ISIT'2013 1569720155

Inferring Neural Connectivity via Measured Delay in Directed Information Estimates

Nima Soltani and Andrea Goldsmith
Department of Electrical Engineering
Stanford University
Stanford, CA, USA
Email: nsoltani@stanford.edu, andrea@wsl.stanford.edu

***Abstract*—Directed information between two neural signals has previously been used to successfully determine if there is a synapse connecting one neuron to the other. However, there are situations in which the directed information method leads to false positives for detecting connections due to the influence of a third neuron. We propose a method that accounts for these cases using the delay-profile of the causally-conditioned entropy rate to find a measured delay range and its intersection with a connection-based delay range. Through simulations in NEURON, we show how this method outperforms connectivity inference based on directed information alone.**

I. Introduction

Interactions between groups of neurons are enabled by the synaptic connections between them, and hence determining this connectivity is of significant interest in neuroscience and related fields. The goal of this work is to propose a tool for inferring neural connections in the brain. Neuronal signaling can be observed in the voltage across the membrane of the cells, known as the membrane potential. The connections we are interested in are chemical synapses that would connect a source neuron N_X to a destination neuron N_Y, and allow N_X to cause inflections in N_Y's membrane potential when N_X undergoes an action potential, also known as a spike. A common way to determine the existence of this synapse is to place electrodes near each neuron and record the voltage of each electrode. One can then perform signal processing on the recorded traces of N_X and N_Y, which are X^n and Y^n respectively, to make conclusions about their connectivity. This can be challenging as on average a neuron N_Y has synaptic connections with about 10000 other source neurons that cause inflections in its membrane potential; we want to determine whether N_X is one of those 10000.

The general theme behind connection inference algorithms is one of Granger causality. Granger causality implies that a process $\mathbf{A}$ causally influences another process $\mathbf{B}$ if the future of $\mathbf{B}$ is better predicted given the causal past of $\mathbf{A}$ than without it [1]. In applying this technique, the recorded traces X^n and Y^n are assumed to be observations of random processes $\mathbf{X}$ and $\mathbf{Y}$, and previous works use generalized linear models (GLMs) to fit the likelihood functions of the recorded traces [2]–[4]. By finding the log-likelihood ratio of scenarios that include and exclude the influence of $\mathbf{X}$ on $\mathbf{Y}$, inferences can be made on the existence of a connection from N_X to N_Y [2]. It has also been shown that this log-likelihood ratio is related to directed information [3]. Directed information is an information theoretic measure used to characterize the capacity of channels with feedback [5]. While GLMs make estimating the joint distributions simple and parameterized, non-parametric models have also been used for the joint distributions to estimate the directed information [6]. Using directed information to make inferences about causal relationships is not restricted to neural connections. For example, it can be used as a method to estimate causal influence in stock markets and also to detect the unknown delay in a communication channel [6].

Inference using directed information often performs well in detecting connections between N_X and N_Y given X^n and Y^n, but not always. There are topologies in which no synapse exists from N_X to N_Y, but due to the influence of a third neuron, N_Z, the directed information between X^n and Y^n will be greater than zero, leading to a false inference that a synaptic connection exists. The two primitive topologies that lead to these scenarios, shown in Figure 1, are the relay (or proxy) topology, in which N_X and N_Y are connected through a relay, and the broadcast (or cascading) topology, in which N_X and N_Y are fed by a common source neuron.

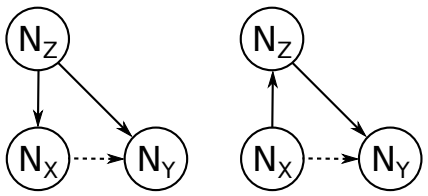

Fig. 1. When using directed information based methods, two topologies of neurons can lead to falsely inferring a connection: the broadcast (left) and relay (right) topologies. The falsely inferred connections are indicated by dashed lines.

