Correctly inferring causal connectivity using optogenetics

Mikkel Elle Lepperød, Konrad Paul Kording

March 25, 2018

Abstract

Neurons interact through spikes and a central objective of neuroscience is measuring the way how neurons causally affect one another. To probe such interactions, scientists often use optogenetics which typically leads to stimulation effects on multiple cells. This then produces a so-called confounding problem - we cannot know which of the stimulated neurons affected the activity of a given target neuron. Here we show how the resulting biases are large and how causal inference techniques, in particular instrumental variables, from econometrics can ameliorate this problem. Our approach utilizes the fact that neurons have absolute refractory times where the stimulation has no influence, and the missing stimulation acts like an effectively random variable which can then be used to disentangle causality. If we can not apply truly random stimulation to a network of neurons then causal inference techniques are needed to identify mechanisms.

1 Introduction

We want mechanism. Seriously. Neuroscience is to a large part experimentally driven.

We use observational studies. But causality is typically wrong. ¡nutshell version of the mehler paper¿

The experiments are extremely hard to control, even with the most accurate measurements of environmental variables and minimal genetic variance. Actually, given the vast amount of neurons and their connections in primitive mammals experiments are to be considered an observational sort. In a quest for experimental control we are now able to measure from thousands of neurons at the same time with micro-second precision by electrophysiology or with high spatial accuracy by two photon imaging. However, the large

amount of unmeasured activity makes estimates of functional connectivity confounded. By confounded we mean that correlations between neuron pairs are largely influenced by the background activity. One example of this confounding is when two neurons A, B are driven by a common input D, while one wants to estimate connectivity between A and C. However only B and C is connected such that B and C are strongly correlated, then since B and C is correlated then A and C is correlated and a regression $C = \beta A$ will give an erroneous $\beta > 0$ due to the the A, B correlation given by their common input D. 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. To estimate causal relationships between neurons, stimulation is appropriate, and if able to stimulate single neurons the ability to estimate causal relationships by regression is within grasp. However, this is experimental challenging and yield very low cell count. Therefore an optogenetic approach is more than often preferable.

Within the bounds of reachable tissue for photons emitted by a light source the decay of light intensity decays proportional to $1/r^2$. Furthermore if the distribution of neurons is homogeneous the probability of affecting a neuron increases in distance by r^2. This essentially makes the distribution of affected neurons flat. When performing simultaneously recording with optogenetic stimulation we again get a confounding problem when estimating the functional relationships between neuron pairs. In this case the confounding is when several neurons see the same stimulus and thus become highly correlated during stimulation yielding statistically indistinguishable functional relationships, this case is similar to the above example where D now is the optogenetic stimulation. Flat distribution of V.

The problem of endogenous regressors is common in microeconomics and in the social sciences, where observed outcomes can be driven by a large number of unknown factors. It is thus largely recognized that special care has to be taken when attempting to infer causal relationships between a putative actor and an outcome.

cite

Let's say that you want to estimate the return β from education x to yearly income y with the regression $y = \beta x + u$, where u is the factors other than schooling that contributes to yearly income. One of the factors in u is a persons ability, however this may also affect schooling and thus the regressor is correlated with the error term and a regression estimate will not estimate the magnitude of causation from schooling. In this case one may use the proximity to a college or university as an instrumental variable (IV) z. There are two criterion that must be fulfilled when choosing an instrumental variable, the instrument must be (1) uncorrelated with the error term, and

cite card 1995, found in cameron

(2) correlated with the regressor x. Then one may for instance estimate β by a two stage ordinary least squares (OLS) where the first stage is an OLS of the regressor and the instrument which is further used in the second stage with the dependent variable y. summary sentence.

cite TSOLS

worry is this really correct? Yes, its . IVs are (under the right assumptions) actually correct.

The IV technique can also be employed if one seeks to estimate the causal connectivity between neuron pairs. In this study we begin by showing how a simple network of three leaky integrate and fire (LIF) neurons can be heavily confounded under the right circumstances when simulating optogenetic targeting as direct current pulses. We then show that using the refractory period as an instrumental variable we are able to produce good estimates of causal connectivity. Furthermore we show that the method works in a large (N=1250) network of LIF neurons. With this data at hand we compare the IV method with a logistic regression.

