Uncovering Network Architecture Using an Exact Statistical Input-Output Relation of a Neuron Model

Safura Rashid Shomali[1,*], Majid Nili Ahmadabadi[2], Seyyed Nader Rasuli[3,4], Hideaki Shimazaki[5]

1 School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran 19395-5746, Iran

2 Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran 14395-515, Iran

3 Department of Physics, University of Guilan, Rasht 41335-1914, Iran

4 School of Physics, Institute for Research in Fundamental Sciences (IPM), Tehran 19395-5531, Iran

5 Graduate School of Informatics, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan

* safura@ipm.ir

Summary

An appealing challenge in Neuroscience is to identify network architecture from neural activity. A key requirement is the knowledge of statistical input-output relation of single neurons in vivo. Using a recent exact solution of spike-timing for leaky integrate-and-fire neurons under noisy inputs balanced near threshold, we construct a unified framework that links synaptic inputs, spiking nonlinearity, and network architecture, with statistics of population activity. The framework predicts structured higher-order interactions of neurons receiving common inputs under different architectures: It unveils two network motifs behind sparse activity reported in visual neurons. Comparing model's prediction with monkey's V1 neurons, we found excitatory inputs to pairs explain the sparse activity characterized by negative triple-wise interactions, ruling out shared inhibition. While the model predicts variation in the structured activity according to local circuitries, we show strong negative interactions are in general a signature of excitatory inputs to neuron pairs, whenever background activity is sparse.

Keywords

network architecture, sparse activity, common inputs, higher-order interactions, leaky integrate-and-fire neuron model, threshold regime, input-output relation, triple-wise interactions, spontaneous activity

Introduction

One interest in neuroscience is to *reveal neural circuitries* using recorded neurons' activities that are constrained by synaptic connections, and the *nonlinear spiking dynamics* of individual neurons. There exist many theoretical studies, all trying to make a connection between *network structure* and *neural activity's statistics* (Hu et al., 2013, 2014; Ocker et al., 2017a; Ostojic et al., 2009; Pernice et al., 2011; Rosenbaum et al., 2017; Trousdale et al., 2012). However, nearly most of them consider a linear response around stationary distribution and do not take account of the nonlinearity in single neuron's input-output relation (but see Ocker et al. 2017b). Such attempts are justified as long as synaptic strengths are weak enough so one single input is quite unlikely to cause the spike of the postsynaptic neuron; however, such an assumption of weak synaptic input is in contrast with the task of revealing network architecture, where we cannot neglect strong synaptic connections.

Here we look at the simplest possible scenario, where the experimentalist records spiking activity of three neurons in vivo, while s/he cannot directly reveal neural connectivity. Will the three neurons spike independently, or show correlations due to possible shared inputs? In the latter case, are such inputs shared between each pair of them, or among three of them all? Are shared inputs excitatory or inhibitory? And finally, does either of the three observed neurons make a direct synaptic connection to another of them? It is interesting that a closed-form analytic relation that addresses the nonlinearity of input-output in a spiking neuron model, lets us answer almost all of these questions.

Recently, we theoretically investigated the statistical dependency of postsynaptic spikes on synaptic inputs for an LIF neuron model, which includes synaptic and membrane dynamics (Shomali et al., 2018). Despite the fact that integration and thresholding mechanism of neuronal firing tend to regularize inter-spike intervals, neuronal spikes in vivo are highly variable (Shadlen and Newsome, 1998; Softky and Koch, 1993); it is suggested to be the result of balanced inputs from excitatory and inhibitory neurons near spiking threshold (Shadlen and Newsome, 1998; Vreeswijk and Sompolinsky, 1996, 1998). We derived the dependency of output spikes on arbitrary synaptic input of interest called a signaling input, when the noisy background inputs are balanced near the spiking threshold (Shomali et al., 2018), a condition verified for in vivo neurons during stimulus presentation (Tan et al., 2014).

By using the analytical model, we examine correlations in simultaneous spiking activity of two and three neurons in a small time window. For this aim, we use the information geometric measures of neural interactions (Nakahara and Amari, 2002), known as pairwise and triple-wise interactions, that explain excess or paucity of synchronous spikes expected from their independent activity. The interaction among three neurons was experimentally examined by Ohiorhenuan et al. in V1 of monkey, in vivo. They found significant negative triple-wise interactions if neurons are spatially close (Ohiorhenuan et al., 2010; Ohiorhenuan and Victor, 2011); it indicates a sig-

nificantly higher probability for simultaneous silence of the three neurons. The same negative triple-wise interactions were reported for nearby neurons in hippocampal slices (Shimazaki et al., 2015).

Our obtained analytical solution allows us to probe the effect of common inputs on the population activity, whether the common input is weak or strong, excitatory or inhibitory. Comparing with the aforementioned experimental observation of negative triple-wise interactions (Ohiorhenuan et al., 2010), we find a non-intuitive architecture of excitatory common input to pairs can explain the observed negative triple-wise interactions.

Materials and Methods

Leaky integrate-and-fire neuron at the threshold regime

Here, we briefly introduce the exact statistical solution of the postsynaptic spike density (Shomali et al., 2018). The neuron receives a synaptic input (a signaling input) on top of background weak synaptic inputs that bring the membrane potential close to the threshold. The background activity is modeled as the Gaussian white noise which may replicate the synaptic inputs to V1 neurons when a visual stimulus is presented (Tan et al., 2014). Below, we bring the stochastic differential equation for membrane potential of the leaky integrate-and-fire (LIF) neuron model, and the solution for the Inter-Spike-Interval (ISI) density when neuron receives the signaling input on top of the Gaussian noise.

We consider an LIF neuron with membrane time constant of τ_{m} and threshold voltage of V_θ, receiving transient signaling input, ΔI, at time τ_b, and independent Gaussian noise, I_0 (Fig 1A):

$$\tau_{\mathrm{m}}\frac{dV}{dt} = -(V - V_r) + I(t), \tag{1}$$

the input current, $I(t)$ is:

$$I(t) = I_0(t) + \Delta I(t, \tau_b), \tag{2}$$

where $I_0(t)$ is the fluctuating background input induced by presynaptic neurons and $\Delta I(t, \tau_b)$ is the transient signaling input. The background input has a mean, $\bar{I}$, and a variance, $2D/\tau_{\mathrm{m}}$, where D is the diffusion coefficient.

