# Introduction

The following report consists of study of effective connectivity in the brain using different approaches when motor imagery tasks are performed. The aim of this study is to infer information flow graphs consistent with all subjects when motor imagery tasks are performed. In this study, I used Transfer Entropy and Granger Causality as a measure of effective connectivity for motor imagery tasks. BCI IV 2a (4 class motor imagery) dataset is used for the analysis.

# Dataset Description

The BCI IV 2a dataset consists of EEG signals from 9 different subjects (C. Brunner). The subjects were asked to perform 4 different motor imagery tasks namely imagination of movement of left hand, right hand, both feet and tongue. Each session comprised of 6 runs, each run consisting of 48 trials (12 trials per each class) yielding a total of 288 trials per session. The signal was recorded using 22 EEG electrodes according to the 10-20 system. It is bandpass filtered between 0.5 Hz and 100 Hz and sampled with 250 Hz.

Data is extracted from 9 channels (Liang, 2016) out of the 22 channels for each subject and is averaged on all trials belonging to the motor imagery action (left/right). Different frequency bands correspond to different actions in motor cortex. Activity in beta range has been suggested to subserve the maintenance of current limb position and strong cortico-muscular coherence has been found in this band (Vicente R, 2011). The data is band passed using 8-30 Hz bandpass filter for most of the experiments expect GC measurement[1].

# Literature

Efforts to characterise the structural connectivity of the brain are complemented by focus on functional and effective connections (Seth AK, 2015). Functional connectivity aims at describing the statistical dependencies between two variables, where as effective connectivity aims at finding a simplest possible circuit which explains the working of observed data.

Over the past decade many studies tried to elucidate network connectivity using effective connectivity measures. (Korzeniewska). One of the most widely used measures to study effective connectivity is Granger Causality. (Granger, Aug., 1969). GC and its extensions help in identifying the linear interactions in the system. However, GC is only well applicable when three prerequisites are met: 1. The interactions between the systems should be approximately linear, 2. Data should have relatively low noise levels (Nalatore H, 2007), 3. Cross talk between the signals has to be low. However, the interactions between the brain signals are usually non-linear. Transfer Entropy is an alternative approach to study effective connections for neuroscience (Gourévitch, 2007) (S. Sabesan, 2009). TE is a model free approach based on information theory and can capture non linearities to a great extent.

[1] Granger Causality Analysis on 8-30 Hz band passed data gave non positive covariance matrix and hence Auto Regressive models can’t be estimated. Thus, data is passed through high pass filter of 5Hz.

## Transfer Entropy

Transfer Entropy is a rigorous derivation of Weiner Causality between two time series data X and Y within information theoretic framework proposed by Schreiber (Schreiber, 2000). We assume that the given time series data can be approximated by a markov process. Thus, we reconstruct a state space of a process with embedding dimension d and delay $\tau$. Let the time series be $X = x_t \; and \; Y = \; y_t$ the delay embedding vector is defined as:

$x_t^d = \left(x_t, x_{t-\tau,} x_{t-2\tau} \dots . x_{t-(d-1)\tau}\right)$ and similarly, for $y_t$.

If dynamics of Y is independent of past of X then, $p(y_{t+1} | y_t^n, x_t^m) = p(y_{t+1} | y_t^n) \quad (1)$

To measure the deviation from this criterion Schreiber measured KL divergence on either side of (1). Thus, transfer entropy is given as:

$$TE(X \rightarrow Y) = \sum\nolimits_{y_{t+1}, y_t^n, x_t^m} p(y_{t+1}, y_t^n, x_t^m) \log\left(\frac{p(y_{t+1} | y_t^n, x_t^m)}{p(y_{t+1} | y_t^n)}\right)$$

The Transfer Entropy is model free and asymmetric i.e $TE(X \rightarrow Y)$ is not equal to $TE(Y \rightarrow X)$ and handles non linearity by definition.