In works that address this situation [2], [3], all other neurons in the generating network are recorded, in the case of Figure 1 this would be N_Z, and through conditioning on these neurons the false positives are eliminated. In practice, however, there is no guarantee that recordings from all related neurons in the network will be available. Specifically the recording associated with the culprit of a falsely detected connection may not be

This work was supported by the NSF Center for Science of Information (CSoI): NSF-CCF-0939370.

available. Thus, we aim to find a better way to determine connectivity, without the necessary assumption that all neurons in the network are observed. To do so, we find the delay range of the directed information transferred from X^n to Y^n using the recordings and then compare it to what the delays should be with the given hypothesized topology, thus allowing us to reject certain false positive elements in an inferred connectivity graph.

In the remainder of the paper, we describe the problem in more detail in Section II, and we discuss our proposed technique for improved connection inference in Section III. We show simulation results for the broadcast and relay cases in Section IV, and we close with concluding remarks and future work in Section V

II. Problem setup

The recorded voltage traces for neurons N_X and N_Y, which are separated by a distance $|N_X N_Y|$, will be referred to as X^n and Y^n, and are assumed to be observations of the jointly stationary ergodic processes, $\mathbf{X}$ and $\mathbf{Y}$. They are assumed to be jointly and separately Markov in the following ways:

$$P\left(X_i, Y_i | X^{i-1}, Y^{i-1}\right) = P\left(X_i, Y_i | X_{i-D}^{i-1}, Y_{i-D}^{i-1}\right) \quad (1)$$

$$P\left(X_i | X^{i-1}\right) = P\left(X_i | X_{i-D}^{i-1}\right) \quad (2)$$

$$P\left(Y_i | Y^{i-1}\right) = P\left(Y_i | Y_{i-D}^{i-1}\right) \quad (3)$$

For the examples given in this paper, $X_i, Y_i \in \{0, 1\}$, but the binary alphabet assumption can be easily extended to larger alphabets. Unless otherwise specified, the goal is to determine if there is a connection from N_X to N_Y.

A. Directed information

Directed information [5] is defined as:

$$I\left(X^n \to Y^n\right) \triangleq \sum_i I\left(X^i; Y_i | Y^{i-1}\right) = H\left(Y^n\right) - \sum_i H\left(Y_i | Y^{i-1}, X^i\right) \quad (4)$$

This can be thought of as the reduction in uncertainty about $\mathbf{Y}$ when given the causal past of $\mathbf{X}$. When directed information is positive, the implication is that we can more accurately predict $\mathbf{Y}$ given its past and the causal past of $\mathbf{X}$ than we can if only $\mathbf{Y}$'s past is used for prediction. In this case therefore there is a causal influence from $\mathbf{X}$ to $\mathbf{Y}$ in the Granger causality sense and a synaptic connection is inferred. Simple situations such as the broadcast and relay topologies, however, have non-zero directed information, even though there is no direct, physical connection between N_X and N_Y. Given that we know the third neuron, N_Z, and have its recorded trace, Z^n, one solution is to condition on the third neuron [3]. The causally-conditioned directed information to calculate is thus:

$$I\left(X^n \to Y^n \| Z^n\right) \triangleq \sum_i I\left(X^i; Y_i | Y^{i-1}, Z^i\right)$$

This conditioning reduces the directed information from X^n to Y^n to zero as needed. This method, however, requires two things: an extra electrode to concurrently measure the third signal, and knowledge of which neuron to use as the third neuron. We therefore look for other methods that can reduce the estimated directed information using only two recorded traces.

B. Neural signal propagation

The classical view towards neurons are that they are made up of input, output and processing areas, known as dendrites, axons and the soma, respectively, and communicate by generating pulses or spikes, called action potentials. It is important to note that propagation of these signals through the neuron is not instantaneous. The passive signal propagation from the dendrites to the soma, and the active signal propagation along the axons to the synapses, take time that can be on the order of milliseconds [7].

If we can estimate an upper bound on the propagation velocities then, given the spacing between the electrodes, we can estimate the minimum propagation delay that a signal from N_X would have to undergo before it can affect N_Y. Also, when an action potential reaches the synapse, it can cause an inflection in the membrane potential of the post-synaptic neuron that can be approximated to last for a finite amount of time. This amount of time depends on the type of receptor that the post-synaptic cell uses. We can use this in addition to an estimate of the maximum propagation delay to come up with a maximum delay. The models associated with these delay estimates are based on the detailed system dynamics as provided in Chapters 2 and 6 of [7].

