (26)$$

where N_c denotes the number of correct predictions, and N_{tot} is the total number of connectivities, and N_w is the number of wrong predictions (number of false positives).

Benchmark algorithms: The baseline methods include traditional two-step methods, which conducted brain source localization first with existing ESI algorithms, such as MNE [53], sLORETA [54], dSPM [55], eLORETA [56] etc., followed by applying the Granger causality analysis on the estimated source signals for source space connectivity estimation. The observation signals were set with sampling rate 100Hz and a group of 10s time series were generated. In order to analyze the impact of the integration of iEEG, several hyperparameters, i.e., the number of partially observable brain regions, different levels of SNR, and complexity of brain networks (which includes the number of activated source regions and the maximum in-degree of each region), were evaluated comprehensively.

Computational resources: All experiments are run on a Windows 11 pro desktop with 32G memory, an Intel i9-12900KF CPU and an A6000 GPU of 48G memory.

A.4 Real Data Description and Preprocessing

The dataset comprises intracranial recordings from 15 epilepsy patients undergoing clinical monitoring for seizure localization while performing a modified Sternberg verbal working memory task. This paradigm temporally segregated encoding, maintenance, and recall phases. The comprehensive dataset includes simultaneous scalp EEG recordings following the 10-20 system, depth electrode iEEG recordings, and the corresponding MNI coordinates with anatomical labels for all intracranial electrodes [34]. Each participant completed multiple experimental sessions, with 50 distinct events per session. Each event epoch consisted of an 8-second recording (sampled at 200 Hz for EEG and 2000 Hz for iEEG), structured as follows: fixation (0-1s), working memory encoding (1-3s), maintenance (3-6s), and response (6-8s). Our analysis specifically targeted the encoding and maintenance phases. To account for potential temporal extension of the auditory encoding process beyond the visual stimulus presentation, we analyzed only the final 2 seconds of the maintenance phase, consistent with methodological considerations outlined in prior research [40].

All iEEG and EEG electrodes were co-registered to the standard 'fsaverage' template [57]. The forward model was calculated using MNE-Python with source spacing set to oct-5, generating over 2000 patches per hemisphere. Forward solutions were computed independently for each subject according to the standard 10-20 montage of their specific electrode configurations. For regional connectivity analysis, source locations were aggregated into anatomical regions defined by the Harvard-Oxford (HO) atlas (48 cortical and 21 subcortical areas) using nearest-neighbor mapping in MNI space, thereby reducing the high-dimensional source space to a more tractable atlas-based representation. iEEG channels were mapped to HO atlas areas based on electrode positions, with signals averaged across channels assigned to identical atlas regions. The iEEG leadfield matrix was constructed as a binary mapping, with unit values representing measured atlas areas and zeros elsewhere. In the absence of empty room recordings, we modeled EEG noise as a multivariate Gaussian distribution (mean=0, standard deviation=1) and iEEG noise as a multivariate Gaussian with reduced variance (mean=0, standard deviation=0.1). Both EEG and iEEG signals were subsequently downsampled to 50 Hz.

A.5 Assessment of model performance with respect to different causal analysis under different SNR

We further use transfer entropy and partial directed coherence for two-stage causal analysis baselines. As can be seen from Table 3 and 4. The proposed method is consistently better than baselines.

Table 3: Evaluation of performance with different levels of scalp EEG SNR. The connectivity estimation for two-step methods is based on transfer entropy.