The TE is estimated using IDTxl (P. Wollstadt, 2018) bivariate analysis. Specifically, the bivariate analysis uses a greedy algorithm to identify the samples in the source's past which contribute significantly to the current value of the target. The estimation depends on the hyperparameter 'max lag sources' which defines the search space for candidate selection. The hyperparameter is set by analysing the autocorrelation as well as mutual information (between time series and its delayed version) plots. The tau (delay between the samples) is set as 2, which is greater than 1.5 (Vicente R, 2011). The search space for both source as well as target are taken to be same.

## Granger Causality

Granger Causality is a powerful metric to analyse functional interactions from time series data. (Seth AK, 2015). G - Causality is based on a simple idea that causes both precede and help predict their effects. G – Causality is estimated by comparing two VAR models. Let X, Y, Z be the time series data and we are interested to measure the information flow between X and Y. First a full VAR model is jointly estimated for all the variables and corresponding prediction errors are calculated. A reduced VAR model is estimated using only (Y, Z) and corresponding errors are estimated. If the prediction error of Y using full regression is considerably low when compared to error in reduced VAR, then X G – Causes Y.

The Granger Causality is estimated using MVGC toolbox. (Barnett L, 2014). The 'VAR' regression model is estimated using 'OLS' method and information criteria is extracted using 'LWR' mode. Model order is estimated using 'AIC' criteria. Alpha is set to be 0.05 for significance tests and 'FDR' is used for multiple hypothesis correction.

# Methods

## Effective Connectivity using Partial Directed Coherence (Liang, 2016)

The main challenge of detecting effective connectivity is that the measured signals have power spectra peaks in different frequency bands. To solve this, Multivariate Empirical Mode Decomposition is utilized to decompose a multi-channel signal into a set of oscillatory signals known as Intrinsic mode functions (IMFs). Due to precise separation of co-existing time scales, IMFs associated with diverse frequency bands can be used to investigate a causal effect between the signals.

MEMD is utilized to extract different frequency components from EEG signals forming Intrinsic mode functions. The effective connectivity is measured by applying partial directed coherence on the extracted IMFs. It is observed that best classification results are obtained when IMF1 or IMF3 alone were taken into consideration. Thus, only IMF1 is considered for EC estimation.

Given a set {X(t)} = $\{x_1(t), x_2(t), x_3(t) \ldots\ldots, x_N(t)\}$of simultaneously observed multichannel EEG signals described by following MVAR process:

$$\sum_{r=1}^{p} \Lambda(\mathrm{r})\mathrm{X}(\mathrm{t}-\mathrm{r}) = E(t) \text{ with } \Lambda(0) = I \qquad (2)$$

E (t) is a vector of multivariate zero-mean uncorrelated white noise process and Λ(r) is the N×N matrices of model coefficients, the dimension p in the model can be estimated using the Akaike criterion while the values of coefficients in Eq. (2) can be obtained by the Nuttall-Strand method, which is reported to be the best one among multivariate autoregressive estimators [9]. In order to investigate the spectral properties of the examined process, Eq. (2) is transformed to the frequency domain.

$$\Lambda(f)X(f) = E(f) \; where \; \Lambda(f) = \sum_{r=0}^{p} \Lambda(r)e^{-j2\pi f\Delta tr}$$

and PDC is defined as

$$\pi_{i<-j}^{2}(f) = \frac{|\Lambda_{ij}(f)|^2}{\sum_{k=1}^{N} |\Lambda_{kj}(f)|^2}$$

Thus, PDC provides influence of xj on xi via comparison of xj on all other variables.

## Effective Connectivity using Normalised Transfer Entropy (Shovon, 2014)

Normalised Transfer Entropy is more sensitive to cognitive tasks when compared to Transfer Entropy. Due to finite size and non-stationarity of EEG data, TE matrices contain large amounts of noise. Noise/bias has been removed from TE by subtracting the average transfer entropy from shuffled version of X to Y over several shuffles. This shuffled transfer entropy is denoted by $< TE_{X_{shuffle} \to Y} >$. Normalised Transfer Entropy is then calculated from X to Y with respect to total information in Y itself.

