Inferring causal connectivity from pairwise recordings and optogenetics

Mikkel Elle Lepperød[1,2], Tristan Stöber[2,3,4], Torkel Hafting[1,2], Marianne Fyhn[2,5], and Konrad Paul Kording[6]

[1]Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway
[2]Centre for Integrative Neuroplasticity, University of Oslo, Oslo, Norway
[3]Department of Informatics, University of Oslo, Oslo, Norway
[4]Department of Computational Physiology, Simula Research Laboratory, Lysaker, Norway
[5]Department of Bioscience, University of Oslo, Oslo, Norway
[6]Department of Neuroscience, University of Pennsylvania, Pennsylvania, USA

November 20, 2018

Abstract

To study how the brain works, it is crucial to identify causal interactions between neurons, which is thought to require perturbations. However, when using optogenetics we typically perturb multiple neurons, producing a confound - any of the stimulated neurons can have affected the postsynaptic neuron. Here we show how this produces large biases, and how they can be reduced using the instrumental variable (IV) technique from econometrics. The interaction between stimulation and the absolute refractory period produces a weak, approximately random signal which can be exploited to estimate causal connectivity. When simulating integrate-and-fire neurons, we find that estimates from IV are better than naïve techniques ($R^2 = 0.77$ vs $R^2 = 0.01$). The difference is important as the estimates disagree when applied to experimental data from stimulated neurons with recorded spiking activity. Presented is a robust analysis framework for mapping out network connectivity based on causal neuron interactions.

1 Introduction

The central goal of neuroscience, arguably, is to understand the mechanisms or causal chains that give rise to activity in the brain, to perception, cognition, and action. Complex systems such as the brain are hard to understand because of the numerous ways the contributing elements may interact internally (Jonas and Kording, 2017). Because of this, it is not sufficient to know the correlations between variables or even be able to predict them. While observing correlations within the system is relatively easy, transitioning from observed correlations to a causal or mechanistic understanding is hard. After all, there can be many ways that the same activities emerge from distinct causal chains (Drton et al., 2011; Peters et al., 2017). Reaching a mechanistic level of understanding in the mammalian brain is incredibly hard as they contain countless neurons (e.g. 86 billion in the human brain (Azevedo et al., 2009)), each of which influences many other neurons. Even if we could record all neurons at the same time, estimating causality and producing a mechanistic understanding would be extremely challenging.

In today's typical studies, we only record from a small subset of all neurons. The data we obtain from such recordings, e.g. from electrophysiology or calcium imaging, is observational, which means that it does not result from randomized perturbations. In such cases, we can never know to which level the observed activity was caused by other observed activity, or by unobserved activity. The activity of the unobserved neurons is thus called confounders. If mechanisms are estimated from observational data in the presence of confounders, the consequence may be large errors and incorrect conclusions (Angrist and Pischke, 2008). Unobserved neural activity confounds estimates of causal interactions and makes it difficult to estimate underlying mechanisms.

Confounding is the big threat to causal validity (Pearl, 2009) irrespective of the use of simple regression techniques or advanced functional connectivity techniques (Stevenson et al., 2008; Honey et al., 2009; Aitchison and Lengyel, 2017; Pfau et al., 2013). To estimate connectivity it is first and foremost important that the used signals reflect cause and effect, therefore we use the term causal connectivity. Consider an example where we want to estimate causal connectivity between two observed, but unconnected neurons, A and C (Fig. 2(a)) by stimulating optogenetically. A third unobserved neuron B, is driven together with A by a common input and they are thus strongly correlated. Furthermore, B and C are connected, thus the input drives C through B. Consequently, A and C are also correlated and the regression $C = \beta A + noise$ will misleadingly conclude a direct interaction when causally

interpreted. In this case, we say the regressor A is endogenous and the regression coefficient β estimates the magnitude of association rather than the magnitude of causation. Naïve regressions in partially observed systems will generally not reveal causality.

A much used method for estimating the output of single neurons is to perform multiple regression analyses (Pillow et al., 2008; Roudi et al., 2009), modeling each neuron with a generalized linear model (GLM). Multiple regression may be a solution to confounding problems as they support "explaining away" background activity (Stevenson et al., 2008). However, this is only a meaningful strategy if most neurons are included in the recordings. Furthermore, only under certain assumptions about nonlinearity or noise sources does a fully observed system become identifiable (Daniusis et al., 2012; Shimizu et al., 2006). This is rarely the case in experimental settings, especially in the mammalian brain. Thus, brain data, will almost never satisfy the criteria needed for identifiability (Pearl, 2009). We thus aim to develop methods that can identify causal connectivity between neuron pairs.

To estimate causal relationships between neurons, stimulating the presynaptic neuron is the gold standard. In fact, a common definition of causality is in terms of the effect of changing one variable in the system, independently of changing other variables – an intervention (Pearl, 2009). If we stimulate single neurons, the ability to estimate causal relationships by regression is within reach. However, this is experimentally challenging and yields low cell count because it requires intracellular, juxtacellular or two-photon stimulation (Pinault, 1996; Lerman et al., 2017; Nikolenko et al., 2007; Emiliani et al., 2015). Because gold-standard perturbations are challenging, it is necessary and highly desirable if causality could be obtained from optogenetic stimulation in combination with neural recordings of large populations of neurons (Boyden et al., 2005; Zemelman et al., 2002).

Interpreting the results from optogenetic stimulation in terms of causal interactions is difficult. In most experimental settings, optogenetic stimulation will affect many neurons simultaneously. Hence, the stimulus will produce a distributed pattern of activity. This distributed pattern of stimulation produces activity which then percolates through the network of neurons. Thus any postsynaptic activity induced by stimulation could in principle come from any of the stimulated neurons, introducing problematic confounders.

For insights into how we may resolve the confounding problem induced by optogenetic stimulation, we may look to other fields that have addressed the problem of endogenous regressors. The inference of causality from observational data is addressed in the fields of statistics (Pearl, 2009), machine

learning (Peters et al., 2017) and econometrics (Angrist and Pischke, 2008). These fields have extensively worked on methods to estimate causality in the face of potential confounding and may offer us clues on how to solve our problems.

A commonly used approach towards causal inference in economics are instrumental variables, invented by Wright (1928); see Appendix B. Let us say that we want to estimate the return β from education x to yearly wages y with the regression $y = \beta x + u$. Here u are the factors other than education that contribute to yearly wages. One of the factors in u is a person's cognitive ability which may also affect education. The regressor x is thus correlated with the error term u. This will imply that the regression estimate of β will not estimate the magnitude of causation from education on wages, but rather, its association. In this case one may use the proximity to a college or university as an instrumental variable (IV) (Card, 1993). This is following the idea that proximity to a college does not affect your cognitive ability but may affect your proclivity to attend college. We expect living in proximity of colleges and universities to give higher probability to attend without affecting other contributing factors to wages such as cognitive ability. Then, in order to attribute the causal effect of education on wages one may calculate the ratio of covariances $\beta = cov(\text{proximity}, \text{wages})/cov(\text{proximity}, \text{education})$. This ratio corrects for the confounding factor.

The IV technique has been used extensively in econometrics and can provide provable unbiased causal estimates given three main assumptions (Angrist and Pischke, 2008). First, the instrument must be decorrelated with the error term. Second, the instrument must be correlated with the regressor. Third, there must be no direct influence of the instrument on the outcome variable, but only an influence through the regressor variable. The validity of these assumptions is central when using the IV approach.

For an instrument to be good, it needs to be unaffected by other variables. In the brain, almost everything is affected by the network state. However, certain variables can be more or less affected. For example, the overall activity of the network is due to slow and strongly nonrandom dynamics. In contrast, the temporal pattern of when a neuron is in a refractory state may be in good approximation random. First, if neurons are spiking according to conditional Poisson distributions, their exact timing conditioned on the network state, will be random. While refractoriness may not be perfectly random, the exact times of spiking are notoriously difficult to predict (Stevenson et al., 2008) suggesting that refractoriness is quite random.

Here we show that the IV technique can be employed if one seeks to estimate the causal connectivity between neuron pairs. We begin by showing

how confounding factors are introduced by conventional optogenetic stimulations. We then simulate this confounding effect in a simple network of three leaky integrate and fire (LIF) neurons. With this simple model we show that by using the refractory period as an IV we are able to distinguish between connected and unconnected neuron pairs. We compare these estimates with a naïve, although widely used, cross-correlation histogram (CCH) method that fails to distinguish respective pairs. We then turn to a simulated network of randomly recurrent connections of excitatory and inhibitory LIF neurons with distributed synaptic weights. With this data at hand, we first calculate the mean squared errors of the IV method and show that it is robust to different simulated network states. We also compare the amount and size of false positive and false negative estimates and goodness of fit on synaptic weights with pairwise assessments using CCH and logistic regression. Finally, we tested the methods on experimental data from extracellular recordings of single unit activity from optogenetically perturbed ensembles of neurons in the hippocampus. The observed differences between the IV and the CCH estimate, underline the importance of considering potential confounding, when estimating connections based on neural activity.