Method	Metrics	SNR=5	SNR=0	SNR=-5	SNR=-10
MNE	Sen	0.255 \pm 0.093	0.255 \pm 0.093	0.255 \pm 0.093	0.255 \pm 0.093
	Acc	0.002 \pm 0.001	0.002 \pm 0.001	0.002 \pm 0.001	0.002 \pm 0.001
DSPM	Sen	0.000 \pm 0.000	0.000 \pm 0.000	0.008 \pm 0.024	0.008 \pm 0.024
	Acc	0.000 \pm 0.000	0.000 \pm 0.000	0.033 \pm 0.100	0.000 \pm 0.001
SLORETA	Sen	0.262 \pm 0.074	0.277 \pm 0.043	0.292 \pm 0.046	0.296 \pm 0.053
	Acc	0.002 \pm 0.000	0.002 \pm 0.000	0.002 \pm 0.000	0.002 \pm 0.001
ELORETA	Sen	0.296 \pm 0.053	0.296 \pm 0.053	0.296 \pm 0.053	0.291 \pm 0.058
	Acc	0.004 \pm 0.001	0.003 \pm 0.001	0.003 \pm 0.000	0.002 \pm 0.000
ALCMV [36]	Sen	0.140 \pm 0.104	0.166 \pm 0.095	0.212 \pm 0.113	0.232 \pm 0.103
	Acc	0.096 \pm 0.064	0.056 \pm 0.033	0.021 \pm 0.013	0.007 \pm 0.005
ASTAR [37]	Sen	0.009 \pm 0.018	0.009 \pm 0.018	0.004 \pm 0.013	0.004 \pm 0.013
	Acc	0.014 \pm 0.031	0.014 \pm 0.031	0.005 \pm 0.014	0.004 \pm 0.011
VSSI-ARD [38]	Sen	0.087 \pm 0.065	0.067 \pm 0.058	0.066 \pm 0.053	0.062 \pm 0.056
	Acc	0.120 \pm 0.085	0.088 \pm 0.121	0.062 \pm 0.089	0.035 \pm 0.060
EEG with 0% of iEEG obs.	Sen	0.707 \pm 0.194	0.592 \pm 0.169	0.461 \pm 0.155	0.402 \pm 0.132
	Acc	0.429 \pm 0.292	0.253 \pm 0.245	0.189 \pm 0.176	0.034 \pm 0.023
EEG with 30% of iEEG obs.	Sen	0.780 \pm 0.136	0.716 \pm 0.115	0.639 \pm 0.120	0.491 \pm 0.103
	Acc	0.404 \pm 0.270	0.501 \pm 0.269	0.545 \pm 0.224	0.595 \pm 0.167

Table 4: Evaluation of performance with different levels of scalp EEG SNR. The connectivity estimation for two-step methods is based on partial directed coherence.

Method	Metrics	SNR=5	SNR=0	SNR=-5	SNR=-10
MNE	Sen	0.306 \pm 0.151	0.217 \pm 0.157	0.150 \pm 0.178	0.115 \pm 0.190
	Acc	0.011 \pm 0.009	0.011 \pm 0.007	0.012 \pm 0.014	0.011 \pm 0.017
DSPM	Sen	0.111 \pm 0.210	0.147 \pm 0.197	0.126 \pm 0.200	0.105 \pm 0.204
	Acc	0.018 \pm 0.030	0.151 \pm 0.136	0.155 \pm 0.122	0.129 \pm 0.178
SLORETA	Sen	0.270 \pm 0.194	0.214 \pm 0.181	0.175 \pm 0.191	0.121 \pm 0.213
	Acc	0.007 \pm 0.006	0.006 \pm 0.003	0.007 \pm 0.004	0.005 \pm 0.006
ELORETA	Sen	0.467 \pm 0.170	0.431 \pm 0.143	0.366 \pm 0.163	0.417 \pm 0.168
	Acc	0.019 \pm 0.011	0.013 \pm 0.006	0.007 \pm 0.002	0.006 \pm 0.002
ALCMV [36]	Sen	0.643 \pm 0.069	0.699 \pm 0.104	0.716 \pm 0.113	0.699 \pm 0.077
	Acc	0.004 \pm 0.001	0.005 \pm 0.000	0.005 \pm 0.001	0.005 \pm 0.001
ASTAR [37]	Sen	0.078 \pm 0.042	0.068 \pm 0.060	0.055 \pm 0.039	0.064 \pm 0.048
	Acc	0.038 \pm 0.023	0.034 \pm 0.033	0.026 \pm 0.018	0.029 \pm 0.022
VSSI-ARD [38]	Sen	0.648 \pm 0.215	0.658 \pm 0.221	0.624 \pm 0.214	0.515 \pm 0.218
	Acc	0.005 \pm 0.001	0.005 \pm 0.001	0.005 \pm 0.001	0.004 \pm 0.001
EEG with 0% of iEEG obs.	Sen	0.707 \pm 0.194	0.592 \pm 0.169	0.461 \pm 0.155	0.402 \pm 0.132
	Acc	0.429 \pm 0.292	0.253 \pm 0.245	0.189 \pm 0.176	0.034 \pm 0.023
EEG with 30% of iEEG obs.	Sen	0.780 \pm 0.136	0.716 \pm 0.115	0.639 \pm 0.120	0.491 \pm 0.103
	Acc	0.404 \pm 0.270	0.501 \pm 0.269	0.545 \pm 0.224	0.595 \pm 0.167