2 Results

data showing synchrounous regular does not work (can be put together with regular stim. pattern)

I think I would choose a couple of examples that emulate what an experimentalist would do (a) periodic slow (bad power not enough data) (b) periodic fast (good power but confounded) (c) random fast (good everything) But I would then also redo all of them with a "weak network" model where therefore the assumptions are pretty well matched an also a "strong network" model where neurons have massive influence on one another (maybe oscillate or something) and hence the weakness assumption that is behind the IV justification breaks down

Consider three neurons (A,B,C) where A and B receives optogenetic stimulation and only B is connected to C Fig. 2(a). The stimulation induces a strong correlation between A and B Fig. 2(a) to (c), and thus the monosynaptic connection between B and C is thus spuriously inherited between A and C confounding the system by rendering the cross correlation histograms (CCHs) between BC and AC statistically indistinguishable as seen in Fig. 2(c) middle and lower right panel respectively.

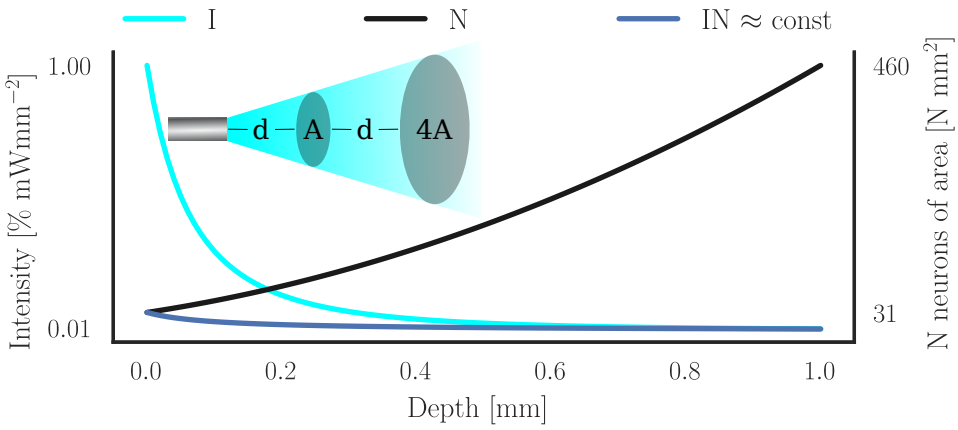


Figure 1: Spatial extent of optogenetic stimulus. Due to scattering and geometric loss the light intensity (I cyan) plotted as percentage of intensity exiting the optogenetic fiber follows approximately an inverse square law z^{-2} where z is the distance from the fiber; see methods. If neurons are uniformly distributed the number of optogenetically affected neurons increase by z^2 (N black) rendering the probability of activating a neuron approximately constant (IN blue).