The ISI density as a function of signaling input's amplitude, A, and spike timing, τ_b, for arbitrary shape of transient signaling input, ΔI, when $t \geq \tau_b$ at the threshold regime, reads (Shomali et al., 2018):

$$\begin{aligned} J(t) = \frac{\sqrt{\kappa\omega}}{\pi\tau_{\mathrm{m}}}\{ &\exp(-\frac{\varphi_+^2}{2})[1 + \sqrt{\frac{\pi}{2\kappa}}\varphi_+ \exp(\frac{\varphi_+^2}{2\kappa}) \times (1 + Erf(\frac{\varphi_+}{\sqrt{2\kappa}}))] - \\ &\exp(-\frac{\varphi_-^2}{2})[1 + \sqrt{\frac{\pi}{2\kappa}}\varphi_- \exp(\frac{\varphi_-^2}{2\kappa}) \times (1 + Erf(\frac{\varphi_-}{\sqrt{2\kappa}}))]\}, \end{aligned} \tag{3}$$

where $Erf(x) = (2/\sqrt{\pi})\int_0^x \exp(-t^2)dt$, $r(t) = \exp(-t/\tau_{\mathrm{m}})$ and $\kappa(t, \tau_b)$, $\omega(t, \tau_b)$ and $\varphi_\pm(t, \tau_b)$

are:

$$\begin{aligned}
\kappa(t,\tau_b) &= (1-r(t)^2)/(1-r(t-\tau_b)^2), \\
\omega(t,\tau_b) &= \frac{r(t-\tau_b)^2(1-r(\tau_b)^2)}{(1-r(t)^2)^3}, \\
\varphi_\pm(t,\tau_b) &= \sqrt{\frac{\tau_{\mathrm{m}}V_\theta^2}{D(1-r(\tau_b)^2)}}\{\pm r(\tau_b) - \int_{\tau_b}^{t}\frac{ds}{\tau_{\mathrm{m}}}\frac{\Delta I(s,\tau_b)}{V_\theta}\}.
\end{aligned} \tag{4}$$

In the absence of signaling input ($t < \tau_b$), the ISI density reduces to the known formula (Tuckwell, 1988; Wang and Uhlenbeck, 1945):

$$J_0(t) = \frac{1}{\tau_{\mathrm{m}}}\sqrt{\frac{2}{\pi}\frac{\tau_{\mathrm{m}}V_\theta^2}{D}\frac{r^2(t)}{(1-r^2(t))^3}}\exp\left[-\frac{\tau_{\mathrm{m}}{V_\theta}^2}{2D}\frac{r^2(t)}{1-r^2(t)}\right]. \tag{5}$$

In the following analysis, we use Eq 3 with a square shape of signaling input given by

$$\Delta I(t,\tau_b) = A \times \begin{cases} 0 & t < \tau_b, \\ 1/\Delta t & \tau_b \le t \le \tau_b + \Delta t, \\ 0 & \tau_b + \Delta t < t. \end{cases} \tag{6}$$

Spiking density of LIF neuron after signaling input arrival

In order to further analyze the effect of the signaling input, it is convenient to consider postsynaptic spike timing relative to the arrival time of the signaling input (Fig 1A, Left). We reset the time origin to the timing of signaling input arrival, and find the density $f(\tau)$ that postsynaptic neuron spikes at τ after the signaling input arrival. Suppose that the last spike of the postsynaptic neuron occurred at τ_b before the arrival of signaling input (Fig 1A, Right). Then the conditional density of postsynaptic spike time, $f(\tau|\tau_b)$, is calculated as (Shomali et al., 2018):

$$f(\tau|\tau_b) = \frac{J(\tau+\tau_b)}{1-\int_0^{\tau_b} J(s)ds}, \tag{7}$$

where the denominator is a normalization term to satisfy $\int_0^\infty f(\tau|\tau_b)d\tau = 1$.

Next, we compute the probability density of the postsynaptic spike timing averaged over various timings of τ_b. Given the fixed time origin at the signaling input arrival, the probability density of the duration τ_b is known to follow the probability of backward recurrence time from the renewal point process theory (Cox, 1962). This density is $p(\tau_b) = \mu[1 - \int_0^{\tau_b} J_0(V_\theta, s)\, ds]$, where μ is the mean firing rate of postsynaptic neuron when there is no signaling input, $\mu = (\int_0^\infty sJ_0(V_\theta, s)\, ds)^{-1}$. By marginalizing Eq 7 with respect to τ_b and using the backward recurrence time, $p(\tau_b)$, we obtain (Shomali et al., 2018):

$$\begin{aligned}
f(\tau) &= \int_0^\infty f(\tau|\tau_b)p(\tau_b)\, d\tau_b \\
&= \mu\int_0^\infty J(\tau+\tau_b)\, d\tau_b.
\end{aligned} \tag{8}$$

