Electrophysiological Brain Network Estimation with Simultaneous Scalp EEG and Intracranial EEG: Inference Algorithm and Applications

Shihao Yang and Feng Liu*

Department of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ 07030, United States.

*Corresponding author(s). E-mail(s): fliu22@stevens.com;

Abstract

Activity in the human brain is composed of complex firing patterns and interactions among neurons and neuronal circuits. The neuroimaging field underwent a paradigm shift over the past decades from mapping tasked evoked brain regions of activations towards identifying and characterizing the dynamic brain networks of coordinating brain regions. Electrophysiological signals are the direct manifestation of brain activities, thus characterizing the whole brain electrophysiological networks (WBEN) can serve as a fundamental tool for neuroscience studies and clinical applications. The electrophysiological networks inferred from electroencephalogram (EEG) source imaging surfers from low accuracy limited by the Restricted Isometry Property (RIP), while the invasive EEG derived electrophysiological networks can only characterize partial brain regions where invasive electrodes reside. In this work, we introduce the first framework for the integration of scalp EEG and intracranial EEG (iEEG) for WBEN estimation with a principled estimation framework based on state-space models, where an Expectation–Maximization (EM) algorithm is designed to inference the state variables and brain connectives 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, which demonstrates the importance of inclusion of iEEG signal for WBEN estimation. For real data with simultaneous EEG and iEEG, we applied the developed framework to understand the information flows of the encoding and maintenance phases during the working memory task. The information flows between the subcortical and cortical regions are delineated, which highlights more significant information flows from cortical to subcortical regions compared to maintenance phases. The results are

consistent with previous research findings, however with the view of the whole brain scope, which underscores the unique utility of the proposed framework.

Keywords: Electrophysiological Brain Networks, State-Space Models, Multimodal neuroimaging, Brain connectome, Working Memory

1 Introduction

Brain connectivity networks consist of complex connections between neurons that carry information throughout the brain and support various cognitive functions in neuroscience [1, 2]. However, estimating and reconstructing these connectomes is a complex task that requires a combination of multiple neuroimaging and computational modeling techniques [3–6]. Many studies have suggested that an accurately inferred brain connectivity map can gain insights into the relationship between brain structure and function, reveal the fundamentals of cognitive processes, and uncover the mechanisms of neuropsychiatric diseases [7–12].

In the past decades, most researchers leveraged functional Magnetic Resonance Imaging (fMRI) for brain functional network modeling and analysis. fMRI is one of the most widely used non-invasive neuroimaging techniques, it indirectly measures the brain activity by measuring changes in blood flow dynamics as well as oxygen levels. fMRI can provide a high spatial resolution of activity. fMRI has been used to characterize the functional connections between different regions of the brain [13–15]. By analyzing fMRI data, researchers can explore the connectomes of the brain under different states and tasks, unveiling the complex relationships between the structure and function of brain networks [15–17]. However, fMRI has limitations on relatively high cost, non-portability, and low temporal resolution, which limit its applications where high temporal solution is necessary [18, 19]. Therefore, the neuroimaging techniques with a temporal resolution of subsecond level, such as Magnetoencephalography (MEG) or Electroencephalogram (EEG) can be used to estimate the WBEN after Electrophysiological Source Imaging (ESI). Studies that using MEG or EEG to reconstruct brain functional networks have leveraged different ESI approaches and connectivity metrics [20–26]. Those frameworks consists of two steps, which is Step (1): EEG/MEG source localization or ESI; Step (2): brain network construction with the estimated source space signal.

The existing ESI framework suffers from low accuracy when it comes to the estimation of whole brain regions due to the ill-posed property and is limited by Restricted Isometry Property (RIP) in theory [27]. Traditionally, neurophysiologically plausible

priors, such as sparsity, have to be incorporated in ESI as the regularization terms to enforce a feasible solution [28–30]. Estimation of the whole brain networks poses more challenges to solving the ESI problem as it further relaxes the sparsity constraint [31]. Another existing problem for the current ESI framework is there is no supervision in the learning process of the latent source space signal, and model validation is conducted only after the reconstruction of the source signal based on either synthetic data or neuroscience knowledge, thus limiting the usability and accuracy of the estimated source signal.

In addition to non-invasive modalities, invasive neuroimaging technologies, such as intracranial Electroencephalography (iEEG) including Electrocorticography (ECoG) and stereoelectroencephalography (sEEG) that place subdural electrodes on the brain surface or penetrating electrodes in the subcortical brain regions can achieve a more accurate connectivity map among the regions of interest within the brain of high temporal resolution, in order to mitigate the noise caused by extracranial measurements [32–34]. However, it is infeasible to record the whole brain's intracranial signal due to the invasive nature of iEEG, which makes the brain only partially observable. As iEEG electrodes can only cover part of the brain, important nodes or links might be missing at other brain regions. For example, in the epilepsy seizure onset zone analysis, Gao et al. used MEG/EEG source imaging and identified the interictal spikes that are missed by the iEEG [35], which showcases the value of a synergy of scalp EEG and invasive iEEG. Another group of researchers suggests most interictal epileptiform discharges arise in brain tissue outside the site of seizure onset and propagate toward it [36]. In such cases, having a WBEN estimation can provide a more comprehensive characterization of signal genesis and propagation at the whole brain scale.

To address the above issues, we take the scalp EEG and iEEG as complementary to each other and believe the fusion of two modalities can provide a delineation of the whole brain's activation/connectivity map[35, 37]. The integration of iEEG and EEG can yield a faithful reconstruction of whole brain electrophysiological connectivity, however, a principled multimodal integration modeling and inference framework has not been explored before. The pipeline of the proposed approach is illustrated in Fig. 1

2 Method

2.1 Basic problem definition

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

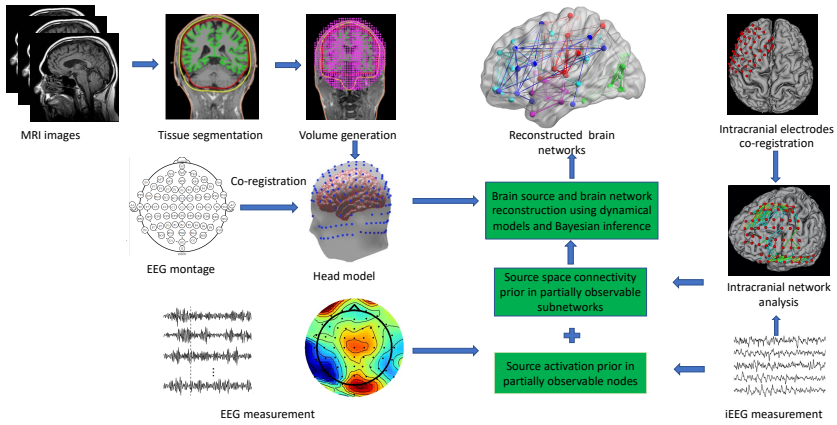

Fig. 1 The overall pipeline of integration of EEG and iEEG for source space brain network reconstruction

$$
\begin{aligned}
x_t &= \sum_{k=1}^{K} \mathbf{\Phi}_k x_{t-k} + n_t,\ t = 1, ..., T,\ x_t \in \mathbf{R}^N \\
y_t &= L x_t + w_t,\ y_t \in \mathbf{R}^M \\
z_t &= C x_t + e_t,\ z_t \in \mathbf{R}^O
\end{aligned} \tag{1}
$$

where N, M, and O are the number of the source regions, EEG electrodes, and iEEG electrodes, respectively. $\mathbf{\Phi}_k \in \mathbf{R}^{N \times N}$ 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, we can see from the model definition that the parameters that need to be estimated are $\mathbf{\Phi}$ and $\mathbf{Q}$.