2 Results

2.1 Optogenetics is not local

Optogenetic stimulation is generally seen as a perturbation that by-and-large affects neurons in proximity (¡ 1mm) of the light source with the effect decreasing with distance. However, this is a misleading way of conceptualizing the spatial effect of stimulation as it depends on multiple factors. Light intensity and opsin density are important as more light and ion channels will cause a stronger effect on each cell. Moreover, the number of potentially stimulated neurons is critical as more neurons will have a larger impact on the overall population activity. Finally, physiological properties of the cells are important as light may have a stronger effect on spiking activity when the membrane potential of the cell is sufficiently close to the firing threshold. The induced effect of optogenetic stimulation as a function of distance should be given by a relation between the four parameters: light intensity, spatial distribution of neurons, distributions of membrane potential across neurons, and the distribution of induced photo currents.

To estimate the light intensity, we calculated the spatial extent of laser light delivered by fiber-optics under plausible experimental conditions according to Aravanis et al. (2007); see Section 4.7. While their experiment

uses blue light, their fits assume no absorption making the equations somewhat general. This modeling of light intensity yield an approximately $1/r^2$ reduction with distance r from the stimulation site Fig. 1 (cyan line). This is explained by the surface of a 3D shell growing with $4\pi r^2$ and photons will be roughly isotropic (but see (Thunemann et al., 2018)) beyond the scattering length Fig. 1 (inset). The same number of photons has to cross each of the spheres around the stimulation location unless they are absorbed or scattered inwards. As a result, the density of photons decreases with distance.

The number of illuminated neurons at a given distance will, however, increase with distance to the stimulation site given that neurons are roughly uniformly distributed in brain tissue Fig. 1 (black line). In fact, it will increase by approximately r^2 with distance. This derives from the same surface scaling as for the 3D shell as for the photon flow. Thus the number of neurons that can be activated increases rapidly with distance.

To estimate the effect of stimulation, the mechanism with which light affects spiking activity needs to be considered. This can largely be characterized by the distribution of membrane potentials across neurons. Surprisingly, this distribution has been observed to be symmetrically distributed and relatively flat (Paré et al., 1998; Destexhe and Paré, 1999; Rudolph and Destexhe, 2006). The expected response from a pulse of light that induces a charge Q should be proportional to the number of affected neurons whose membrane potential sit within a Q/C range of the threshold (C is the capacitance). Given that the distribution of membrane potentials is relatively flat (the density close to the threshold is generally within an order of magnitude of the density of its mode) suggests that the spiking response to a perturbation for any neuron is roughly proportional to the induced photo current. The peak amplitude of the photo current relates approximately logarithmically to the light intensity (Wang et al., 2007); see Section 4.7. Assuming that opsins are evenly distributed across neurons, the induced photo current will not be proportional to light intensity - it will fall slower. Based on this, we calculate the overall stimulation effect to be the product of the number of neurons in a spherical slice and the peak amplitude photo current. This product actually increases with distance (up to the distance where absorption becomes important) Fig. 1 (blue line). In other words, there is more activation at 500um than at 100um. Thus, optogenetic stimulation utilizing single photon activation does not produce a localized effect.

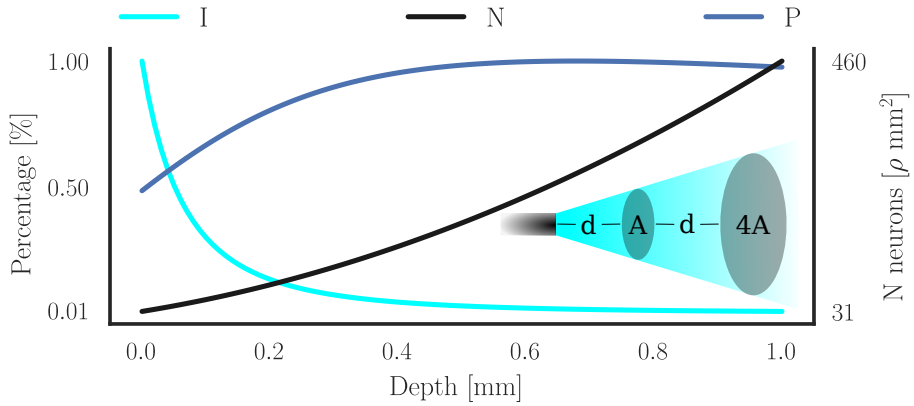


Figure 1: **Spatial extent of optogenetic stimulus**. Due to scattering and geometric loss the light intensity (I, cyan line) with an intensity of $10mW/mm^2$ exiting the optogenetic fiber follows approximately an inverse square law r^{-2} where r is the distance from the fiber. If neurons are uniformly distributed, the number of affected neurons in a spherical slice increases by r^2 (N, black line). The total photo current (P, blue line) calculated as the sum over neurons of peak amplitude photo current in a spherical slice thus increases with distance due to the nonlinear relation between light intensity and photo current, depicted as percentage of maximum.

2.2 Confounding as a problem for the estimation of causal effects

When we stimulate many neurons at the same time, and observe a postsynaptic neuron to be active after our stimulation, it is hard to know which of the stimulated neurons produced the activity. To illustrate such confounding effects we simulated a network comprised of three neurons (A,B, and C) Fig. 2(a). The neurons receive Poisson spike trains and have Gaussian white noise added to the membrane potential. Neurons were also interacting, where spikes of neuron B increase the probability of firing for neuron C, but there were no other interactions. Finally, we allowed simulated optogenetic stimulation (current pulse) to affect neurons A and B (but not C). We thus have a simple system for exploring questions of causality.

After running the simulation, the peri-stimulus time histogram of the stimulated neurons (Fig. 2(b)) shows the result of both the stimulation itself (suppressed for visibility) and the neuron's refractory period Fig. 2(c) (AA, BB). Since the stimulation affects A and B simultaneously, it induces a strong correlation between A and B Fig. 2(c) (AB). This further generates a strong correlation between A and C, confounding the system by rendering the cross-correlation histograms (CCHs) between BC and AC both statistically significant ($p_{\text{fast}} < 0.001, p_{\text{diff}} < 0.001$; see Section 4.2). A naïve reading of this result may suggest causal influences of both A and B on C.

Even though the correlation peak between B and C is larger than between A and C, due to correlated spikes outside the periods with stimulation, one may imagine a situation where only A and C is measured, giving rise to a false prediction that they are connected. Alternatively, A may have a stronger response to the optogenetic stimulation, in which case we may even have a stronger A-C correlation than B-C correlation. If stimulation affects multiple neurons simultaneously, there is a real confounding problem.