A.6 Assessment of model performance using false positive rate

To address ghost interactions, which mainly represent false positive connections in brain network analysis, we calculated the false positive rate (FPR) of different methods. As shown in Table 5, the proposed method achieved significantly lower FPR compared to other benchmark algorithms, demonstrating superior specificity in identifying true brain network connections while effectively suppressing spurious interactions caused by signal leakage.

Table 5: False positive rate of different methods under varying observation noise levels. The connectivity estimation for two-step methods is based on Granger causality.

Method	Metric	SNR _{obs} =5	SNR _{obs} =0	SNR _{obs} =-5	SNR _{obs} =-10
MNE	FPR	0.154 \pm 0.273	0.156 \pm 0.260	0.145 \pm 0.263	0.151 \pm 0.262
dSPM	FPR	0.129 \pm 0.276	0.137 \pm 0.266	0.139 \pm 0.265	0.139 \pm 0.263
eLORETA	FPR	0.256 \pm 0.278	0.265 \pm 0.272	0.272 \pm 0.275	0.258 \pm 0.271
sLORETA	FPR	0.146 \pm 0.271	0.147 \pm 0.263	0.146 \pm 0.263	0.148 \pm 0.262
ALCMV [36]	FPR	0.643 \pm 0.225	0.700 \pm 0.236	0.732 \pm 0.242	0.736 \pm 0.249
ASTAR [37]	FPR	0.172 \pm 0.071	0.228 \pm 0.101	0.188 \pm 0.085	0.165 \pm 0.092
VSSI-ARD [38]	FPR	0.548 \pm 0.176	0.522 \pm 0.179	0.468 \pm 0.197	0.371 \pm 0.194
EEG with 0% of iEEG obs.	FPR	0.037 \pm 0.091	0.038 \pm 0.089	0.028 \pm 0.061	0.071 \pm 0.085
EEG with 30% of iEEG obs.	FPR	0.007 \pm 0.005	0.008 \pm 0.005	0.003 \pm 0.003	0.002 \pm 0.002

A.7 Assessment of model performance with respect to network architectural complexity

Impact of brain network complexity on the WBEN estimation: The benefit of integrating iEEG and scalp EEG is further validated with different network complexity levels. The network complexity level is designed with varied number of activated source regions, and varied value of in-degree of node that effects the number of edges in the brain networks. The number of activated source regions is set to be 10, 15 and 20, while the in-degree was varied from 1 to 4 with other parameters being the same as the previous experiments.

Table 6: Evaluation of performance with different number of activated source regions

Method	Metrics	#Brain Regions=10	#Brain Regions=15	#Brain Regions=20
MNE	Sen	0.162 \pm 0.120	0.251 \pm 0.191	0.154 \pm 0.156
	Acc	0.008 \pm 0.004	0.013 \pm 0.008	0.011 \pm 0.006
dSPM	Sen	0.152 \pm 0.134	0.254 \pm 0.195	0.208 \pm 0.212
	Acc	0.007 \pm 0.003	0.012 \pm 0.009	0.010 \pm 0.005
sLORETA	Sen	0.165 \pm 0.123	0.272 \pm 0.189	0.206 \pm 0.214
	Acc	0.008 \pm 0.004	0.013 \pm 0.008	0.010 \pm 0.005
eLORETA	Sen	0.383 \pm 0.135	0.390 \pm 0.169	0.389 \pm 0.146
	Acc	0.010 \pm 0.005	0.009 \pm 0.005	0.011 \pm 0.005
EEG with 0% of iEEG observation	Sen	0.461 \pm 0.155	0.350 \pm 0.087	0.398 \pm 0.096
	Acc	0.189 \pm 0.176	0.118 \pm 0.095	0.191 \pm 0.091
EEG with 30% of iEEG observation	Sen	0.639 \pm 0.120	0.518 \pm 0.110	0.547 \pm 0.093
	Acc	0.545 \pm 0.224	0.500 \pm 0.172	0.552 \pm 0.182