We now formalize these delay models for communication between neurons. First, we define C to be the set of closed intervals on the positive reals, $C = \{[a, b] \,|a, b \in R^+, a \leq b\}$. We then define the delay range mapping function, $f : R^+ \to C$. Given a particular distance d representing the separation between two neurons, this function returns the *predicted delay range for information transfer* D_p for which one neuron's activity can affect the other's activity. Specifically for neurons N_X and N_Y separated by a distance of $|N_X N_Y|$, given activity in one neuron at time t, this neuron can cause an effect in the other during the time interval $\{\tilde{t} | \tilde{t} - t \in f(|N_X N_Y|)\}$.

There are numerous ways to find the delay range mapping function f. While there are closed-form solutions for propagation velocities in the axon and the dendrites [7], they are typically based on simplified models. If we have specific types of neurons in mind, more detailed models of neurons can be simulated, such as those in [8]. By simulating pairs of neurons with varying sizes of axons and dendrites, we can find estimates of the delay for different neuron spacings, thus giving an estimate for f. Alternatively, if we had access to brain slices and had labels for true connectivity, we could find an empirical estimate for f by stimulating the neurons and measuring the delay.

III. Connection Inference

We now propose a method of determining which elements of X^i influence the distribution of Y_i using a modified version of the directed information rate or, more specifically, of the

causally-conditioned entropy rate, first defined in [9]. We define the *delay-profile* of the causally-conditioned entropy rate for sources satisfying (1), (2) and (3) as:

$$H_{Y\|X}(j) \triangleq \begin{cases} \frac{1}{n}\sum_i H\left(Y_i|Y_{i-D}^{i-1}, X_{i-D}^{i-j}\right), & \text{if } j \in 0,\ldots,D \\ \frac{1}{n}\sum_i H\left(Y_i|Y_{i-D}^{i-1}\right), & \text{otherwise} \end{cases} \tag{5}$$

This modifies the original causally-conditioned entropy rate by conditioning each entropy term in the summation on less and less of the most recent history of the process $\mathbf{X}$ with increasing time delay j. Writing the problem in this form leads to a compact representation for the directed information rate:

$$\begin{aligned} &\frac{1}{n}I(X^n \to Y^n) \\ &= \frac{1}{n}\sum_i H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}, X^i) && (6) \\ &= \frac{1}{n}\sum_i H(Y_i|Y_{i-D}^{i-1}) - H(Y_i|Y_{i-D}^{i-1}, X_{i-D}^i) && (7) \\ &= H_{Y\|X}(D+1) - H_{Y\|X}(0) && (8) \end{aligned}$$

where (6) follows from (4), (7) follows from (1) and (3), and (8) follows from (5). Also, as j increases, $H_{X||Y}(j)$ monotonically increases as the entropy terms in the summation of (5) are conditioning on fewer and fewer elements of X^n. Combining these two ideas, for time delays $0 \le a < b \le D+1$,

$$\begin{aligned} &H_{Y||X}(b) - H_{Y||X}(a) \\ &= \frac{1}{n}\sum_i H(Y_i|Y_{i-D}^{i-1}, X_{i-D}^{i-b}) - H(Y_i|Y_{i-D}^{i-1}, X_{i-D}^{i-a}) \\ &\le \frac{1}{n}\sum_i H(Y_i|Y_{i-D}^{i-1}) - H(Y_i|Y_{i-D}^{i-1}, X_{i-D}^i) \\ &= \frac{1}{n}I(X^n \to Y^n) \end{aligned}$$

with equality holding if $a = 0$ and $b = D+1$. Thus the difference between the delay-profile evaluated at different delays corresponds to a measure upper bounded by the directed information rate. We would like to find what indices $0 \le a < b \le D+1$ contain the bulk of the information content of the processes. Firstly if $I(X^n \to Y^n) = 0$, then similar to [3], [6], we conclude that there is no connection from N_X to N_Y. If the directed information is non-zero then, given $\epsilon > 0$, we find the maximum a and the minimum b such that

$$\begin{aligned} H_{Y||X}(a) - H_{Y||X}(0) &< \frac{\epsilon}{n}I(X^n \to Y^n) \\ H_{Y||X}(D+1) - H_{Y||X}(b) &< \frac{\epsilon}{n}I(X^n \to Y^n) \end{aligned}$$

Thus the interval $D_m \triangleq [a, b]$ is a central range of delays which contains at least a fraction $1-2\epsilon$ of the directed information between the processes, and we call this the *measured delay range of information transfer* from $\mathbf{X}$ to $\mathbf{Y}$.

