

2017-07-15

Advanced Correlation Grid: Analysis and Visualisation of Functional Connectivity among Multiple Spike Trains

Masud, MS

<http://hdl.handle.net/10026.1/9506>

10.1016/j.jneumeth.2017.05.016

Journal of Neuroscience Methods

Elsevier BV

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

This is the author's accepted manuscript. The final published version of this work: Advanced Correlation Grid: Analysis and Visualisation of Functional Connectivity among Multiple Spike Trains (the version of record) is published by Elsevier in the Journal of Neuroscience Methods, 15th July 2017, Pages 78-101 available at: [<http://dx.doi.org/10.1016/j.jneumeth.2017.05.016>].

This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

Article Details

- Accepted: 11th May 2017
- Available online : 12th May 2017
- Embargo Period: 18 months

Advanced Correlation Grid: Analysis and Visualisation of Functional Connectivity among Multiple Spike Trains

Mohammad Shahed Masud, Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka-1000, Bangladesh.

Roman Borisyyuk, Liz Stuart, Centre for Robotics and Neural systems, Plymouth University, Plymouth, UK.

Abstract

Background

This study analyses multiple spike trains (MST) data, defines its functional connectivity and subsequently visualises an accurate diagram of connections. This is a challenging problem. For example, it is difficult to distinguish the common input and the direct functional connection of two spike trains.

New method

The new method presented in this paper is based on the traditional pairwise cross-correlation function (CCF) and a new combination of statistical techniques. First, the CCF is used to create the Advanced Correlation Grid (ACG) correlation where both the significant peak of the CCF and the corresponding time delay are used for detailed analysis of connectivity. Second, these two features of functional connectivity are used to classify connections. Finally, the visualization technique is used to represent the topology of functional connections.

Results

Examples are presented in the paper to demonstrate the new Advanced Correlation Grid method and to show how it enables discrimination between (i) influence from one spike train to another through an intermediate spike train and (ii) influence from one common spike train to another pair of analysed spike trains.

Comparison with existing methods

The ACG method enables scientists to automatically distinguish between direct connections from spurious connections such as common source connection and indirect connection whereas existing methods require in-depth analysis to identify such connections.

Conclusions

The ACG is a new and effective method for studying functional connectivity of multiple spike trains. This method can identify accurately all the direct connections and can distinguish common source and indirect connections automatically.

Keywords: Functional connectivity, Correlation, Analysis, Visualisation

1 Introduction

The brain receives, processes, and transmits information regarding a particular stimulus through stereotyped electrical discharges called action potentials, or spikes. The signals which come from the stimulus, are transformed into sequences of spikes, at an early stage of processing within the central nervous system. Spike trains are the starting point for most of the processing performed by the brain (Kandel, 2000; Dayan & Abbott, 2001). Characterizing the relationship between the stimulus and the spike trains is an important issue in Neuroscience as

it underpins how the brain works in response to the stimulus. Many studies have been performed into this relationship between stimulus and spike trains (Espinosa & Gerstein, 1988; Gerstein & Kirkland, 2001; Gochin et al., 1990, 1991; Eggermont, 1991; Lindsey et al., 1992c; Vaadia et al., 1995; Wilson & McNaughton, 1994; Skaggs & McNaughton, 1996; Li et al., 1999; Shannon et al., 2000; Louie & Wilson, 2001; Pillow et al., 2008).

In addition to the relationship between stimulus and spike trains, it is also important to understand the functional connectivity between spike trains in response to a particular stimulus. This is another challenging problem within Neuroscience which could benefit from statistical methods to analyse multiple spike trains (Brown et al., 2004; Haslinger et al., 2013). In order to study the functional connectivity of the spike trains, it is essential to assess the spiking activity of multiple single neurons recorded simultaneously.

In Neuroscience, the Cross-Correlation Function (CCF) is a widely used measure of functional connectivity between spike trains (Perkel et al., 1967). The CCF has been applied to many neural systems in order to make powerful inferences about functional connectivity. Fundamentally, it is a statistical technique used to test the independence of two spike trains using the theory of stochastic point processes. This technique is also applied to assess oscillation, propagation delay, effective connection strength, synchronization, and the spatiotemporal structure of a network (Konig et al., 1995; Brown et al., 2004; Pillow et al., 2008; Nicolic et al., 2012).

In order to make inferences from the CCF, Brillinger (1976) introduced a normalization technique for the CCF using a confidence interval. Thus, peaks exceeding the confidence interval of the CCF are considered to be significant. A peak in a CCF indicates that there is a high probability that a spike in one spike train is caused by a spike in another spike train with some time delay involved. The significant peak in the CCF indicates that the null hypothesis on independence of two spike trains is not supported by the data and should be rejected. Consequently, there is an influence from one spike train to another. However, the interpretation of this influence, in terms of functional connectivity, is challenging. This is due to the fact that this influence can be considered to be (i) a direct connection between two spike trains, (ii) the result of some common source to both spike trains or (iii) an indirect connection, defined as a connection via some intermediate neuron.

There are several methods in the literature for the analysis of multiple spike trains (for example, Pillow et al., 2008; Stevenson, 2008; Grün & Rotter, 2010; Kriener et al., 2009; Masud & Borisyuk, 2011; Reimer et al., 2012; Jovanović & Rotter, 2016). One such method is the correlation grid (Stuart et al., 2005). The correlation grid is a visualization technique used to analyse the synchronous firings of simultaneously recorded multiple spike trains. The fundamental idea of this technique is to arrange spike trains into clusters that are functionally connected and display them in a symmetrical grid. A measure of distance, based on normalized CCF of two spike trains is used to perform the cluster analysis.

The correlation grid has been successfully used for the study of functional connectivity. However, the correlation grid cannot automatically distinguish between direct and spurious (both indirect and common source) connections. The aim of this paper is to present a statistical method called the ‘Advanced Correlation Grid (ACG)’ to analyse the functional connectivity of a large number of spike trains (15 – 1000 spike trains) using the CCF. The main advantage of the ACG method is that ACG makes it possible to define an accurate diagram of functional connections. More specifically, the ACG method can reliably differentiate direct connections from spurious (indirect and common source) connections using an automatic algorithm.

Section 2 of this paper describes the CCF in detail. Then Section 3 reviews the original correlation grid. Following this, Section 4 describes functional connectivity and Section 5 describes the ACG in detail. Sections 6 and 0 present case studies to report the operational of the ACG in detail. The case studies use data generated by ELIF (Enhanced Leaky Integrate and Fire) model (Borisyuk, 2002). The first case study consists of a small set of fifteen spike trains. In this set, all the connections have medium strength of influence with one exception, a single connection with very strong influence. The second case study consist of a large set of fifty spike trains in which all the connection strengths are of medium influence. The effectiveness of the ACG method is presented in Section 8. In order to study the accuracy of the method different scenarios of spike train data set such as same strength of influence, low noise and high noise are considered. In this section, the result of the functional connectivity obtained by the AVG method is compared to an existing called the Cox method. Section 9 presents the application of ACG to the experimental data recorded from the visual cortex of the cat. Conclusions of the work are presented in Section 10. Finally, Section 11 presents the description, dynamics and the parameter values for the ELIF generator.

2 Cross Correlation Function

The CCF algorithm (Masud et al., 2011) is applied to a pair of spike trains A and B where it is assumed that these spike trains are stationary. One spike train is arbitrarily assigned to be the target spike train and the other becomes the reference spike train. A correlation window is defined as $(2 * u + 1)$ bins of short time intervals h , where both h and u are values selected by the investigator. For each spike, on the reference spike train, the correlation window is positioned such that its centre is directly aligned with the current spike. Thus, there are u bins to the left and right of the current spike; the correlation window is effectively centred over that spike. Refer to Figure 1, where $h = 1\text{ms}$ and $u = 2$ for the purpose of presenting the algorithm only. For each spike on the reference spike train (B), the counting function $n_{AB}(v)$ counts and accumulates the number of times that spikes on the target train (A) coincide with the current reference spike. Thus, the counting function $n_{AB}(v)$ is calculated over the recording time T .

In order to test the independence of two spike trains, Brillinger (1976) proposed the estimate $\hat{\rho}_{AB}(v) = \sqrt{\hat{p}_{AB}(v)/\hat{p}_A\hat{p}_B}$, where $\hat{p}_{AB}(v) = n_{AB}(v)/2hT$, $\hat{p}_A = n_A/T$ and $\hat{p}_B = n_B/T$. This normalises the counting function $n_{AB}(v)$ accordingly. Here, n_A and n_B denote the number of spikes in the spike trains A and B , respectively.

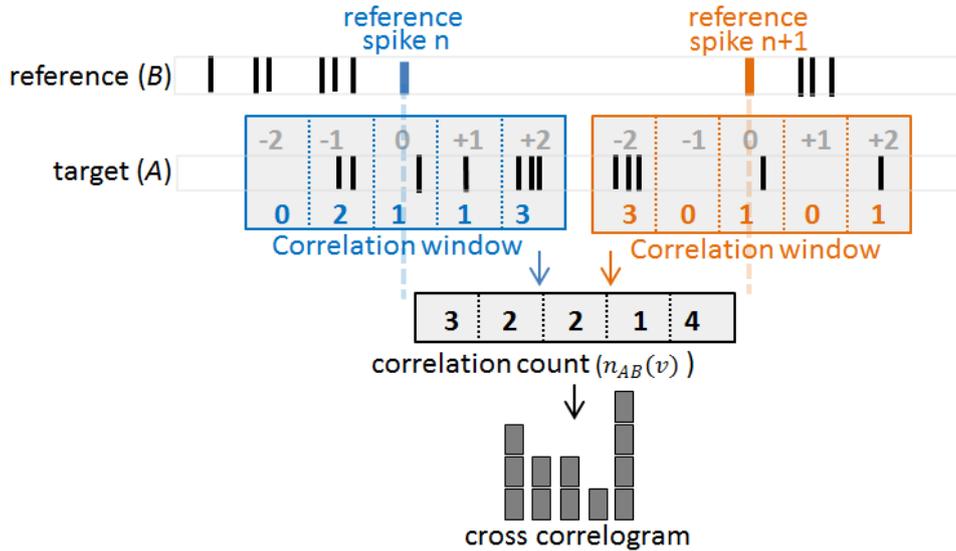


Figure 1: Visual representation of the CCF algorithm for two spike trains A (the target spike train) and B (the reference spike train). The correlation count function denotes the summation of coincident spikes for reference spikes n and $n+1$ only

For a large sample size the random variables $\hat{\rho}_{AB}(v)$ are independent and their distribution is the normal with mean $m = \sqrt{\hat{p}_{AB}(v)/\hat{p}_A\hat{p}_B}$ and standard deviation $s = 1/(2\sqrt{2hT\hat{p}_A\hat{p}_B})$.

Thus, when spike trains A and B are independent, the mean of $\hat{\rho}_{AB}(v)$ is equal to one, since $\hat{\rho}_{AB}(v) = \hat{\rho}_A \hat{\rho}_B$.

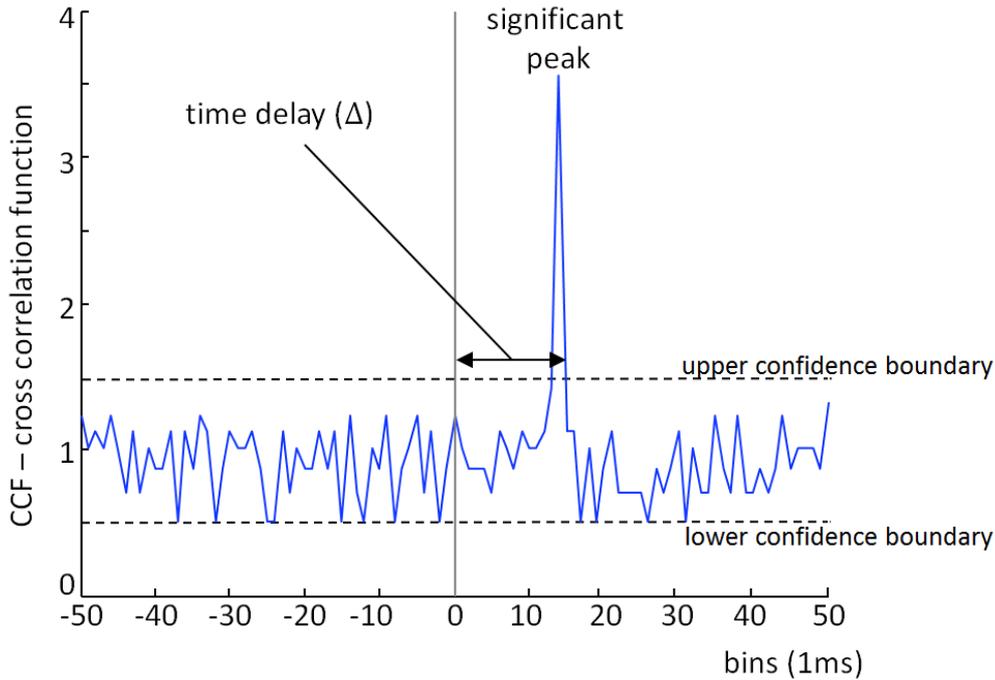


Figure 2: A sample cross-correlation function of two spike trains which depicts the confidence interval, significant peak and time delay (Δ).

The null hypothesis H_0 states that the two spike trains are independent. An alternative hypothesis H_1 is that there is dependence, at least for some time shift (bin), between spike trains. To test this hypothesis, the CCF values for all bins are considered. If these values are sufficiently small (inside the confidence interval) then the data does not contradict the H_0 hypothesis. The method used to calculate the confidence interval for testing this hypothesis was defined by Brillinger (1979). The boundaries of the confidence interval at the significance level α are plotted by two horizontal lines at levels $1 \pm Q_{cr}^\alpha / (2\sqrt{2hT\hat{\rho}_A\hat{\rho}_B})$, where Q_{cr}^α is the critical value of the normal distribution corresponding to the significance level α . If H_0 is correct then all values of the CCF should fall inside the confidence interval and the estimated value of the CCF ($\hat{\rho}_{AB}(v)$) must be zero. If some value of the CCF exceeds the upper boundary of the confidence interval, then the null hypothesis H_0 must be rejected. Thus, it is concluded that the two spike trains are not independent. The peak is defined by the values of the normalised CCFs which lie outside the confidence interval. Each of these peaks is characterised by the corresponding bin which defines the position of the peak. The bin is selected in order to maximise deviation from the upper boundary of the significance interval. This peak is referred

to as the significant peak. Note that if there is more than one significant peak in the cross-correlation function, then the highest significant peak is considered to be the main peak.

A sample cross-correlation function of two spike trains is shown in Figure 3. The highest value of the CCF, which exceeds the upper boundary of the confidence interval, is assessed as being a measure of the strength of influence that is exerted from spike train *A* onto spike train *B*. Furthermore, the corresponding time shift is assessed as being a time delay (Δ) in the propagation of the spike from *A* to *B* (Nikolić, 2007). Note that the value of the main peak in the CCF is subsequently used to (i) quantify the distance between pairs of spike trains in a clustering algorithm and (ii) represent the density of shading in the corresponding cell of the correlation grid.

3 The Correlation Grid

The Correlation Grid (Stuart et al., 2005) was developed to provide an organised yet abstract overview of a large number of CCFs. Each of the underlying CCFs represents the synchrony of a pair of spike trains during a specific time period. Therefore, the Grid provides an overview using a simple two dimensional grid which enables investigators to identify groups of “associated” neurons. Associated neurons are those which have a high probability of generating spikes, approximately simultaneously or within a constant time delay, of each other.