Define the unknown parameters as $\theta = \{\mathbf{\Phi}, \mathbf{Q}\}$. Then the log-likelihood can be written in the form:

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

The log-likelihood of the model can be defined as

$$logp(y_{1:T}, z_{1:T}|\theta),\ \theta = \{\mathbf{\Phi}, \mathbf{Q}\} \tag{3}$$

where $\mathbf{\Phi}$ and $\mathbf{Q}$ are the parameters for the model, and its definition is the same as before.

Since the number of the observations is far more than that of the source, the inverse estimation problem is highly ill-posed. To alleviate this phenomenon and simplify the problem, we add a regularization term based on the assumption that the connection from all other regions to a given region is sparse. In this case, we can define the regularized maximum log-likelihood of parameters for the model as :

$$\theta^* = \arg\max_{\theta}\ logp(y_{1:T}, z_{1:T}|\theta) - \lambda \sum_{n=1}^{N} \|\mathbf{\Phi}_{n,:}\|_1 \tag{4}$$

where $\Phi_{n,:}$ is the nth 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 [38].

2.2 An EM estimation framework

In the domain of statistical methodology, the Expectation–Maximization algorithm serves as an iterative computational approach meticulously employed for the purpose of ascertaining the utmost likelihood, whether it pertains to local optima or the acme of posterior estimations (MAP) of parameters within the context of statistical models. These models, notably contingent upon latent variables concealed from direct observation, underscore the necessity for such intricate methodologies. The EM iterative procedure is characterized by its cyclical execution, entailing two pivotal stages: firstly, the Expectation (E) step, whereby a function is crafted to encapsulate the expected value of the log-likelihood, thoughtfully computed utilizing the prevailing parameter estimate. Secondly, the Maximization (M) step, during which a judicious parameterization is derived, endeavoring to maximize the anticipated log-likelihood cultivated during the antecedent E step. Subsequently, these diligently derived parameter estimates assume a central role in illuminating the distributional characteristics of the latent variables in preparation for the ensuing E step within this iterative loop. Since

the data distribution in the source domain is unknown, we can apply the Expectation-Maximization(EM) framework that takes the source state as the latent variable 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 will firstly illustrate the E-step, by starting from the Equation 2 defined above and using the facts that the observations of EEG and iEEG are conditional independent on the source state x, we can rewrite and derive the log-likelihood by introducing source state $x_{1:T}$ as

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

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

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

and

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

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 also be obtained in a similar way based on the linear dynamical model from Equation 1. and utilizes the presumption that $\mathbf{Q}$ is a diagonal covariance matrix with the items $\{\sigma_i^2, i = 1, ..., N\}$ on its diagonal, then we have

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

where $\hat{\mathbf{x}}_i = [x_{i,K+1}, x_{i,K+2}, ..., x_{i,T}]^T$, $\underline{\mathbf{X}} = [[x_{1,K+1}, x_{1,K+2}, ..., x_{1,T}]^T, ..., [x_{1,1}, x_{1,2}, ..., x_{1,T-K}]^T, ..., [x_{N,K+1}, x_{N,K+2}, ..., x_{N,T}]^T, ..., [x_{N,1}, x_{N,2}, ..., x_{N,T-K}]^T]$, and $\hat{\phi}_i = [[\mathbf{\Phi}_k]_{i,j}]^T$ for $k = 1, ..., K$, $j = 1, ..., N$. By substituting Equations 6-8 into 5, we can reformulate the formula and take the expectation to get the Q-function for EM as

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

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

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

where the bracket superscript represents the jth 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 [39]. In order 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 Equation 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, we need to find the optimal $\theta = \{\mathbf{\Phi}, \mathbf{Q}\}$ that maximizes the Q-function defined above. Since the number of observations is far less than that of the source regions which makes the inverse problem ill-posed. We then introduce the l_1 regularization on $\mathbf{\Phi}$ based on the premise that the functional connection among a given region and others possesses sparse properties to reduce the difficulty of problem-solving. Thus, we can have the equation of the maximization with the form

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

which can be solved easily by using the Fast Adaptive Shrinkage/Thresholding Algorithm (FASTA). In this way, we successfully build up the EM framework to estimate the source.

2.3 FIS for source density estimation

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 Equation 1. Taking into account computing performance factors, we can define a fusion observation by combining y and z together. The intuition is that we can think of EEG and iEEG as one measurement, but the measuring noise between two types of electrodes has no correlation, and the conditional density of merged observation is also a Gaussian. Thus, we redefine observations as

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

where

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

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

Then the log-likelihood problem can be transformed to Equation 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 the $p(x_{1:T}|y^{+}_{1:T}, \theta)$. And we 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 [39]. Then, we can apply the FIS 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[38], 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, we need to redefine the augmented source state as $\underline{\mathbf{x}}_t = [x_t; x_{t-1}; ...; x_{t-K+1}], \underline{\mathbf{x}}_t \in \mathbf{R}^{(KN)}$ and the augmented dynamic model as Equation 22 to transform VAR(K) problem into a VAR(1) one.

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

where

$$
\begin{aligned}
\underline{\mathbf{\Phi}} &= \begin{pmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \dots & \mathbf{\Phi}_{(K-1)} & \mathbf{\Phi}_K \\ \mathbf{I}_{(M\times M)} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(M\times M)} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_{(M\times M)} & \mathbf{0} \end{pmatrix}, \qquad \underline{\mathbf{\Phi}} \in \mathbf{R}^{KN\times KN} \\
\underline{\mathbf{n}}_t &= \begin{pmatrix} n_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \underline{\mathbf{n}}_t \in \mathbf{R}^{KN}, \qquad \underline{\mathbf{n}}_t \sim \mathcal{N}(0, \underline{\mathbf{Q}}) \\
\underline{\mathbf{Q}} &= \begin{pmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}
\end{aligned}
\tag{17}
$$

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. We can also have the augmented version of merged observations as

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

where

$$
\underline{\mathbf{L}}^+ = \begin{pmatrix} \mathbf{L}^+ & \mathbf{0} \end{pmatrix}, \qquad \underline{\mathbf{L}}^+ \in \mathbf{R}^{M\times KN} \tag{19}
$$

In this way, we can calculate the $p(\underline{\mathbf{x}}_t | y_{1:T}^+, \theta)$ via FIS if the observations as well as θ are known.