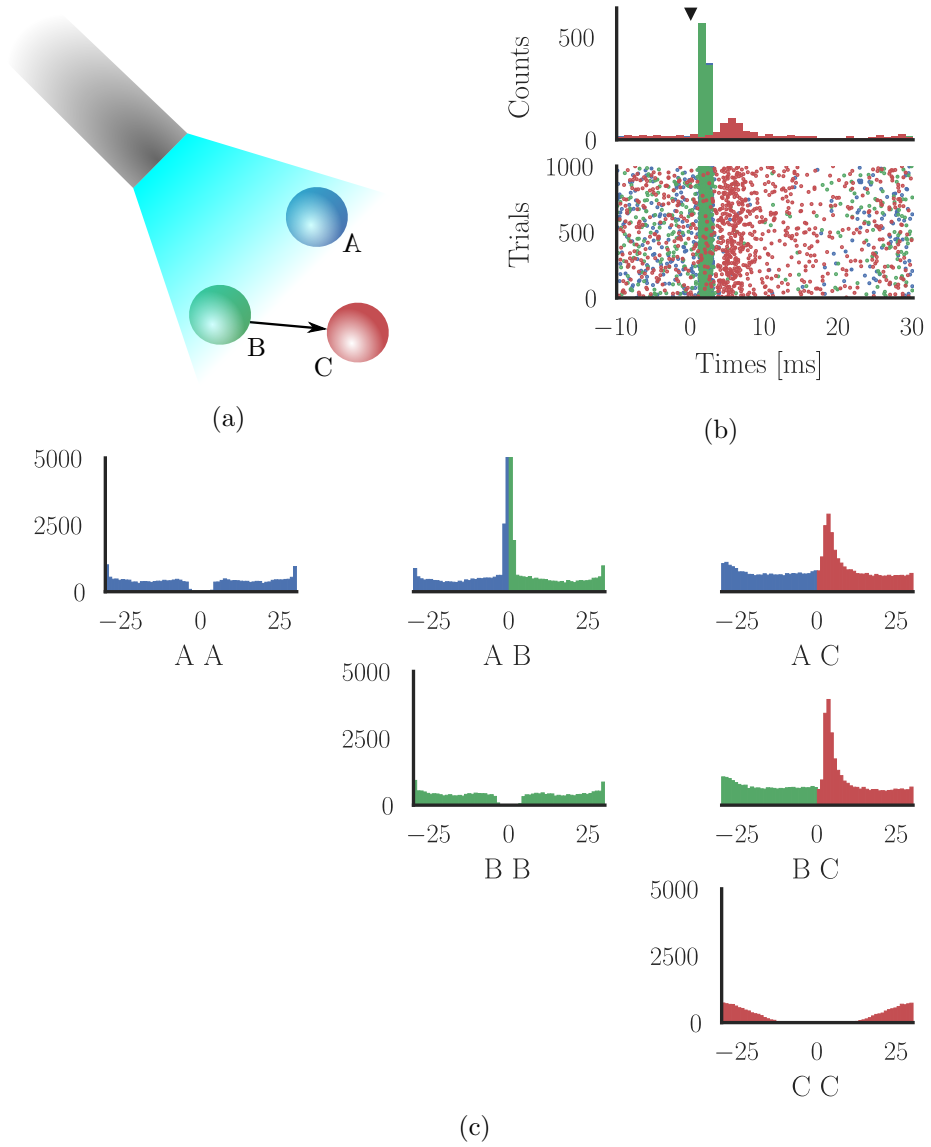


Figure 2: **Optogenetic stimulation induces spurious correlations**. Simple network containing three neurons shows stimulation configuration with blue laser light and the connections with arrows (a). The neurons A and B are stimulated in 1000 trials and the corresponding peristimulus time-histogram are shown in (b) upper panel with a raster plot in the lower panel. Cross-correlation histograms (CCHs) are shown in (c) where axes represent time lag in ms and counts of coincident spikes in bins of 1 ms.

2.3 Instrumental variables to resolve confounding

In order to estimate the actual influence of stimulation of a neuron on postsynaptic neurons, we need to distinguish the influence of one stimulated neuron from the influence of another stimulated neuron. We would thus need something that affects the stimulation effect separately across neurons. Arguably, refractoriness is such a variable. If a neuron is in its absolute refractory period, then no amount of stimulation will make it spike. This gives us an interesting way of inferring causality, by comparing the network state between a time when a neuron is able to spike and when the neuron is unable to spike when the stimulation hits.

Instrumental variables require the existence of a variable (e.g. refractoriness) that affects one variable of interest (the presynaptic neuron), but affects the rest of the network (including the postsynaptic neuron) only through that variable. This independent influence then allows quantifying the influence of the variable of interest on the rest of the network. In our case, the refractory states of a neuron is in good approximation independent on short time scales (see Discussion for caveats). It affects the influence of stimulation on the presynaptic neuron (Fig. 3(a)). The trials where a stimulus is unable to elicit a spike due to the refractory state can then be used to identify causal effect on the putative postsynaptic neuron.

We can now investigate if the use of an instrumental variable gives a better estimate of connectivity strength than simply analyzing the lagged correlations by means of the CCH calculated with Eq. (6) (Fig. 3(b, c) dashed lines). We use the IV estimator given by Eq. (4) on the three neuron system (Fig. 3(b, c) solid lines). It converges to the correct causal conclusions that the weights $w_{BC} = 0.2$ and $w_{AC} = 0$ as opposed to the CCH method which falsely concludes that $w_{AC} \approx 0.1$. For such a simple system, it produces meaningful estimates of the causal interactions between neurons.

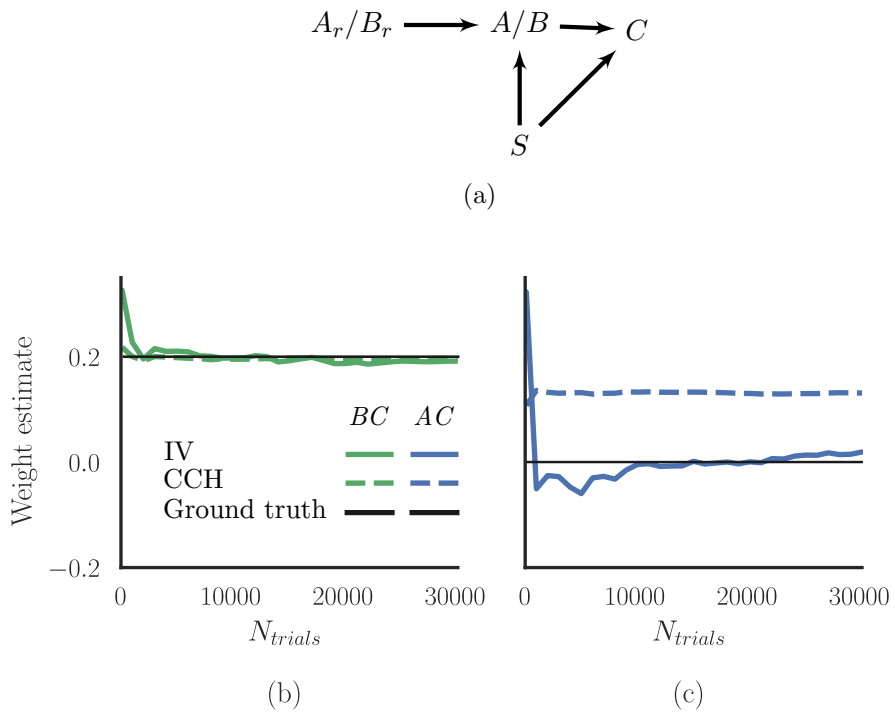


Figure 3: **Instrumental variable estimation (IV) of connectivity**. (a) during instrumental variable estimation we use a variable that is assumed to be random (here refractoriness) which influences a variable of interest (here spiking) and to use this influence to infer the causal interaction of that variable on other variables (here spiking of A or B onto C). A popular estimation approach for IVs, the Wald technique, correctly estimates causal connectivity in the A, B, C system using the refractory period. The path diagram in (a) shows the associations between neurons A, B or C, the stimulation S, and the IV as A_r or B_r (r for refractory). The IV estimator calculated by Eq. (4) converges to $\hat{\beta}_{BC} \approx 0.2, \hat{\beta}_{AC} \approx 0$ after approximately 5000 trials as seen in (b, c). Black line indicates ground truth and the results of the cross-correlations are showed as dotted lines. Note the difference between the IV estimate and CC method in (c).

2.4 Larger simulated networks

Interacting neurons in a biological network exhibit inhibition and interact in many ways. To evaluate the IV method in a more meaningful setting we simulated a recurrent neural network consisting of 1250 randomly connected LIF neurons where 250 had inhibitory synapses. The network was tuned to be in an asynchronous regime; see Fig. 7(a) with log-normally distributed synaptic weights according to patch-clamp experiments (Sayer et al., 1990; Mason et al., 1991); see Fig. 7(c) and Table 1 for parameters. Furthermore, we selected 800 excitatory neurons for stimulation and gave each neuron a random spatial distance from the simulated optogenetic stimulus (Fig. 4(a, b)). The stimulus intensity was then set according to Eq. (17) with a maximum of 8 pA, and was constant throughout the trials. The trial onset had a temporal Poisson distribution with period 100 ms and was further clipped between 100-150 ms. For weight estimates, we randomly selected among the excitatory population 100 stimulated neurons and 100 nonstimulated neurons.

To compare the IV and CCH methods we calculated the mean squared error of the weight estimates as a function of the number of trials and hit rate (Fig. 4(c, d)); see methods section for details. The IV estimator's precision decreases similarly in three different settings with varying amounts of relative inhibition g while CCH remain constant shown in Fig. 4(c) on a logarithmic scale. The slopes of the MSE for IV was found to be $-0.53, -0.48, -0.31$ for $g = 9.9, 4.4, 3.0$ respectively. Furthermore, the MSE increases with hit rate for both IV and CCH, where hit rate $= 1$ indicates that the stimulation induces a spike for each trial.

We then compared the IV estimator which exploits the refractory period, with the CCH method given by Eq. (6) which ignores network confounding. To get a good comparison between refractory and non-refractory states we required a maximum of 90% hit rate. To indicate the amount of connections that are falsely attributed to a non-zero weight, we calculated the amount of false positives. This was given as the percentage of estimated synapses larger than 0.05 where the true weight was 0, finding 99.4% for CCH and 0.2% for the IV estimator; see Fig. 5(a). In addition, we compared the size of the estimates at false positive instances and found that the CCH method have significantly higher median than IV (p=0, difference (Δ) = 0.164, permutation resampling (Wassermann, 2006)). The IV approach, while not being perfect, thus considerably outperforms the CCH approach.