We now have two delay ranges generated from different data. The predicted delay range, D_p, was generated based on physiological constraints on communication, whereas the measured delay range, D_m, was generated from the observations X^n and Y^n. If $D_m \nsubseteq D_p$, we conclude that we cannot explain the measured behavior using our physiological constraints, and hence a connection cannot exist.

Previous techniques boiled down to comparing two hypotheses depending on the directed information: either $\mathbf{Y}$ was generated independently of the causal past of $\mathbf{X}$, corresponding to directed information being zero, or $\mathbf{Y}$ has some sort of correlation with the causal past of $\mathbf{X}$, corresponding to directed information being positive. The difference between our proposed method and previous work is that we expand the second hypothesis. Not all correlations are created equal. If the $D_m \nsubseteq D_p$, then we cannot explain the observation of non-zero directed information by a direct synapse as this hypothesis is precluded by our physiological constraints. If $D_m \subseteq D_p$, however, then we have more certainty in the existence of a direct connection.

A. Example

As an example, we will apply our technique to a simplified version of a pair of neurons, N_X and N_Y, whose observed voltage traces follow the following distribution:

$$X_i \sim \text{i.i.d. Bernoulli}(p_X) \tag{9}$$

$$Y_i = \begin{cases} 1, & \text{if } Y_{i-1} = 0, X_{i-2} = 1 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

We are not aware of Y_i's dependence on X_{i-2} a priori, and would like to know if there is a connection from N_X to N_Y. The placement of the probes indicates that the neurons are a distance of 500 μm apart, and we have a delay range mapping f, which gives us $D_p = f(500\mu\text{m}) = [2\text{ms}, 4\text{ms}]$. We conservatively assume that the system is 5-th order Markov. By observing for long enough n we would see that for any unbiased estimator of the conditional entropy rate, the delay-profiles will converge to:

$$\begin{aligned} H_{Y||X}(j) &= \begin{cases} 0, & \text{if } j \in 0,1,2 \\ p_X H(X), & \text{otherwise} \end{cases} \\ H_{X||Y}(j) &= H(X) \end{aligned}$$

The result can be seen in Figure 2, for $p_X = 0.5$. For $H_{Y||X}$, any $\epsilon > 0$ gives us $D_m = [2\text{ms}, 3\text{ms}]$ (shown as a horizontal interval). Since $D_m \subseteq D_p$, we can still explain the observations by a connection, and thus we conclude a connection is likely to exist from N_X to N_Y. For the other case, we see $H_{X||Y}$ is constant, corresponding to a directed information of 0, meaning the process $\mathbf{X}$ was generated independently of the causal past of $\mathbf{Y}$, and hence we conclude that a connection is unlikely to exist from N_Y to N_X.

B. Remaining False Positive Regions

While this method reduces the occurrences of false positives in connection inference, it is no panacea. Consider the simple broadcast and relay scenarios of Figure 1 in which we can measure the voltages of two unconnected neurons, N_X and N_Y, located at $(-1, 0)$ and $(1, 0)$ respectively, and there exists a third neuron N_Z which connects to N_X and N_Y, whose