Next, according to [40] one can define

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

as the mean, covariance matrix as well as the cross-covariance matrix for any given t_1 and t_2. Then we can fit the model with the FIS framework.

In the forward filtering step, we start with the initial value obtained via VAR. Then we can do for $t = 0, 1, 2, ..., T-1$

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

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

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

Then we can calculate the cross-covariance matrix according to Equation 20, and finally, simply extract the first N rows of $\underline{\mathbf{x}}_{t|T}$ and the N order submatrix of the upper

left corner of the matrix $\mathbf{B}_{t_1+1,t_2|T}$ for $\forall t_1, t_2 = 1, 2, ..., T$, which are exactly the first- and second-order moment of $p(x_t|y^+_{1:T}, \theta), \forall t$, to finalize the E-step in Section 2.2.

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 $\theta^{(j+1)} = \underset{\theta}{\operatorname{argmax}}\{\mathcal{Q}(\theta; \theta^{(j)}) + \lambda \sum_{i=1}^N \|\mathbf{\Phi}^{(j)}_{i,:\ |\theta}\|_1\}$ using FASTA.

5: $j = j + 1$

6: **end while**

Ensure: Optimal parameter θ^*

3 Results

In this study, we first conducted experiments on simulated data and evaluated relevant indicators and results, and then we tested the method on a dataset collected from a real Sternberg verbal working memory task to explore the connectivity maps between cortical and subcortical regions during the encoding and maintenance phases. The experiment setups as well as the evaluation metrics for both simulated and real-data experiments, will be elaborated first, followed by the results and analysis, respectively.

3.1 Simulated Experiment

3.1.1 forward model

To generate synthetic EEG data, we used a real head model to compute the leadfield matrix. The T1-MRI images were introduced from FreeSurfer [41] templates. The brain tissue segmentation and source surface reconstruction were conducted using FreeSurfer. Then a three-layer boundary element method (BEM) head was built based on the reconstructed surfaces. A 128-channel BioSemi EEG cap layout was used, and the EEG channels were co-registered with the head model and further validated on the MNE-Python toolbox [42]. The source space was parcelled and clustered according

to the ico-1 spacing standard that contains 42 sources in each hemisphere, with 84 sources combined, resulting in a leadfield matrix L with a dimension of 128 by 84.

3.1.2 Data Generation and Preprocessing

Next, to generate a simulated time series of a causal series. We referred to the famous Berlin Brain Connectivity Benchmark (BBCB) [43] and slightly simplified it by using the autoregression idea based on K randomly generated state transition matrices $\mathbf{\Phi}$ to generate the source signal, and to ensure the source signal does not diverge, each component in $\mathbf{\Phi}$ should 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 pass bandwidth $[0.1\text{Hz}, 40\text{Hz}]$.

With the generated source signal, we can easily generate the observations by multiplying the leadfield matrix with the source signal and adding the channel-wise correlated random Gaussian noise according to a given signal-to-noise ratio (SNR). It is worth noting that, if the partial observations are introduced, the leadfield matrix should be augmented as illustrated in 2.3, and based on the assumption that the measurement noise of iEEG should not only be independent of that of the EEG but also have much higher SNR level [44]. Thus the iEEG noise needs to be generated separately under a relatively high SNR, and we can obtain the intact observation noise correlation matrix by then. In this way, all the inputs for the proposed framework are ready.

3.1.3 Evaluation Metrics

In the simulated experiment, the ground truth of the connectivity network is defined based on the generated state transition matrix $\mathbf{\Phi}$. The easiest way to represent if there is a link between region i and j is by looking at the summation of K $\mathbf{\Phi}$s and see if the component with index (i, j) or/and (j, i) is non-zero. To make the problem better interpretive, we set $K = 1$, thus the source signal is similar to a Markov process where the state at time t is only dependent on the state at time $t - 1$. In this case, the transition matrix $\mathbf{\Phi}$ itself can also work as the correlation matrix for the source signal. Therefore, if both true network $\mathbf{\Phi}$ and estimated $\tilde{\mathbf{\Phi}}$ have non-zero value on the sample component, we counted it as a good prediction otherwise, a bad one. To evaluate the performance, we defined the sensitivity and accuracy of the connectivity

links estimation as

$$
\begin{aligned}
Sensitivity &= \frac{number\ of\ good\ predictions}{total\ number\ of\ links} \\
Accuracy &= \frac{number\ of\ good\ predictions}{number\ of\ good\ predictions\ +\ number\ of\ bad\ predictions}
\end{aligned}
\tag{23}
$$

Moreover, the connectivity prediction can be visualized via BrainNet Viewer[45] to provide a more intuitive comparison.

3.1.4 Simulation Setups and Results Analysis

The simulation was conducted in three different manners to thoroughly analyze the effectiveness while introducing iEEG observations in brain network inference. In all experiment setups, the benchmarks were set to the traditional two-step method, which does source localization by taking classical algorithms, such as MNE, sLORETA, etc., first and then applying the Granger causality analysis for each permutation of the estimated source signals. As mentioned before, the evaluation of connectivity is based on the state transition matrix, thus, to further sparsify and robustify the estimated connectivity, we drop all the absolute correlation values that are less than a given threshold, in this simulation, 0.01 was set as the threshold. The observation signals were set with sampling rate 100Hz and a group of 10s time-length sequences with 1000 time steps were generated while the first one second of data were dropped to ensure the stabilization. In order to comprehensively analyze the impact of introducing iEEG, several hyperparameters, i.e., partial observations proportion, SNR, and causal network complexity (which includes the number of activated source regions and the maximum in-degree of each region), were evaluated within a given range. Then we will illustrate each simulation experiment as follows.