3.1 Creating the Correlation Grid

For a given dataset, of n spike trains, all pair wise CCFs are generated and normalised using the Brillinger method, and the main peaks $c_{i,j}$ ($i = 1, \dots, n, j = 1, \dots, n$) are calculated for all pairs. Finally, the results are displayed as an n -by- n symmetrical grid of grey scale cells, representing the strength of influence between all pairs of spike trains (Figure 3). Thus, the magnitudes of main peaks are encoded from white, representing a non-significant peak, to black, representing the largest peak in the grid. The user has the flexibility to view ‘all peaks’ or solely significant peaks. Significant peaks are those that exceed the higher bound of the confidence interval.

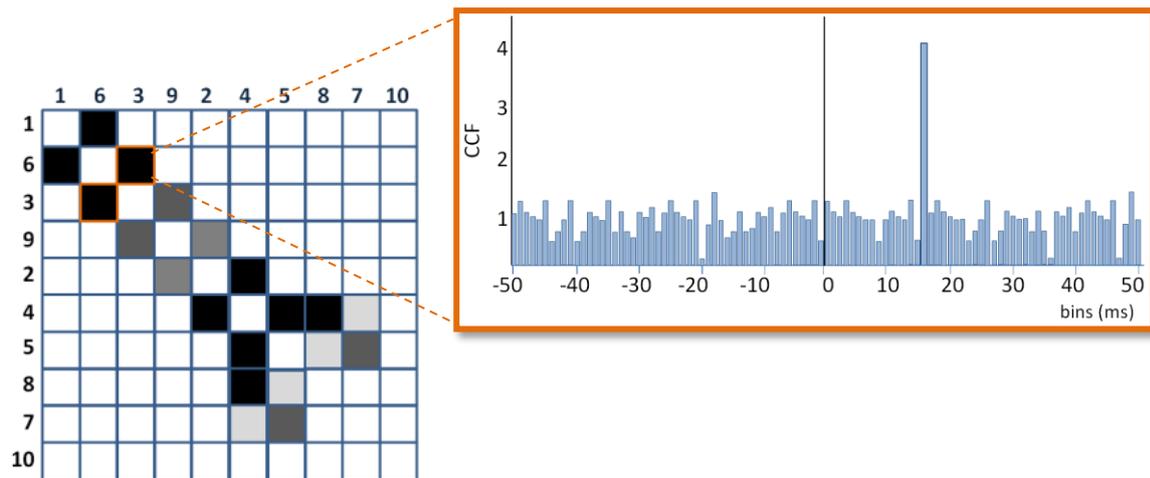


Figure 3: A typical Correlation Grid where the number of spike trains n is 10. In this example, only significant peaks are represented (encoded by the depth of grey scale) in the display. The highlighted portion of the display (shown in orange) shows the CCF of spike trains 3 and 6. Note that the grid has been reordered correlation grid. Note that all CCF's were generated using a bin size of 1 millisecond and a window size of 100ms.

Additionally, it is useful to reorder the correlation grid in order to highlight the inherent relationships between multiple spike trains. The method used to accomplish this reordering is the furthest neighbour cluster analysis. A detailed description of the creation of the Correlation Grid and the clustering algorithm is available in the original paper (Stuart et. al, 2005).

4 Functional connectivity

It has been established (Brown et al., 2004) that identifying the connectivity between neurons is crucial to our understanding of how the brain works. It is useful to distinguish the structural (physical) connection between two neurons (electrical gap junction or synaptic connection) from the functional connection which means that there is a relationship (statistical dependence/correlation) between spiking activities of these neurons (Sporns, 2007). This paper concentrates on functional connectivity of multiple spike trains. The goal is to develop an effective statistical method for finding functional connections (statistical dependencies) and visualisation of the resulting connection diagram. This approach is based on using the traditional cross correlation function. Although the CCF has been successfully used to identify functional connectivity in neuronal assemblies (Stuart et al., 2005), these investigations were mostly limited to a small group of neurons. The subsequent introduction of the Correlation Grid enabled investigators to manage larger numbers of neurons. However, as the scale of investigations increased, it became increasingly difficult to distinguish between different types of connections (Dahlhaus et al., 1997; Eichler et al., 2003; Makarov et al., 2005; Nykamp, 2005; Stevenson et al., 2008; Park et al., 2008; Nedungadi et al., 2009). The three different

types of connection between two spike trains are defined as (i) a direct connection (ii) an indirect connection and (iii) a ‘common source’ connection. A ‘direct connection’ is a connection whereby one neuron modulates the firing pattern of another neuron directly. An ‘indirect connection’ is defined as a spurious functional connection due to connectivity via an additional intermediate neuron. Finally, a ‘common source’ connection is defined as a spurious connection due to an influence from the common source to both neurons in a pair. In order to derive the accurate topology of functional connections between a group of neurons (spike trains), in addition to direct connections it is essential that all spurious connections, both common source and indirect connections, are identified. The remainder of this paper presents a new combination of statistical techniques for the automated statistical solution to this problem.

4.1 Classification of functional connection

When there are no significant peaks in a CCF, it is simple to deduce that there is no connection between the two neurons. Alternatively, when a CCF has a significant peak, then its interpretation is more complicated. When a peak exists, it can be deduced that there is some influence from one neuron to another. It is unclear whether this influence is the result of direct coupling or spurious connectivity. Historically, breakthroughs have shown that two additional measurements from the CCF can be used to distinguish between direct and spurious connections. These are (i) the height of the highest significant peak (ρ) (Aertsen et al., 1989) and (ii) the corresponding time delay (Δ) (Nikolić, 2007). The three different types of connection are shown in Figure 4.

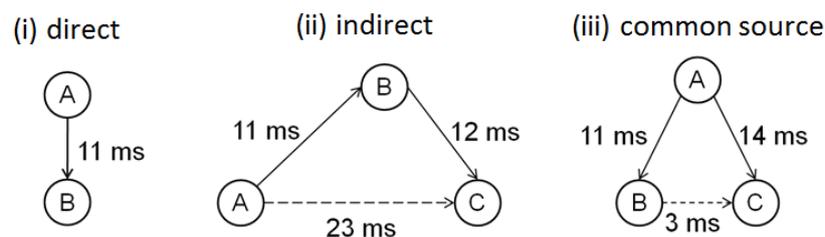


Figure 4: Schematic diagram of (i) Direct connection where neuron A is directly connected to neuron B, (ii) Indirect connection where neuron A is connected to neuron C through neuron B and, (iii) Common Source connectivity, where neuron A is connected to both neuron B and C.

Figure 4(i) shows an example of a direct connection from neuron A to neuron B. This connection has a delay of 11 milliseconds in spike propagation. Figure 4(ii) shows an example of an indirect connection from neuron A to neuron B with a delay of 11 milliseconds. Additionally, it shows a time delay of spike propagation from neuron B to neuron C which is 12 milliseconds. Note that in this case, an indirect connection exists from neuron A to neuron

C with a delay of 23 milliseconds. When compared to direct connections, a relatively lower value for the significant peak and a relatively longer value for the time delay are important characteristics for identifying indirect connections.

Figure 4(iii) shows a ‘common source’ connectivity diagram of 3 neurons. Neuron *A* influences both neurons *B* and *C*, with a delay in spike propagation of 11 and 14 milliseconds, respectively. Note that in this case, a common source connection exists from neuron *B* to neuron *C* with a delay of 3 milliseconds. When compared to direct connections, a relatively lower value for the significant peak and a relatively shorter value for the time delay are important characteristics for identifying ‘common source’ connections.

5 Description of ACG

The Advanced Correlation Grid (ACG) is a modification to the original correlation grid aimed at differentiating between direct connections and spurious (indirect and common source) connections. This advanced functionality is achieved using an automated algorithm. Each of the five steps of ACG is described in detail.

5.1 Step 1 - Calculation of the CCF

Let us consider n simultaneously recorded neurons and $k = (n^2 - n)/2$ is the number of CCFs. To test the hypothesis of independence of two spike trains, k independent tests are run for the n spike trains. However, note that in statistics a multiple comparison problem occurs when a set of simultaneous statistical tests are performed. This problem occurs when all the tests are considered as a family and the significance level α is set for the entire family of statistical tests. For this reason, the error of incorrectly rejecting the null hypothesis is more likely to occur. Nevertheless, several statistical techniques have been developed to overcome this known problem.

The Bonferroni correction (Abdi, 2007) is one such method that can be used to address the problem. The Bonferroni correction is based on the idea that when testing a set of dependent or independent hypotheses, the significance level α should be adjusted according to the number of tests performed. Thus, if the significance level for a set of k simultaneous tests is considered to be α , the significance level for each individual test will be α/k .

When applying the Bonferroni correction to the k pairwise CCFs, the upper and lower boundaries of the confidence interval are calculated for each pairwise CCF. Any peak that exceeds the upper boundary of the confidence interval is considered to be significant.

Significant peaks can be found on both the positive and negative side of the CCF. Here, a significant peak on the positive side of the CCF is considered to be a measure of the dependence of one spike train on another. If there are several significant peaks, then the highest significant peak is considered to be the measure of influence strength. Subsequently, all of the highest significant peaks ρ_{ij} ($i, j = 1, 2, \dots, n$), $i \neq j$ and the corresponding time delays Δ_{ij} ($i, j = 1, 2, \dots, n$), $i \neq j$ are calculated for n spike trains. Non-significant peaks ρ_{ij} of the pairwise CCF support the null hypothesis (namely that the spike trains are independent). Therefore, there is no connection from neuron i to neuron j and so ρ_{ij} is not distinguishable from zero. These non significant peaks are not included in the analysis.

5.2 Step 2 – Detection of Outliers

In order to detect direct connections within the neuronal assemblies, it is necessary to distinguish these connections from the spurious connections (common source and indirect). To achieve this goal three groups (direct, common source and indirect connection) are identified using cluster analysis. In the cluster analysis two measurements are considered namely significant peak and time shift obtained from the CCF's. Within neuronal assemblies, some very strong synaptic connections may exist between neurons. The corresponding CCFs may show some very large significant peaks which can be considered to be outliers as they deviate significantly from the other significant peaks. These “outlier connections” can be ascribed to come from direct connections.

When outliers are present, cluster analysis may lead to an incorrect conclusion. For example, cluster analysis produces three clusters: (i) significant peaks having outlier with moderate time shifts (ii) significant peaks with moderate and small time shifts (ii) significant peaks with either small or large time shifts. The problem here is that cluster analysis may result in some direct connections being wrongly classified as either common source or indirect connections. To avoid this misleading conclusion, outlier connections are identified first and classified as direct connections. The remaining significant peaks associated with time shifts are used to classify the cluster of direct, the cluster of common source and the cluster of indirect connections. Note that whilst these outlier connections are not used in step 3 of this process, they are crucial to step 4, when the topology of the neuronal assembly is verified.

5.2.1 The Z-score

Among several outlier labelling methods, a commonly used method is the Z-score. The Z-score is defined as

$$Z_i = \frac{x_i - \bar{x}}{sd}$$

where $X_i \sim N(\mu, \sigma^2)$, that is X_i is normally distributed with mean μ and variance σ^2 and \bar{x} , and sd , are the sample mean and sample standard deviation of data, respectively. The basic idea of the Z-score rule is that if X follows a normal distribution with mean μ and variance σ^2 , then Z follows a standard normal distribution with a mean of zero and a variance of one. Any Z-scores that exceed three (in absolute values) are generally considered to be outliers. According to Schiffler (1988), a maximum Z-score is dependent on the sample size, and it is computed as $(n - 1)/\sqrt{n}$, where n is the sample size. Since no Z-score exceeds 3 in a sample size less than or equal to 10, the Z-score method is not very good for outlier labelling, particularly in small data sets (Iglewicz and Hoaglin, 1993). Another limitation of the Z-score rule is that the standard deviation can be inflated by a few or even a single observation having an outlier value. Thus, it can cause a masking problem.

Two estimators are used in the Z-Score, the sample mean and sample standard deviation. These estimators can be affected by a few outlier values or by even a single outlier value. To avoid this problem, another outlier labelling method known as modified Z-score can be used.

5.2.2 The modified Z-score

In the modified Z-score, the absolute deviations of the sample observations from the median, are calculated. The median of these deviations (MAD) is used instead of the standard deviation and the median is used instead of mean of the sample, respectively (Iglewicz and Hoaglin, 1993), such that

$$MAD = median\{|x_i - \tilde{x}|\}$$

where \tilde{x} is the sample median. The modified Z-Score (M_i) is computed as

$$M_i = \frac{0.6745(x_i - \tilde{x})}{MAD}$$

where $E(MAD) = 0.6745 \sigma$ for large data. Iglewicz and Hoaglin (1993) suggested that observations are labelled as outliers when $|M_i| > 3.5$. Applying the modified Z-score to the significant peaks ρ_{ij} , obtained from the pairwise cross-correlation function, identifies any outliers that exist. Within the scope of this study, only outliers above the upper boundary were investigated. Only the outliers that lie above the upper boundary are of interest as they represent outliers with very strong connections which can be considered to be direct connections. For this reason, significant peaks are labelled outliers when $M_i > 3.5$.

5.3 Step 3 - Hierarchical Cluster analysis

All of the non-outlier significant peaks ρ_{ij} and the corresponding time delays Δ_{ij} are used to classify functional connections. For a set of significant peaks ρ_{ij} , that do not have outliers with corresponding delays Δ_{ij} , the typical scatter plot is shown in Figure 5.

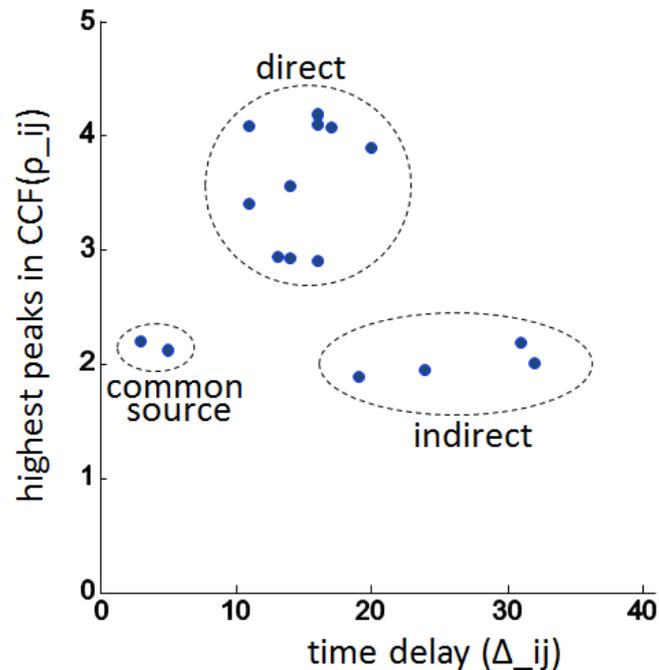


Figure 5: An example of a typical scatter plot of a set of significant peaks ρ_{ij} and the time delays Δ_{ij} where the classification of direct connection, indirect connection and common source are indicated by circles.

From this figure, it can be deduced that the direct connections are identified as the group of peaks with high significant peak and moderate time delay. Common source connections can be identified as those with low significant peaks and short time delays. Finally, indirect connections are characterised by low significant peak and large time delay. A hierarchical clustering algorithm is applied to classify the set of significant peaks ρ_{ij} and time delays Δ_{ij} .

The hierarchical clustering algorithm is used to find clusters of similar objects within a data set. A hierarchical cluster analysis begins by calculating the “mathematical” distances among objects in the data set. A variety of measures can be used to calculate the distance between significant peaks in the CCF. Note that for data that exhibits linear relationships, the Euclidean distance is a widely adopted measure. Initially, each object is in a cluster of its own. Clustering begins by finding the two clusters which are most similar i.e. closest, based on the Euclidean distance between them. Once these two clusters are identified, they are merged into a single cluster. Note that the characteristics of this new cluster are based on a combination of the objects in the cluster.

This procedure, of combining two clusters and merging their characteristics, is repeated until all the objects have been merged into a single large cluster. The results of hierarchical clustering are typically displayed in the form of a two dimensional diagram known as a dendrogram (or tree diagram). Within the dendrogram, branches in the tree represent clusters. A variety of linkage methods can be used to determine the order in which clusters merge together such as single linkage, complete linkage and average linkage.