It might be that modeling refractory periods in the context of a naïve regression estimates connectivity equally well as the IV method. We thus

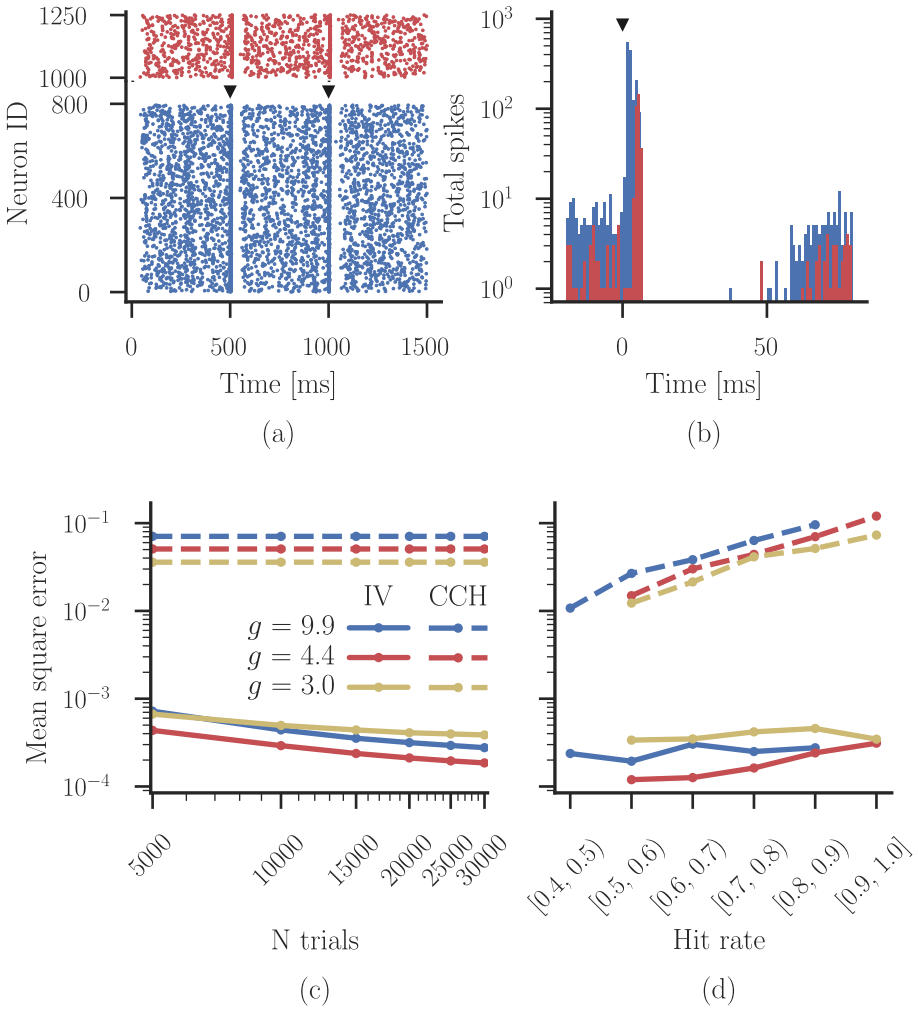


Figure 4: **Mean square error (MSE) of IV and CCH estimators in a network of two populations.** (a) Raster plots showing inhibitory neurons (upper panel, red) and excitatory neurons (lower panel, blue) stimulated with varying intensity with the strongest at lower neuron number. (b) Histogram of all neurons, where time zero indicates stimulation onset (marked), stimulated excitatory neurons drive inhibition which then silents the entire network. (c, d) The IV- and CCH estimators are evaluated for the recurrent neural network at three different amounts of relative inhibition g as a function of (c) number of trials and (d) hit rate.

performed a logistic regression Fig. 5a denoted LOGIT. Here, we show that LOGIT performs worse than IV (and even CCH) illustrating the advantage of using the refractory period as an instrumental variable (p=0, $\Delta = 0.577$, permutation resampling). To further evaluate the methods, we calculated false negatives as instances where the true weight is non-zero but estimated to be zero in Fig. 5(b) shows that the CCH and IV estimators perform equally well on that measure (p=0.59, $\Delta = 0.017$, permutation resampling), while the LOGIT has no false negatives. Finally, we wanted to evaluate the estimated weights as a function of true weights shown in Fig. 5(b,c) after 30000 trials. The IV estimator yields a good prediction ($R^2 = 0.77$), while the CCH method mainly estimates the strength of stimulation while true weights are poorly estimated ($R^2 = 0.01$). Utilizing refractory periods as an instrumental variable considerably improves the estimations.

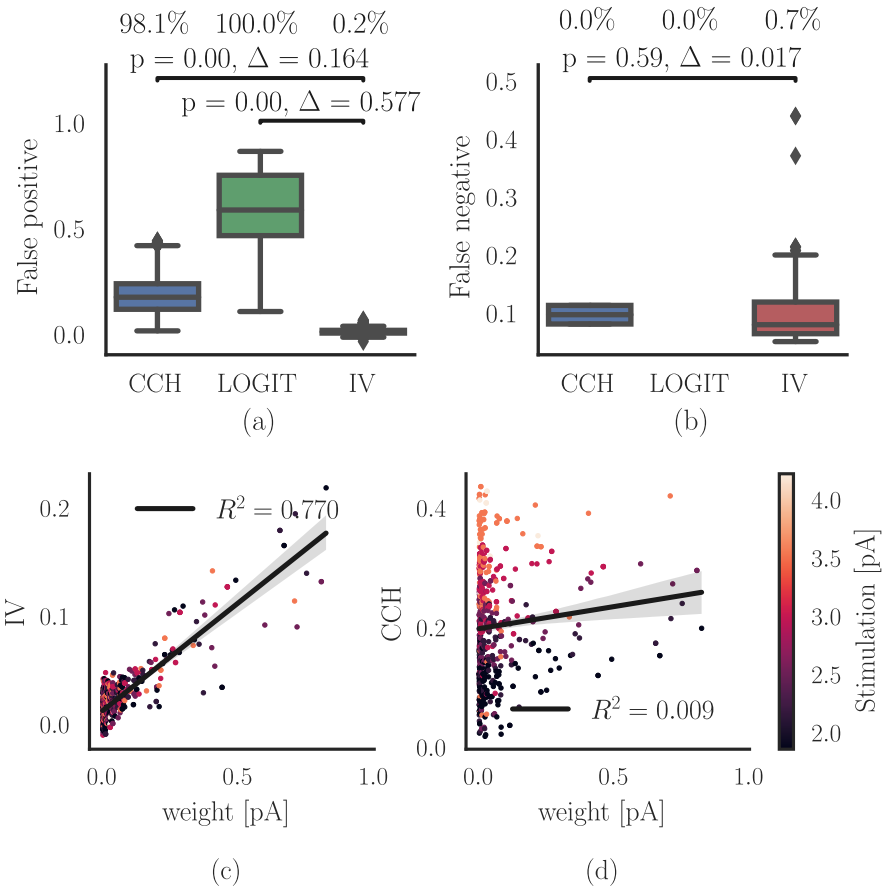


Figure 5: **False estimates and goodness of fit**. False positives are shown in (a) for the CCH method, logistic regression (LOGIT) and the IV estimator. False negatives for CCH and IV are shown in (b). Positive estimates of weight as a function of true weight are scattered for the IV estimator in (b) and CCH in (c), color coded by the size of perturbation intensity. Shaded area shows the 95 % confidence interval calculated by boot-strapping. Data were obtained from simulation of model 1 Table 1.

2.5 IV and CCH estimates from hippocampal spike recording with concurrent optogenetic stimulation

To test the feasibility of the IV approach on *in vivo* data, we compared the CCH and IV estimators on openly available extracellular single unit recordings from two mice with light-pulse stimulations of CA1 pyramidal neurons (English et al., 2017). We investigated connections from presynaptic units, that exhibited a significant increase in their firing rate upon stimulation, with putative postsynaptic units that did not respond to the stimulation (English et al., 2017). These experiments were not optimally designed for the use of the IV method to infer causal connectivity. Because of a low stimulation hit rate, we used an IV window of 7.5 ms, larger than the absolute refractory period of pyramidal cells of about 4 ms (as estimated from autocorrelograms, data not shown). This allows a proof-of-principle evaluation of the IV method on experimental data.

We want to know to what degree the IV method can be used on experimental data and if it gives different results. Indeed, when we apply the IV method, Fig. 6, it seems to work robustly. Although we find that the results from IV are somewhat correlated with those from CCH (Pearson correlation coefficient 0.34, p-value of non-correlation 0.07, on median values after bootstrapping), there are considerable differences between the methods. In many cases, the confidence bounds of the two methods are truly non-overlapping Fig. 6(a,b), suggesting that the differences can not only be explained by noise. These preliminary results illustrate that the IV approach may be applicable to address a broad range of causality estimation problems in neuroscience.