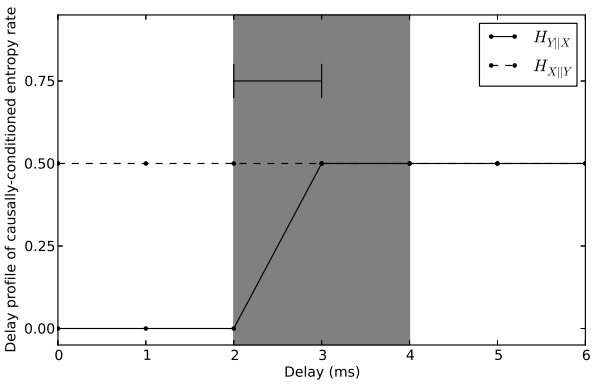

Fig. 2. Delay profile of the causally-conditioned entropy rate (5), for the example scenario with $\mathbf{X}$ and $\mathbf{Y}$ distributed as (9) and (10), with $p_X = 0.5$. The horizontal range shown is D_m for $H_{Y||X}$. The shaded areas correspond to D_p. In this case a connection is inferred from N_X to N_Y, but not from N_Y to N_X.

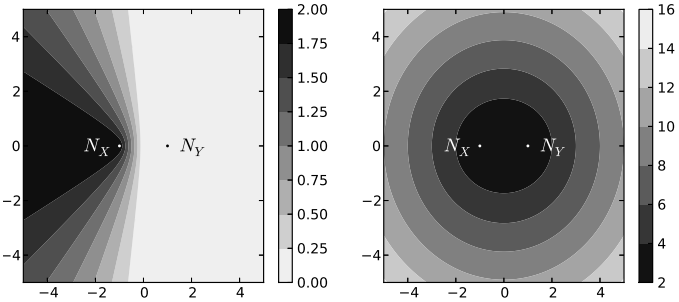

Fig. 3. Filled contour plots used to show where the proposed method still has false positives in a simplified scenario. On the left, the shading represents ΔD_m for the broadcast case, and on the right it represents ΣD_m for the relay case. In both cases the darker regions correspond to false positive areas in more accurate predicted delay range.

voltage we cannot observe. Given fixed locations for N_X and N_Y, shown in Figure 3, we illustrate which locations of the third neuron, N_Z, can still cause false positives in the proposed method. For simplicity, we assume the measured delay range D_m for two connected neurons spaced a distance d apart will be the singleton $\{kd\}$, $k > 0$. Initially we will also assume this for the delay range mapping f, in that $f(d) = \{kd\}$, so $D_p = \{kd\}$. Without loss of generality, we assume that $k = 1$, meaning $D_{m,\mathbf{X}\to\mathbf{Y}} = 2$.

For the broadcast scenario, a false positive will still occur if $\Delta D_m \triangleq D_{m,\mathbf{Z}\to\mathbf{Y}} - D_{m,\mathbf{Z}\to\mathbf{X}} \subseteq D_{p,\mathbf{X}\to\mathbf{Y}}$. In Figure 3 (left), we plot ΔD_m for different locations of N_Z. For $D_{p,\mathbf{X}\to\mathbf{Y}} = 2$ this would correspond to N_Z being located to the left of N_X anywhere on the line through N_X and N_Y. As we expand $D_{p,\mathbf{X}\to\mathbf{Y}}$ to include not just a point, but intervals containing smaller delays (i.e. $D_{p,\mathbf{X}\to\mathbf{Y}} = [kd-\epsilon, kd]$), the false positive areas become hyperbolic regions with foci at N_X and N_Y.

Similarly, for the relay scenario, a false positive will still occur if $\Sigma D_m \triangleq D_{m,\mathbf{X}\to\mathbf{Z}} + D_{m,\mathbf{Z}\to\mathbf{Y}} \subseteq D_{p,\mathbf{X}\to\mathbf{Y}}$. In Figure 3 (right), $D_{p,\mathbf{X}\to\mathbf{Y}} = 2$ would correspond to N_Z being located anywhere along the line connecting N_X and N_Y. By expanding $D_{p,\mathbf{X}\to\mathbf{Y}}$ to include longer delays (i.e. $D_{p,\mathbf{X}\to\mathbf{Y}} = [kd, kd+\epsilon]$), the false positive regions turn into ellipses with foci at N_X and N_Y.