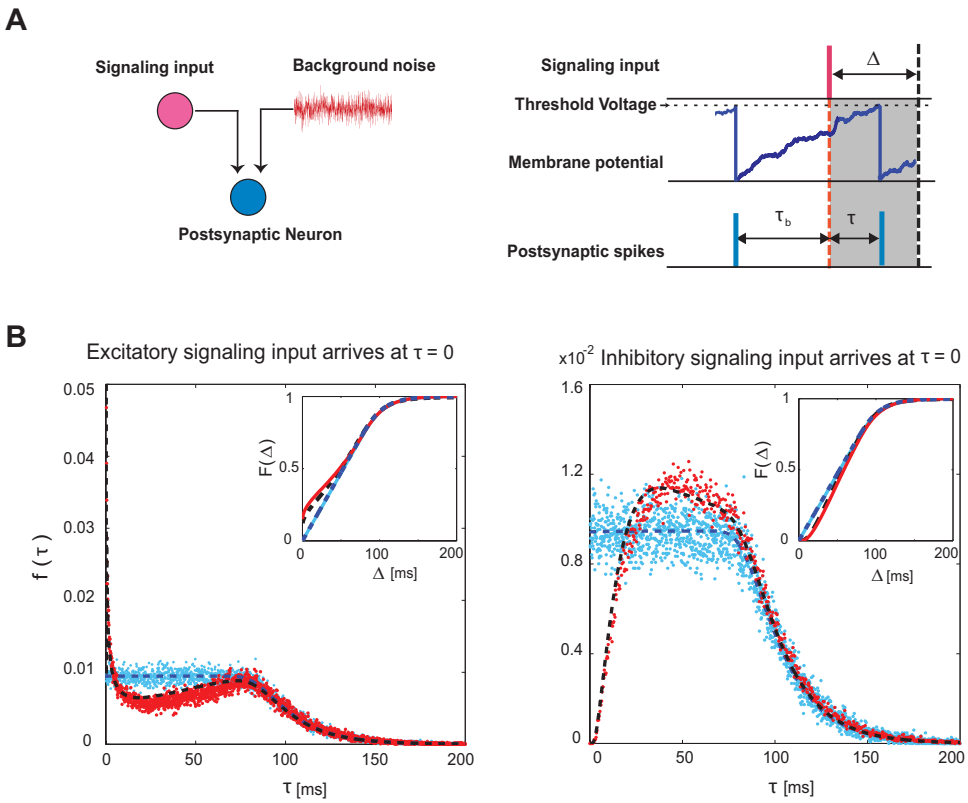

Fig 1. Analysis of a postsynaptic neuron receiving signaling input on top of the background noise. (**A**) Left: A schematic model of postsynaptic neuron driven by background Gaussian noise and a transient signaling input. Right: Timing of postsynaptic spikes and the arrival time of signaling input. Postsynaptic neuron generates spike (blue tick) and resets, signaling input (red tick) arrives at τ_b after last spike of postsynaptic neuron; it changes the voltage trajectory of membrane potential upon arrival. The postsynaptic spike occurs at τ after arrival time of the signaling input. The gray shaded area of width Δ indicates the observation time window during which the probability of postsynaptic activity pattern is computed. (**B**) The probability density of the first spike at time τ after arrival (Eq 8) of excitatory (left) and inhibitory (right) signaling input. The simulation results (red dots) and the analytical solutions (dashed black line, Eq 8) are well matched. The same results for signaling input with zero amplitude are depicted; the analytic (dashed blue line) and simulation results (blue dots) well coincide. Inset: The probability of spiking within Δ after signaling input arrival, Eq 9. The parameters are: $V_\theta = 20\,\mathrm{mV}$, $\tau_m = 20\,\mathrm{ms}$, $A = 5\,\mathrm{mVms}$, and the diffusion coefficient is $D = 0.74\,(\mathrm{mV})^2\mathrm{ms}$. The values of the parameters are chosen using the physiologically plausible range (McCormick et al., 1985).

Note that when the amplitude of the signaling input, A, is diminished to zero, $J(V_\theta, t) = J_0(V_\theta, t)$ and $f(\tau)$ simplifies to $f_0(\tau) = \int_\tau^\infty \mu J_0(V_\theta, s)\, ds$.

Now we can address what is the probability of having one or more spikes in a specific time window, Δ. It is given by a cumulative density function of $f(\tau)$. It would read:

$$F_A(\Delta) = \int_0^{\Delta} f(\tau) d\tau, \tag{9}$$

where the subscript A indicates that $F_A(\Delta)$ is the function of amplitude of the signaling input. In the absence of signaling input, we have $F_0(\Delta) = \int_0^{\Delta} f_0(\tau) d\tau$. Based on the probability of spiking within specific window Δ, $F_A(\Delta)$, one can find the probability of patterns considering the assumption that neurons are conditionally independent. This calculation is described in the next section.

Figure 1B shows the marginalized spiking density at time τ after signaling input arrival for squared shape input (Eq 6, dashed black lines). Early spiking (small τ) after the arrival of excitatory (inhibitory) input is highly more (less) probable, compared to no-signaling input case (dashed blue lines). However, the spiking densities with or without the signaling input are virtually identical at sufficiently large τ; indicating the short-lasting effect of the signaling input. Accordingly, the cumulative distribution functions, $F_A(\Delta)$, (Fig 1 insets) with or without the signaling input differ for small Δ, but are indistinguishable for large Δ. This result implies that we cannot discern the presence of signaling input if we use large time window.

Results

Neural interactions as a function of signaling input and network architecture

Here, the motivation to investigate the interactions among three neurons comes from some experimental studies (Ohiorhenuan et al., 2010; Ohiorhenuan and Victor, 2011) which recorded the activity of three neurons simultaneously (Fig 2). For two neurons, there is one possible architecture of shared input: A common input to both of them. For three neurons, however, it could be either (i) a shared input among three of them (red connections in Fig 2A), or (ii) one or more shared inputs to each pair among them (green connections). Assuming symmetry, the former one leads to *star architecture* or common input to trio (Fig 2B, left) while the latter one makes the *triangle architecture* or common inputs to pairs (Fig 2B, right). In the following sections, we firstly investigate pairwise interaction among two neurons, when they receive a common signaling input on top of the noisy inputs; then we extend it to triple-wise interaction of neurons. To formally investigate the neuronal correlation including higher order interactions, we use the information geometric measure of correlations (Amari, 2009b; Martignon et al., 2000; Nakahara and Amari, 2002). The higher-order interaction measured by this framework indicates the neuronal interactions that cannot be expected by the lower order statistics.

I. Pairwise interaction in the presence of a common signaling input

Firstly, we analyze the interaction of two postsynaptic neurons receiving a common signaling input in addition to independent background noise (Fig 3A, Top). The middle panel of Fig 3A illustrates the timing of postsynaptic spikes before and after the common input arrival. By binning their spike sequences with a Δ time window (Fig 3A, Middle, and Left), we compute

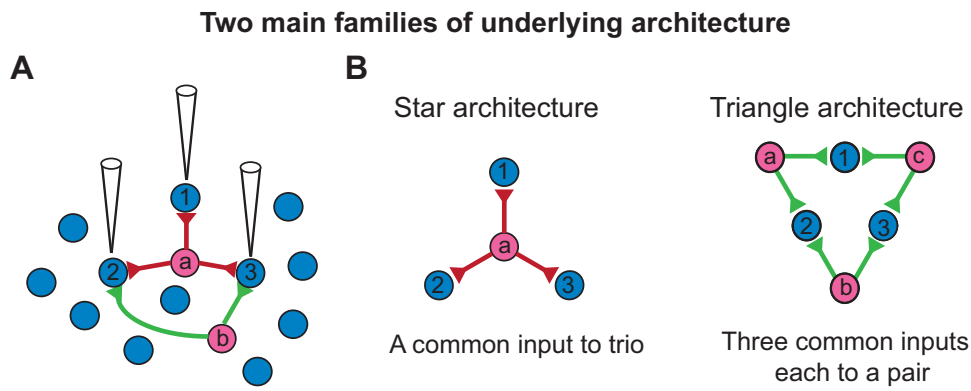

Fig 2. The activity of three postsynaptic neurons are recorded simultaneously. (**A**) The neurons (blue circles) would act independently in the absence of any *common input*. A common input (pink circle) could be an input to three of them (red connections), or at least two of them (green connections). Assuming symmetry, we would have two main families of architectures: one common to trio (star architecture, left), or three commons each to a pair (triangle architecture, right).

probabilities of activity patterns of postsynaptic neurons. We assume that neurons sparsely receive a random common input with firing rate λ in a Poissonian fashion. We then, segment the spike sequences using bins aligned at the onset of common input. Let $x_i = \{0, 1\}$ $(i = 1, 2)$ be a binary variable, where $x_i = 1$ represents that the ith neuron emitted one or more spikes in the bin, and $x_i = 0$ represents that the neuron was silent.