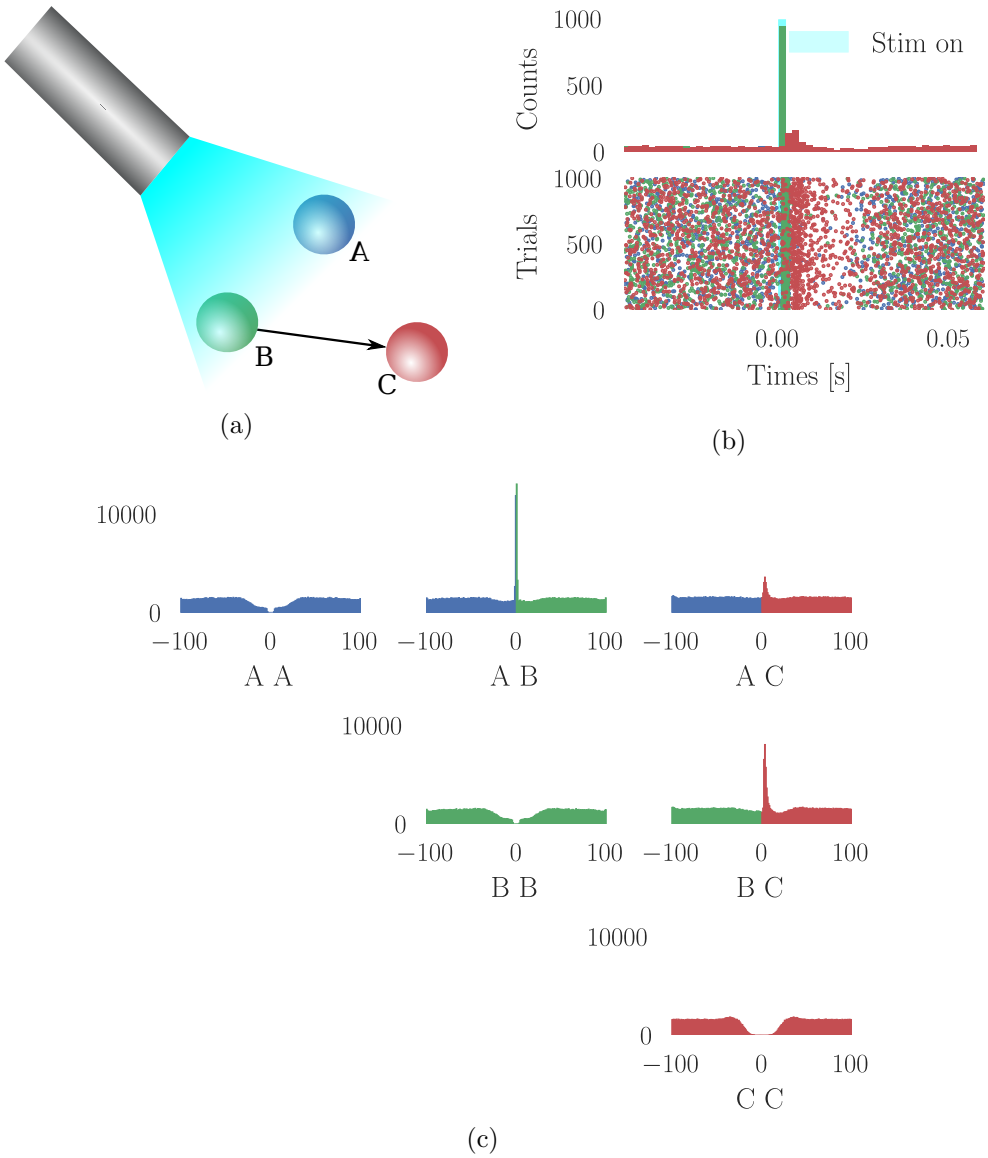
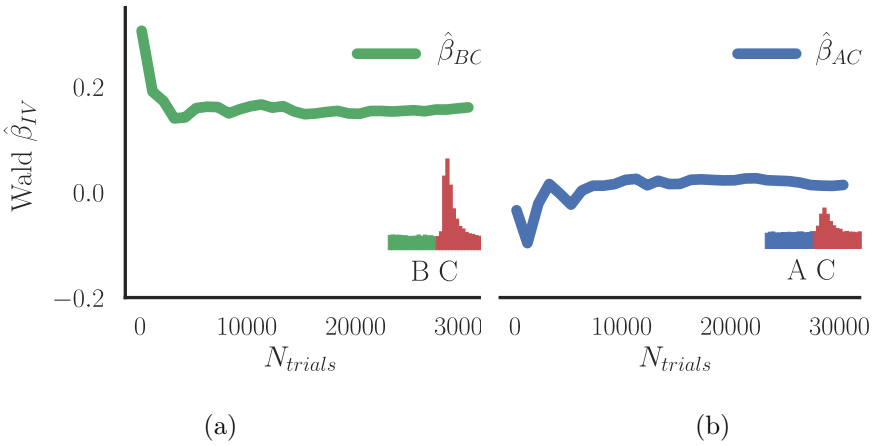


Figure 2: Sketch of a simple network containing three neurons depicted in red stimulated with blue laser light due to optogenetic activation. The neurons A and B are stimulated 30000 trials and the histogram of all trials are shown in (b) upper panel with the corresponding raster plot for each trial in the lower panel. Horizontal and vertical axes in correlation histograms represents time lag in ms and counts of coincident spikes in bins of 1 ms in (c)



(a) (b)

Figure 3: Wald estimator correctly estimates causal connectivity. The confounded network of three neurons (A,B,C) Fig. 2(a) is disentangled with the Wald estimator quickly converging to $BC \approx 0.2, AC \approx 0$. Insets represent high resolution zoom of cross correlation histograms of BC and AC respectively where horizontal and vertical axes in correlation histograms represents time lag in ms and counts of coincident spikes in bins of 0.1 ms for (a) and (b).