IV. Simulations

In order to show the effectiveness of this method, we created simulations in NEURON [10], which is a simulation software used to model activity of neurons and networks of neurons by means of compartmental models. The model we simulated consisted of three identical neurons connected in the broadcast and relay topology shown in Figure 1. The neurons were designed to have background excitation and inhibition to keep the membrane potential fluctuating. For all neurons, the excitatory background spiking activity was initiated at the dendrites by a Poisson process with rate 10 Hz, and the inhibitory background spiking activity was initiated at the soma by a Poisson process with rate 10 Hz. The model was simulated for 10 minutes, corresponding to $n = 600000$ when sampled at 1 KHz. In both simulations, $|N_X N_Z| = 587$ μm, $|N_Y N_Z| = 837$ μm, and $|N_X N_Y| = 837$ μm. For the delay range mapping function, we assume $f(d) = d/0.6$ $\frac{\text{m}}{\text{s}} + [0, 3\text{ms}]$, corresponding to a propagation velocity of 0.6 $\frac{\text{m}}{\text{s}}$ and a delay spread of 3 ms.

The delay profiles for the broadcast topology can be seen in Figure 4, for $\epsilon = 0.05$. The horizontal interval lines correspond to D_m, and the gray areas correspond to D_p. For the true connections (N_Z to N_X and N_Z N_Y), $D_m \subseteq D_p$, and hence we correctly infer a connection exists. For the N_X to N_Y inference, since $|N_X N_Z| < |N_Y N_Z|$, once N_Z generates a spike, N_X tends to spike before N_Y does, and if we were to use directed information alone, we would falsely infer a connection exists. However, because $D_m \nsubseteq D_p$, the proposed method rejects this scenario.

The relay scenario delay profiles can be seen in Figure 5, for $\epsilon = 0.05$. It is clear that the proper connections (N_X to N_Z and N_Z to N_Y) are detected, since $D_m \subseteq D_p$ in these cases. In the case of $N_X \to N_Y$, because of the connection from N_Z to N_Y, N_X typically spikes before N_Y does, and thus using directed information alone leads to a false positive. However, because $D_m \nsubseteq D_p$, we reject the possibility of a connection from N_X to N_Y.

V. Conclusions and Future Work

We have proposed a method for improving neural connectivity inferences using the delay profile of the causally-conditioned entropy rate and a predicted delay range. We consider the case when only the recordings from two neurons are available. The false positives in detecting connections are reduced with our method by looking at the intersection of the predicted delay range, which is given a priori based on the hypothesized topology, and the measured delay range, which is estimated from the delay profile of the causally conditioned entropy rate of the neural recordings. If the intersection is not equal to the measured delay range, then we are able to identify and reject false positive scenarios.

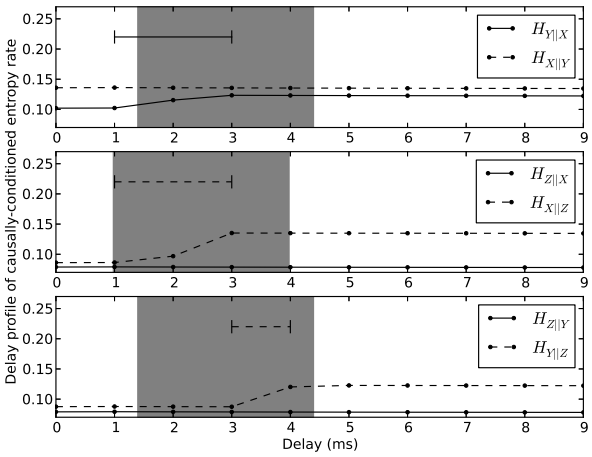

Fig. 4. Delay profile of the causally-conditioned entropy rate (5) with $\epsilon = 0.05$ for all pairs of variables in the broadcast topology. The horizontal ranges shown are the measured delay ranges, corresponding to the delay profile that is non-constant, meaning non-zero directed information. The shaded areas correspond to the predicted delay range. If the measured delay ranges lie inside the predicted delay ranges, then a connection is inferred.