We denote by $P(x_1, x_2)$ the probability mass function of the binary activity patterns of the two postsynaptic neurons. Here $P(1, 1)$ and $P(0, 0)$ are the probabilities that both neurons are, respectively, active and silent within Δ. Similarly, $P(1, 0)$ is the probability that neuron 1 emits one or more spikes while neuron 2 is silent during Δ; $P(0, 1)$ represents the opposite situation. The probability mass function is represented in the form of an exponential family distribution as

$$P(x_1, x_2) = \exp\left(\theta_1^{\mathrm{p}} x_1 + \theta_2^{\mathrm{p}} x_2 + \theta_{12}^{\mathrm{p}} x_1 x_2 - \psi\right), \tag{10}$$

where $(\theta_1^{\mathrm{p}}, \theta_2^{\mathrm{p}}, \theta_{12}^{\mathrm{p}})$ are the canonical parameters, and ψ is a log-normalization parameter. In particular, θ_{12}^{p} is an information geometric measure of pairwise interaction (Amari, 2001, 2009b; Nakahara and Amari, 2002). Accordingly, pairwise interaction parameter is computed as:

$$\theta_{12}^{\mathrm{p}} = \log \frac{P(1, 1)P(0, 0)}{P(1, 0)P(0, 1)}. \tag{11}$$

If the binary activity of two neurons are independent, then $\theta_{12}^{\mathrm{p}} = 0$.

For two neurons, we construct the probabilities of the activity patterns by using the probability of spike occurrence in a bin Δ (Eq 9). Since two neurons receive common inputs, they

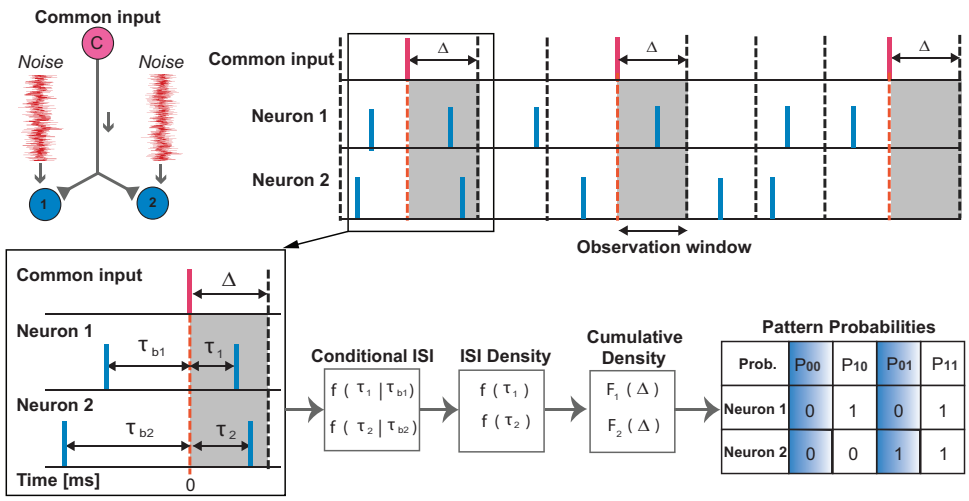

Fig 3. Analysis of two neurons receiving common signaling input on top of background noise. Top, left: A schematic model of two postsynaptic neurons (Neuron 1 and Neuron 2, blue circles) driven by independent noise and a common signaling input (pink circle). Top, right: Spike trains of two postsynaptic neurons (blue spikes) receiving common signaling inputs (red spikes) with rate λ. The gray shaded area of width Δ indicates the observation time window during which we compute the probabilities of postsynaptic activity patterns. Bottom, left: Timing of postsynaptic spikes relative to the arrival time of the common input. The last spike of Neuron 1 (Neuron 2) has occurred τ_{b1} (τ_{b2}) before the arrival of common input, and their next spikes happen at τ_1 (τ_2) after that. Bottom, middle: The conditional ISI density after input arrival is calculated by Eq 7. Marginalizing over previous spike (τ_b), the probability of spiking after input arrival (ISI density) is achieved (Eq 8). The next step is calculating the cumulative density function (Eq 9) that is the probability of having one or more spikes within window Δ. Bottom, right: Based on $F_A(\Delta)$ and the fact that neurons are independent, the probability of having a particular pattern of spikes for two neurons is obtained. Four possible binary activity patterns $(00, 01, 10, 11)$ of two postsynaptic neurons and their associated probabilities (P_{ij}, $i, j \in \{0, 1\}$). The '1' denotes occurrence of at least one spike within the Δ time window, whereas the '0' represents the silence of neuron within this window.

are conditionally independent. Thus using Eq 9:

$$P_A(x_1, x_2) = P_A(x_1)P_A(x_2) = \prod_{i=1}^{2} F_A(\Delta)^{x_i}(1 - F_A(\Delta))^{1-x_i}, \tag{12}$$

where A stands for the amplitude of common input current ($A = 0$ represents the absence of common input). Since the rate of common input is λ, it is applied at $\lambda\Delta \times 100\%$ of the bins whereas it is absent in $(1-\lambda\Delta)\times 100\%$ of the bins. Accordingly, we consider the spike sequences of two postsynaptic neurons as a mixture of the two conditions, with or without the common