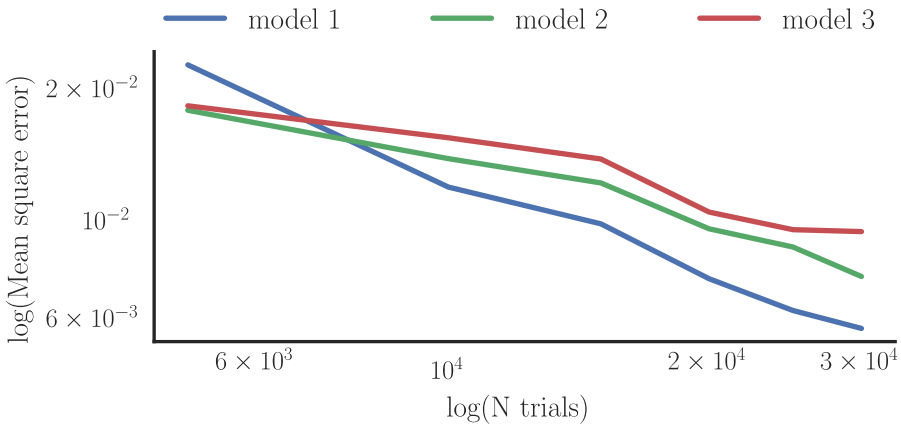


Figure 4: Mean square error (MSE) of Wald estimator in a recurrent neural network of three different parameter settings. The MSE as a function of number of trials shown on a logarithmic scale for asynchronous network activity when varying the relative inhibition yielding higher firing rate.

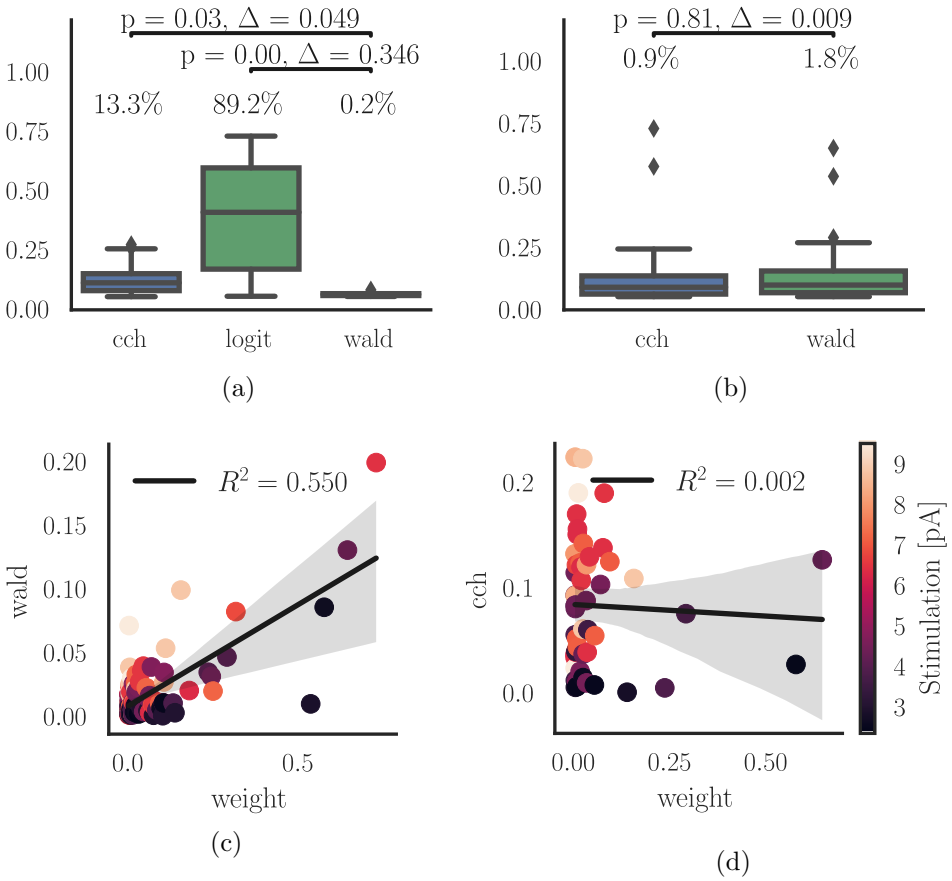


Figure 5: False estimates and goodness of fit. Panel 5(a) show false positives for the cross correlation histogram (cch) method, logistic regression (logit) and the Wald IV estimator (wald). False negatives for cch and wald is shown in 5(b). Positive estimates of weight as a function of true weight is scattered for the wald estimator in 5(c) and cch in 5(d) color coded by the size of perturbation intensity.