Impact of Partial Observations Proportion: We first validated how the proportion of partial observations can affect the performance in connectivity estimation. In this case, the independent variable is partial observations proportion, namely the coverage rate, and its range was set from 0% to 50%. The SNR for EEG and iEEG observations were set to -5dB and 30dB, respectively. The network complexity was set with the number of activations as 10 and in-degree as 2. For each parameter combination, 10 repetitive runs were applied to ensure that the results were accurate and reliable. The statistics are shown in Table 1

It is worth noting that as partial observation proportion increases, both sensitivity and accuracy on connectivity estimation show an increasing trend, and the increase, especially for accuracy, is significant at 30%, which proves that when iEEG is introduced, and an appropriate amount of activated area is observed, performance can be significantly improved. A visualization example for 30% partial observation is shown in Figure 2

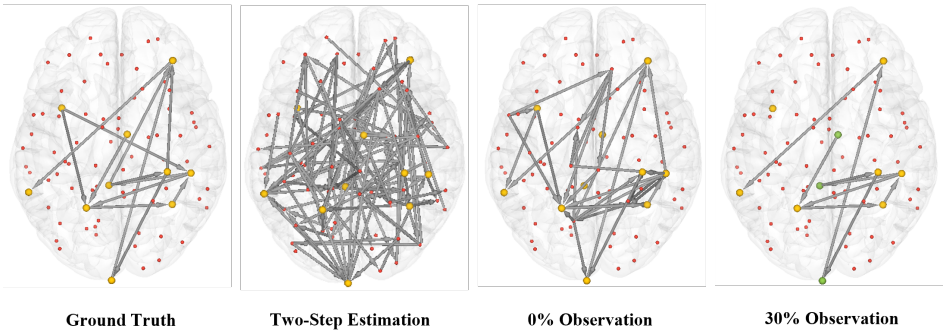

Fig. 2 Visualization for network estimation in different methods, where the activated patch centers are highlighted with yellow color, and the iEEG observed patch centered are highlighted with green color. For two-step methods, the eLORETA that has the best performance is taken for comparison.

Table 1 Partial Observations Proportion Impact Evaluation

Metrics	0%	10%	20%	30%	40%	50%	60%
Sensitivity	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
Accuracy	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 Observation SNR: The SNR of the observations is another key factor. As mentioned before, here we assume the SNR in the iEEG signal is far greater than that of the EEG signal, plus the iEEG signal should be stable and accurate enough with a relatively high SNR, whereas the EEG signal is much inaccurate and less stable. Therefore, we simply set the iEEG SNR to 30dB as in the previous experiment and changed the SNR for EEG to see how accurate iEEG measurement can improve the estimation performance under different disturbance levels of EEG observations. The SNR for the EEG signal was set from -10dB to 10dB, meanwhile, the iEEG observation proportion was set to 30%, and the network complexity configuration is the same as the previous experiment in that the number of activations is 10 and the in-degree is 2. Table 2 shows the statistical result for this experiment.

As we can see from the results, the classical two-step method usually does not perform well when SNR is low, when SNR increases, even if some algorithms, such

as eLORETA can capture nearly half of the correct connectivities, the accuracy is still unsatisfying. By contrast, for the dynamic model method when SNR for measurement reaches a relatively high level, the performance of both using and not using iEEG observation has a similar performance. However, when SNR is decreasing, metrics for both cases show a downward trend. But we can find that when SNR is less than zero, the accuracy for no iEEG observation case drops significantly, indicating that in the absence of iEEG observation, a large number of incorrect links in the network are estimated, whereas, with the help of iEEG, the accuracy is relatively stable that is no significant decline in performance, which further prove that by introducing iEEG on can estimate the brain connectivity much accurate, especially under low-quality extracranial measurement circumstances. Therefore, we can conclude that when measurement noise becomes severe, introducing a small proportion of iEEG will not significantly improve correct estimations, i.e. sensitivity, but it can better constrain the incorrect estimated links or, in other words, can achieve a much higher accuracy.

Table 2 SNR Impact Evaluation

Method	Metrics	SNR=-10	SNR=-5	SNR=0	SNR=5	SNR=10
MNE	Sensitivity	0.149 ± 0.159	0.152 ± 0.134	0.203 ± 0.144	0.199 ± 0.144	0.191 ± 0.149
	Accuracy	0.006 ± 0.003	0.007 ± 0.003	0.011 ± 0.007	0.01 ± 0.008	0.01 ± 0.007
dSPM	Sensitivity	0.149 ± 0.159	0.152 ± 0.134	0.203 ± 0.144	0.199 ± 0.144	0.191 ± 0.149
	Accuracy	0.006 ± 0.003	0.007 ± 0.003	0.011 ± 0.007	0.01 ± 0.008	0.01 ± 0.007
sLORETA	Sensitivity	0.149 ± 0.159	0.152 ± 0.134	0.203 ± 0.144	0.199 ± 0.144	0.191 ± 0.149
	Accuracy	0.006 ± 0.003	0.007 ± 0.003	0.011 ± 0.007	0.01 ± 0.008	0.01 ± 0.008
eLORETA	Sensitivity	0.363 ± 0.143	0.383 ± 0.135	0.412 ± 0.119	0.442 ± 0.101	0.481 ± 0.095
	Accuracy	0.01 ± 0.005	0.01 ± 0.005	0.012 ± 0.006	0.013 ± 0.006	0.015 ± 0.007
iEEG Observation 0%	Sensitivity	0.402 ± 0.132	0.461 ± 0.155	0.592 ± 0.169	0.707 ± 0.194	0.815 ± 0.156
	Accuracy	0.034 ± 0.023	0.189± 0.176	0.253 ± 0.245	0.429 ± 0.292	0.582 ± 0.331
iEEG Observation 30%	Sensitivity	0.491 ± 0.103	0.639 ± 0.120	0.716 ± 0.115	0.780 ± 0.136	0.908 ± 0.059
	Accuracy	0.595 ± 0.167	0.545 ± 0.224	0.501 ± 0.269	0.404 ± 0.270	0.559 ± 0.247

Table 3 Number of Nodes Impact Evaluation

Method	Metrics	#Nodes=10	#Nodes=15	#Nodes=20
MNE	Sensitivity	0.152 ± 0.134	0.242 ± 0.186	0.148 ± 0.144
	Accuracy	0.007 ± 0.003	0.012 ± 0.008	0.011 ± 0.005
dSPM	Sensitivity	0.152 ± 0.134	0.254 ± 0.195	0.208 ± 0.212
	Accuracy	0.007 ± 0.003	0.012 ± 0.009	0.01 ± 0.005
sLORETA	Sensitivity	0.152 ± 0.134	0.254 ± 0.195	0.208 ± 0.212
	Accuracy	0.007 ± 0.003	0.012 ± 0.009	0.01 ± 0.005
eLORETA	Sensitivity	0.383 ± 0.135	0.39 ± 0.169	0.389 ± 0.146
	Accuracy	0.01 ± 0.005	0.009 ± 0.005	0.011 ± 0.005
iEEG Observation 0%	Sensitivity	0.461 ± 0.155	0.350 ± 0.087	0.398 ± 0.096
	Accuracy	0.189± 0.176	0.118 ± 0.095	0.191 ± 0.091
iEEG Observation 30%	Sensitivity	0.639 ± 0.120	0.518 ± 0.110	0.547 ± 0.093
	Accuracy	0.545 ± 0.224	0.500 ± 0.172	0.552 ± 0.182