Table 7: Evaluation of performance with different values of node in-degree

Method	Metrics	In-degree=1	In-degree=2	In-degree=3	In-degree=4
MNE	Sen	0.117 \pm 0.043	0.162 \pm 0.120	0.230 \pm 0.197	0.304 \pm 0.259
	Acc	0.005 \pm 0.002	0.008 \pm 0.004	0.014 \pm 0.011	0.016 \pm 0.011
dSPM	Sen	0.121 \pm 0.037	0.152 \pm 0.134	0.423 \pm 0.294	0.504 \pm 0.296
	Acc	0.005 \pm 0.002	0.007 \pm 0.003	0.011 \pm 0.010	0.011 \pm 0.009
sLORETA	Sen	0.131 \pm 0.058	0.165 \pm 0.123	0.428 \pm 0.292	0.511 \pm 0.289
	Acc	0.005 \pm 0.002	0.008 \pm 0.004	0.011 \pm 0.010	0.011 \pm 0.010
eLORETA	Sen	0.196 \pm 0.098	0.383 \pm 0.135	0.505 \pm 0.189	0.630 \pm 0.139
	Acc	0.005 \pm 0.003	0.010 \pm 0.005	0.010 \pm 0.007	0.009 \pm 0.006
EEG with 0% of iEEG observation	Sen	0.570 \pm 0.194	0.461 \pm 0.155	0.401 \pm 0.198	0.422 \pm 0.241
	Acc	0.228 \pm 0.234	0.189 \pm 0.176	0.149 \pm 0.201	0.109 \pm 0.168
EEG with 30% of iEEG observation	Sen	0.797 \pm 0.129	0.639 \pm 0.120	0.546 \pm 0.143	0.422 \pm 0.203
	Acc	0.697 \pm 0.248	0.545 \pm 0.224	0.498 \pm 0.159	0.410 \pm 0.215

Based on performance summarized in Table 6 and Table 7, the performance of the two-step methods is usually limited, with high false positive predictions. By integrating iEEG and scalp EEG in the WBEN estimation, both *Acc* and *Sen* showed improved performance compared with the comparison methods. The experiments further highlight the benefits of leveraging simultaneously recorded scalp EEG and iEEG for WBEN estimation.

A.8 Intra- and Inter-subject analysis

We conducted a comprehensive analysis of connectivity pattern consistency to validate the reliability of our derived brain connections across different subjects and experimental sessions using Pearson correlations.

Intra-subject consistency: Encoding phase: 0.539 ± 0.062 , Maintenance phase: 0.511 ± 0.043 .
Inter-subject consistency: Encoding phase: 0.355 ± 0.037 , Maintenance phase: 0.353 ± 0.059 .

The results reflect the expected balance between network stability and task-specific adaptability based on EEG data. The moderate intra-subject correlations are consistent with the functional flexibility of brain networks during working n-back memory tasks (each time different words are presented to the participants). Individual connectivity patterns are preserved across sessions while allowing for task-related adaptations to different verbal stimuli. The inter-subject consistency reflects a biologically meaningful balance [58]. It demonstrates the presence of shared functional networks underlying verbal working memory while preserving the individual neural diversity that characterizes brain function [59]. This moderate correlation (compared to random correlation (≈ 0) between two vectors of 4761 dimensions (directed edges from a 69 by 69 matrix)) indicates our method successfully captures both the common neural mechanisms required for the task and the individual network signatures that contribute to cognitive variability across subjects.