3 Discussion

Spike time methods works well when network is desynchronized, small amplitude stimulation with close range. Our method opens up for the possibility of strong perturbations giving larger yield. We are making the network results quite hard as all synapses have the same time constant such that multiple neurons are affecting the same cell at the same time for each stimulation

Discuss distribution of V. Discuss 2p stimulation. Conclude that there is absolutely no reason to hope things will work with 1p.

No way. But cite Buszaki.

easily usable in the context of many experiment. Potentials for future improvement.

IVs on multiple scales.

Broad opportunity, cite RDD, cite diff-in-diff etc, broad overview of the promise of econometrics for neuro, matching

Multiple illuminations, designed sticky opsins

Pairwise Wald vs GLM multiple neurons

Inhibitory connections may show up as negative Wald, but needs proper thinking

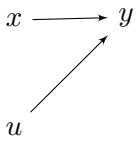
4 Methods

4.1 Instrumental variable estimation

A simple approximation of the connectivity strength between putative sender and receiver neurons x and y respectively can be to ignore long distance excitation and simply calculate the relation between the spike times in x and y with a regression model given by

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

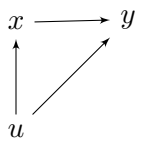
Here y is the dependent variable, x is the explanatory variable, β is the slope of the x, y curve and u is an unknown error term. This system follows the causal path:



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. 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 downstream neuron. Another problem is if the network state affects both the probability of a neuron to fire and also the probability of downstream 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. In general, if multiple neurons are stimulated synchronously our estimates will be off, potentially massively so as seen in the equation

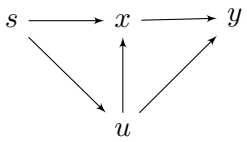
$$y = \beta x + u(x). \tag{2}$$

Following the causal path



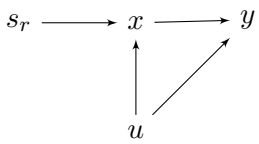
We now have the relation $\frac{\mathrm{d}y}{\mathrm{d}x} = \beta + \frac{\mathrm{d}u}{\mathrm{d}x}$. To get at causality we thus require some stimulation that only highlight the activity in y caused by x, however optogenetic stimulation is not specific enough as the photons will activate parts of the network activity u. To have a remedy, we need something that can distinguish between different neurons that are stimulated. We thus require some instrument s_r which is (1) uncorrelated with the network u and (2) is correlated with the regressor x. The graph with stimulation is illustrated by

cite the book or whatever



The previous graph represents a confounded stimulation, however we may use the fact that a neuron that has fired just before the stimulation will be in absolute refractory period and hence have $x_s = 0$. This introduces times

where the spike from one of the stimulated neurons is missing. Neurons have low firing probability once we use small bins and are very noisy at spiking at small time scales. Thus if we assume that (1) the stimulation pattern is random (2) the network activity u is asynchronous and irregular we may use the refractory period as an instrumental variable illustrated in the following graph



this will not work if you stimulate non random since then you may correlate u with sr, u is still driving x but asynchronously so sr is still uncorrelated with u.

Here s_r represents times where the sender neuron is refractory during stimulation. This is then an estimator that compares the downstream activity when a given neuron is non-refractory with the downstream activity when it is, thus removing the confounding. The true β is given by

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

Since our instrument s_r is binary we may use the Wald estimator to estimate β_{IV} by cite wald

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

Here $\bar{y}_s$ is the average number of trials where stimulating x resulted in a response in y and $\bar{y}_{s_r}$ is the average number of trials where an unsuccessful stimulation resulted in a response in y.

However the estimate of the refractory period s_r may be biased, as the network state may differ between times when a neuron is refractory and times when it is not. This issue we can also correct for by calculating β for baseline bins with no stimulation and no stimulation during absolute refractory period. This will linearly correct for the differences in network state that are due to refractory periods while maintaining the causal validity of Eq. (4).