To apply cluster analysis to the set of significant peaks ρ_{ij} and time delays Δ_{ij} , both measurements need to be normalized, so that the values of significant peaks and time delays are in the range between 0 and 1. This normalization is carried out to ensure that the values of significant peaks and time delays are not affected by differences in the scales of measurement. A clustering algorithm is applied to these normalized significant peaks and time delays. In the clustering algorithm, the distance between pairs of observations is calculated using the Euclidean distance and the average linkage is used for calculating the distance between two clusters. The average linkage is used due to its reduced sensitivity to outliers. The aim of the clustering algorithm is to divide the connections into three clusters: direct connections, common source and indirect connections. Note that for a data set with two or more variables, distances are significantly affected by differences in the scales of measurement used. Therefore, it is good practice to transform each of the variables to ensure their scales are similar.

5.4 Step 4 – Verifying the classification of connections

At this stage in the process, the CCF of each neuron pair is available. The CCF has been used to ascertain whether a significant peak exists and if it does, to quantify any time delay. Each significant peak is characterised by the size of the peak (ρ) and the time delay (Δ). Subsequently, outliers are detected in the resultant set of significant peaks with associated time delays. Recall that these outliers are ascribed to direct connections and so they are crucial to this part of the process. It is both these outlier connections in conjunction with the cluster of direct connections that are used to restructure the neural assembly. The final assembly is displayed using an n-by-n matrix of functional connectivity. In this matrix of functional connectivity, each row represents the target neuron and each column represents the reference neuron.

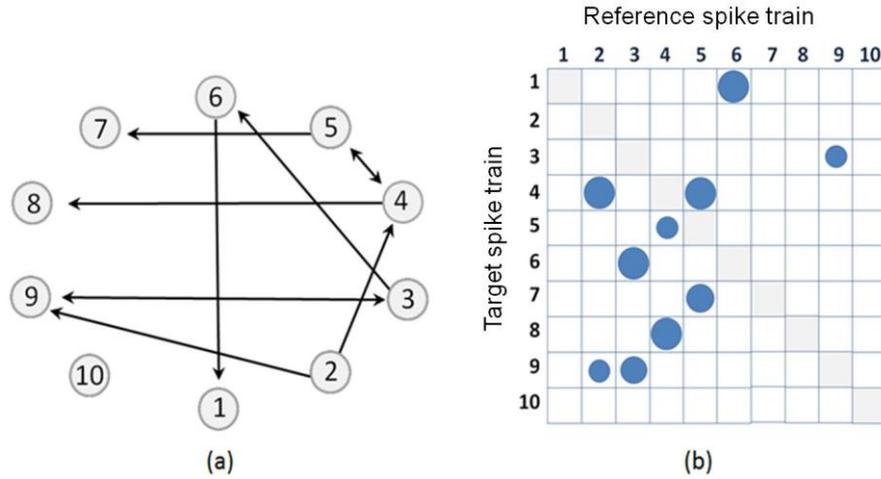


Figure 6: (a) Connection scheme of the ten neurons. (b) An example of a direct functional connectivity matrix of these ten neurons obtained by CCF.

Figure 6(a) shows a sample topology for a data set of ten neurons generated by ELIF model (see Appendix 1). Figure 6(b) depicts the corresponding matrix of functional connectivity, also known as the Advanced Correlation Grid (ACG). In this grid, a circle in the cell (i, j) indicates that there is a direct connection from neuron j to neuron i . Additionally, the radius of the circle represents the normalized strength of connection between this pair of neurons, which is proportional to the height of the significant peak in the corresponding CCF. From this grid, it is clear that ten direct functional connections exist.

In the final step of the ACG algorithm, all of the significant connections in the common source cluster are investigated in detail. This is to verify that these connections are the result of common sources of input. Full details of this verification are given in Appendix 1. Once verification of the common source connections is complete, verification of the indirect connections is performed. Full details of this verification are given in Section 5.4.2. Note for clarity, let us refer to a group of three neurons as a triplet.

5.4.1 Common source connections

Initially, a common source connection is selected and then assessed. The aim of this assessment is to verify that a common source connection exists from neuron i to neuron j . Whilst neurons i and j remain fixed, all possible triplets (i, j, k) , $k = 1, 2, \dots, n$ are identified. Additionally, the corresponding triplet time delays $(\Delta_{ij}, \Delta_{ki}, \Delta_{kj})$ are calculated from the CCF. As shown in Figure 7, time delays $(\Delta_{ki}, \Delta_{kj})$ indicate the delay of a direct connection from neuron k to neurons i and j , respectively.

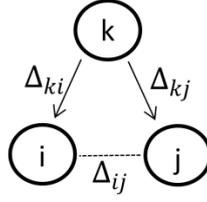


Figure 7: Triple $t(i,j,k)$ denotes three neurons i , j and k and their connectivity. The connection (i,j) represents the possibility of a common source connection existing between i and j .

The time delay Δ_{ij} indicates the delay of the connection from neuron i to neuron j . The connection from neuron i to neuron j is considered to be a common source connection if it satisfies the following criteria:

$$\Delta_{ij} = |\Delta_{ki} - \Delta_{kj}| \quad (1)$$

This verification procedure is repeated for all of the significant connections in the common source cluster.

5.4.2 Indirect connections

Initially, an indirect connection is selected and assessed. The aim of this assessment is to verify that an indirect connection exists from neuron i to neuron j .

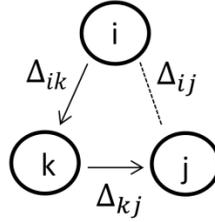


Figure 8: Triple $t(i,j,k)$ denotes three neurons i , j and k and their connectivity. The connection (i,j) represents the possibility of an indirect connection existing between i and j .

Whilst neurons i and j remain fixed, all possible triplets (i,j,k) , $k = 1,2,\dots,n$ are identified. Additionally, the corresponding triplet time delays $(\Delta_{ij}, \Delta_{ki}, \Delta_{kj})$ are calculated from the CCF (see Figure 13). The connection from neuron i to neuron j is considered to be an indirect connection if it satisfies the following criteria:

$$\Delta_{ij} = \Delta_{ik} + \Delta_{kj} \quad (2)$$

This verification procedure is repeated for all of the significant connections in the indirect connections cluster.

5.4.3 Tolerance limits

To assess the prescribed tolerance (ϵ) in time delays (given in milliseconds), the common source connections criteria (1) and the indirect connections criteria (2) can be rewritten as shown in (3) and (4):

$$|\Delta_{ki} - \Delta_{kj}| - \epsilon \leq \Delta_{ij} \leq |\Delta_{ki} - \Delta_{kj}| + \epsilon \quad (3)$$

and

$$\Delta_{ik} + \Delta_{kj} - \epsilon \leq \Delta_{ij} \leq \Delta_{ik} + \Delta_{kj} + \epsilon \quad (4)$$

Note that both the verified common source and indirect connections obtained from (3) and (4) are subsequently displayed in the Advance Correlation Grid. The assumptions which have been formulated here are in line with current presented (Nikolić 2007). Note that these connections are encoded using a different colour (see Figure 12) so they can be easily distinguished.

6 Case Study 1 - functional connectivity of fifteen neurons

This case study uses a data set generated using the ELIF model. The dataset is made up of fifteen neurons, generated over a period of 30,000 milliseconds using the connection architecture shown in Figure 9(a) with the connection strengths visualised in Figure 9(b). From this figure, it is clear that a total of sixteen connections exist between all of the neurons. Note that the connection from neuron #5 to neuron #13 has the greatest strength. Figure 10 depicts the raster plot of spiking activity. The purpose of this case study is to illustrate the effectiveness of ACG and to demonstrate how it can be used to derive topologies based solely on spike train data sets. It is crucial to reiterate that only the spike trains are made available; the schemes of connections as well as all neuron characteristics are hidden. Note that in this research, self-coupling is not used to deduce functional connectivity of the fifteen spike trains.

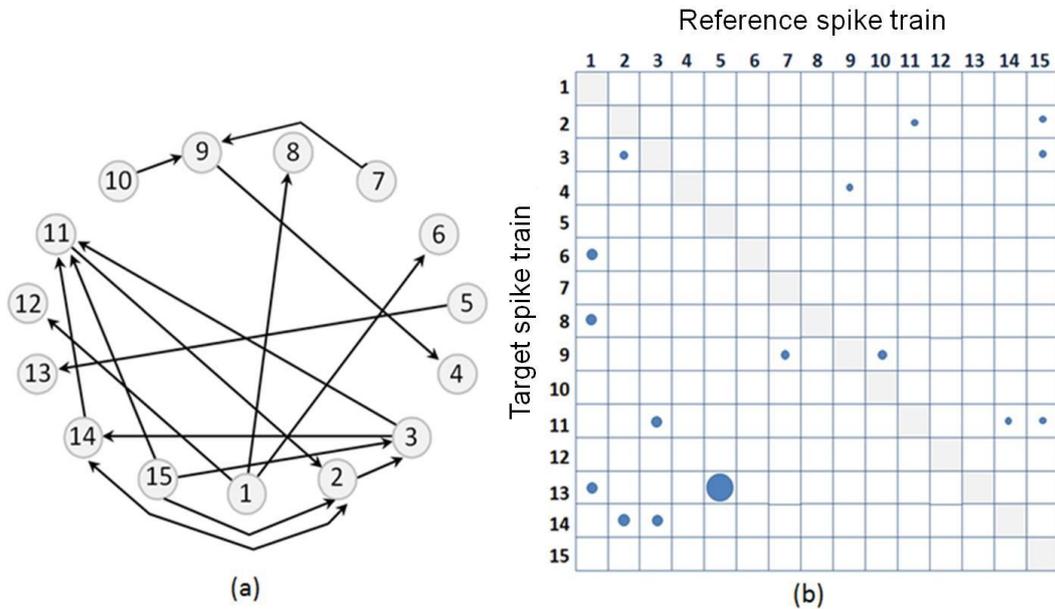


Figure 9: (a) Connection scheme of fifteen neurons. There are sixteen non zero connections which are shown by arrows. (b) Actual connection strengths of the sixteen non zero connections in the matrix format.

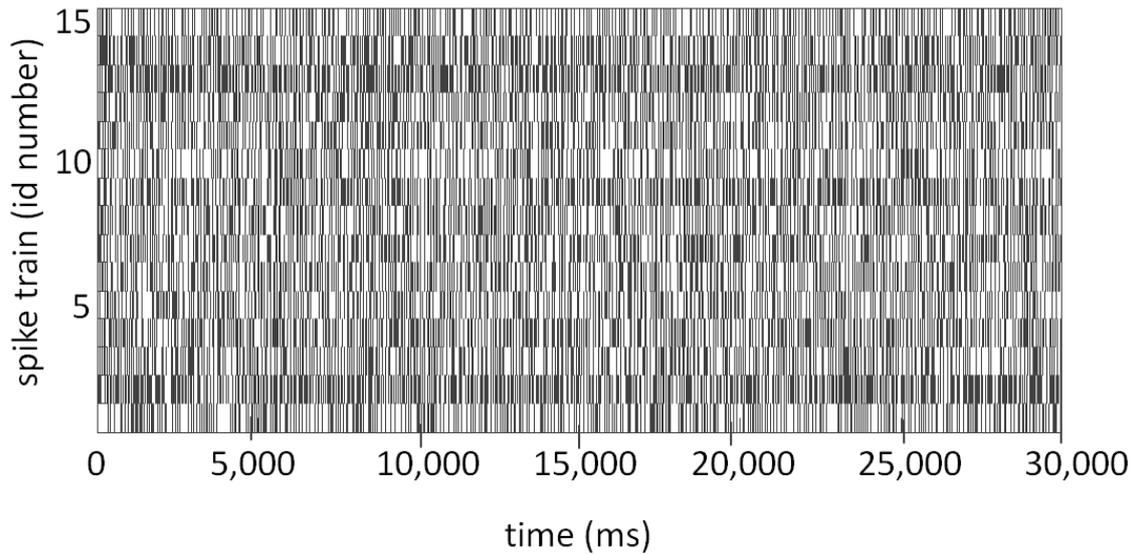


Figure 10: Raster plot of the fifteen spike trains, generated over a period of 30,000 milliseconds, with the connection architecture shown in Figure 9(a).

6.1 Calculation of CCF

For this study made up of fifteen spike trains, a total of $(15^2 - 15)/2 = 105$ pairwise CCFs are calculated. A bin size of 1 millisecond and a correlation window of 100 milliseconds are used. In order to test the independence of the pairwise spike trains, the level of significance $\alpha = 0.05$ is used alongside the Bonferroni correction. Thus, a connection is deemed to be significant if a peak in the CCF exceeds the upper boundary of the confidence interval. At this stage in the process, a total of 25 significant connections are identified for all fifteen spike trains (refer to Table 1).

Reference	1	1	1	2	2	2	3	3	5	6
Target	6	8	12	3	11	14	11	14	13	8
Peak	4.37	4.57	3.05	3.50	1.64	3.26	2.73	3.23	6.52	2.44
Time delay	13	15	11	12	27	14	12	12	12	2
Reference	7	7	9	10	10	11	12	12	12	14
Target	4	9	4	4	9	2	4	6	8	11
Peak	1.72	3.57	3.27	1.74	3.46	2.57	1.69	1.90	1.92	2.98
Time delay	23	11	12	25	13	13	23	2	4	13
Reference	14	15	15	15	15					
Target	15	2	3	11	14					
Peak	1.73	2.71	3.57	3.73	2.14					
Time delay	40	12	14	11	26					

Table 1: Significant connections of the fifteen neurons with peak and time delay. Connections are indicated from reference spike train to the target spike train.

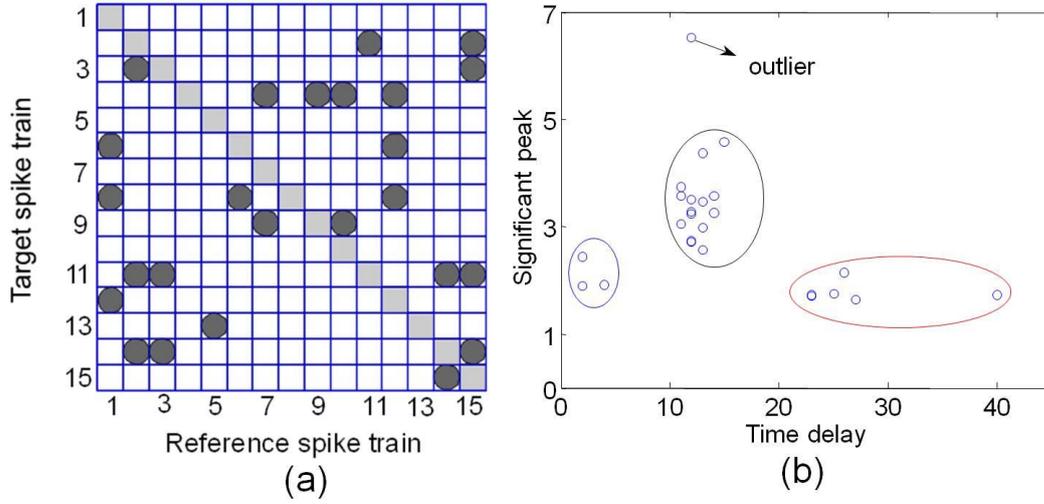


Figure 11: (a) Significant connections obtained from pairwise CCF analysis. A total of 25 significant connections are obtained from the fifteen neurons. (b) Scatter plot of these 25 significant connections showing the significant peak with corresponding time delay.

Table 1 shows all the significant connections of the fifteen neurons alongside their corresponding peak heights and time delays. The height of significant peaks ranges from 1.64 to 6.52. The time delay ranges from 2 milliseconds to 40 milliseconds. These significant connections are also visualised in Figure 11(a) where each significant connection is represented by a filled, circular “icon”. The direction of the significant connection is deduced to be from the reference spike train to target spike train. For example, in Figure 11(a), neuron #1 has significant connections to neurons #6, #8, and #12 as shown by the three icons in the first column.