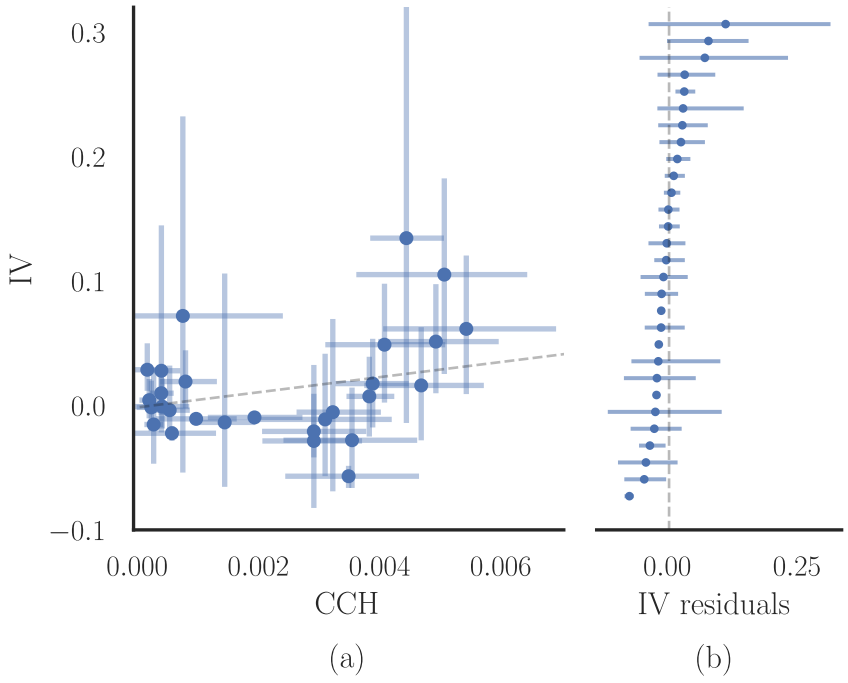


Figure 6: **IV and CCH estimates on hippocampal recordings.** IV and CCH estimates for connections from optogenetically activated to non-activated units that showed a significant peak in the CCH (English et al., 2017). (a) Blue dots are median values and error bars indicate the 95% confidence intervals calculated by bootstrapping ($n = 1000$). Dashed line represents linear fit of median values, which we show because CCH estimates may be linear biased. For better visualization, we omitted one estimated connection at $CCH = 0.01$ and $IV = 0.36$. (b) Sorted IV residuals and errorbars in relation to the linear fit (vertical line). In many cases, 95% confidence intervals do not overlap with linear fit.

3 Discussion

Here we have asked if the refractory period of neurons can be used as an instrumental variable to reverse engineer the causal flow of activity in a network of simulated neurons. We have found that this approach performs considerably better than the naïve method. We have found that neither naïve linear nor naïve logit models produce reliable estimates of connectivity between neuron pairs. The IV approach effectively reverse engineers causality by looking at the response that is missing because of refractoriness which effectively allows better estimates of causal effects. When applied to real data, we obtain robust estimates that differ from those of naïve estimators.

At the moment, we have no ground-truth data set at hand to test our technique and compare with other approaches. Ideally, we would have known causal effects from single-cell stimulation (e.g. from two-photon optogenetics) to establish causal effects. Such data should contain many randomly distributed, short and intensive stimulation trials combined with traditional optogenetics, designed in a way where refractoriness matter. Such a dataset, to the best of our knowledge, is currently not available and prevents us from testing how good our estimator would work on experimental data. Future experiments are needed to obtain reliable insights.

For the refractory period to be a good instrument, it is necessary that it is not overly affected by the network activity. This will clearly be problematic in many cases. After all, network activity affects neuron activity and hence refractoriness. However, there are multiple scenarios where refractoriness will be a good instrument. For example, if we have balanced excitation and inhibition, we may expect largely independent refractory states of individual neurons. If a neuron biophysically implements something like conditional Poisson spiking, its refractory states will be random conditioned on the network state. Importantly, we may expect the phase of a neuron to be far more random than the activity of the network as a whole.

The randomness of refractory times is the one factor which makes or breaks the IV approach. Even if neurons' refractory states are strongly correlated during normal network operation, there may be ways of randomizing refractoriness. First, it would help to use a task and situation where neurons are as uncorrelated as possible. Second, we may use a set of conditioning pulses of stimulation to increase independence of refractory states. Giving one burst of stimulation which is strong enough to elicit multiple spikes from each neuron may effectively randomize their phases (Ermentrout et al., 2008). Third, we may utilize chemical, behavioral, or molecular perturbations to produce a good instrumental variable. For example, we may be

able to construct intracellular oscillators that are unaffected by neural activities or constructs that force a neuron into quiescence at random times. In neuroscience there has been no effort yet to produce good instrumental variables, so there may be many possibilities for improvements.

One popular way of estimating causal effects is fitting generalized linear models (GLMs) to simultaneously recorded neuron activities (Pillow et al., 2008; Roudi et al., 2009). GLMs are multiple nonlinear regressions and require multiple neurons to perform well. In fact, if activity from all neurons were recorded, GLMs might be sufficient to estimate causal connections. However, complete recordings are not possible in the mammalian brain, especially not in primates, where recordings include only a very small subset of the neurons involved in the actual computation. When using GLMs one may accuratly estimate latency distributions and sequences of spikes from individual neurons. These ideas should, arguably, be merged with IV approaches. One of the strengths of the IV estimator presented here is that it only requires one pair to be recorded because we can utilize the randomness of the refractory periods along with random stimulations. Under those assumptions, the IV estimator can produce actual causal estimates.

The main problem with optogenetic stimulation, when used to infer connectivity, is its non-local property. This is due to the inverse relation between changes in light intensity and affected number of neurons combined with a logarithmic relation between light intensity and photocurrent Wang et al. (2007). In addition, the distribution of membrane potentials across neurons is relatively flat (Destexhe and Paré, 1999; Rudolph and Destexhe, 2006; Paré et al., 1998) making neurons highly sensitive to perturbations. One could however, imagine situations where optogenetic activation was more local. If for example, the membrane potential distributions were skewed with the mode far from threshold, a very strong stimulus would be required for a neuron to elicit spikes. There could also be other ways of making optogenetic stimulation more local. For example, if one engineered opsins or brain tissue that are more light absorbent (e.g. by ubiquitously producing melanin) one could stimulate more locally. How to engineer more localized stimulation is an important problem when causally interrogating a system.

Very weak laser pulses in noisy networks might mainly elicit spikes in very few close-by neurons in each trial (English et al., 2017). However, the stimulus will still affect the membrane potential of many neurons further away, some of which will spike. Therefore, weak stimulation does not remove the principal problem of correlation based techniques. After all, the network still acts as a confounder and, if anything, the weak stimulation will reduce the statistical power of the approach. Lowering stimulation amplitudes does

not appear to be a way of obtaining meaningful causal estimates.

There are many techniques for causal inference, most of which are largely unknown to the field of neuroscience, and are based on approximating randomness in a seemingly regular world. In many cases, one could use regression discontinuity designs in a spiking system (Lansdell and Kording, 2018; Imbens and Lemieux, 2008). Moreover, one could use a difference in difference approach (Abadie, 2005). Matching approaches (Stuart, 2010; King and Nielsen, 2016), can be used when comparing similar network states and their evolution over time. In general, neuroscience is in a quest for causal interpretations, we should therefore be able to benefit considerably by utilizing techniques that are popular in the field of causal inference.

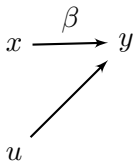
4 Methods

4.1 Instrumental variable estimation

A simple approximation of the connectivity strength between a presynaptic neuron x and postsynaptic neuron y can be to ignore external excitation and simply calculate the relation between the spike times in x and y with a regression model given by

$$y = \beta x + u. \tag{1}$$

Here y is the dependent variable, x is the explanatory variable, β is the effect of x on y and u is an unknown error term. Equation (1) follows from the causal path diagram (Wright, 1921, 1923)

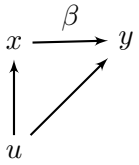


Assuming that changes in spike times y are described by βx i.e. $\frac{\mathrm{d}y}{\mathrm{d}x} = \beta$ for spike times $x \forall x, y \in C^1$. One problem with this idea is that in a confounded system, perfectly correlated neurons will give statistically indistinguishable β. In the extreme case where two neurons are both made to fire every time they are stimulated, they will have the same weights according to Eq. (1). After all, during stimulation $y = 1$ for both, even if only one of them drives the postsynaptic neuron. Another problem is if the network state affects both the probability of a neuron to fire and also the probability

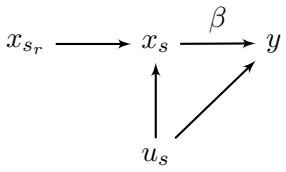
of postsynaptic neurons to fire. In this case, the network state can induce a correlation which will make the estimation highly biased. Arguably, the network state will, in all realistic models, have a dramatic influence on all neurons and the regression model is better described by

$$y = \beta x + u(x). \tag{2}$$

Corresponding to the following path diagram



Here we have the relation $\frac{\mathrm{d}y}{\mathrm{d}x} = \beta + \frac{\mathrm{d}u}{\mathrm{d}x}$. To get at causality we thus require some stimulation that only highlights the activity in y caused by x, disassociating x from u. However, the optogenetic stimulation is not specific to x and will activate parts of the network activity u. Let us assume that the stimulus renders only a subset of u correlated with x, namely u_s (s denotes stimulated). To disassociate x_s from u_s we need something that can distinguish between different neurons that are stimulated. We thus require some instrument x_{s_r} which is (1) uncorrelated with the network u, (2) is correlated with the regressor x_s and (3) not correlated with y (Angrist and Pischke, 2008). We assume that the neurons are independent at small time scales and that stimulation additionally randomize membrane potential individually in neurons. We may thus use the fact that a neuron that has fired just before the stimulation will be in an absolute refractory state and hence have $x_{s_r} = 0$ independently of u, where the subscript s_r denotes stimulation during refractory state. This introduces times where the spike from one of the stimulated neurons are missing. Thus we may use the refractory period as an instrumental variable, as illustrated with the following path diagram



Here x_{s_r} represent times where the presynaptic neuron is refractory during stimulation. The true β is given (Wright, 1928) by

$$\beta_{IV} = \frac{\mathrm{d}y}{\mathrm{d}x_{s_r}} / \frac{\mathrm{d}x}{\mathrm{d}x_{s_r}} \tag{3}$$

This is then an estimator that compares the postsynaptic activity when a given neuron is non-refractory with the postsynaptic activity when it is refractory, thus removing the confounding.