Note that if we assume that the stimulation is so strong that the network effect u_{ns} has a negligible effect on x_s and that due to synaptic and axonal delay the effect of u_s is unable to affect x_s in it's respective stimulus response time. Then a conventional method like logistic regression will work.

4.2 Statistical testing on 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 [English et al., 2017] and the difference from the CCH convoluted with a hollow Gaussian [Stark and Abeles, 2009].

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

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 defined in [English et al., 2017] as

$$P_{trans} = \frac{1}{n} \sum_{m=4ms}^{8} CCH(m) - \lambda_{Gauss}(m), \tag{6}$$

where n is the number of spikes detected in the presynaptic neuron.

4.3 Simulated network

To simulate a recurrent network of 1000 excitatory and 250 inhibitory neurons we used the leaky integrate and fire (LIF) model given by

refer to parameter table

$$\tau \dot{V}_i(t) = -(V_i(t) - E_L) + RI_i(t). \tag{7}$$

When the membrane potential V_i of neuron i reaches a threshold θ an action potential is emitted and reset to the leak potential E_L followed by an absolute refractory period τ_{rp}. The membrane time constant is represented

by τ and $I_i(t)$ denotes the synaptic currents arriving at the soma modeled as a sum of alpha functions given by

$$RI_i(t) = \sum_j J_{ij} \sum_k t \exp(-(t - t_j^k)/\tau_{syn}). \tag{8}$$

Different synapses are denoted by $j = 1, ..., C + C_{ext}$ with post synaptic potential (PSP) amplitude (synaptic efficacy) denoted as J_{ij}. When neuron j emits it's k'th spike at time t_j^k it arrives at neuron i at $t = t_j^k + D$ where D denotes a transmission delay. Furthermore all neurons are driven by an external Poisson drive.

make sure delay is correctly inserted

Synaptic efficacies were log normally distributed with a minimum and maximum amplitude at $0.05mV$ and $2.05mV$ respectively within bounds of experimental findings [Sayer et al., 1990, **?**] as shown in Section 4.3.

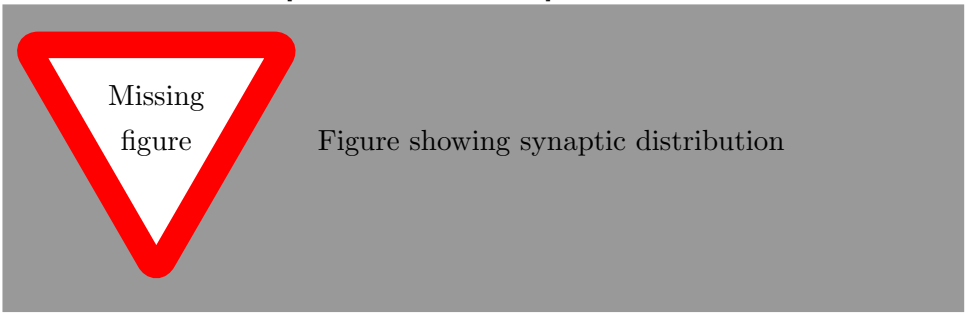
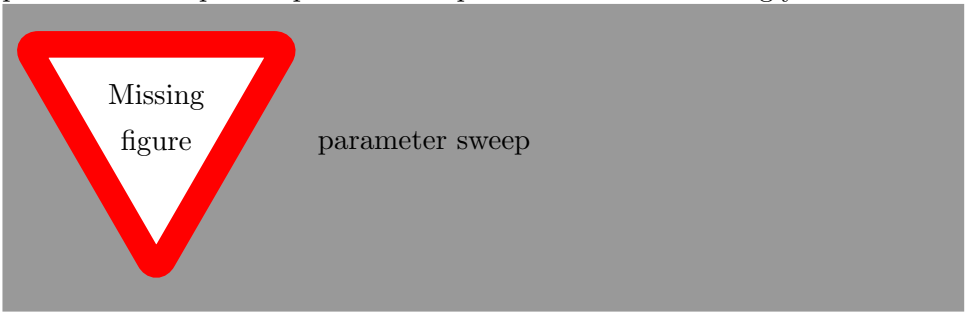
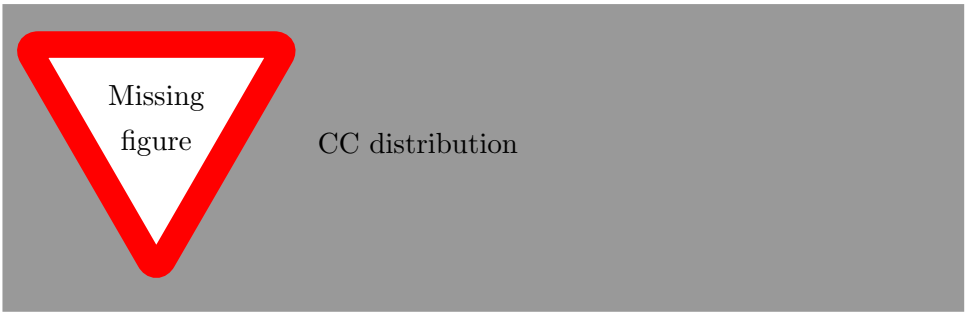