$$NTE_{X \to Y} = \frac{TE_{\,X \to Y} - < TE_{X_{shuffle} \to Y} >}{H(y_{t+1}|y_t)}$$

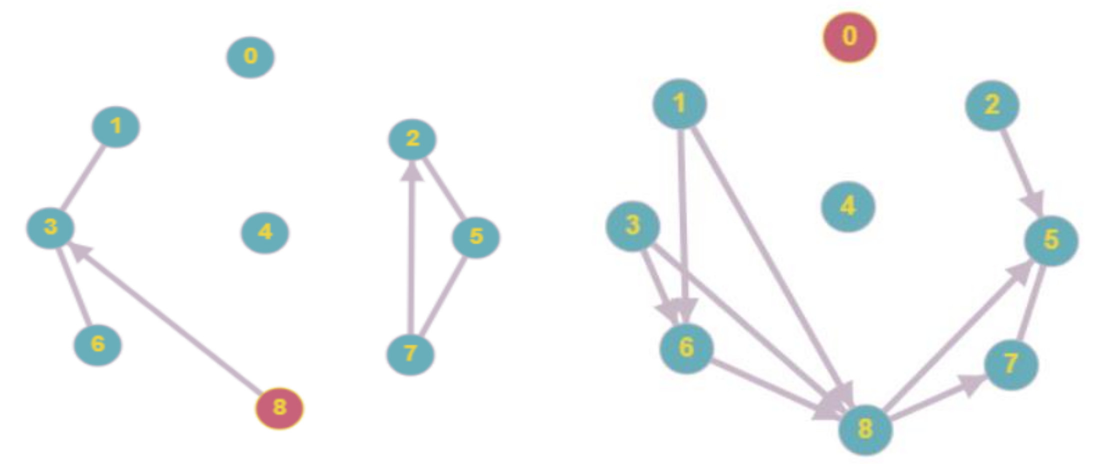
where $H(y_{t+1}|y_t)$ is the conditional entropy of y at t+1 given its value at t. Thus, NTE is in range of [0,1].

The embedding dimension in TE is estimated by observing the autocorrelation and mutual information (between the time series and its delayed version) plots. The tau(delay) is set to be 2. $< TE_{X_{shuffle} \rightarrow Y} >$ is estimated over 30 shuffled versions of X (Neymotin, 2011). The conditional entropy is estimated using pyitlib module.

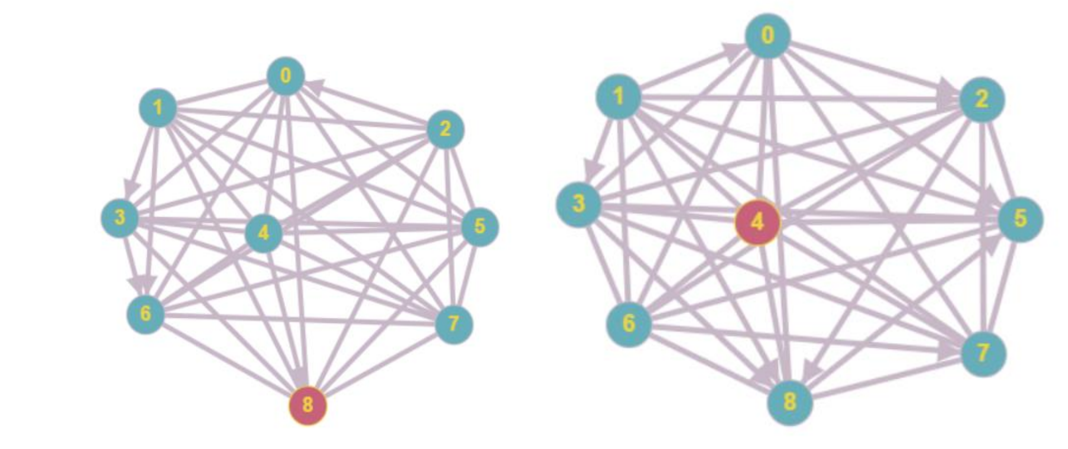
## Results and Discussion

The top 10 Effective Connections identified from Partial Directed Coherence method are plotted and compared with EC measured by Transfer Entropy.