Since our instrument x_{s_r} is binary we may calculate the IV (or more precisely Wald) estimator (Wald, 1940) β_{IV} by

$$\hat{\beta}_{IV} = \frac{\bar{y}_s - \bar{y}_{s_r}}{\bar{x}_s - \bar{x}_{s_r}} = \bar{y}_s - \bar{y}_{s_r} \tag{4}$$

Here $\bar{y}_s$ is the average number of trials where successfully stimulating x resulted in a response in y and $\bar{y}_{s_r}$ is the average number of trials where an unsuccessful stimulation of x resulted in a response in y. The successful stimulations of x are denoted x_s and thus $\bar{x}_s \equiv 1$. Conversely x_{s_r} denotes unsuccessful stimulations of x i.e. stimulations of x during its refractory state and $\bar{x}_{s_r} \equiv 0$.

To utilize the refractory period as an IV on the simulated data we first picked out one window of 4 ms for each of the presynaptic and postsynaptic neuron with a latency relative to stimulation time of 0 and $\tau_{syn} + D$ ms (see Eq. (10)) respectively. By classifying each window for each trial whether x contained a spike we obtained the two arrays y_s and y_{sr}.

We found some negative values of the IV estimator which were largely suppressed by requiring a hit rate $< 90\%$. Hit rates larger than 90% happens mainly at strong stimulation intensity which lower statistical power in the IV or can lead to correlated refractory times. We hypothesize that strong stimulations can lead to synchrony induced by the stimulation (Ermentrout et al., 2008). This hypothesis was strengthened by observing that the negative values did not occur when the stimulation intensity was set to zero (data not shown). Furthermore, the simulated neural network introduces much response overlap due to synapses having identical synaptic time constants and transfer delays. This can interfere with inference since multiple neurons are affecting the same cell at the same time for each stimulation. However, this is less likely to occur in biological networks, which have high variability of synaptic properties and where firing patterns are sparser. These biological aspects would most likely work to the advantage of the IV method. Independence of refractoriness would be further improved, if in addition a clever stimulation routine was implemented such that the distribution of stimulation strength varies spatially from trial to trial.

4.2 Cross correlation histogram

The statistical tests giving the probabilities p_{diff} and p_{fast} were done according to Stark and Abeles (2009); English et al. (2017). Briefly, to test if the cross correlation histogram (CCH) peak was significant we employed two tests. By using the Poisson distribution with a continuity correction (Stark and Abeles, 2009) given by Eq. (5) we calculated p_{diff} by comparing the peak in positive time lag with the maximum peak in negative time lag, called p_{causal} in English et al. (2017). The probability p_{fast} represents the difference between CCH and it's convolution with a hollow Gaussian kernel (Stark and Abeles, 2009). These two measures of significance were required to be < 0.01 and given by

$$p(N|\lambda(m)) = 1 - \sum_{k=0}^{N-1} \frac{e^{-\lambda(m)}\lambda(m)^k}{k!} - \frac{e^{-\lambda(m)}\lambda(m)^N}{2N!}. \tag{5}$$

Here λ represents the counts at bin m and N is the number of bins considered. To estimate the connection weight between pairs we used the spike transmission probability first defined in English et al. (2017) as

$$p_{trans} = \frac{1}{n} \sum_{m=3ms}^{6ms} CCH(m) - \lambda_{Gauss}(m), \tag{6}$$

where n is the number of spikes detected in the presynaptic neuron and $\lambda_{Gauss}(m)$ is the CCH count convolved with a hollow Gaussian kernel at bin m.

4.3 Logistic regression

To utilize the refractory period without using it as an IV we estimated synaptic weights using a logistic regression. To do this we first picked out one window of 4 ms for each of the presynaptic and postsynaptic neuron with a latency relative to stimulation time of 0 and $\tau_{syn} + D$ ms (see Eq. (10)) respectively. By classifying each window for each trial whether it contained a spike we obtained two binary arrays, the regressor x and the dependent variable y where we want to estimate the probability $P(y = 1|x)$ by fitting the parameters $\boldsymbol{\beta}$ such that

$$y = \begin{cases} 1 & \text{if } \beta_0 + \beta_1 x + u > 0 \\ 0 & \text{else} \end{cases} \tag{7}$$

where u is an error term. Further, we used the logit link function such that the the probability giving the proxy for synaptic weight is given by

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \tag{8}$$

The model was fitted using the python package scikit-learn (Pedregosa et al., 2011)

4.4 Simulated network

To simulate a recurrent network of excitatory and inhibitory neurons we used NEST (Gewaltig and Diesmann, 2007) to and the LIF model given by

$$\frac{\mathrm{d}V_m^i}{\mathrm{d}t} = -\frac{(V_m^i - E_L)}{\tau_m} + \frac{I_{syn}^i(t)}{C_m}. \tag{9}$$

When the membrane potential V_m^i of neuron i reaches a threshold V_{th} an action potential is emitted and V_m^i reset to the leak potential E_L followed by an absolute refractory period τ_{ref}. The membrane time constant is represented by τ_m and $I_{syn}^i(t)$ denotes the post synaptic current (PSC) for neuron i modeled as a sum of alpha functions given by

$$I_{syn}^i(t) = \sum_{j=1}^{C} J_j \alpha(t - t_j - D), \tag{10}$$

where t_j denotes an incoming spike through synapse j at delay D and C is the number of incoming synapses on neuron i. The PSC amplitude is given by J_j and the alpha function is given by

$$\tau_{syn}\alpha(t) = te^{-\frac{t}{\tau_{syn}}} H(t). \tag{11}$$

Here τ_{syn} denotes the synaptic integration time constant and H is the Heaviside step function. All neurons were driven by an external Poisson process with rate $rate_p$.

Synaptic weights were log-normally distributed such that the increase in membrane potential V_m^i due to one spike were restricted to lie between $V_{syn} = 0.0mV$ and $V_{syn} = 2.0mV$ based on experimental findings (Sayer et al., 1990; Mason et al., 1991). The synaptic distribution is shown in Fig. 7(c) where the inhibitory PSC amplitude is given by $J_{in} = gJ_{ex}$ where J_{ex} denotes the excitatory synaptic weight.

To find suitable parameters yielding asynchronous activity we measured the population correlation coefficient given by

$$\langle CC \rangle_{pop} = \left\langle \left\langle \frac{h_i - \langle h_i \rangle}{std(h_i)} \frac{h_j - \langle h_j \rangle}{std(h_j)} \right\rangle \right\rangle_{pop}, \tag{12}$$

where h is the spike time histogram with binsize at $5ms$ for neuron i, j and $\langle \cdot \rangle$ is the mean operator. The distribution of CC is shown in Fig. 3 which were found by performing several parameter sweeps picking three parameter sets which mainly differed in firing rate (data not shown).

To further evaluate the network state we calculated the coefficient of variation (CV) of the population given by

$$\langle CV \rangle = \left\langle \frac{std(ISI_i)}{\langle ISI_i \rangle} \right\rangle_{pop}, \tag{13}$$

where ISI denotes the inter-spike interval of neuron i. We were unable to have the network showing an irregular state; see Fig. 7(b) partly to the finite synaptic integration time constant $\tau_{syn} = 1ms$. To verify that indeed this was due to τ_{syn} we performed several simulations with lower τ_{syn} obtaining $\langle CV \rangle_{pop} > 1$ (data not shown). It would likely be easier to achieve irregular network state if synapses were conductance based (Kumar et al., 2008). However, we settled with current based synapses as we were mainly interested in achieving an asynchronized state ($\langle CC \rangle_{pop} < 0.01$).