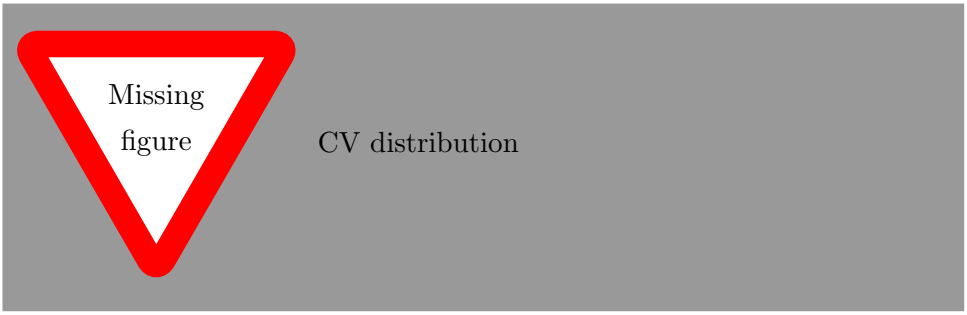
Figure showing synaptic distribution

To find suitable parameters yielding asynchronous activity as measured by the population correlation coefficient shown in Fig. 3 we performed several parameter sweeps and picked three parameter sets accordingly



parameter sweep





small perturbations on noisy networks will not yield a correlation between AC because the variance in stimulus response time in A will not be explained by the variance in spike times of C. Small variations of onset times in A relative to C will over many trials will wash out correlations between AC and at the same time strengthen correlations between BC. The only way to get a strong correlation between AC is to impose a short and strong stimulus effectively making AB spike coincidentally.

4.4 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, ideal diffusing media given by

$$T = \frac{1}{Sz + 1}. \tag{9}$$

Here T denotes a transmisison fraction, $S = 10.3mm^{-1}$ is the scattering coefficient for mice [Aravanis et al., 2007] and z is the distance from a light source [Ho et al., 2017]. Further we combined diffusion with geometric loss assuming that absorption is negligible as in [Aravanis et al., 2007] and com-

puted the intensity as presented in Fig. 1 by

$$\frac{I(z)}{I(z=0)} = \frac{\rho^2}{(Sz+1)(z+\rho)^2} \tag{10}$$

where

$$\rho = r\sqrt{\left(\frac{n}{NA}\right)^2 - 1}, \tag{11}$$

and $r = 100\mu m$ is the radius of the optical fiber, $NA = 0.37$ is the numerical aperture of the optical fiber and $n = 1.36$ is the refraction index for gray matter [Ho et al., 2017].

To estimate the distribution of light intensity on affected neurons we assumed a neuron density of $10^5 N mm^{-3}$ and foo und the volume of a cut cone that could contain 950 "source" excitatory neurons which were found to yield the depth $z = 0.175mm$. The last 50 of the excitatory neurons was used as the "target" population which together with the inhibitory neurons were not perturbed directly by the light stimulus.

In order to keep the stimulus model as simple as possible we used a perturbation strength of $10pA$ which was found suitable by investigating the percentage of successful stimulations to be around 50%.

References

[Aravanis et al., 2007] 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.

[English et al., 2017] 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.

[Ho et al., 2017] Ho, A. H. P., Kim, D., and Somekh, M. G. (2017). *Handbook of photonics for biomedical engineering.*

[Sayer et al., 1990] 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.

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

Todo list