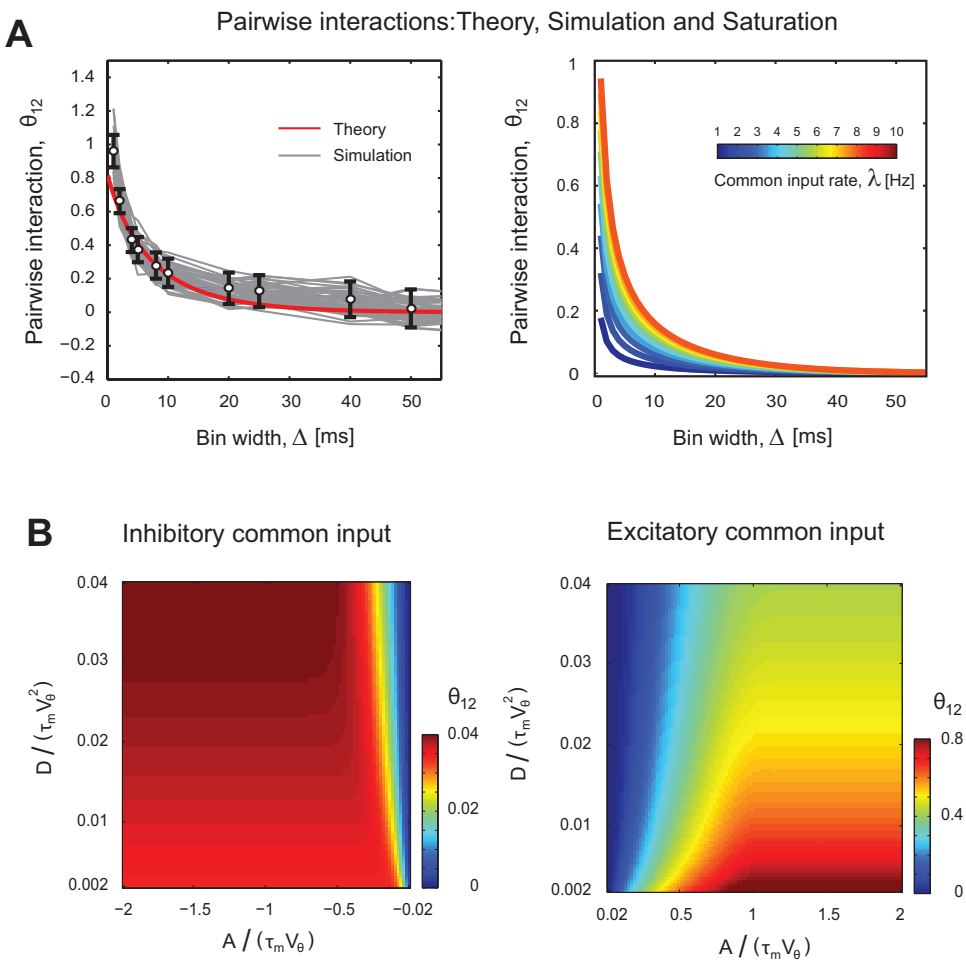

Fig 4. Analysis of the interaction between two neurons receiving common signaling input on top of background noise. (**A**) Interaction of two postsynaptic neurons (θ_{12}^{p}, Eq 11) as a function of bin size, Δ. The parameters are: $\tau_m = 20\,\mathrm{ms}$, $V_\theta = 20\,\mathrm{mV}$, $A = 5\,\mathrm{mV\ ms}$ and $D = 0.74(\mathrm{mV})^2\mathrm{ms}$. Left: The pairwise interaction computed from simulated spike sequences (gray lines: 50 individual trials each containing about 2500 spike occurrence of common input; dots and error bars: mean$\pm$standard deviation) and the analytic result of the mixture model (red line, Eq 11) for common input rate $\lambda = 5\,\mathrm{Hz}$. Right: Analytical value of θ_{12}^{p} as a function of Δ for different common input rates, λ. (**B**) The pairwise interaction as a function of two scaled parameters, $D/(\tau_m V_\theta^2)$ and $A/(\tau_m V_\theta)$ for excitatory (right) and inhibitory (left) common inputs with $\lambda = 5\,\mathrm{Hz}$ and $\Delta = 5\,\mathrm{ms}$.

input. Hence the probabilities of the activity patterns are given by

$$P_{\mathrm{mix}}(x_1, x_2) = (\lambda\Delta)P_A(x_1, x_2) + (1 - \lambda\Delta)P_0(x_1, x_2). \tag{13}$$

The pairwise interaction predicted by the mixture model is compared with simulated spike

sequences (Eq 11, 13). Figure 4A displays the interaction for different bin sizes, Δ, when the two neurons receive common excitatory input. The pairwise interaction predicted by the mixture model agrees with the simulation results (Left panel, red and gray lines respectively). θ^{p}_{12} has an increasing trend with the rate of common input (Fig 4A, Right). However, if we increase the bin size, the probability of having one or more spikes within Δ increases and saturates to 1 (Fig 3B, Inset) regardless of presence or absence of the signaling input. It means $F_A(\Delta)/F_0(\Delta) \to 1$, which results in vanishing pairwise interaction as the bin size increases. In sum, we can hardly discriminate between the presence and absence of common input if we use a large bin size.

We examine pairwise interactions by changing two independent parameters, scaled amplitude of the signaling input $A/(\tau_m V_\theta)$ and scaled variability of the noisy background input $D/(\tau_m V_\theta^2)$ (Fig 4B). As expected, the pairwise interactions are positive for both common excitatory and inhibitory inputs. However, the interactions are significantly weaker in the inhibitory case. It implies that it is difficult to observe the effect of inhibitory common input for this range of postsynaptic firing rate and that the strong pairwise interactions are the indicator of having common excitatory input.

For each value of normalized diffusion coefficient, $D/(\tau_m V_\theta^2)$, there exists a critical normalized amplitude for common excitatory input, $A/(\tau_m V_\theta) \sim 1$. From that critical value on, postsynaptic neuron's spiking density, and consequently pairwise interaction, does not change anymore (Fig 4B, Right). The saturation value of pairwise interaction is inversely correlated with normalized diffusion coefficient; since higher normalized diffusion coefficient disperses the voltage trajectories of membrane potential, the probability of spiking after common input arrival decreases. Similar behavior for θ^{p}_{12} is observed for inhibitory input (Fig 4B, Left). Nevertheless, in contrary to the common excitatory input case, pairwise interaction and $D/(\tau_m V_\theta^2)$ are directly correlated.

II. Higher-order interaction among three neurons receiving common signaling inputs

We now extend the analysis of neural interactions to three neurons. In the exponential form, the probability mass function of binary activity patterns for three neurons is written as

$$P(x_1, x_2, x_3) = \exp\left(\sum_{i=1}^{3} \theta^{\mathrm{t}}_i x_1 + \sum_{i<j} \theta^{\mathrm{t}}_{ij} x_i x_j + \theta^{\mathrm{t}}_{123} x_1 x_2 x_3 - \psi\right). \tag{14}$$

If $\theta^{\mathrm{t}}_{123}$ (the triple-wise interaction parameter) is 0, the distribution reduces to the pairwise maximum entropy model, i.e., the least structured model that maximizes the entropy given that the event rates of individual neurons and joint event rates of two neurons are specified (Cover and Thomas, 1991). That is, a positive (negative) triple-wise interaction indicates that the three neurons generate synchronous events more often (less) than the chance coincidence expected from event rates of individual neurons and their pairwise correlations. From this equation, the triple-wise interaction among three neurons for exponential family of probability mass function, is calculated as (Amari, 2009a; Nakahara and Amari, 2002):

$$\theta^{\mathrm{t}}_{123} = \log \frac{P(1,1,1)P(1,0,0)P(0,1,0)P(0,0,1)}{P(0,0,0)P(0,1,1)P(1,0,1)P(1,1,0)}. \tag{15}$$

In the first architecture (star architecture), three neurons receive a single common signaling input at the same time (Fig 2B, Left). The conditional probability of activity patterns given a common input to all three neurons is $P_A(\mathbf{x}) = \prod_{i=1}^{3} F_A(\Delta)^{x_i}(1 - F_A(\Delta))^{1-x_i}$. Similarly, the probability mass function of three neurons without any common input is obtained as $P_0(\mathbf{x}) = \prod_{i=1}^{3} F_0(\Delta)^{x_i}(1 - F_0(\Delta))^{1-x_i}$. Thus, we model spike occurrence by a mixture of the two conditions, with or without common input, as:

$$P(\mathbf{x}) = \lambda\Delta P_A(\mathbf{x}) + (1 - \lambda\Delta)P_0(\mathbf{x}). \tag{16}$$