Table 4 In-degree of Nodes Impact Evaluation

Method	Metrics	In-degree=1	In-degree=2	In-degree=3	In-degree=4
MNE	Sensitivity	0.121 ± 0.037	0.152 ± 0.134	0.221 ± 0.19	0.298 ± 0.271
	Accuracy	0.005 ± 0.002	0.007 ± 0.003	0.014 ± 0.012	0.015 ± 0.011
dSPM	Sensitivity	0.121 ± 0.037	0.152 ± 0.134	0.423 ± 0.294	0.504 ± 0.296
	Accuracy	0.005 ± 0.002	0.007 ± 0.003	0.011 ± 0.01	0.011 ± 0.009
sLORETA	Sensitivity	0.121 ± 0.037	0.152 ± 0.134	0.423 ± 0.294	0.504 ± 0.296
	Accuracy	0.005 ± 0.002	0.007 ± 0.003	0.011 ± 0.01	0.011 ± 0.009
eLORETA	Sensitivity	0.196 ± 0.098	0.383 ± 0.135	0.505 ± 0.189	0.63 ± 0.139
	Accuracy	0.005 ± 0.003	0.01 ± 0.005	0.01 ± 0.007	0.009 ± 0.006
iEEG Observation 0%	Sensitivity	0.57 ± 0.194	0.461 ± 0.155	0.401 ± 0.198	0.422 ± 0.241
	Accuracy	0.228 ± 0.234	0.189 ± 0.176	0.149 ± 0.201	0.109 ± 0.168
iEEG Observation 30%	Sensitivity	0.797 ± 0.129	0.639 ± 0.120	0.546 ± 0.143	0.422 ± 0.203
	Accuracy	0.697 ± 0.248	0.545 ± 0.224	0.498 ± 0.159	0.41 ± 0.215

Impact of Causal Network Complexity: Lastly, we verified the effect of introducing iEEG observations on the experiment under different causal network complexity circumstances. As mentioned before, the network complexity consists of the activated source number, namely the number of nodes in the network, and the in-degree for each node that affects how many edges are in the network. To mitigate the coupling phenomenon that may influence the results, the two independent variables were changed separately. The number of nodes was changed from 10 to 20, while the in-degree was adjusted from 1 to 3. The other parameters are similar to previous settings in that the iEEG observation proportion was set to 30%, and the SNR for EEG and iEEG observations were set to -5dB and 30dB, respectively.

According to the result, for two-step methods, the performance is usually unsatisfying, and even though the sensitivity is increased while the network complexity increases, they usually suffer from low accuracy or incorrect connectivity prediction. According to our method, we can conclude that whether it is increasing the number of nodes or increasing the number of links to increase the complexity of the network would result in performance degradation. However, due to the existence of iEEG, in most cases, both sensitivity and accuracy, especially the latter, are better than the case where iEEG is missing. This result can be another proof that a more accurate connectivity estimation could be achieved while the iEEG can be collected with EEG simultaneously. The evaluation details can be found in Table 3 and 4.

3.2 Real Data Experiment - Cortical-subcortical Information Flow During Verbal Working Memory Task

Working Memory (WM) is commonly associated with learning, understanding, executive functioning, information processing, intelligence, and problem-solving in humans and a variety of animals from infancy to old age [46]. Maintaining content in WM requires communication within an extensive network of brain regions [47]. In this section, we applied the proposed method to estimate and analyze the brain network between cortical and subcortical regions at different WM processes, particularly the encoding and maintenance process, in a verbal WM task for epilepsy patients.

3.2.1 Data Description and Preprocessing

The dataset was collected from 15 subjects during a verbal working memory task, where subjects were epilepsy patients undergoing intracranial monitoring for localization of epileptic seizures. Subjects performed a modified Sternberg task in which the encoding of memory items, maintenance, and recall were temporally separated. The dataset includes simultaneously recorded scalp EEG with the 10-20 system, iEEG recorded with depth electrodes, waveforms, and the MNI coordinates and anatomical labels of all intracranial electrodes [48]. Each subject contains several sessions, while for each session, there are 50 events. Considering the possibility of similar experimental results distribution, for subjects that contain more than 4 sessions, we discard the excess parts to make sure there are enough samples for each subject and reduce the redundant computation. Moreover, there is an 8s recording with 200 Hz and 2000 Hz sampling rates for EEG and iEEG in an event, respectively, where 0-1 s represents the fixation phase, 1-3 s is the encoding phase in WM, 3-6 s records the maintenance phase in WM, and 6-8s is for response phase. In this paper, we only focus on the encoding and maintenance phases, and due to the possibility that the encoding phase could exceed the expected time interval, which is caused by the auditory encoding process that may extend past the visual stimulus in the experiment, only the last 2 s of maintenance should be preserved [47].

Since there is no MRI data for each subject, thus a template head MRI from FreeSurfer was applied. We then used MNE-Python to calculate the forward model based on the 'fsaverage' template [49] over the oct-5 source space, where each hemisphere resulted in about 2052 patches. Since the electrodes for each subject are different, the forward solution was calculated independently according to the standard 10-20 montage for the electrodes used by each subject. To analyze the connectivity

among different functional regions, we introduced the Harvard-Oxford (HO) atlas that contains 48 cortical and 21 subcortical structural areas derived from anatomical structural data and segmentations, and all the source locations from the forward model were aggregated into each atlas area by following the nearest neighbor principle in MNI space. In this way, we transformed the original source space with high dimensions into a lower-dimensional atlas source space. Meanwhile, we averaged the columns in the EEG leadfield matrix based on the source aggregation results for the sources that were merged into the same atlas area. Similar processing steps were applied to iEEG data. The iEEG channels were firstly mapped onto HO atlas areas based on the electrodes' positions, and measurement signals were averaged in accordance with the channels that were assigned to the same atlas area by then. The row of leadfield matrix for iEEG that represents the mapping from source signal to electrode measurement was set as a unit vector with a value of 1 at the assigned atlas area and 0 otherwise for each electrode. Since there are no empty room recordings for the EEG and iEEG device from the dataset, in this work, we assume the noise of EEG as a multivariate Gaussian with 0 mean and standard deviation 1, and iEEG as a multivariate Gaussian with 0 mean and standard deviation 0.1. Lastly, both EEG and iEEG signals were down-sampled at a 50 Hz sampling rate.

3.2.2 Results

In the experiment, we used an order-2 VAR model for the EM estimation framework mentioned in section 2.2. We estimated the regional connectivity of brain activity during encoding and maintenance phases separately for each human participant's tasks. Then we took the sum over the estimated state transition matrices for each task and obtained the aggregated result by averaging all task-leveled state transition matrices according to different phases. To ensure the stability of the result, we only keep the connection from region i to j that contains signal power no less than 10% of the signal power from region j to itself, i.e., $\underline{\mathbf{\Phi}}(i,j) > 0.1 * \underline{\mathbf{\Phi}}(j,j)$. Moreover, we further merged the HO atlas regions in the cortical area into lobe-leveled granularity, while caudate, putamen, and pallidum were aggregated as basal ganglia to pursue robust results further.