6.2 Outlier detection (step 2)

An outlier connection is easily identified from neuron #5 to neuron #13. In Table 1, the peak of this connection is given as 6.52 with a time delay is 12 milliseconds. Therefore, this is ascribed to be a direct connection. Figure 11(b) shows a scatter plot of all 25 significant connections including the single outlier at the top of the plot.

6.3 Cluster analysis (step 3)

At this stage, the clustering algorithm is applied to the remaining 24 significant connections in order to distinguish between direct, indirect and common source connections.

Reference	1	1	1	2	2	2	3	3	6	7
Target	6	8	12	3	11	14	11	14	8	4

Peak	4.37	4.57	3.05	3.50	1.64	3.26	2.73	3.23	2.44	1.72
Time delay	13	15	11	12	27	14	12	12	2	23
Type	D	D	D	D	I	D	D	D	CS	I
Reference	7	9	10	10	11	12	12	12	14	14
Target	9	4	4	9	2	4	6	8	11	15
Peak	3.57	3.27	1.74	3.46	2.57	1.69	1.90	1.92	2.98	1.73
Time delay	11	12	25	13	13	23	2	4	13	40
Type	D	D	I	D	D	I	CS	CS	D	I
Reference	15	15	15	15						
Target	2	3	11	14						
Peak	2.71	3.57	3.73	2.14						
Time delay	12	14	11	26						
Type	D	D	D	I						

Table 2: Classification of the remaining 24 significant connections of fifteen neuron where *D* denotes direct connections, *I* denotes indirect connections and *CS* denotes common source connections.

All the remaining 24 significant peaks and corresponding time delays are normalized (within the range 0 and 1) and clustering is applied to these normalized values.

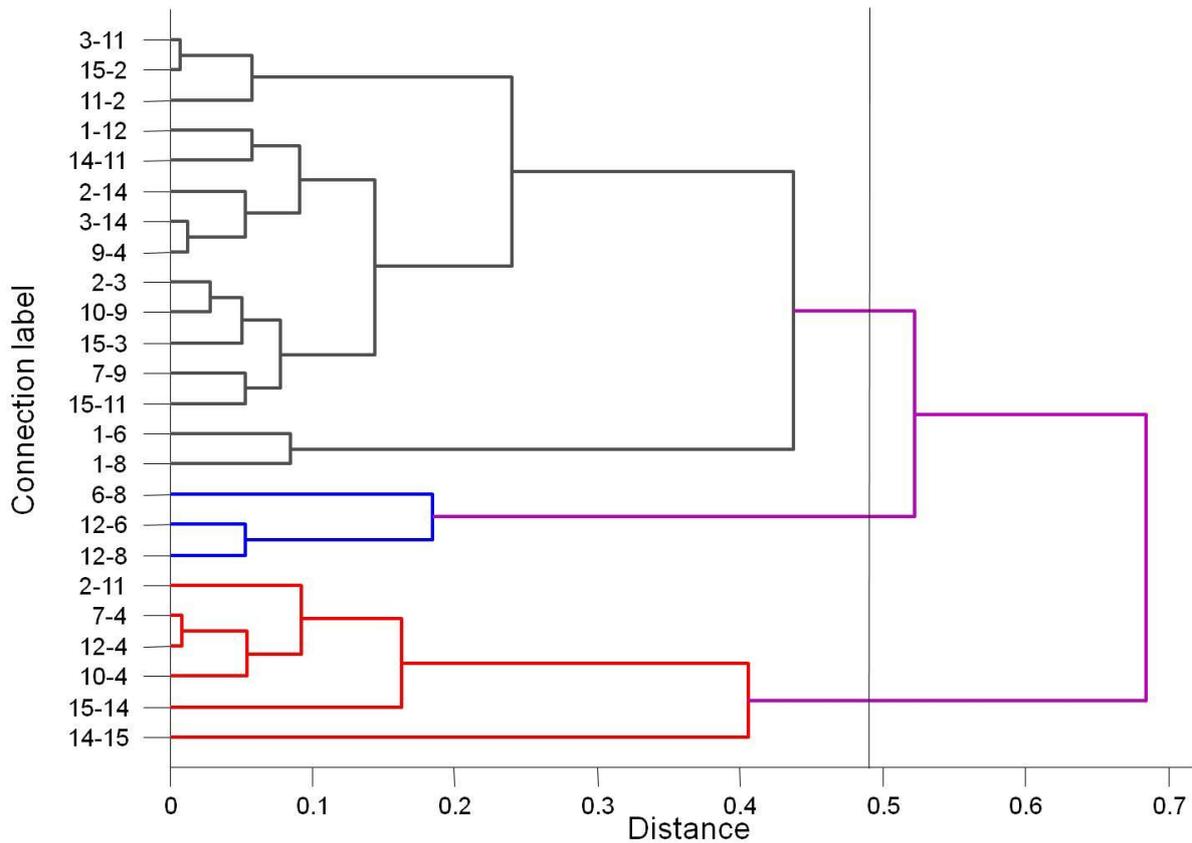


Figure 12: Dendrogram of 24 significant connections. Three clusters are indicated by different colours with the connection labels. The grey colour indicates direct connection, the blue colour indicates common source and red colour indicates indirect connection.

Since the goal is the identification of direct connections, indirect connections and common source connections, a vertical line is drawn on the dendrogram (Figure 12) in such a way that it clearly distinguishes the three clusters. With reference to Table 2, it can be deduced that connections within Cluster #1 have high significant peaks and moderate time delays so they are considered to be direct connections. Additionally, connections within Cluster #2 have low significant peaks and short time delays so they are ascribed as being the result of a common source. Finally, the connections in Cluster #3 have low significant peaks and longer time delays and so it is deduced that they are the result of indirect connections.

6.4 Verification of connections (step 4)

6.4.1 Common source connections

In order to verify the common source connections of Cluster #2, all relevant “three neuron” groups are identified. Relevant groups are defined as those that exhibit the architecture of common source connections as well as including a connection from either (i) neuron #6 to neuron #8, or (ii) neuron #12 to neuron #6 or (iii) neuron #12 to neuron #8 (refer to Figure 12 for those connections within Cluster #2).

To clarify this part of the process, let us identify the first relevant group given in Table 2. The first connection is from neuron #1 to neuron #6. The second is from neuron #1 to neuron #8. Since, the input to both neuron #6 and neuron #8 are from neuron #1, this neuron group (1, 6, 8), also known as a triplet, is potentially relevant. Subsequently, its relevance is confirmed since the common source connection from neuron #6 to neuron #8 is one of the three connections in Cluster #2. In the relevant triplet (1, 6, 8), neuron #1 influences both neurons #8 and #6 with corresponding time delays 15 milliseconds and 13 milliseconds. Thus, using equation (1), the connection from neuron #8 to neuron #6 is identified as a common source connection since the time delay for this connection is 2 milliseconds. Recall that a short delay characterises common source connections.

Similarly, the triplet (1, 8, 12) is the next potentially relevant group. Given that the connection from neuron #8 to neuron #12 is one of the three connections in Cluster #2, this group is relevant. Thus, the connection from neuron #12 to neuron #8 is a common source connection since the time delay is only 4 milliseconds.

Triplet (1, 6, 12) is relevant so relevant delays are assessed. The connection from neuron #6 to neuron #12 is a common source connection since the time delay is only 2 milliseconds.

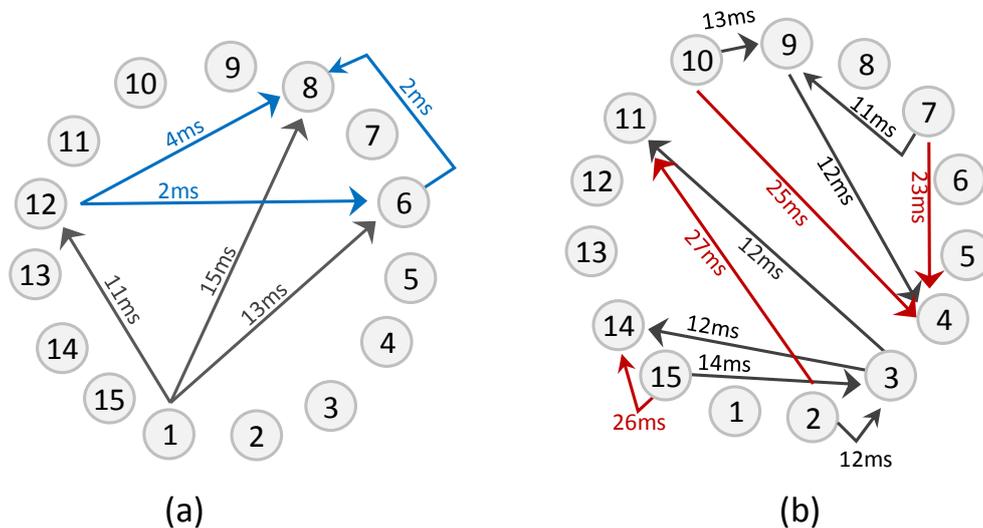


Figure 13: (a) All common source connections (blue) obtained from the common source cluster in Table 2, (b) Some of the Indirect connections (red) obtained from the cluster of indirect connections in Table 2.

Note that the next connection in the table is from neuron #2 to neuron #3. As this does not overlap with any previous connections, the next connection, from neuron #2 to neuron #11, is assessed. Since, the input to both neuron #3 and neuron #11 are from neuron #2, this triplet (2,

3, 11) is again potentially relevant. However, in this case, its relevance is not confirmed as the connection from neuron #3 to neuron #11 is not one of the connections in Cluster #2.

Since all three common source connections have been identified, this part of the process is complete. This resultant data is visualised in Figure 13. In the figure the blue lines indicate the common source connection and the red lines indicate the indirect connections. These two colours are used throughout this paper to understand the common source and indirect connections.

6.4.2 Indirect connections

All connections within Cluster #3, the cluster of indirect connections, need to be verified. A similar procedure is adopted where every relevant triplet is assessed. Relevant groups are defined as those that (i) exhibit the architecture of indirect connections and (ii) include a connection that is within the group of indirect connections, Cluster #3. Again, once a relevant group is identified, delays are investigated to verify the connection as indirect.

It transpires that Cluster #3, actually incorporates four correctly classified connections as well as two incorrectly classified connections. Initially, let us focus on the four correctly classified indirect connections, refer to Figure 13(b). These indirect connections are from neuron #2 to neuron #11, neurons #7, #10 and #12 to neuron #4, neuron #14 to neuron #15 and neuron #15 to neuron #14 (Table 2). Investigation of Figure 13(b) reveals the triplets (#2, #3, #11), (#4, #7, #9), (#4, #9, #10) and (#3, #14, #15).

In the triplet (#2, #3, #11), neuron #2 influences neuron #3 with a delay 12 milliseconds and neuron #3 influences neuron #11 with a delay 12 milliseconds. Based on Equation 2, the connection from neuron #2 to neuron #11 is an indirect connection with neuron #3 defined as the intermediate neuron since the delay for the connection from neuron #2 to propagate to neuron #11 is 27 milliseconds which is relatively long. Connections from neuron #7 to neuron #4 and connection from neuron #10 to neuron #4 have already been deduced to be indirect. For both these connections neuron #9 acts as an intermediate neuron. Similarly, connection from neuron #15 to neuron #14 is indirect. For this connection neuron #3 acts as an intermediate neuron. More importantly, all these three indirect connections have long time delays: 23 milliseconds, 25 milliseconds and 26 milliseconds, respectively.

Now, focus on the two connections incorrectly classified as indirect connections from (i) neuron #12 to neuron #4, and, (ii) neuron #14 to neuron #15. These two connections lie

within the indirect cluster of indirect connections. Investigation of Figure 13(b) shows that by trialling an indirect connection from neuron #12 to neuron #4, there is no possible triplet to which this indirect connection could belong, as neuron #12 is otherwise unconnected. Therefore, this connection cannot be verified as indirect.

Further, to verify the connection from neuron #14 to neuron #15, a triplet (#3, #14 and #15) must exist. In this case, such a triplet does indeed exist. However, it does not exhibit the required architecture of an indirect connection from neuron #14 to neuron #15. Therefore, this connection cannot be verified as indirect. From these two incorrectly classified indirect connections, it can be concluded that initially the connections are classified as indirect, based on their peaks and time delays. However, for a connection to be classified as indirect, the architecture must also exhibit the classical architecture for an indirect connection, as shown in Figure 4(ii). Therefore, before a connection is verified as being indirect, its architecture must also be assessed to ensure it meets the architectural criteria. This is also true for common source connection.

6.4.3 The Advanced Correlation Grid

Classifying all the significant connections into direct, indirect and common source connections, the functional connectivity of the fifteen neurons is derived. This connectivity is automated and more clearly visualised using the Advanced Correlation Grid, shown in Figure 14.

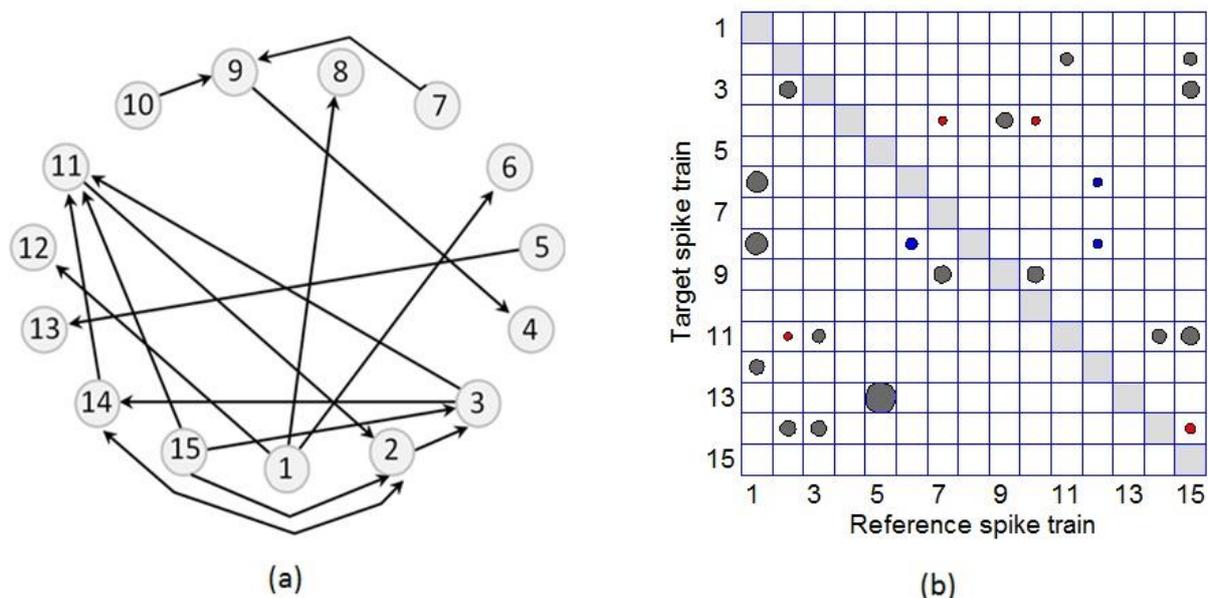


Figure 14: (a) Direct connections obtained from the ACG which is identical to the connection scheme shown in Figure 9(a). (b) A single grid showing the Direct connection (grey), the common source (blue) and indirect (red) connections. The radius of the circle shows the strength of connection.

The grid in Figure 14(b) shows the functional connectivity of the direct connections (shown in grey). This succinctly reveals the correct connectivity of the diagram used to generate the fifteen neurons (Figure 9(a) with corresponding connection strengths Figure 9(b)). It also shows the connections which are not present in the connectivity diagram but are due to common source or indirect coupling. More importantly, the common source and indirect connections have relatively small radii compared to the radius of the direct connections shown in grey.