In the second architecture (triangular architecture), each pair of three neurons receives common signaling input from an independent presynaptic neuron with frequency λ (Fig 2B, Right). The first, second, and third common inputs project to neuron 1 and 2, neuron 2 and 3, and neuron 1 and 3, respectively. The three common inputs are independent from each other, and happen with an equal frequency, λ. When the first common input is active but the two other common inputs are silent, the pattern probabilities of three postsynaptic neurons are given as $P_A^1(\mathbf{x}) = [\prod_{i=1}^{2} F_A(\Delta)^{x_i}(1 - F_A(\Delta))^{1-x_i}] \times [F_0(\Delta)^{x_3}(1 - F_0(\Delta))^{1-x_3}]$. This situation happens in $(\lambda\Delta)(1 - \lambda\Delta)^2 \times 100\%$ of the bins. The probabilities of activity patterns at which neurons receive from the second (third) common input, $P_A^2(\mathbf{x})$ ($P_A^3(\mathbf{x})$), are given similarly to this equation. Here we note that the common inputs may be simultaneously applied at the same bin because of their independence. Namely, two common inputs coincide at $(\lambda\Delta)^2(1 - \lambda\Delta) \times 100\%$ of the bins. The pattern probability in the bins at which common inputs 1 and 2 coincide is given by $P_A^{12}(\mathbf{x}) = [\prod_{i=1,3} F_A(\Delta)^{x_i}(1 - F_A(\Delta))^{1-x_i}] \times [F_{2A}(\Delta)^{x_2}(1 - F_{2A}(\Delta))^{1-x_2}]$. Similarly, we define $P_A^{23}(\mathbf{x})$ and $P_A^{13}(\mathbf{x})$ for the bins at which the common inputs 2 and 3, and the common inputs 1 and 3 coincide, respectively. Finally, all common inputs coincide at $(\lambda\Delta)^3 \times 100\%$ of the bins, at which the pattern probability is given by $P_A^{123}(\mathbf{x}) = \prod_{i=1,2,3} F_{2A}(\Delta)^{x_i}(1 - F_{2A}(\Delta))^{1-x_i}$. Thus, the parallel spike sequences are modeled as a mixture of these probability mass functions,

$$\begin{aligned} P(\mathbf{x}) = (1 - \lambda\Delta)^3 P_0(\mathbf{x}) + \sum_{i=1}^{3} (\lambda\Delta)(1 - \lambda\Delta)^2 P_A^i(\mathbf{x}) + \\ \sum_{i<j} (\lambda\Delta)^2(1 - \lambda\Delta) P_A^{ij}(\mathbf{x}) + (\lambda\Delta)^3 P_{2A}^{123}(\mathbf{x}). \end{aligned} \tag{17}$$

The triple-wise interaction parameters computed from simulated spike sequences of postsynaptic neurons are compared with the theoretical predictions, using the mixture models (Fig 5A and B, Left). Activities of neurons that receive a simultaneous common excitatory input (star architecture) are characterized by positive triple-wise interactions (Fig 5A, Left) whereas activities of neurons that receive independent common excitatory inputs to pairs (triangular architecture) are characterized by negative triple-wise interactions (Fig 5B Left). Figure 5A and B, Right, show a decreasing trend for triple-wise interaction as the bin size increases; by the same reason given for pairwise interaction (Fig 4A, Right). The dependency of the triple-wise interaction on the common input rate is also shown in right panels in Fig 5A and B.

Figure 5C shows triple-wise interactions under star (Top) and triangular (Bottom) architectures for excitatory (Right) and inhibitory (Left) common input as a function of scaled diffusion coefficient $D/(\tau_m V_\theta^2)$ and scaled amplitude $A/(\tau_m V_\theta)$. A single common excitatory input, in the