The estimated dynamic network for encoding and maintenance phases is shown in Figure. 3. The results showed a significant difference in the network between the two phases. In the encoding phase, the results revealed that there are mainly connections from the cortical to subcortical regions. Specifically, information flows from

the frontal, temporal (including regions that are related to auditory processing), and parietal lobes to the thalamus and basal ganglia. On the other hand, the connections that start from the subcortical region mainly point to its internal areas. The results showed a direction from the basal ganglia and hippocampus to the thalamus in the encoding phase. Besides, we can also find there exist connections that start from the amygdala to the basal ganglia and thalamus. By contrast, during the maintenance phase, the main difference lay in the connections originating from subcortical regions compared to the encoding phase. A significant reverse direction from the thalamus and basal ganglia to the frontal and parietal lobes can be found, and there is also a connection from the hippocampus to the temporal lobe (including the auditory processing regions) in this phase.

3.2.3 Discussion

To our knowledge, this is the first inference framework of the whole brain connectivity map during the encoding and maintenance phases of the Sternberg WM task based on simultaneously collected EEG and iEEG. Our study not only helps to infer the brain-wide connectivity network for the WM process, but it can also be applied to the research of other similar problems. It provides new insight into the importance of a clinically integrated collection of iEEG and EEG information, as well as an in-depth exploration of the integration modeling and inference framework

The results of this study showed the activation and dynamic network of cortical and subcortical regions during a verbal working memory task. The results not only showed the same cortical regions activated in the verbal WM task as in traditional studies, such as frontal regions, but also presented subcortical-related connections, which have been demonstrated in previous related studies.

Firstly, a similar result was found in the information flow between cortical and subcortical areas during the encoding and maintenance phases that related to an auditory encoding process in the WM task as in the previous study, which showed the information flows from the temporal lobe to the hippocampus in the encoding phase. A reverse direction appears in the maintenance phase [47]. Although, in our estimation, there is no significant result revealing the connection from the temporal lobe to the hippocampus in the encoding phase, which is likely caused by the position deviation during atlas region aggregation in the preprocessing stage, we still find the information flow from the temporal lobe to other memory processing regions that are

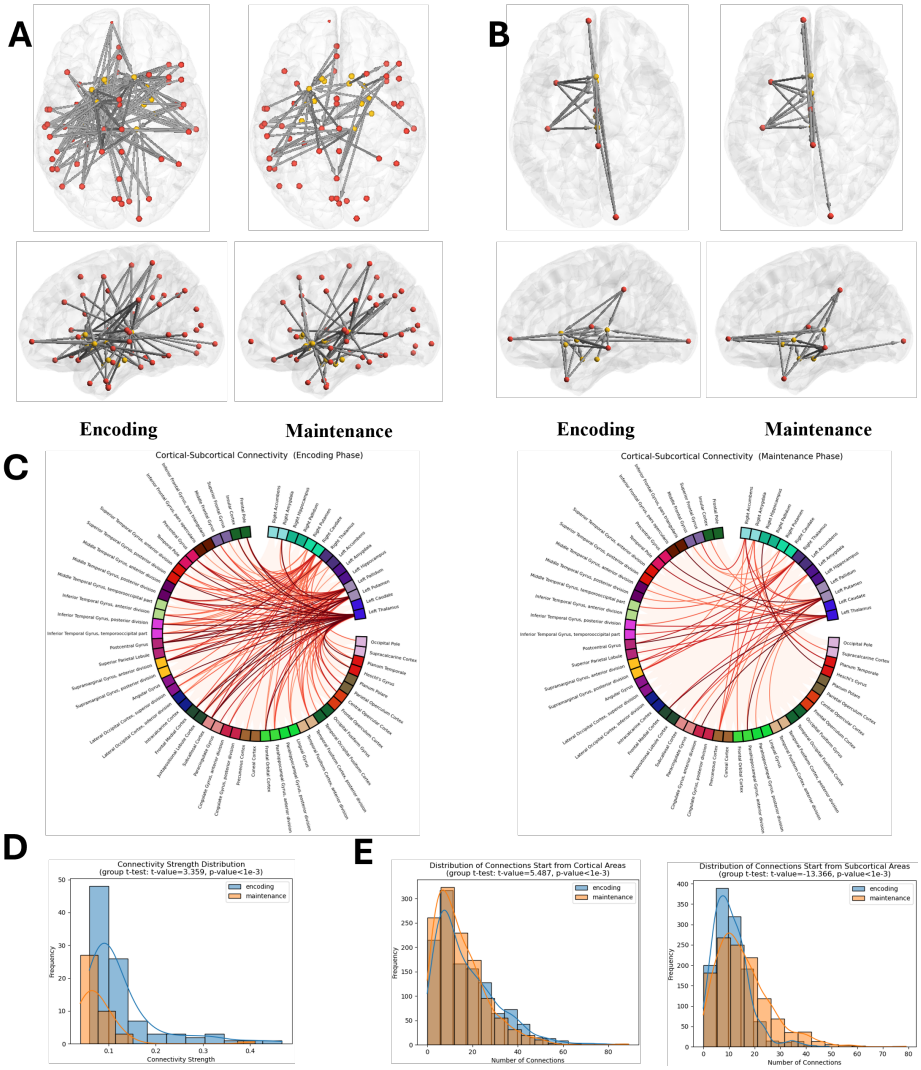

Fig. 3 Cortical-subcortical connectivity estimation during encoding and maintenance phases in the verbal working memory task. Sub-figure A shows the brain connectivity network in two WM phases according to the HO atlas, while B is the network of aggregated lobe-leveled regions used for macroscopic expression. C is the circular brain connectivity graph in HO atlas parcellation for encoding and maintenance phases, respectively. D shows the distribution of absolute connectivity strength between cortical and subcortical regions during encoding and maintenance phases based on the averaged results, and a group t-test was applied with a t-value of 3.359 and $p < 0.001$. Lastly, E presents the distribution of connection numbers during two phases according to the event-leveled results, where the figure on the left represents the distribution of the number of the information flows starting from the cortical regions, and a group t-test gave a t-value of 5.487 and $p < 0.001$ between the two phases, whereas the right one presents the information flows start from the subcortical regions and with a t-value of -13.336 and $p < 0.001$ for group t-test.

adjacent to hippocampus, e.g., caudate, which has also been proved to have synergy with hippocampus in the WM process [50]. While in the maintenance phase, the same reverse information flow from the hippocampus to the temporal lobe was shown in our estimation.

Since it is generally believed that WM is resource-limited, therefore, it is reasonable that the brain would strive to enhance WM in order to facilitate attention, temporary storage, and manipulation of target material [51, 52]. The results obtained by our method are consistent with previous studies, both indicating that the basal ganglia in verbal WM play a role in enhancing goals and inhibiting distracting factors, especially the caudate part is continuously activated during the encoding stage.

Previous research employing fMRI only to investigate language generation in healthy individuals has revealed not only activity in the right caudate nucleus and putamen but also activation in related regions within the left hemisphere [53]. These findings suggested that left basal ganglia activity plays a role in augmenting language processing in the dominant hemisphere, while right basal ganglia activity serves to suppress potentially disruptive activation in the non-dominant right frontal cortex. The same phenomenon is presented in our results, which suggest that bilateral basal ganglia is involved in activities during different phases of WM.