7 Case Study 2 - functional connectivity of fifty neurons

Section 6 demonstrates that in the case of very strong connections and a small set of neurons, the ACG is effective for identifying functional connectivity. In this section, the ACG is further tested using a data set with a large number of neurons with connections of medium strength. Thus, a set of fifty neurons is generated for a period of 20,000 milliseconds using the connection architecture shown in Figure 15 with the connection strengths shown in Figure 16: Connection strengths of the fifty non zero connections in the matrix format. There are a total of fifty connections throughout the data set and all connections have medium strength.

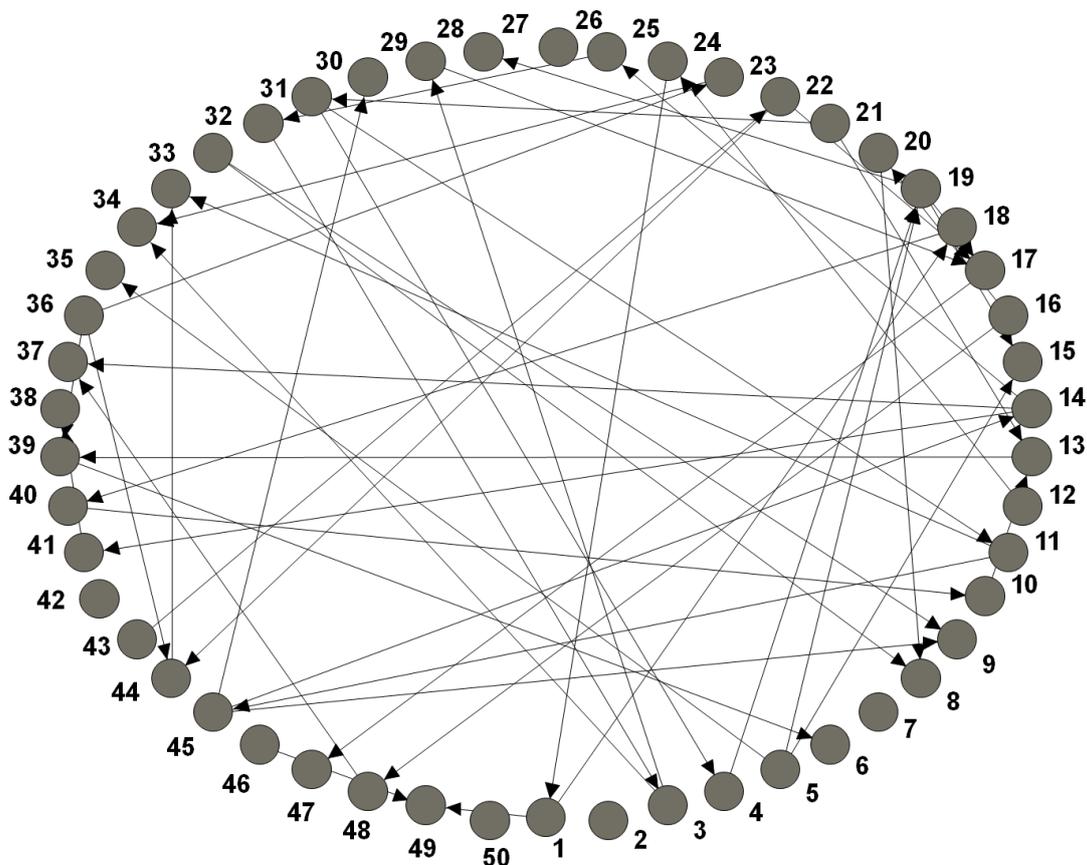


Figure 15: Connection scheme of 50 neurons with 50 connections which are shown by arrows

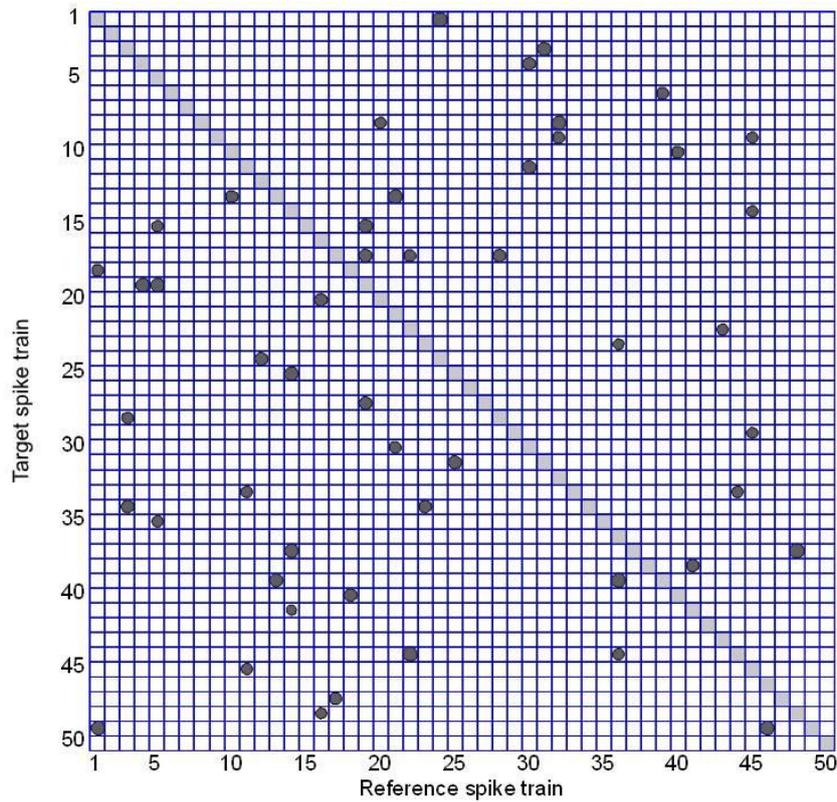


Figure 16: Connection strengths of the fifty non zero connections in the matrix format.

7.1 Calculation of CCFs (step 1)

A total of $(50^2 - 50)/2 = 1225$ pairwise CCF are calculated with a bin size of 1 millisecond and a correlation window of 100 milliseconds.

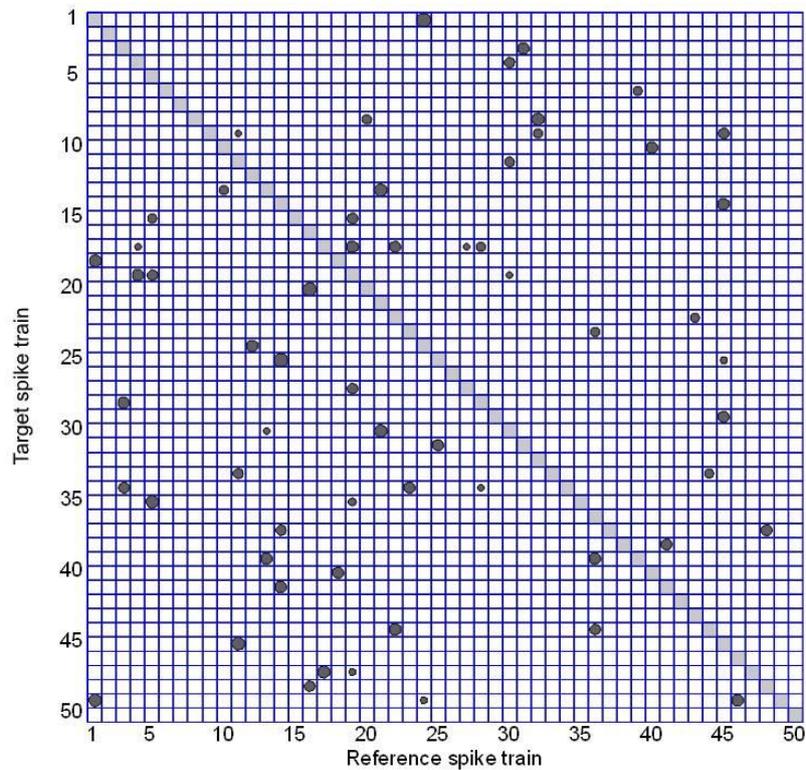


Figure 17: Significant connections obtained from pairwise cross-correlation analysis

To test the independence of the pairwise spike trains the level of significance $\alpha = 0.05$ is applied with the Bonferroni correction. Subsequently, a total of 60 significant connections are identified as shown in Figure 17.

In this figure, significant connections are denoted using a filled, circular “icon” and the direction of connection is designated as being from the reference spike train to the target spike train.

7.2 Outlier detection (step 2)

Step 2 is applied. In this case, no outliers are identified.

7.3 Cluster analysis (step 3)

Step 3 is applied. The results of cluster analysis are shown in Figure 18. Please refer to Table 3 in Appendix 2 for full results of this analysis.

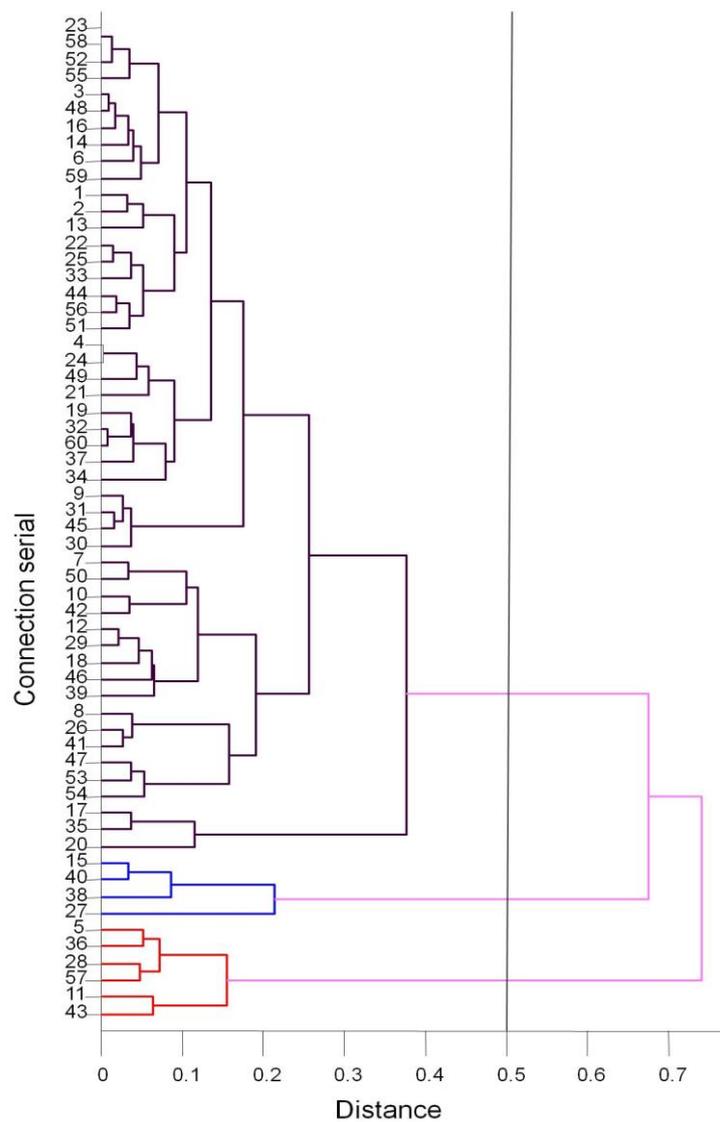


Figure 18: Dendrogram of 60 significant connections. Three clusters are indicated by different colours.

7.4 Verification of connections (step 4)

7.4.1 Verification of common source connections

Step 4 identifies the four connections: neuron #13 to neuron #30, neuron #19 to neuron #35, neuron #27 to neuron #17 and neuron #28 to neuron #34. Table 3 shows that all these four connections are common source connections.

7.4.2 Verification of indirect connections

Investigation of Table 3 reveals that all these six connections are verified as indirect connections.

7.4.3 Conclusions

Classification and subsequent verification of all significant connections into direct, indirect and common source clusters has been achieved.

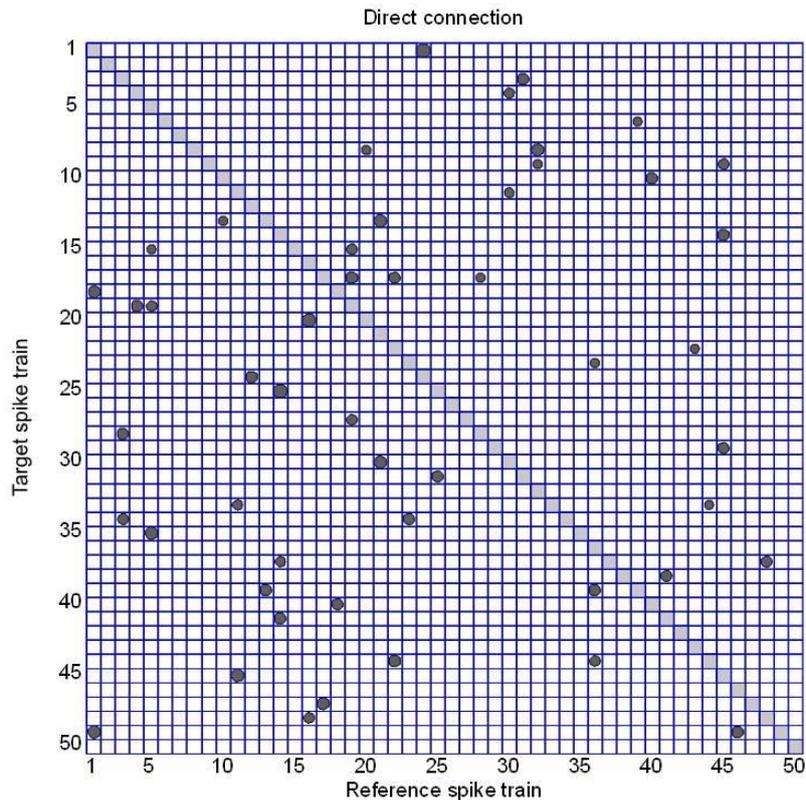


Figure 19: Direct connections which are exactly the same as at the connection scheme shown in Figure 15. Radius of the circle shows the strength of connection estimated from CCF and is proportional to the strength of connection in Figure 16.

The resultant functional connectivity of the fifty neurons is visualised using two grids (Figure 19 and Figure 20). The first grid reveals the correct connectivity used to generate the fifty neurons (Figure 15) with corresponding connection strengths (Figure 16). Note that the radii of the circles depict the estimated strength of connections.

This is very similar to the radii of the circles in the diagram of connections (Figure 16). The other grid in Figure 20, shows the connections which are not present in the connectivity diagram but are due to common source or indirect coupling. Moreover, the common source and indirect connections have small radii relative to the direct connections; this means that they are spurious connections.