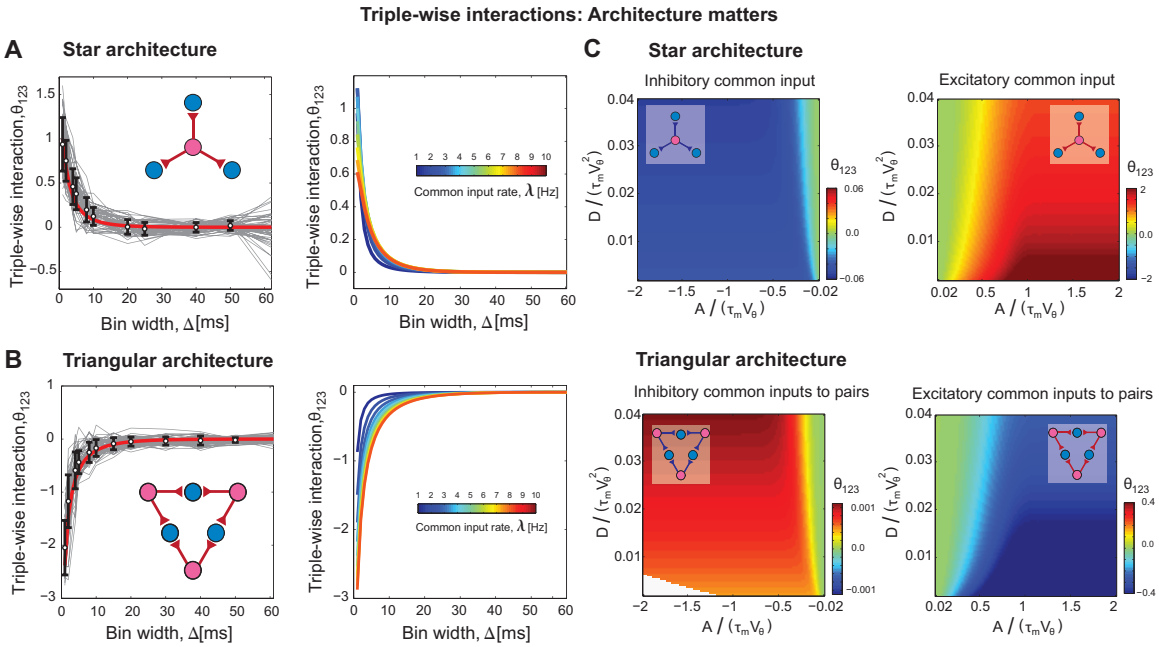

Fig 5. Comparing $\theta^{\mathrm{t}}_{123}$, the triple-wise interaction of 3 LIF neurons, in the two leading architectures. (**A**) The interaction when the postsynaptic neurons, blue ones, are in a star architecture receiving a single excitatory $A = 5\,\mathrm{mV}$ ms common input at the same time. The parameters are: $\tau = 20\,\mathrm{ms}$, $V_\theta = 20\,\mathrm{mV}$, $A = 5\,\mathrm{mV}$ ms and $D = 0.74\,(\mathrm{mV})^2\mathrm{ms}$. **Left**: The triple-wise interaction computed from simulated spike sequences (gray lines: 50 individual trials each containing about 2500 spike occurrence of common input; dots and error bars: mean$\pm$standard deviation) and the analytic result of the mixture model (red line, Eq 15) for $\lambda = 5\,\mathrm{Hz}$. **Right**: Analytical value of $\theta^{\mathrm{t}}_{123}$ as a function of Δ for different common input rates, λ. (**B**) The triple-wise interaction when the postsynaptic neurons, blue ones, are in a triangular architecture. Each pair of postsynaptic neurons receives an independent common excitatory input. The parameters are as in (A). **Left**: The triple-wise interaction as a function of Δ. **Right**: Analytical value of $\theta^{\mathrm{t}}_{123}$ as a function of Δ for different common input rates, λ. (**C**) Triple-wise interactions of 3 neurons receiving common excitatory or inhibitory inputs under star (Top panel) and triangular (Bottom panel) architectures. $\theta^{\mathrm{t}}_{123}$ is represented as a function of two scaled parameters: $D/(\tau_m V_\theta^2)$ and $A/(\tau_m V_\theta)$, while other parameters are: $\lambda = 5\,\mathrm{Hz}$ and $\Delta = 5\,\mathrm{ms}$. White color in bottom left panel shows numerically indeterminate region due to very small diffusion coefficient and strong inhibition. The right-hand side figures are for excitatory and the left-hand side ones are for common inhibitory inputs.

star architecture, significantly increases the probability that all three neurons spike in the observation time window of Δ, $P(1,1,1)$; whereas a single common inhibitory input increases the probability of the reverse pattern, $P(0,0,0)$. This simply changes the sign of $\theta^{\mathrm{t}}_{123}$ in Eq 15. In the triangular architecture, with common excitatory input, however, each common input causes postsynaptic spike for two neurons and does not drive the other one. This primarily increases

$P(1,1,0)$ (or any other of its permutations) and attenuate the fraction (Eq 15); it means negative $\theta^{\mathrm{t}}_{123}$. For common inhibitory input, the probability of the reversed pattern, $P(0,0,1)$ (or any other permutations) increases; this results in larger nominator in Eq 15 and positive triple-wise interaction. These results demonstrate that not only the type of common input (excitation or inhibition) but also the underlying architecture (star or triangular) determines the sign of triple-wise interactions. This helps us to relate the ubiquitously observed negative triple-wise interaction in activities of cortical neurons (Ohiorhenuan et al., 2010; Ohiorhenuan and Victor, 2011; Shimazaki et al., 2012; Yu et al., 2011) to either star architecture with inhibitory input or triangular structure with excitatory input. Notably, in both architectures, the triple-wise interactions for the common inhibitory inputs are significantly weaker, *i.e.* harder to measure experimentally, than those expected for common excitatory inputs.

Network structure and common input type can be determined from neural activity in vivo: Comparison with experimental data

Above observations raise a question: Is it possible to determine the type of common input and the underlying architecture from the event activity of neuronal population? Figure 6 shows the first-order, pairwise and triple-wise interactions of 3 neurons in star or triangular architecture, receiving either common excitatory or inhibitory inputs. The interactions are calculated for some different values of scaled amplitude of common input ($A/(\tau_m V_\theta)$) and diffusion coefficient ($D/(\tau_m V_\theta^2)$) which cover the physiologically plausible range. Figure 6A shows that the same θ^{t}_1 may be realized by different combinations of $A/(\tau_m V_\theta)$ and $D/(\tau_m V_\theta^2)$ for two architectures and common input type; thus it would be impossible to identify the underlying architecture and the type of common input only from the first-order parameter.

Nevertheless, we found that the four combinatorial cases occupy distinct regions in the θ^{t}_{12} versus $\theta^{\mathrm{t}}_{123}$ plane (Fig 6B). Thus, in principle, by investigating the interaction parameters, it is possible to identify the underlying architecture and the type of common inputs (excitation or inhibition) to three homogeneous LIF neurons. However, within each of the conditions (except for star architecture with common excitatory inputs), the parameters overlap, which makes it impossible to identify the underlying parameters such as the input's amplitude or diffusion coefficient from the interaction parameters. Finally, both pairwise and triple-wise interactions are considerably weak when the neurons receive inhibitory inputs. It means, substantially sparser population activity due to the higher-order interaction can be generated if they are produced from common excitatory inputs under the triangular architecture, rather than common inhibitory inputs under the star architecture.

In Fig 7, we compare experimental data with our theoretical predictions. The previous study (Ohiorhenuan and Victor, 2011) investigated the relation between triple-wise interaction (Eq 15) of three neurons (ordinate) and an average marginal pairwise interactions (Eq 11) of neuron pairs in the group (abscissa). Their analysis showed that many of local neurons that are separated within 300 μm exhibited positive pairwise and negative triple-wise interactions. According to the analysis in the previous section, there are two alternative hypotheses to explain such interactions: common excitatory inputs with the triangular architecture or inhibitory inputs with the star architecture. Between the two, we find that triangular architecture with common excitatory inputs can reproduce the empirical interactions. The solid black line shows