Table 1




[2]




[3]

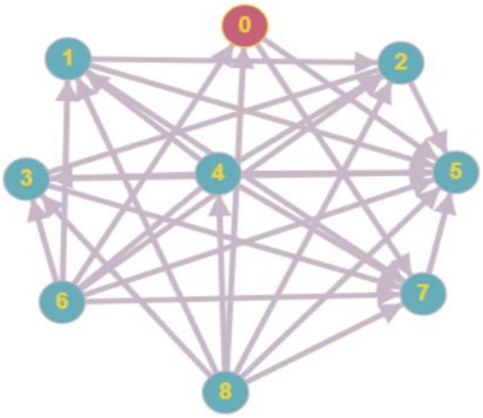
Table 1 shows that the most of the connections estimated by the Partial Directed Coherence are identified by the Transfer Entropy measurement also. The connections indeed vary a lot with subjects due to different baselines and experimental conditions. Hence, effective connections by Partial Directed Coherence are not suitable for inferring effective connections consistent with all subjects

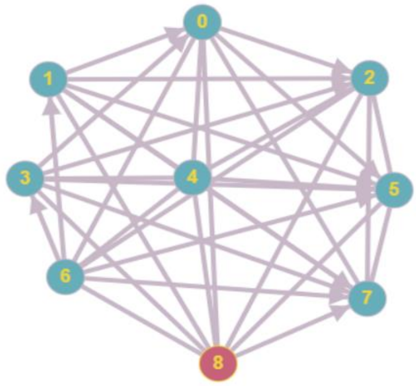
---

[2] Top 10 effective connections estimated by Partial Directed Coherence for subject 6

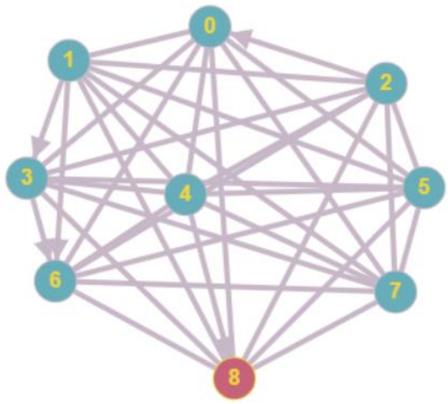
[3] Effective Connections estimated by Transfer Entropy for subject 6

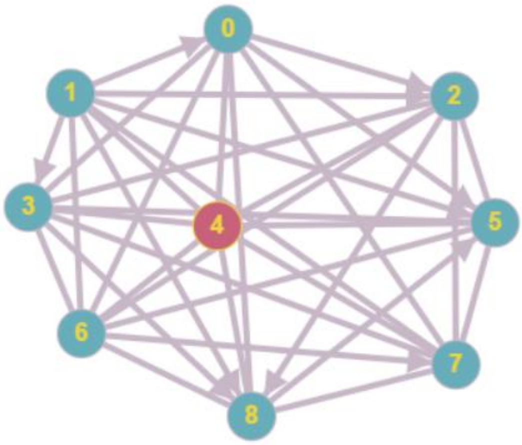
Table 2



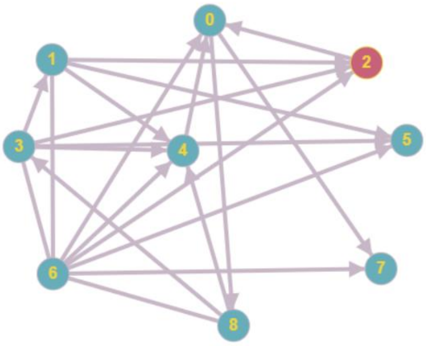



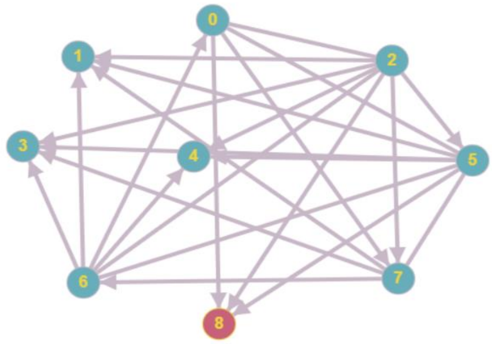

[4]