4.5 Calculating the mean square error

The conditional probability of neuron y firing given a spike from neuron x denoted $P(y|x)$ is related to the connection strength w_{xy} and the background activity. Since LIF neurons integrate linearly and have fixed thresholds we expect $P(y|x)$ to be proportional to the connection strength w_{xy}, i.e. $P(y|x)$ w_{xy}. To calculate the mean squared errors we thus normalized w_{xy} to the range $[0, max(IV_{xy})]$ or $[0, max(CCH_{xy})]$ to calculate the MSE of the IV and CCH estimators respectively by

$$MSE(w) = \frac{1}{N} \sum (\hat{w} - \bar{w})^2).$$

Here $\bar{w}$ is the normalized weight and $\hat{w}$ is the estimated weight.

4.6 Calculating goodness of fit and false estimates

The strength of the perturbations was at the maximum very large and led to many instances where the hit-rate was above 90%. This represents extreme

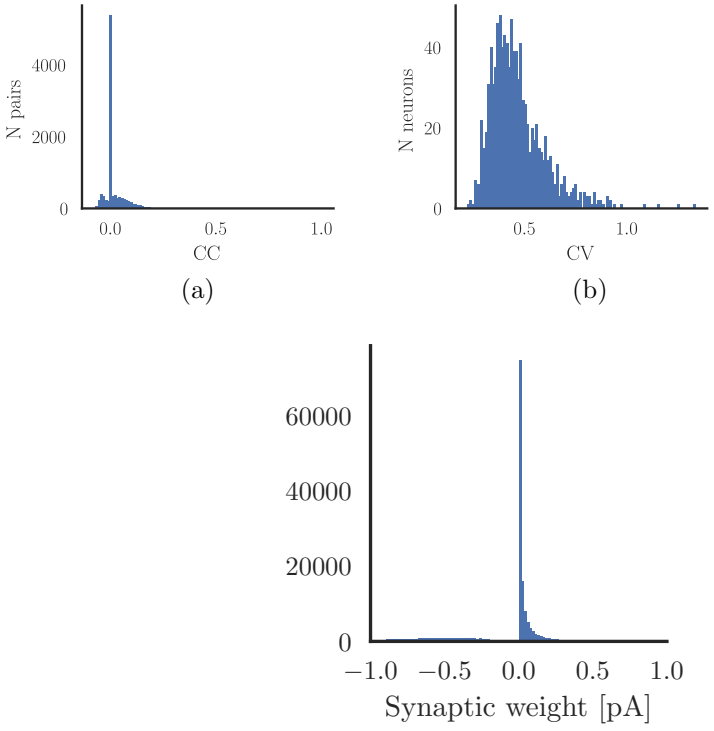


Figure 7: Network state

name	model 1	model 2	model 3	units
$N_{neurons}$	1250			
Δt	0.1			
N_{ex}	1000			
N_{in}	250			
eta	0.9			
$rate_p$	3694.26			Hz
V_{reset}	0			mV
V_m	0			mV
E_L	0			mV
t_{ref}	2			ms
τ_m	20			ms
V_{th}	20			mV
C_m	1			pF
V_{syn}	0.2			mV
g	9.9	4.4	3	
V_{syn}^{high}	2.05			mV
V_{syn}^{low}	0.05			mV
var_{syn}	0.5			mV^2
τ_{syn}^{in}	1			ms
τ_{syn}^{ex}	1			ms
$delay$	1.5			ms
eps	0.1			
C_{ex}	100			
C_{in}	25			
J_{in}	0.88727	0.394342	0.26887	pA
J_{ex}	0.0896232			pA
J_{high}^{ex}	0.918638			pA
J_{low}^{ex}	0.0224058			pA
J_{high}^{in}	0.918638			pA
J_{low}^{in}	0.0224058			pA
$time_{simulation}$	3685312			ms
$rate_{in}$	8.69	10.56	12.7	Hz
$rate_{ex}$	6.5	9.22	11.5	Hz

Table 1: Simulation parameters of three different models.

name	model 1	model 2	model 3	units
$stim_N^{in}$	0			
$stim_N^{ex}$	800			
$stim_{amp}^{in}$	0			pA
$stim_{amp}^{ex}$	10			pA
$stim_{duration}$	2			ms
$stim_{period}$	100			ms
$stim_{max}^{period}$	150			ms
$density$	7514			Nmm^{-3}
S	10.3			mm^{-1}
NA	0.37			
r	0.1			μm
n	1.36			
n_{Hill}	0.76			
K	0.84			
$depth$	0.7			mm

Table 2: Stimulation parameters of three different models.

experimental conditions with very high light intensity and would yield large errors Fig. 4. To see how the IV and CCH estimators compared we thus selected source neurons that had strictly less hit-rate than 90%. This led to the maximum perturbation strength of $5pA$.

4.6.1 Goodness of fit

The goodness of fit was indicated by the R^2 value of a linear regression calculated by ordinary least squares (Seabold and Perktold, 2010) of the relation between estimated values (IV or CCH) with the true weight.

4.6.2 False positives

We calculated false positives from a subset N_{sub} of the pairs N where the true weight was $w = 0$. Then a false positive was defined as an estimate that were larger than 0.05. The percentage of false positives was then calculated by $N_{sub}(x > 0.05)/N_{sub}$.

4.6.3 False negatives

We calculated false negatives from a subset N_{sub} of the pairs N where the true weight was $w = 0$. Then a false negative was defined as an estimate that had a difference from the true weight which were larger than 0.05. The percentage of false negatives was then calculated by $N_{sub}(|x - w| > 0.05)/N_{sub}$.

4.7 Perturbation intensity

In order to replicate an optogenetic experiment we modeled transmission of light through brain tissue with the Kubelka-Munk model for diffuse scattering in planar, homogeneous media (Ho et al., 2017) given by

$$T = \frac{1}{Sr + 1}. \tag{14}$$

Here T denotes a transmission fraction, S is the scattering coefficient for mice (Aravanis et al., 2007) and r is the distance from a light source. Further we combined diffusion with geometric loss assuming that absorption is negligible as in Aravanis et al. (2007) and computed the intensity as presented in Fig. 1 by

$$\frac{I(r)}{I(r = 0)} = \frac{\rho^2}{(Sr + 1)(r + \rho)^2} \tag{15}$$

where r is the distance from the optical fiber and

$$\rho = \frac{d}{2}\sqrt{\left(\frac{n}{NA}\right)^2 - 1}. \tag{16}$$

Here d is the diameter of the optical fiber, NA is the numerical aperture of the optical fiber and n is the refraction index for gray matter (Ho et al., 2017); see numerical values for parameters in Table 2.

To estimate the distribution of light intensity on stimulated neurons we distributed 795 neurons uniformly in 10 spharical slices in the range $[0, 1mm]$ which had a radius given by the cone shaped light; see 1 inset.

To further estimate the peak amplitude photo current we used the Hill equation fitted by parameters found in Wang et al. (2007) given by

$$P = I_{max}\frac{I^n}{K^n + I^n} \tag{17}$$

Here, $I_{max} = 642pA$ is the maximum curren, $n = 0.76$ is the Hill coefficient and $K = 0.84mW/mm^2$ represents the half-maximal light sensitivity of the

ChR2. We further used the light intensity I given by Eq. (15) multiplied by an initial intensity of $10mW/mm^2$.

Since the model neurons are not scaled to mimic "realistic" values in their membrane potential we set the maximum stimulation strength to $8pA$ which was found suitable by investigating the percentage of successful stimulations to be $< 100\%$. We then selected 100 of the excitatory neurons that were not stimulated as the "target" population which together with the inhibitory neurons were not perturbed directly by the light stimulus.

4.8 Application to hippocampal recordings

We applied the IV method to CA1 recordings of two mice[1]. These recordings came from silicon probes of four shanks with thirty two channels and had integrated μLEDs implanted in hippocampal region CA1. Mice expressed channelrhodopsin-2 under the control of an excitatory neuron-specific promoter, CaMKII::ChR2 (English et al., 2017). Each animal was recorded on several days, while mice freely behaved in their home cage. Each recording lasted more than 3 hours. Spikes were then sorted by the experimenter using kilosort (Pachitariu et al., 2016). During analysis, we treated each recording day independently. During each session, sinusoidal and pulse stimulations lasting 10 or more milliseconds were applied with different intensities. The experimenter determined whether a unit was optogenetically stimulated by applying two criteria (English et al., 2017). First, the number of spikes during each stimulation was noted and compared to the number of spikes in the same interval but two seconds before the stimulation. A unit was considered optogenetically labeled, if the p-value of a Wilcoxon ranksum non-parametrical test of means was below 10^{-10}. Second, it was required that the absolute number of spikes during stimulation is on average 50% larger compared to the number of spikes in same the interval, but two seconds before stimulation.

We considered only sessions containing pulse-stimulations. At each shank, we grouped similar stimulation intensities. By cross-correlating stimulation onset time with spikes, we quantified the required time for each intensity group to significantly increase the firing rate above baseline. We first binned spikes with 3 ms bin-width and then calculated the baseline by taking the mean of the cross-correlogram for negative time lags, -45 to -1.5 ms. Finally, we applied a significance threshold of 0.01 on the probability to obtain the observed or a higher count per bin of the poisson distribution with con-

[1] `https://buzsakilab.nyumc.org/datasets/McKenzieS/`

tinuity correction (Abeles, 1982). For each labeled neuron, those intensities were selected that caused a significant increase in firing probability within 7.5 ms. A labeled unit was only considered a putative presynaptic partner if the number of significant pulse stimulation events exceeded 2500. In total, connections between 17 putative presynaptic and 86 putative postsynaptic units were calculated.