Besides, previous work also proved that not only the basal ganglia but also the thalamus is active during the encoding and maintenance phases of a verbal WM task [51], and our result was consistent with the previous study, where the basal ganglia, especially the caudate, and thalamus are active during the encoding and maintenance phase, and we found there is an information flow from frontal, temporal and parietal lobes to the thalamus and basal ganglia in the encoding phase and a reverse directional flow in the maintenance phase.

Moreover, previous studies have suggested lots of evidence of the amygdala contributing to memory processing, for it plays an important role in the state-dependent encoding of exploratory behavior [54], and contains a higher proportion of response-specific concept cells compared to the hippocampus [55]. We also found that the amygdala is active in our results, particularly during the encoding process, which is similar to the previous work documented that the amygdala was highly activated in the verbal WM task [56].

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 sub-figures D and E of Figure 3, we found that 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 during 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. A set of group t-tests was applied as proof for the difference significance of the findings as stated in the annotation of Figure 3.

Declarations

The authors declares no competing interests.

4 Conclusion

In this study, we proposed a method to model the dynamic network of the brain by integrating simultaneously measured scalp EEG and iEEG, which is solved based on the EM framework. Unlike many previous brain network estimation approaches based on fMRI, this method only uses scalp EEG and iEEG, which provide the possibility for real-time sampling and analysis in clinical settings. The proposed methodology enjoys the advantage of information complementary between two modalities. With the appropriate deployment of iEEG, even if the EEG signal has a low SNR level, one can still reconstruct the brain connectivity map with higher accuracy and sensitivity compared to previous works that used scalp EEG or iEEG signal only. Numerical experiments demonstrated that the proposed method performs well under high network complexity cases and yields good robustness with low SNR levels. on a real Sternberg verbal working memory experimental data, the proposed method not only obtained similar results to other studies using the same data cohort but also had strong consistency with the findings of previous related studies. Therefore, this work provides insights into the simultaneous acquisition of EEG and iEEG signals in future research related to brain network analysis and clinical lesion localization.

References

[1] Sporns, O.: The human connectome: a complex network. Annals of the new York Academy of Sciences **1224**(1), 109–125 (2011)

[2] Bassett, D.S., Gazzaniga, M.S.: Understanding complexity in the human brain. Trends in cognitive sciences **15**(5), 200–209 (2011)

[3] Wager, T.D., Smith, E.E.: Neuroimaging studies of working memory. Cognitive, Affective, & Behavioral Neuroscience **3**(4), 255–274 (2003)

[4] Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nature reviews neuroscience **10**(3), 186–198 (2009)

[5] Rubinov, M., Sporns, O.: Complex network measures of brain connectivity: uses and interpretations. Neuroimage **52**(3), 1059–1069 (2010)

[6] Van Den Heuvel, M.P., Pol, H.E.H.: Exploring the brain network: a review on resting-state fmri functional connectivity. European neuropsychopharmacology **20**(8), 519–534 (2010)

[7] Supekar, K., Menon, V., Rubin, D., Musen, M., Greicius, M.D.: Network analysis of intrinsic functional brain connectivity in alzheimer's disease. PLoS computational biology **4**(6), 1000100 (2008)

[8] Bressler, S.L., Menon, V.: Large-scale brain networks in cognition: emerging methods and principles. Trends in cognitive sciences **14**(6), 277–290 (2010)

[9] Telesford, Q.K., Simpson, S.L., Burdette, J.H., Hayasaka, S., Laurienti, P.J.: The brain as a complex system: using network science as a tool for understanding the brain. Brain connectivity **1**(4), 295–308 (2011)

[10] Deco, G., Kringelbach, M.L.: Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. Neuron **84**(5), 892–905 (2014)

[11] Fornito, A., Bullmore, E.T.: Connectomics: a new paradigm for understanding brain disease. European Neuropsychopharmacology **25**(5), 733–748 (2015)

[12] Heuvel, M.P., Sporns, O.: A cross-disorder connectome landscape of brain dysconnectivity. Nature reviews neuroscience **20**(7), 435–446 (2019)

[13] Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A.: Neurophysiological investigation of the basis of the fmri signal. nature **412**(6843), 150–157

(2001)

[14] Heeger, D.J., Ress, D.: What does fmri tell us about neuronal activity? Nature reviews neuroscience **3**(2), 142–151 (2002)

[15] Logothetis, N.K.: What we can do and what we cannot do with fmri. Nature **453**(7197), 869–878 (2008)

[16] Biswal, B., Zerrin-Yetkin, F., Haughton, V.M., Hyde, J.S.: Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magnetic resonance in medicine **34**(4), 537–541 (1995)

[17] Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., *et al.*: Correspondence of the brain's functional architecture during activation and rest. Proceedings of the national academy of sciences **106**(31), 13040–13045 (2009)

[18] Kim, S.-G., Richter, W., Uğurbil, K.: Limitations of temporal resolution in functional mri. Magnetic resonance in medicine **37**(4), 631–636 (1997)

[19] Bandettini, P.A.: The spatial, temporal, and interpretive limits of functional mri. Neuropsychopharmacology: The fifth generation of progress: An official publication of the American College of Neuropsychopharmacology, 345 (2002)

[20] Fallani, F.D.V., Astolfi, L., Cincotti, F., Mattia, D., Tocci, A., Salinari, S., Marciani, M., Witte, H., Colosimo, A., Babiloni, F.: Brain network analysis from high-resolution eeg recordings by the application of theoretical graph indexes. IEEE Transactions on Neural Systems and Rehabilitation Engineering **16**(5), 442–452 (2008)

[21] De Haan, W., Pijnenburg, Y.A., Strijers, R.L., Made, Y., Flier, W.M., Scheltens, P., Stam, C.J.: Functional neural network analysis in frontotemporal dementia and alzheimer's disease using eeg and graph theory. BMC neuroscience **10**, 1–12 (2009)

[22] Boersma, M., Smit, D.J., Bie, H.M., Van Baal, G.C.M., Boomsma, D.I., Geus, E.J., Waal, H.A., Stam, C.J.: Network analysis of resting state eeg in the developing young brain: structure comes with maturation. Human brain mapping **32**(3), 413–425 (2011)

[23] Tewarie, P., Hillebrand, A., Schoonheim, M.M., Dijk, B.W., Geurts, J.J., Barkhof, F., Polman, C., Stam, C.J.: Functional brain network analysis using minimum spanning trees in multiple sclerosis: an meg source-space study. Neuroimage **88**, 308–318 (2014)

[24] Zhang, X., Lei, X., Wu, T., Jiang, T.: A review of eeg and meg for brainnetome research. Cognitive neurodynamics **8**, 87–98 (2014)