[5]




[6]

[4] Effective Connections estimated by NTE with threshold at 0.25 for subject 6
[5] Effective Connections estimated by Transfer Entropy for subject 6
[6] Effective Connections estimated by Granger Causality for subject 6

Table 3 Distance Metric

| | |
|---|---|
| Distance for Left MI (TE and NTE@0.025) of Subject 6 is 0.83 | Distance for Right MI (TE and NTE@0.025) of Subject 6 is 0.24 |
| Distance for Left MI (TE and NTE@0.03) of Subject 6 is 1.03 | Distance for Right MI (TE and NTE@0.03) of Subject 6 is 0.41 |
| Distance between left and Right (NTE@0.025) of Subject 6 is 0.65 | Distance between left and Right (NTE@0.03) of Subject 6 is 0.78 |

The normalised transfer entropy values for all the subjects lie in [0,1]. I analysed the NTE values by thresholding at 0.025 and at 0.03. Table 2 shows that the effective connections estimated by NTE (0.025 as threshold) share a lot of visual similarities with effective connections estimated by TE. Despite GC being effective in capturing linear dependencies, most of the connections captured by GC are also identified by NTE. The results also imply that NTE is able to capture information consistent to different subjects.

Table 3 indicates the graph edit distances between the left and right MI graphs of subject 6. The results indicate that NTE@0.025 produces more consistent information flow graphs and can help in observing the effective connections.

## Future Work

The future work aims at generating information flow graphs consistent with all the subjects using NTE as metric and comparing graph similarities using other metrics. This helps us in analysing prominent information flows in brain during motor imagery tasks and comparing it with standard literature.

## References


Barnett L, S. A. (2014). The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *J Neurosci Methods*(10.1016/j.jneumeth.2013.10.018), 50-68.

C. Brunner, R. L.-P. (n.d.). *Bnci-horizon-2020.* Retrieved from BCI competition 2008 - graz data set a.: http://bnci-horizon-2020.eu/database/data-sets

Gourévitch, B. a. (2007). Evaluating information transfer. *Neurophysiol*, 97, 2533–2543.

Granger, C. W. (Aug., 1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 424-438.

Korzeniewska, A. M. (n.d.). Determination of information flow direction among brain structures. *J. Neurosci. Methods*, 125, 195–207.

Liang, S. &.-S.-M.-A. (2016). Discrimination of motor imagery tasks via information flow pattern of brain connectivity. *Technology and health care : official journal of the European Society for Engineering and Medicine*, 24.

Nalatore H, D. M. (2007). Mitigating the effects of measurement noise on Granger causality. *Phys Rev E Stat Nonlin Soft Matter Phys*.

Neymotin, S. &. (2011). Synaptic information transfer in computer models of neocortical columns. *Journal of computational neuroscience*, 69-84.

P. Wollstadt, J. T.-Z. (2018). IDTxl: The Information Dynamics Toolkit xl: a Python package for the efficient analysis of multivariate information dynamics in networks. *Journal of Open Source Software, 4(34)*, 1081.

S. Sabesan, L. B. (2009). Information Flow and Application to Epileptogenic Focus Localization From Intracranial EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 17*(3), 244-253.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 461–464.

Seth AK, B. A. (2015). Granger causality analysis in neuroscience and neuroimaging. *J Neurosci, 35(8)*, 3293-3297.

Shovon, M. H. (2014). Towards a Cognitive Metric using Normalized Transfer Entropy. *ASE BigData/SocialInformatics/PASSAT/BioMedCom 2014 Conference, Harvard University.*

Vicente R, W. M. (2011). Transfer entropy--a model-free measure of effective connectivity for the neurosciences. *J Comput Neurosci.* , 30(1):45-67.