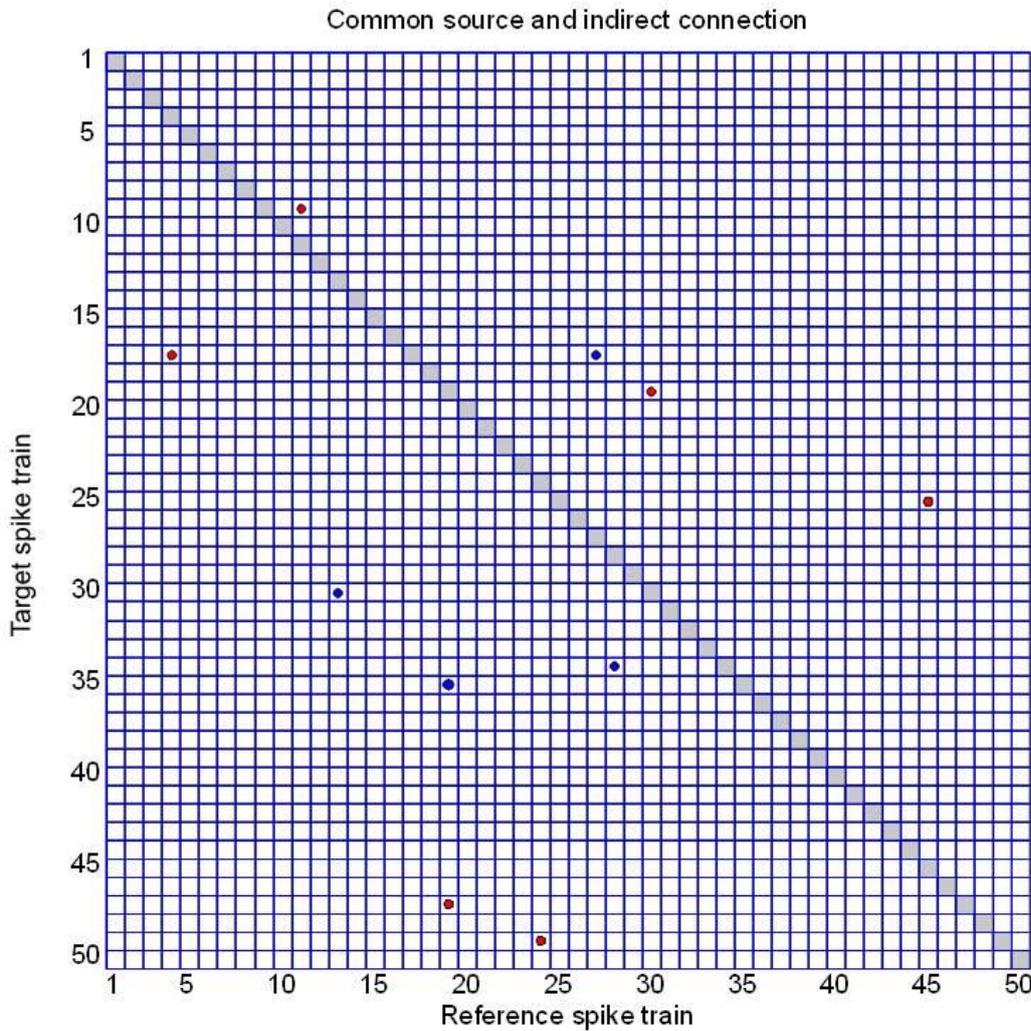


Figure 20: Common source (blue) and indirect (red) connections. Radius of the circle shows the strength of connection and the connections are small relative to the direct connection.

8 Effectiveness of the method

In this section, the ACG method is compared to an existing technique of functional connectivity. Comparisons of the two techniques are made as the data set is varied.

8.1 Comparison of ACG and Cox method

To evaluate the performance of the ACG method, it is compared with an existing method called the Cox method (Masud et al., 2011). The Cox method was selected as it is an effective

technique for identifying functional connectivity. This method provides statistical estimates which measure the influence strengths and their confidence intervals from multiple reference spike trains to the selected target spike train. If any confidence interval contains zero, then it is deduced that there is no functional connectivity from that reference spike train to the target spike train. Otherwise, it is deduced that there is functional connectivity from that reference spike train to the target spike train. Selecting another target spike train and repeating the calculation of the influence strengths from the reference spike trains enables researchers to find all functional connections among multiple spike trains.

A data set consisting of fifteen neurons was generated over a period of 20,000 milliseconds, was generated using the ELIF model. It used the connection architecture shown in Figure 21(a) with the connection strengths visualised in Figure 21(b). From these figures, it is clear that fifteen connections exist.

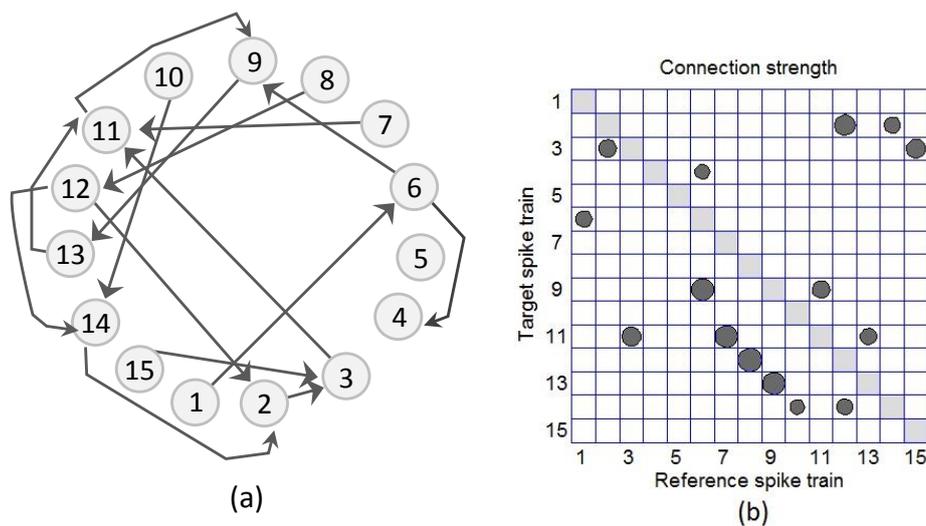


Figure 21: (a) Connection scheme of fifteen neurons. There are fifteen non zero connections which are shown by arrows. (b) Connection strengths of the fifteen non zero connections in the matrix format.

The functional connectivity of these fifteen spike trains was evaluated using both methods (results shown in Figure 22). From the figure, it is clear that the direct connections identified by both the ACG method (Figure 24(a)) and the Cox method (Figure 24(b)) were correct when compared to the connection scheme (Figure 21(b)). However, in addition, the ACG method identified six connections as indirect connections (Figure 22(c)). Though both methods find functional connectivity correctly, the ACG method demonstrates some advantages over the Cox method. Firstly, the connection strengths calculated using the ACG method are more accurate than the connection strengths obtained using the Cox method. For

example, the strength of the connection from neuron#12 to neuron#2, neuron#14 to neuron#2 and neuron#15 to neuron#3 obtained by the ACG method are more accurate than the strengths of connection obtained by the Cox method in comparison to the connection strengths used in the connection scheme. Secondly, the ACG method is easier to apply than the Cox method. The Cox method requires the manual selection of the correct influence function and then analysis to decide the characteristic times of decay and rise of the postsynaptic potential. As a result of this, the ACG method requires less time to identify functional connectivity than the Cox method. Due to this speedup, the ACG method can be more effectively employed as a screening method for data.

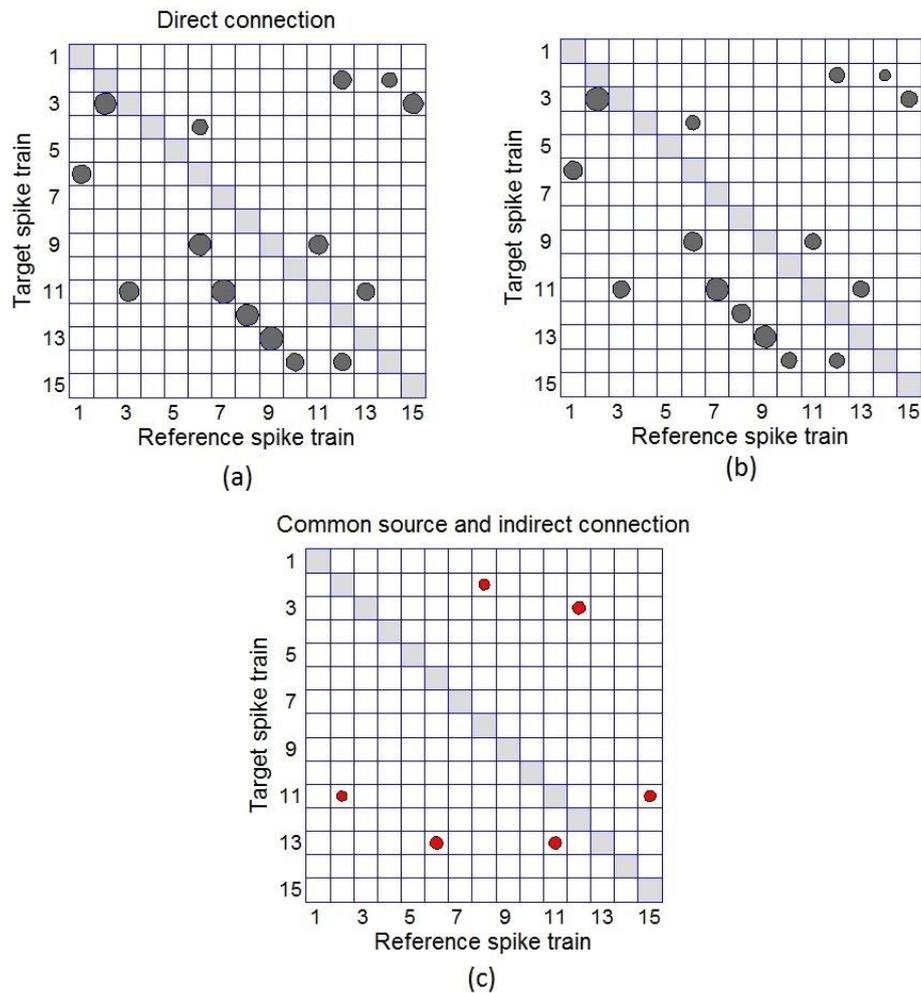


Figure 22: (a) Direct connections obtained from ACG method. (b) Connections obtained from the Cox method. (c) Common source and indirect connections obtained from ACG method.

8.2 Investigations using high and low noise levels

The final investigations reported in this paper relate to variations in the levels of noise in the data sets. In order to assess the ACG method, a random noise component was added to the membrane potential in the ELIF model. Two sets of spike trains were generated for a period of 20000 milliseconds using the same connection architecture as shown in Figure 21(a). In the

first, and second, data set the average amplitude of the noise was 5.1 (high noise), and 3.1 (low noise), respectively. The average value of the threshold is 14.2, therefore high noise is characterised by 36.1% of threshold variation and the low noise is characterised by only 22.0%.

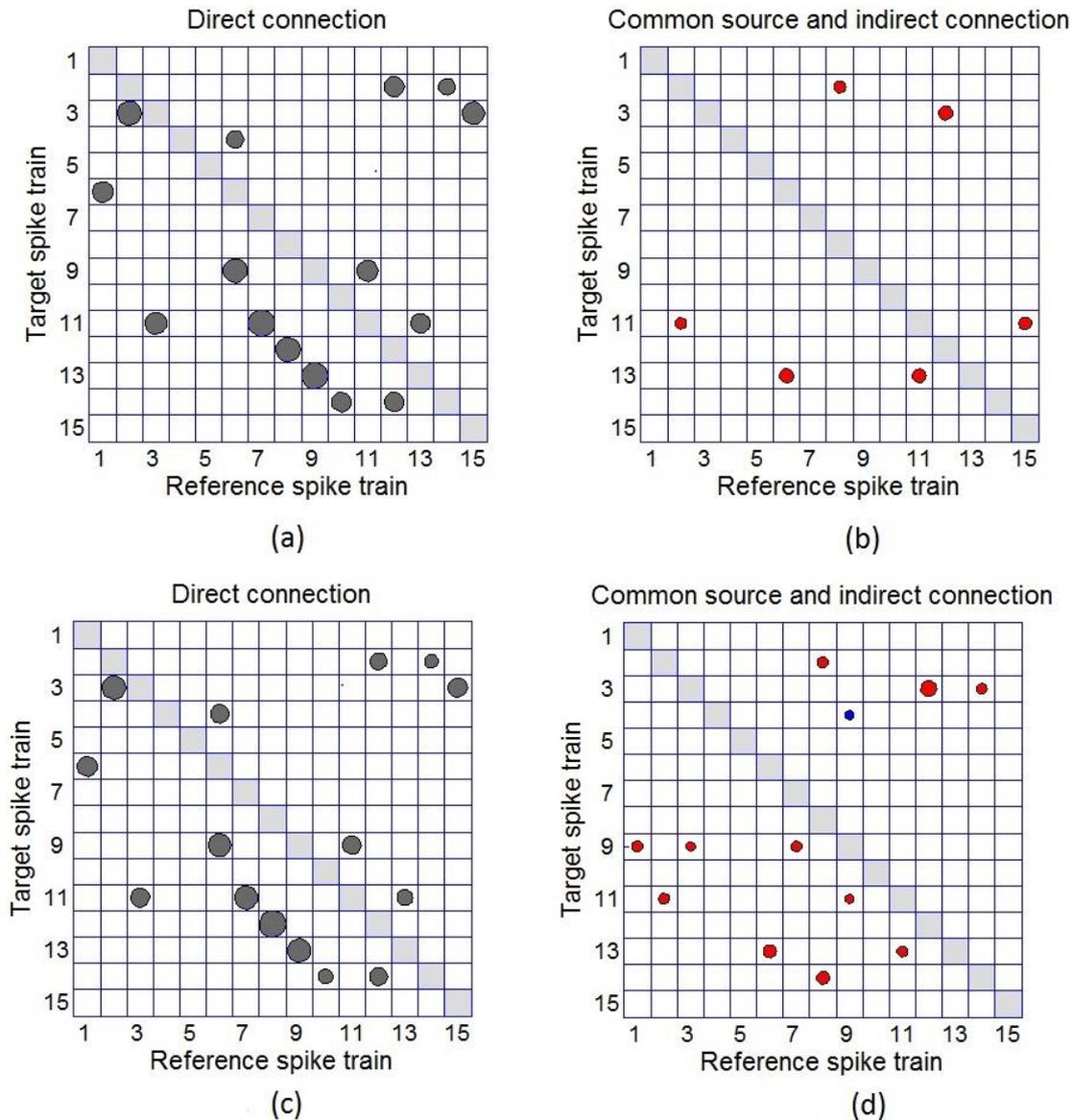


Figure 23: Functional connections of 15 spike trains with high and low noise levels (a) Direct connections obtained from ACG method using the high noise data set (b) Common source and indirect connections obtained from the ACG method using the high noise data set (c) Direct connections obtained from ACG method using the low noise dataset (d) Common source and indirect connections obtained from ACG method using the low noise data set.

Figure 23 shows the functional connectivity of the 15 spike trains obtained by the ACG method for high and low noise. The ACG method accurately identified all the direct connections in both the data set with high levels of noise (Figure 23(a)) and the one with low levels of noise (Figure 23(c), refer to Figure 21(a)). In the case of high noise, the ACG method identified six connections as indirect as shown in Figure 23(b) and for low noise it identified 12 indirect

connections and 1 common source connection (Figure 23(d)). The indirect connections obtained in high and low noise are the same although there are extra indirect connections in the low noise condition. The overall conclusion is that the ACG method is sufficient to manage high and low levels of noise. Based on the Researcher's experience of analysing spike trains with both the ACG and Cox methods, both of these methods work well for high levels of noise and the breaking values of noise amplitude are approximately the same.

The final investigation assessed the effectiveness of the ACG method when the standard deviation of the noise was varied. For this purpose, multiple data sets of spike trains were generated. The data sets all had the same mean but different standard deviation of noise over the period of 20000 milliseconds and the same connection architecture as shown in Figure 21(a). Results showed that increasing the standard deviation of the noise resulted in fewer direction connections being identified by the ACG method. Further research into this is underway.

9 Deriving functional connectivity from experimental data

To identify the functional connectivity of multiple spike trains, the ACG method is applied to a set of experimental data recorded from the visual cortex of the cat (Nikolić, 2007; Schneider et al., 2006). The experimental condition includes application of six different stimuli (different orientations of the moving grid). Each stimulus is repeated 20 times resulting in 120 applications of all stimuli. The order of stimuli presentations is random. Thus for 120 stimuli the spiking activity of 32 channels is recorded. From each channel one spike train is selected – the one with a medium firing rate. This spike train is used to prepare six spike trains corresponding to six stimuli. The twenty time intervals (each of six seconds duration), where the stimulus 1 is presented, have been selected to represent a total interval (120 seconds) of the application of stimulus 1. All spikes from this interval are considered continuously despite the gaps between the intervals of stimulus 1 representation. There are 32 channels in the experiment which results in 32 spike trains for this stimulus. The same operation has been repeated for stimulus 2: all subintervals of six seconds corresponding to application of stimulus 2 have been selected, considered continuously and all spikes have been taken for analysis of the functional connectivity. The same selection is done for each channel and 32 spike trains have been constructed for stimulus 2. Repeating this operation for other stimuli, six sets of 32 simultaneous spike trains were constructed. Each set corresponds to application of one stimulus.