We calculated the CCH based connection strength, p_{trans}, according to (English et al., 2017) and section 4.2. First, spiketrains of optogenetically labeled, putative presynaptic as well as unlabeled, putative postsynaptic units were binned with $0.4ms$ binwidth. The CCH has been convolved with a partially hollow gaussian kernel, with $10ms$ standard deviation and a hollow fraction of 0.6. We selected a time window for p_{fast} and p_{trans} to be between 0.8 and $2.8ms$. The reference negative time window for p_{diff} was -2 to $0ms$. We further applied the same significance level of 0.01 applied to p_{fast} and p_{diff}.

For computing the IV estimate, we used a window size of $7.5ms$ and a fixed synaptic delay of $1ms$.

IV and CCH estimates were separately bootstrapped with a sample size of 1000. For CCH estimates, we employed the additive property of the CCH function. We subdivided spiketrains of a session in in $100s$ segments, and calculated the CCH for each segment individually. We then randomly selected segments with replacement, to match the full length of the session. Estimates from CCH were calculated on the sum of segments. For the IV estimates, we detected, for each optogenetic stimulation event, wether pre- and postsynaptic units are active. Further, we randomly selected stimulation events with replacement to match the original number of stimulations and calculated the IV estimate on the random sample.

5 Acknowledgments

This research was partly funded by the National Institutes of Health (NIH) grant R01MH103910 and the University of Pennsylvania. In addition, the Research Council of Norway Grant 231248 and the University of Oslo. T.S (PhD fellow), M.F. (PI) are part of the Simula-UCSD-University of Oslo Research and PhD training (SUURPh) program, an international collaboration in computational biology and medicine funded by the Norwegian Ministry of Education and Research. We thank Sam McKenzie, Daniel Fine English from the lab of György Buzsáki for sharing and explaining the optotrode data.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1):1–19.

Abeles, M. (1982). Quantification, smoothing, and confidence limits for single-units' histograms. *Journal of neuroscience methods*, 5(4):317–325.

Aitchison, L. and Lengyel, M. (2017). With or without you: predictive coding and bayesian inference in the brain. *Current opinion in neurobiology*, 46:219–227.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Aravanis, A. M., Wang, L. P., Zhang, F., Meltzer, L. A., Mogri, M. Z., Schneider, M. B., and Deisseroth, K. (2007). An optical neural interface: in vivo control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *J. Neural Eng.*, 4(3):S143–S156.

Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., Lent, R., Herculano-Houzel, S., et al. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541.

Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G., and Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. *Nature neuroscience*, 8(9):1263.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. Technical report, National Bureau of Economic Research.

Daniusis, P., Janzing, D., Mooij, J., Zscheischler, J., Steudel, B., Zhang, K., and Schölkopf, B. (2012). Inferring deterministic causal relations. *arXiv preprint arXiv:1203.3475*.

Destexhe, A. and Paré, D. (1999). Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *Journal of neurophysiology*, 81(4):1531–1547.

Drton, M., Foygel, R., and Sullivant, S. (2011). Global identifiability of linear structural equation models. *The Annals of Statistics*, pages 865–886.

Emiliani, V., Cohen, A. E., Deisseroth, K., and Häusser, M. (2015). All-optical interrogation of neural circuits. *Journal of Neuroscience*, 35(41):13917–13926.

English, D. F., McKenzie, S., Evans, T., Kim, K., Yoon, E., and Buzsáki, G. (2017). Pyramidal Cell-Interneuron Circuit Architecture and Dynamics in Hippocampal Networks. *Neuron*, 96(2):505–520.

Ermentrout, G. B., Galán, R. F., and Urban, N. N. (2008). Reliability, synchrony and noise. *Trends in neurosciences*, 31(8):428–434.

Gewaltig, M.-O. and Diesmann, M. (2007). Nest (neural simulation tool). *Scholarpedia*, 2(4):1430.

Ho, A. H. P., Kim, D., and Somekh, M. G. (2017). *Handbook of photonics for biomedical engineering*. Springer Netherlands.

Honey, C., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.-P., Meuli, R., and Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6):2035–2040.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2):615–635.

Jonas, E. and Kording, K. P. (2017). Could a Neuroscientist Understand a Microprocessor? *PLoS Comput. Biol.*, 13(1):1–24.

King, G. and Nielsen, R. (2016). Why propensity scores should not be used for matching. *Copy at http://j. mp/1sexgVw Download Citation BibTex Tagged XML Download Paper*, 378.

Kumar, A., Schrader, S., Aertsen, A., and Rotter, S. (2008). The high-conductance state of cortical networks. *Neural Comput.*, 20(1):1–43.

Lansdell, B. and Kording, K. (2018). Spiking allows neurons to estimate their causal effect. *bioRxiv*, page 253351.

Lerman, G. M., Gill, J. V., Rinberg, D., and Shoham, S. (2017). Two photon holographic stimulation system for cellular-resolution interrogation of olfactory coding. In *Optics and the Brain*, pages BrM3B–5. Optical Society of America.

Mason, a., Nicoll, A., and Stratford, K. (1991). Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro. *J. Neurosci.*, 11(January):72–84.

Nikolenko, V., Poskanzer, K. E., and Yuste, R. (2007). Two-photon photostimulation and imaging of neural circuits. *Nature methods*, 4(11):943.

Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., and Harris, K. D. (2016). Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *BioRxiv*, page 061481.

Paré, D., Shink, E., Gaudreau, H., Destexhe, A., and Lang, E. J. (1998). Impact of spontaneous synaptic activity on the resting properties of cat neocortical pyramidal neurons in vivo. *Journal of neurophysiology*, 79(3):1450–1460.

Pearl, J. (2009). *Causality*. Cambridge university press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Peters, J., Janzing, D., and Schölkopf, B. (2017). *Elements of causal inference: foundations and learning algorithms*. MIT Press.

Pfau, D., Pnevmatikakis, E. A., and Paninski, L. (2013). Robust learning of low-dimensional dynamics from large neural ensembles. In *Advances in neural information processing systems*, pages 2391–2399.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995.

Pinault, D. (1996). A novel single-cell staining procedure performed in vivo under electrophysiological control: morpho-functional features of juxtacellularly labeled thalamic cells and other central neurons with biocytin or neurobiotin. *Journal of neuroscience methods*, 65(2):113–136.

Roudi, Y., Tyrcha, J., and Hertz, J. (2009). Ising model for neural data: model quality and approximate methods for extracting functional connectivity. *Physical Review E*, 79(5):051915.

Rudolph, M. and Destexhe, A. (2006). On the use of analytical expressions for the voltage distribution to analyze intracellular recordings. *Neural computation*, 18(12):2917–2922.

Sayer, R. J., Friedlander, M. J., and Redman, S. J. (1990). The time course and amplitude of EPSPs evoked at synapses between pairs of CA3/CA1 neurons in the hippocampal slice. *J. Neurosci.*, 10(3):826–836.

Seabold, S. and Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A. (2006). A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(Oct):2003–2030.

Stark, E. and Abeles, M. (2009). Unbiased estimation of precise temporal correlations between spike trains. *J. Neurosci. Methods*, 179:90–100.

Stevenson, I. H., Rebesco, J. M., Miller, L. E., and Körding, K. P. (2008). Inferring functional connections between neurons. *Current opinion in neurobiology*, 18(6):582–588.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1):1.

Thunemann, M., Ness, T. V., Kilic, K., Ferri, C. G., Sakadzic, S., Dale, A. M., Fainman, Y., Boas, D. A., Einevoll, G. T., and Devor, A. (2018). Does light propagate better along pyramidal apical dendrites in cerebral cortex? In *Optics and the Brain*, pages JW3A–56. Optical Society of America.

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*, 11(3):284–300.

Wang, H., Peca, J., Matsuzaki, M., Matsuzaki, K., Noguchi, J., Qiu, L., Wang, D., Zhang, F., Boyden, E., Deisseroth, K., et al. (2007). High-speed mapping of synaptic connectivity using photostimulation in channelrhodopsin-2 transgenic mice. *Proceedings of the National Academy of Sciences*, 104(19):8143–8148.

Wassermann, L. (2006). All of nonparametric statistics. *New York*.

Wright, P. G. (1928). *Tariff on animal and vegetable oils*. Macmillan Company, New York.

Wright, S. (1921). Correlation and causation. *Journal of agricultural research*, 20(7):557–585.

Wright, S. (1923). The theory of path coefficients a reply to niles's criticism. *Genetics*, 8(3):239.

Zemelman, B. V., Lee, G. A., Ng, M., and Miesenböck, G. (2002). Selective photostimulation of genetically charged neurons. *Neuron*, 33(1):15–22.