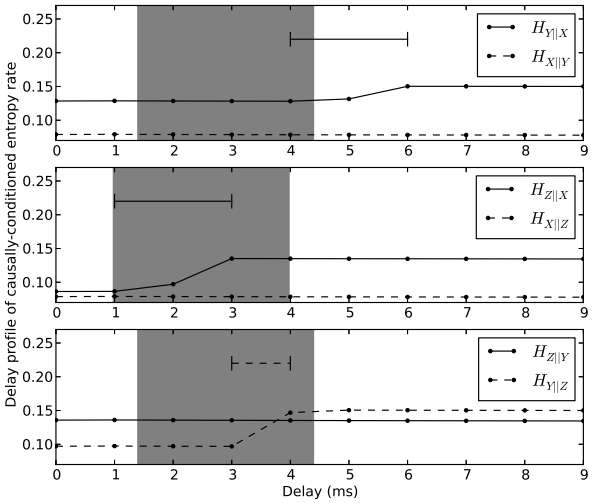

Fig. 5. Delay profile of the causally-conditioned entropy rate (5) with $\epsilon = 0.05$ for all pairs of variables in the relay topology. The horizontal ranges shown are the measured delay ranges, corresponding to the delay profile that is non-constant, meaning non-zero directed information. The shaded areas correspond to the predicted delay range. If the measured delay ranges lie inside the predicted delay ranges, then a connection is inferred.

We still cannot claim that we definitely determine that a connection exists from N_X to N_Y, as there is a subset of the broadcast and relay topologies that still lead to false positives. Our method does, however, eliminate a significant portion of false positives.

Our method uses hard limits on the predicted delay range, but it is possible to extend this work to have prior probabilities on different delay ranges, allowing for a "soft" connection detection. We are also working to extend this work to look at using other prior information that the experimenter might have, such as the expected excitatory or inhibitory effect of the synapse.

While this method was used as a way of determining whether two neurons were connected, it can be generalized to any situation where there is a physical system in which there is a causal delay between the variables, which is not necessarily the delay inferred from the observed random sequences. Applications of directed information for causality inference can be thought of as a hypothesis testing problem [11]. We are working on extending this hypothesis testing scenario for biological and other systems in which we have statistical delay models, similar to the models described in this paper for neural connectivity.

ACKNOWLEDGMENT

The authors would like to thank Stephen Baccus and Todd Coleman for their helpful discussions, and the anonymous reviewers for their constructive feedback.

REFERENCES

[1] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, Aug. 1969.

[2] S. Kim, D. Putrino, S. Ghosh, and E. N. Brown, "A granger causality measure for point process models of ensemble neural spiking activity," *PLoS Computational Biology*, vol. 7, no. 3, p. e1001110, 2011.

[3] C. J. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos, "Estimating the directed information to infer causal relationships in ensemble neural spike train recordings," *Journal of computational neuroscience*, vol. 30, no. 1, pp. 17–44, 2011.

[4] K. So, A. C. Koralek, K. Ganguly, M. C. Gastpar, and J. M. Carmena, "Assessing functional connectivity of neural ensembles using directed information," *Journal of Neural Engineering*, vol. 9, no. 2, p. 026004, 2012.

[5] J. L. Massey, "Causality, feedback and directed information," in *Proceedings of the 1990 International Symposium on Information Theory and its Applications (ISITA)*, 1990.

[6] J. Jiao, H. H. Permuter, L. Zhao, Y.-H. Kim, and T. Weissman, "Universal estimation of directed information," *Submitted to IEEE Transactions on Information Theory*, 2012, http://arxiv.org/pdf/1201.2334v2.

[7] C. Koch, *Biophysics of Computation: Information Processing in Single Neurons*, ser. Computational Neuroscience. New York: Oxford University Press, 1999.

[8] M. L. Hines, T. Morse, M. Migliore, N. T. Carnevale, and G. M. Shepherd, "Modeldb: A database to support compuational neuroscience," *Journal of Computational Neuroscience*, vol. 17, no. 1, pp. 7–11, July-August 2004.

[9] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, Swiss Federal Institute of Technology, Zurich, 1998.

[10] M. L. Hines and N. T. Carnevale, "Neuron: A tool for neuroscientists," *The Neuroscientist*, vol. 7, no. 2, pp. 123–135, April 2001.

[11] H. H. Permuter, Y.-H. Kim, and T. Weissman, "Interpretations of directed information in portfolio theory, data compression, and hypothesis testing," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3248–3259, June 2011.