For the application of ACG only, the second stimulus (chosen randomly) consists of 32 spike trains. The raster plot of the 32 spike trains, shown in Figure 24, reveals that three spike trains (#4, #5 and #29) have significantly higher spiking rates than the other spike trains. These higher rates are outside the physiological range and they probably appear due to some damage of neurons by inserted electrodes. To maintain similarity in terms of spiking rate of the spike trains, these three spike trains are excluded from analysis. All the remaining 29 spike trains have similar spiking patterns. Note that the spiking rates of these 29 spike trains are high over time interval 78000ms - 95000ms.

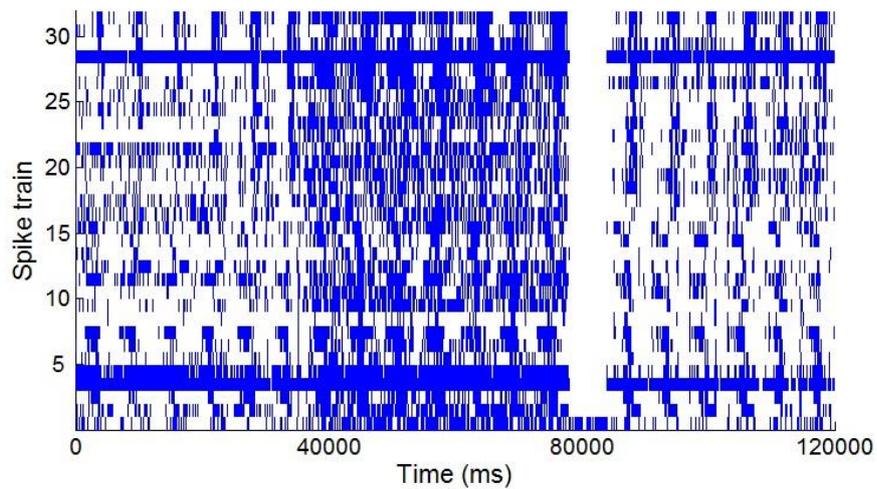


Figure 24: Raster plot of 32 spike trains of stimulus 1. Spike trains #4, #5 and #29 have high spiking rates and are not considered for analysing functional connectivity.

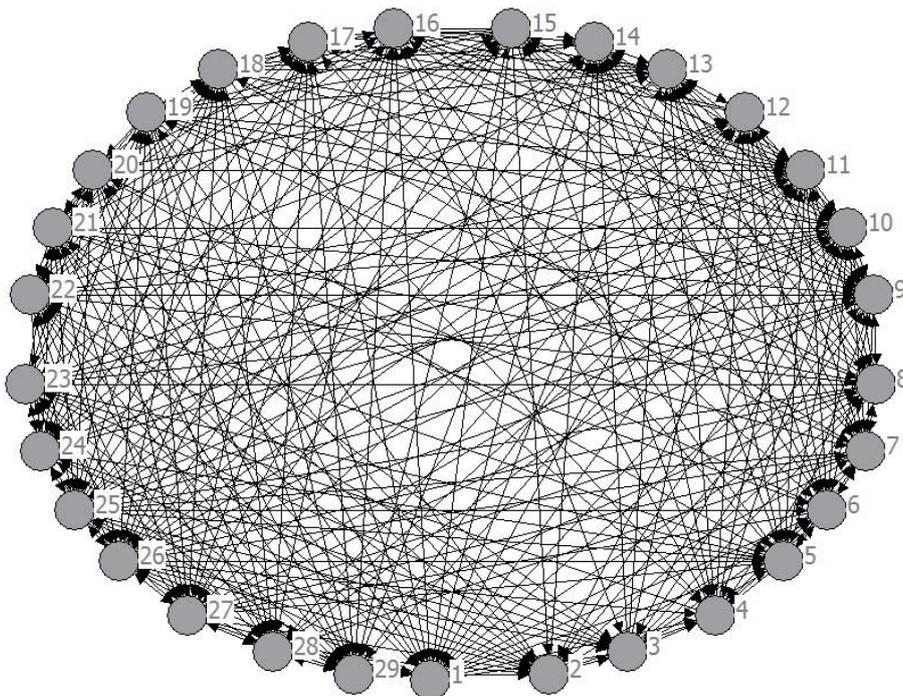


Figure 25: Direct connections obtained by the ACG method for the real data are denoted by the arrows.

For this set of 29 spike trains, a total of $(29^2 - 29)/2 = 406$ pairwise CCF were calculated with a bin size of 1ms and a correlation window of 100 milliseconds. A clustering algorithm is applied to the significant connections obtained by the CCF in order to distinguish direct connections from spurious ones (i.e., indirect connections and common source). After clustering, 442 connections are identified as direct connections. These direct connections are shown by arrows (Figure 25) and matrix format (Figure 26).

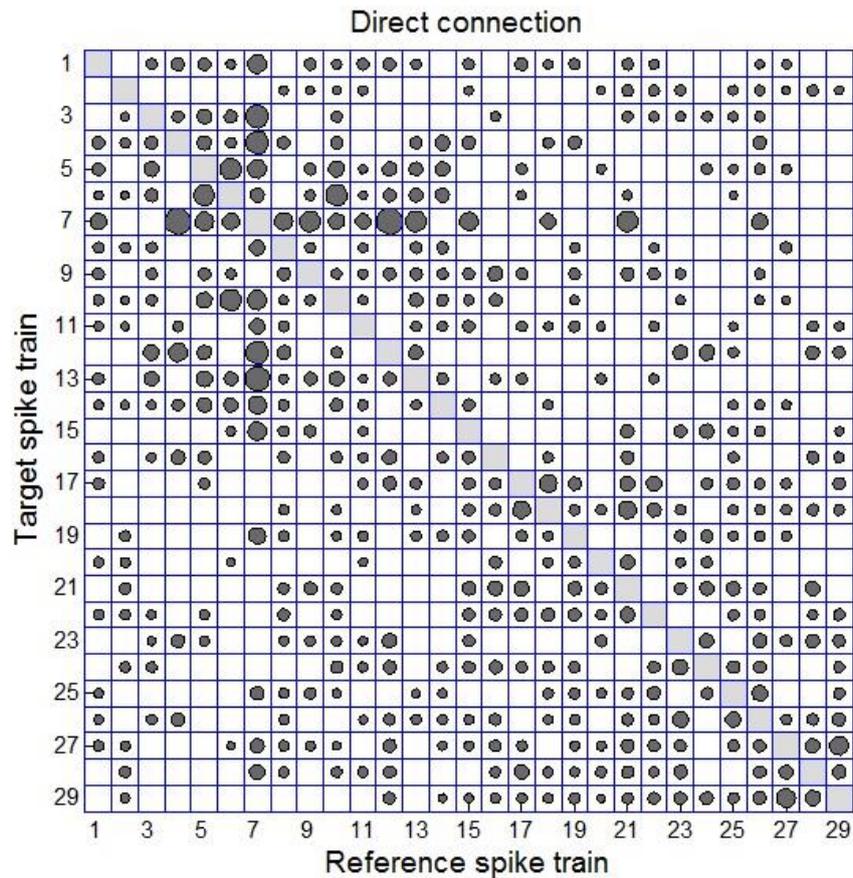


Figure 26: Direct connections obtained using the ACG method applied to real data are shown using a matrix format

Among the 442 connections, 16 connections have relatively strong connection strengths. These connections are mostly related to spike train #7; the incoming and outgoing connections of spike trains #7 are relatively high. Spike train #26 has 21 outgoing connections and spike train #27 has 21 incoming connections which are the highest among 29 spike trains. All other connections have medium strength. There are 208 pairs of connections where both spike trains have functional connectivity to each other. For example the pair (#7, #13); where there is a connection from spike train #7 to spike train #13 and vice versa.

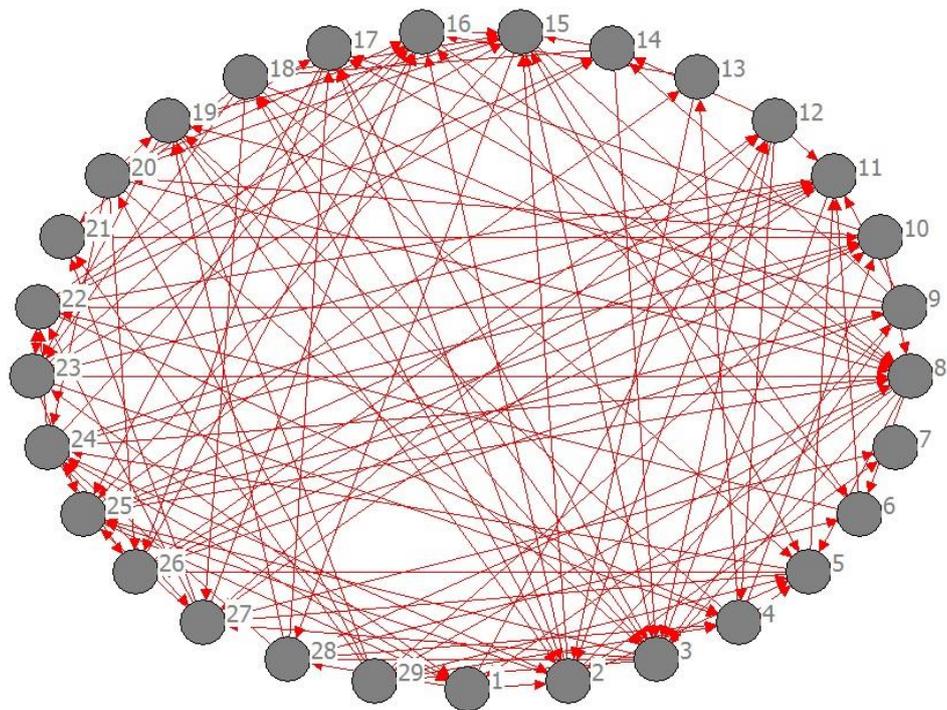


Figure 27: Indirect connections identified by the ACG method applied to real data are indicated by the arrows.

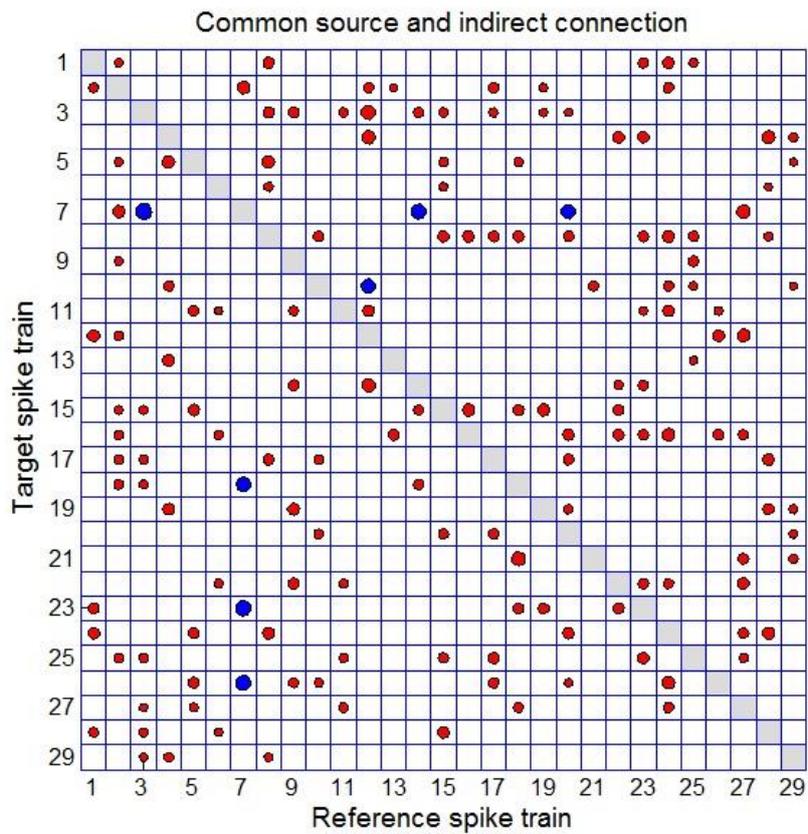


Figure 28: Common source and indirect connections obtained using the ACG method applied to real data are shown using a matrix format.

There are 150 indirect connections and 7 common source connections in the real data set of spike trains. These indirect connections are shown by the arrows in Figure 27. Both the common source and indirect connections are shown in Figure 28. Observe that the common source connections have higher strengths of influence compared to the indirect connections. Spike train #8 has 10 outgoing indirect connections and spike train #2 has 10 incoming indirect connections which are the highest among 29 spike trains. There are 81 pairs of indirect connections where the two spike trains have functional connectivity to each other.

The main goal of this example is to illustrate how the ACG method can be used for analysing simultaneously recorded spike trains in real experiments under variable stimulation. These results are very promising; they confirm that the method can identify both direct and indirect connections. For example, we can speculate that spike train #7 plays a key role of the most influential hub in this neural circuit.

An additional set of results, which will be reported in a separate publication, relates to a study on how the diagram of identified connections for some stimulus will change under stimulation by another stimulus. Assuming that the anatomical connections are the same under different stimulation protocols, it is possible to identify how the functional connectivity changes under the application of different stimuli. It is anticipated that combining the results of the ACG method, for finding diagrams of functional connectivity, with the calculation of an established structural graph theory measure (e.g. distributions of in- and out-degrees) will enable Researchers to reveal details of stimulation mechanism in neuronal circuits.

10 Conclusions

The Advanced Correlation Grid is an innovative technique for studying functional connectivity of multiple neurons. This technique can be used to identify the unknown structure of functional connectivity between neurons. When compared to the original correlation grid (Stuart et al., 2005), the ACG has the advantage that it can automatically distinguish between direct functional connectivity and spurious connections (common source and indirect connections). This method can also identify the direction of functional connectivity, another limitation of correlation grid, where the connections are considered as symmetric connections. This method is also less computer-intensive and it can be applied to a large set of neurons with diverse connections strengths.

Application of this method is shown for two sets of neurons: a small set of fifteen neurons with a very strong strength of influence and a large set of fifty neurons with medium strength of influence. Note that prior to analysis, the functional connectivity is unknown.

The subsequent use of the ACG, was successful in the derivation of an accurate assembly of neurons. This method accurately identified all the direct connections as well as the connection strengths. At the same time this method can distinguish correctly the common source and indirect connections, thus proving, that this is a very effective method for identifying functional connectivity.

Another possible application of the ACG method relates to finding brain motifs (Sporns and Koetter 2004). For example, in the analysis of motifs including three elements (i.e. three spike trains), it is important to distinguish between (i) the motif with common input from one element to two others from (ii) the motif where the common input connections also include a direct connection between elements receiving common input. Our study shows that the ACG method is well positioned for reliable distinguishing between different motifs.

Although we have demonstrated here that the ACG method has several important advantages, obviously there are still some drawbacks. For example, despite the fact that the CCF can detect negative connectivity, the ACG method cannot analyze this case as yet. Therefore the use of the ACG method is limited to the case of excitatory connections only. Such drawbacks are under investigation. For example, some obvious improvement in calculation of the cross-correlation function, similar to the method published in paper (Nikolić et al., 2012) could help improve the analysis of spike trains with a strong regular spiking component. Another example for further development of the ACG method relates to using more advanced statistical techniques for improved measurements of statistical significance. For example, using the method of false discoveries (Benjamini and Hochberg, 1995) could also prove to be highly beneficial.