A.9 Visualization of connectivity estimation in real data experiment

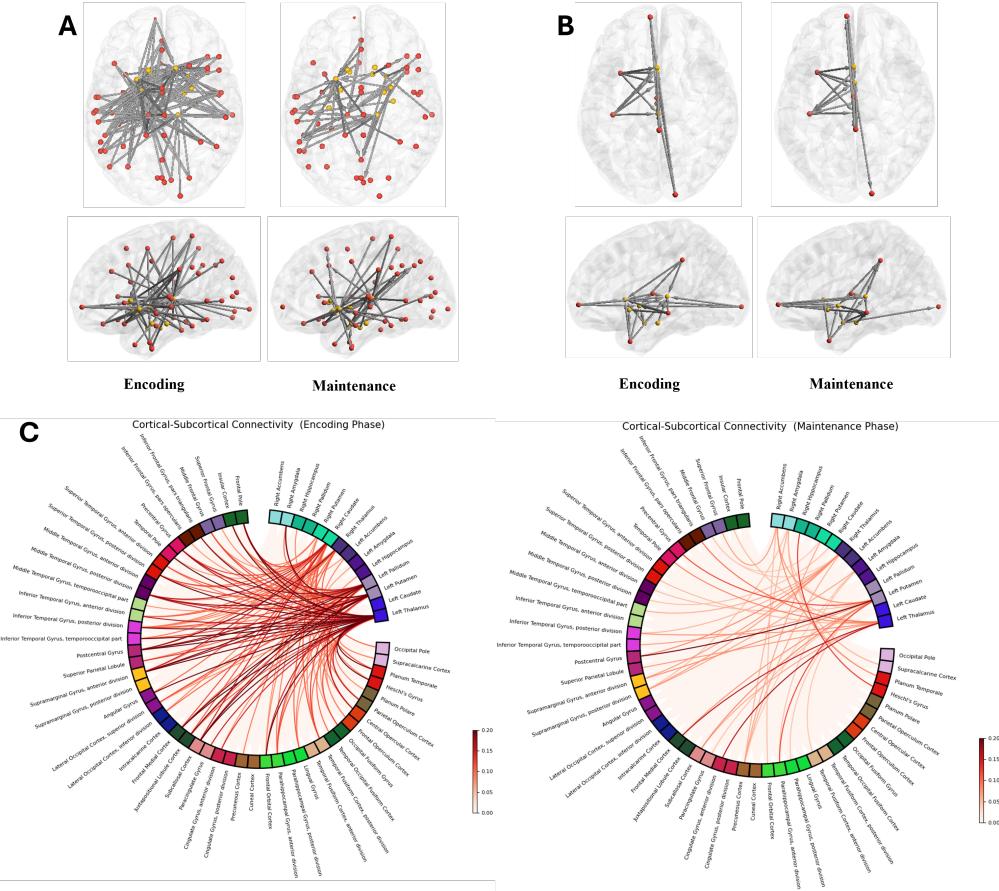


Figure 6: A visualization example of cortical-subcortical connectivity estimation in working memory. Sub-figure **A** shows the brain connectivity network in two WM phases according to the HO atlas, while **B** is the network of aggregated lobe-leveled regions used for macroscopic expression. **C** is the circular brain connectivity graph in HO atlas parcellation for encoding and maintenance phases, respectively.

A.10 Approximated computational complexity

The computational complexity of the proposed method is dominated by the FIS in the E-step. The forward Kalman filtering requires matrix operations at each time step, including the covariance update and matrix inversion, both involving $O(N^3)$ operations, where N is the number of source regions.

To align with established work on brain network analysis using fMRI [60], we also adopted an atlas-based approach where we estimate region-to-region connectivity. In this work, we used the Harvard-Oxford atlas regions ($N = 69$). The computational cost for the atlas-based analysis will remain stable regardless of the different dimensions of the source space. Practically, the brain network estimation is computationally feasible (about 3 minutes wall clock for 1000 time points).

A.11 Limitations

Although this method is a one-step approach to modeling source-space connectivity, false-positive connections may still occur near true connections due to ‘signal leakage’ during source reconstruction [61, 62]. This problem occurs because estimating thousands of sources from data collected by only a hundred electrodes is inherently under-determined. As a result, residual signal leakage between locations will inevitably influence the source data, causing false connectivities to appear as unintended artifacts of true interactions between source pairs. Moreover, the exact locations of these sources may

be inaccurately determined. This issue, known as "ghost interaction," can be mitigated by a novel method that organizes connections into hyperedges based on their adjacency in signal mixing [63].

A.12 Safeguards

The code and result produced from our developed framework can give a delineation of electrophysiological brain networks which can be used for surgical decision. However, epileptologists and neurosurgeons should use this as a augmented tool for decision making and should double check the result with other brain imaging modalities.