[25] Yin, Z., Li, J., Zhang, Y., Ren, A., Von Meneen, K.M., Huang, L.: Functional brain network analysis of schizophrenic patients with positive and negative syndrome based on mutual information of eeg time series. Biomedical Signal Processing and Control **31**, 331–338 (2017)

[26] La Foresta, F., Morabito, F.C., Marino, S., Dattola, S.: High-density eeg signal processing based on active-source reconstruction for brain network analysis in alzheimer's disease. Electronics **8**(9), 1031 (2019)

[27] Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on information theory **52**(2), 489–509 (2006)

[28] Haufe, S., Nikulin, V.V., Ziehe, A., Müller, K.-R., Nolte, G.: Combining sparsity and rotational invariance in EEG/MEG source reconstruction. NeuroImage **42**(2), 726–738 (2008)

[29] Gramfort, A., Thirion, B., Varoquaux, G.: Identifying predictive regions from fMRI with TV-L1 prior. In: Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop On, pp. 17–20 (2013). https://doi.org/10.1109/PRNI.2013.14 . IEEE

[30] Liu, F., Wang, L., Lou, Y., Li, R.-C., Purdon, P.L.: Probabilistic structure learning for eeg/meg source imaging with hierarchical graph priors. IEEE Transactions on Medical Imaging **40**(1), 321–334 (2020)

[31] Candes, E.J.: The restricted isometry property and its implications for compressed sensing. Comptes rendus mathematique **346**(9-10), 589–592 (2008)

[32] Duncan, D., Duckrow, R.B., Pincus, S.M., Goncharova, I., Hirsch, L.J., Spencer, D.D., Coifman, R.R., Zaveri, H.P.: Intracranial eeg evaluation of relationship

within a resting state network. Clinical Neurophysiology **124**(10), 1943–1951 (2013)

[33] Burns, S.P., Santaniello, S., Yaffe, R.B., Jouny, C.C., Crone, N.E., Bergey, G.K., Anderson, W.S., Sarma, S.V.: Network dynamics of the brain and influence of the epileptic seizure onset zone. Proceedings of the National Academy of Sciences **111**(49), 5321–5330 (2014)

[34] Bernabei, J.M., Arnold, T.C., Shah, P., Revell, A., Ong, I.Z., Kini, L.G., Stein, J.M., Shinohara, R.T., Lucas, T.H., Davis, K.A., *et al.*: Electrocorticography and stereo eeg provide distinct measures of brain connectivity: implications for network models. Brain Communications **3**(3), 156 (2021)

[35] Shu, S., Luo, S., Cao, M., Xu, K., Qin, L., Zheng, L., Xu, J., Wang, X., Gao, J.-H.: Informed meg/eeg source imaging reveals the locations of interictal spikes missed by seeg. NeuroImage **254**, 119132 (2022)

[36] Smith, E.H., Liou, J.-y., Merricks, E.M., Davis, T., Thomson, K., Greger, B., House, P., Emerson, R.G., Goodman, R., McKhann, G.M., *et al.*: Human interictal epileptiform discharges are bidirectional traveling waves echoing ictal discharges. Elife **11**, 73541 (2022)

[37] Hosseini, S.A.H., Sohrabpour, A., Akçakaya, M., He, B.: Electromagnetic brain source imaging by means of a robust minimum variance beamformer. IEEE Transactions on Biomedical Engineering **65**(10), 2365–2374 (2018)

[38] Soleimani, B., Das, P., Karunathilake, I.D., Kuchinsky, S.E., Simon, J.Z., Babadi, B.: Nlgc: Network localized granger causality with application to meg directional functional connectivity analysis. NeuroImage **260**, 119496 (2022)

[39] Anderson, B.D., Moore, J.B.: Optimal Filtering. Courier Corporation, ??? (2012)

[40] Johansen, S.: Likelihood-based Inference in Cointegrated Vector Autoregressive Models. OUP Oxford, ??? (1995)

[41] Fischl, B.: Freesurfer. Neuroimage **62**(2), 774–781 (2012)

[42] Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S.: MNE software for processing MEG and EEG data. Neuroimage **86**, 446–460 (2014)

[43] Haufe, S., Ewald, A.: A simulation framework for benchmarking eeg-based brain connectivity estimation methodologies. Brain topography **32**, 625–642 (2019)

[44] Ball, T., Kern, M., Mutschler, I., Aertsen, A., Schulze-Bonhage, A.: Signal quality of simultaneously recorded invasive and non-invasive eeg. Neuroimage **46**(3), 708–716 (2009)

[45] Xia, M., Wang, J., He, Y.: Brainnet viewer: a network visualization tool for human brain connectomics. PloS one **8**(7), 68910 (2013)

[46] Cowan, N.: Working memory underpins cognitive development, learning, and education. Educational psychology review **26**, 197–223 (2014)

[47] Dimakopoulos, V., Mégevand, P., Stieglitz, L.H., Imbach, L., Sarnthein, J.: Information flows from hippocampus to auditory cortex during replay of verbal working memory items. elife **11**, 78677 (2022)

[48] Dimakopoulos, V., Stieglitz, L., Imbach, L., Sarnthein, J.: "Dataset of Intracranial EEG, Scalp EEG and Beamforming Sources from Epilepsy Patients Performing a Verbal Working Memory Task". https://doi.org/10.18112/openneuro.ds004752.v1.0.1

[49] Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis: I. segmentation and surface reconstruction. Neuroimage **9**(2), 179–194 (1999)

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

[51] Moore, A.B., Li, Z., Tyner, C.E., Hu, X., Crosson, B.: Bilateral basal ganglia activity in verbal working memory. Brain and language **125**(3), 316–323 (2013)

[52] Chang, C., Crottaz-Herbette, S., Menon, V.: Temporal dynamics of basal ganglia response and connectivity during verbal working memory. Neuroimage **34**(3), 1253–1269 (2007)

[53] Crosson, B., Benefield, H., Cato, M.A., Sadek, J.R., Moore, A.B., Wierenga, C.E., Gopinath, K., Soltysik, D., Bauer, R.M., Auerbach, E.J., *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 **9**(7), 1061–1077 (2003)

[54] Fustiñana, M.S., Eichlisberger, T., Bouwmeester, T., Bitterman, Y., Lüthi, A.: State-dependent encoding of exploratory behaviour in the amygdala. Nature **592**(7853), 267–271 (2021)

[55] Kamiński, J., Sullivan, S., Chung, J.M., Ross, I.B., Mamelak, A.N., Rutishauser, U.: Persistently active neurons in human medial frontal and medial temporal lobe support working memory. Nature neuroscience **20**(4), 590–601 (2017)

[56] Li, J., Cao, D., Yu, S., Xiao, X., Imbach, L., Stieglitz, L., Sarnthein, J., Jiang, T.: Functional specialization and interaction in the amygdala-hippocampus circuit during working memory processing. Nature Communications **14**(1), 2921 (2023)