As any other numerical technique, the ACG method is based on some assumptions. For example, using this method, it is assumed that all essential spike trains have been recorded and are part of the analysis process. From a neurobiological point of view, in the case of a microelectrode array recording, this assumption can be considered to be valid. Of course, it is true that there are cases where unrecorded neurons, with powerful influences upon the recorded spike trains, can change the result of analysis.

Appendix 1

Enhanced Leaky Integrate and Fire model

The description of the Enhanced Leaky Integrate and Fire (ELIF) model follows the paper by Borisjuk (2002). A discrete-time version of the model neuron is used with the time increment set equal to 1 millisecond. The state of each neuron at the time t is characterised by both a threshold and the total potential, which is the sum of postsynaptic potentials and the noise. When the value of the total potential reaches the threshold, the neuron generates a spike. Subsequently, the spike propagates to other neurons with a time delay. The diagram of connection should be defined as well as connection strengths, time delays, and time decays of postsynaptic potentials. When the spike reaches another neuron, the postsynaptic potential sharply increases up, or down, depending on whether the spike is from an excitatory neuron, or inhibitory neuron, respectively. The value of the connection strength defines the magnitude of this increase. If there are no incoming spikes, the postsynaptic potential exponentially decays to the resting potential. After spike generation, the neuron is unable to generate another spike during an absolute refractory period. When this period expires, the threshold gets the highest value and then exponentially decays to the asymptotic threshold value. This decay is used to model a relative refractory period. To model spontaneous background activity, random noise is added to the membrane potential. The amplitude of the noise decays exponentially over time. A normally distributed random variable, with zero mean and a fixed variance, is added to the noise at each time step. The noise is an independent random process for each element. If the amplitude of noise is large enough, then the element can be spontaneously active even without influences from other neurons.

Dynamics of the ELIF model

The dynamics of the ELIF are governed by the following equations:

a) The threshold:

$$r(t + 1) = (r_{max} - r_{\infty})exp(-(t - t_{sp})/\alpha_{th}) + r_{\infty}$$

where r_{max} is the maximum value of the threshold

r_{∞} is the asymptotic threshold value when $t \rightarrow \infty$

α_{th} is the threshold decay rate

t_{sp} is the last spike time before t .

b) The post-synaptic potential for the input of the neuron:

$$PSP^j(t+1) = PSP^j(t)exp(-1/\alpha_{PSP}^j) + a,$$

$$a = \begin{cases} w^j, & \text{if } t_{sp}^j + \tau^j = t + 1 \\ 0, & \text{otherwise} \end{cases}$$

where w^j is the connection strength, which is positive for the excitatory connections and negative for the inhibitory connections

τ^j is the time delay

α_{PSP}^j is the PSP decay rate of the j th neuron

t_{sp}^j is the time of last spike of the j th neuron before t .

c) The noise:

$$N(t+1) = N(t)exp(-1/\alpha_N) + \xi, \quad \xi \in N(0, \sigma)$$

where α_N is the decay rate of noise

ξ is a random variable which exhibits a normal distribution

d) The soma's membrane potential:

$$V(t+1) = V_{AHP}exp(-(t - t_{sp})/\alpha_V)$$

where V_{AHP} is the value of hyperpolarization after a spike

α_V is the decay rate of the soma's membrane potential

t_{sp} is the time of the last spike before t .

e) The total potential:

$$P(t+1) = \sum_j PSP^j(t+1) + N(t+1) + V(t+1) + I_{ext}(t+1)$$

where I_{ext} is the value of the external input.

f) Spike generation:

If $P(t+1) > r(t+1)$, then $t_{sp} = t + 1$.

g) The absolute refractory period:

There is no spike generation for the time interval $t \in (t_{sp}, t_{sp} + ref)$.

Parameters for spike train generation by ELIF model

An ELIF model can be simulated using software freely available from the following website: <http://www.tech.plymouth.ac.uk/infovis>. To run the simulation, the parameters of ELIF neurons and their coupling should be specified.

Neural parameters describe the parameters of each neuron. The parameters are: a) Maximum value of the threshold, b) Threshold decay rate, c) Asymptotic threshold value, d) Amplitude

of the noise (i.e. the standard deviation of the normally distributed random variable), e) Noise decay rate, f) Initial value of after spike hyperpolarisation, g) Soma's membrane potential decay rate, h) External input, i) Absolute refractory period, and j) Type of the neuron since this software generates activities for both types of neurons (0- non-pacemaker, 1 – pacemaker).

Connection parameters contain the parameters describing non-zero connections between neurons. The parameters are: a) List of the numbers of those neurons which send their connections to the current neuron, b) Connection strengths for these connections (positive for excitatory connection and negative for inhibitory), c) Decay rates of postsynaptic potential for each connection respectively and d) Time lag of spike propagation for each incoming connection (milliseconds).

11 Appendix 2

Table 3 shows the 60 significant connections of the 50 neurons alongside their corresponding peaks and time delays. The peaks of the significant connections range from 1.93 to 4.58 and the time delays range from 1 to 30 milliseconds

Connection serial	1	2	3	4	5	6	7	8	9	10
Reference	1	1	3	3	4	4	5	5	5	10
Target	18	49	28	34	17	19	15	19	35	13
Peak	3.90	3.99	3.69	3.47	2.10	3.70	2.90	3.37	4.03	3.09
Time delay	13	13	13	15	29	14	15	12	18	14
Connection serial	11	12	13	14	15	16	17	18	19	20
Reference	11	11	11	12	13	13	14	14	14	16
Target	9	33	45	24	30	39	25	37	41	20
Peak	1.93	3.24	4.09	3.81	2.05	3.75	4.55	3.29	3.68	4.36
Time delay	26	17	13	13	1	13	16	18	16	18
Connection serial	21	22	23	24	25	26	27	28	29	30
Reference	16	17	18	19	19	19	19	19	20	21
Target	48	47	40	15	17	27	35	47	8	13
Peak	3.36	3.92	3.60	3.48	3.88	3.38	2.50	2.19	3.18	3.99
Time delay	14	15	13	15	15	11	6	30	17	17
Connection serial	31	32	33	34	35	36	37	38	39	40
Reference	21	22	22	23	24	24	25	27	28	28

Target	30	17	44	34	1	49	31	17	17	34
Peak	3.98	3.59	3.87	3.60	4.58	2.20	3.64	2.05	3.08	2.05
Time delay	18	16	16	18	15	28	15	4	18	2
Connection serial	41	42	43	44	45	46	47	48	49	50
Reference	30	30	30	31	32	32	36	36	36	39
Target	4	11	19	3	8	9	23	39	44	6
Peak	3.31	3.19	2.07	3.78	3.93	3.12	3.02	3.72	3.40	6
Time delay	11	14	25	14	18	16	11	13	16	16
Connection serial	51	52	53	54	55	56	57	58	59	60
Reference	40	41	43	44	45	45	45	45	46	48
Target	10	38	22	33	9	14	25	29	49	37
Peak	3.79	3.56	2.92	2.87	3.49	3.83	2.32	3.60	3.79	3.57
Time delay	15	13	11	12	13	14	30	13	12	16

Table 3: This table specifies the entire set of 60 significant connections alongside their peaks and delays. These 60 significant connections are the results of analysing the spike trains from the original assembly of fifty neurons. Connections are directed from the reference spike train to the target spike train.

Acknowledgements

Sincere thanks to Professor Danco Nicolic, Max-Planck Institute for Brain research, Frankfurt, Germany for providing experimental recordings of multiple spike trains.

This work was supported by CARMEN Project funded by the Engineering and Physical Sciences Research Council (UK) [contract number EP/E002331/1].

References

- Abdi, H. *Encyclopedia of Measurement and Statistics*, N.J. Salkind (Ed.), Thousand Oaks, Sage Publications. The Bonferroni and Sidak corrections for multiple comparisons (<http://www.utdallas.edu/~herve/Abdi-Bonferroni2007-pretty.pdf>). 2007.
- Aertsen, A. M. H. J., Gerstein, G. L., Habib, M. K. and Palm, G. *Journal of Neurophysiology*. Dynamics of neuronal firing correlation: Modulation of effective connectivity. 1989, 61, 900-917.
- Benjamini Y. and Hochberg Y.. *Journal of the Royal Statistical Society: Series B (Methodological)*. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. 1995, 57: 1, 289-300.

- Borisyuk, R. *BioSystems*. Oscillatory activity in the neural networks of spiking elements. 2002, 67, 3-16.
- Brillinger, D. R. *The American Mathematical Monthly*. Measuring the association of point processes: a case history. 1976, 83(1),16-22.
- Brillinger, D. R. *Selecta Statistica Canadiana*. Confidence intervals for crosscovariance function. 1979, 1-16.
- Brown, E. N., Kass, R. E. and Mitra, P. P. *Nature Neuroscience*. Multiple neural spike train data analysis: state-of-the-art and future challenges, 2004, 7:5, 456-461.
- Dayan, P. and Abbott, L. F. *Theoretical neuroscience: computational and mathematical modeling of neural systems*, Cambridge: The MIT Press, 2001.
- Dahlhaus, R., Eichler, M. and Sandkuhler, J. *Journal of Neuroscience Methods*. Identification of synaptic connections in neural ensembles by graphical models. 1997, 77, 93-107.
- Eggermont, J. J. *Journal of Neurophysiology*. Neuronal pair and triplet interactions in the auditory midbrain of the leopard frog. 1991, 66:5, 1549-1563.
- Eichler, M., Dahlhaus, R. and Sandkuhler, J. *Biological Cybernetics*. Partial correlation analysis for the identification of synaptic connections. 2003, 89, 289-302.
- Espinosa, I. E. and Gerstein, G. L. *Brain Research*. Cortical auditory neuron interactions during presentation of 3-tone sequences: effective connectivity. 1988, 450(1-2), 39-50.
- Gerstein, G. L. and Kirkland K. L. *Neural Networks*. Neural assemblies: technical issues, analysis, and modelling. 2001, 14(6-7), 569-598.
- Gochin, P. M., Gerstein, G. L. and Kaltenbach, J. A. *Brain Research*. Dynamics of temporal properties of effective connections in rat dorsal cochlear nucleus. 1990, 510, 195-202.
- Gochin, P., Miller, E., Gross, C. and Gerstein, G. *Experimental Brain Research*. Functional interactions among neurons in inferior temporal cortex of the awake macaque. 1991, 84, 505-516.
- Grün S. and Rotter S (Eds.) *Analysis of parallel spike trains*. Springer Series in Computational Neuroscience. 2010.
- Haslinger R, Pipa G, Lewis LD, Nikolić D, Williams Z and Brown E. *Neural Comput*. Encoding through patterns: regression tree-based neuronal population models. August 2013, 25(8), 1953-93.

- Iglewicz, B. and Hoaglin, D. *How to detect and handle outliers*, ASQC Quality Press, 1993.
- Jovanović S., Rotter S. *PLoS Comput Biol*. Interplay between Graph Topology and Correlations of Third Order in Spiking Neuronal Networks. 2016, 12:6, e1004963.
- Kandel, E. R. *Principles of neural science*. McGraw-Hill, 2000.
- Konig, P., Engel, A., Roelfsema, P. and Singer, W. *Neural Computation*. How precise is neuronal synchronization. 1995, 7(3), 469-485.
- Kriener B, Helias M, Aertsen A and Rotter S. *Journal of Computational Neuroscience*. Correlations in spiking neuronal networks with distance dependent connection. 2009 27(2):177-200.
- Lindsey, B. G., Hernandez, Y. M., Morris, K. F. and Shannon, R. *Journal of Neurophysiology*. Functional connectivity between brain stem midline neurons with respiratory-modulated firing rates. 1992c, 67:4, 890-904.
- Li, Z., Morris, K. F., Baekey, D. M., Shannon, R. and Lindsey, B. G. *Journal of Neurophysiology*. Multimodal medullary neurons and correlational linkages of the respiratory network. 1999, 82, 188-201.
- Louie, K. and Wilson, M. A. *Neuron*. Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. 2001, 29, 145-156.
- Makarov, V. A., Panetsosa, F. and de Feob, O. *Journal of Neuroscience Methods*. A method for determining neural connectivity and inferring the underlying network dynamics using extracellular spike recordings. 2005, 144, 265-279.
- Masud, M.S and Borisyuk, R. *Journal of Neuroscience Methods*. Statistical Technique for Analysing Functional Connectivity of Multiple Spike Trains. 2011, 196, 201-219.
- Nedungadi, A. G., Rangarajan, G., Jain, N. and Ding, M. *Journal of Computational Neuroscience*. Analyzing multiple spike trains with nonparametric granger causality. 2009, 27, 55-64.
- Nikolić, D. *Journal of Computational Neuroscience*. Non-parametric detection of temporal order across pairwise measurements of time delays. 2007, 22:1, 5-19.
- Nikolić D., Mures R.C. and Feng W., Singer W. *European Journal of Neuroscience*. Scaled correlation analysis: a better way to compute a cross-correlogram. 2012, 1–21.
- Nikolić D., R.C. Mureşan, W. Feng and W. Singer. *European Journal of Neuroscience*. Scaled correlation analysis: A better way to compute a cross-correlogram. 2012, 35(5), 742-62.

- Nykamp, D. Q. *SIAM Journal on Applied Mathematics*. Revealing pairwise coupling in linear-nonlinear networks. 2005, 65:6, 2005-2032.
- Park, I., Paiva, A. R. C., DeMarse, T. B. and Principe, J. C. *Journal of Neuroscience Methods*. An efficient algorithm for continuous time cross correlogram of spike trains. 2008, 168, 514-523.
- Perkel, D. H., Gerstein, G. L. and Moore, G. P. *Biophysical Journal*. Neuronal spike trains and stochastic point processes II. simultaneous spike trains. 1967, 7, 419-440.
- Pillow J.W., Shlens J., Paninski L., Sher A., Litke A. M., Chichilnisky E. J. and Simoncelli E.P. *Nature*. Spatio-temporal correlations and visual signalling in a complete neuronal population. 2008, 454, 995–999.
- Reimer I.C.G., Staudea B., Ehmb W., Rotter S. *Journal of Neuroscience Methods*. Modeling and analyzing higher-order correlations in non-Poissonian spike trains. 2012, 208, 18–33.
- Schiffler RE. *The American Statistician*. Maximum Z Score and outliers. 1988, 42:1, 79-80.
- Schneider, G., Havenith, M. N. and Nikolić, D. *Neural Computation*. Spatiotemporal structure in large neuronal networks detected from cross-correlation. 2006, 8, 2387-2413.
- Shannon, R., Baekey, D. M., Morris, K. F., Li, Z. and Lindsey, B. G. *Journal of Physiology*. Functional connectivity among ventrolateral medullary respiratory neurones and responses during fictive cough in the cat. 2000, 525.1, 207-224.
- Skaggs, W. E. and McNaughton, B. L. *Science*. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. 1996, 271:5257, 1870-1873.
- Sporns O and Kötter R. PLoS Biol. Motifs in Brain Networks. 2004, 2(11): e369.
- Sporns O. *Scholarpedia*. Brain connectivity. 2007, 2(10):4695.
- Stevenson, I. H., Rebesco, J. M., Miller, L. E. and Kording, K. P. *Current Opinion in Neurobiology*. Inferring functional connections between neurons. 2008, 18, 582-588.
- Stuart, L., Walter, M. and Borisyuk, R. *BioSystems*. The correlation grid: analysis of synchronous spiking in multi-dimensional spike train data and identification of feasible connection architectures. 2005, 79, 223-233.

- Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H. and Aertsen, A. *Nature*. Dynamics of neural interactions in monkey cortex in relation to behavioural events. 1995, 373, 515-518.
- Wilson, M. A. and McNaughton, B. L. *Science*. Reactivation of hippocampal ensemble memories during sleep, 1994, 265, 676-679.