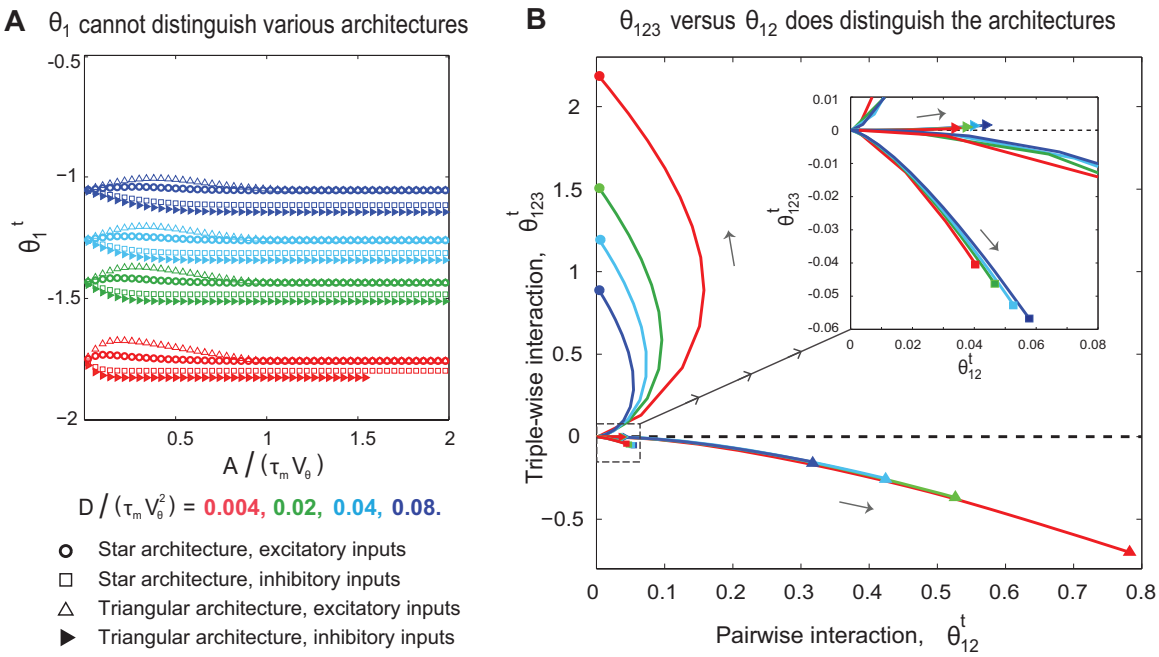

Fig 6. Natural parameters among population activity of 3 LIF neurons receiving common inputs in star and triangular architectures. (**A**) The natural parameter for individual neurons, θ_1^{t}, in star and triangular architectures with common excitatory or inhibitory inputs (four cases are shown by four symbols in the bottom) as a function of $A/(\tau_m V_\theta)$. The color code represents four different scaled diffusion coefficients, $D/(\tau_m V_\theta^2)$. (**B**) The triple-wise interaction versus pairwise interaction for star and triangular architectures with common excitatory and inhibitory inputs. The symbols and the color code are as in (A). The arrows indicate the increasing direction of the scaled amplitude parameter (i.e. $A/(\tau_m V_\theta)$) and the filled symbols show the saturation points. The inset illustrates the interactions in the neighborhood of origin. Comparing (A) and (B), it is clear that θ_1^{t} can hardly be used to distinguish between star and triangular architectures, whereas the two dimensional map of $\theta_{123}^{\mathrm{t}}$ versus θ_{12}^{t} can distinctly separate different architectures and even the type of common inputs especially when the interactions are not very weak. Fixed parameters are $\Delta = 5\,\mathrm{ms}$ and $\lambda = 5\,\mathrm{Hz}$.

theoretical interactions, assuming the triangular architecture with common excitatory inputs, using a very small value of $D/(\tau_m V_\theta^2) = 4 \times 10^{-8}$ for different $A/(\tau_m V_\theta)$. We also consider asymmetric architecture of common excitatory inputs to pairs: two common inputs are shared between two pairs of three neurons. We observe for the asymmetric case of triangular architecture, the dashed black line (Fig 7) is achieved for the tiny value of scaled diffusion parameter ($D/(\tau_m V_\theta^2) = 4 \times 10^{-8}$). Note that the interactions expected for larger and more physiologically plausible values of $D/(\tau_m V_\theta^2)$ occupy in-between solid black line and dashed black line when three common excitatory inputs shared between pairs (the symmetric architecture). For the asymmetric common inputs to pairs in triangular architecture, the region is placed between

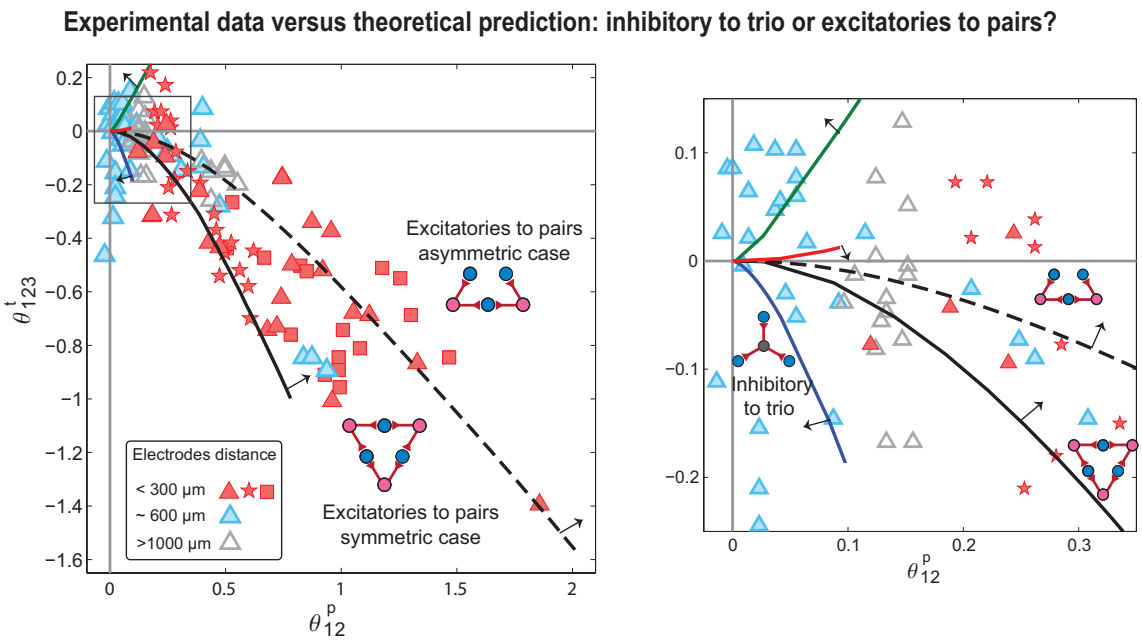

Fig 7. Comparison of theoretical predictions about neural interactions with experimental data extracted from (Ohiorhenuan and Victor 2011, Figure 4b). The data is recorded from V1 neurons of Macaque Monkeys using tetrodes. They analyzed the relation between triple-wise interaction (Eq 15) of three neurons (ordinate) and an average marginal pairwise interaction (Eq 11) of neuron pairs in the group (abscissa). Red and blue filled symbols represent interactions of neurons within 300 and 600 μm vicinity respectively while unfilled grey symbol shows interaction of neurons at > 1000 microns distance. The black line shows interactions theoretically obtained by assuming triangular architecture with common excitatory inputs to each pairs, using a tiny value of $D/(\tau_m V_\theta^2) = 4 \times 10^{-8}$ for different $A/(\tau_m V_\theta)$. Similarly, the blue line is the prediction for star architecture with common inhibitory inputs using the largest value $D/(\tau_m V_\theta^2) = 4$, that we could compute without numerical instability. The dashed black line shows the boundary for asymmetric triangle architecture when two common excitatory inputs are shared between two pairs of neurons, using the same tiny value of scaled diffusion for symmetric case. The right panel magnifies near the origin of the left figure. Here the green (red) line shows the lower (upper) boundaries for star architecture with common excitatory inputs (triangular architecture with common inhibitory inputs). Fixed parameters are $\Delta = 5\,\text{ms}$ and $\lambda = 5\,\text{Hz}$. The data reveals that nearby neurons (red filled symbols) receive pair of excitatory inputs in triangular architecture rather than inhibitory inputs in star architecture.

dashed black line and the abscissa. Thus, in general, triangular architecture with common excitatory inputs is likely to reproduce most of the empirical interactions found in neurons in a local circuitry. To the contrary, it is unlikely for star architecture with common inhibitory input to explain the strong negative interactions (see the blue line; Note that interactions expected for smaller values of $D/(\tau_m V_\theta^2)$ occupy in-between this line and the ordinate).

Although this picture seems valid for common inputs' architecture among three neurons, the question arises whether considering recurrent connectivity for three neurons can change this concluding result in Figure 7. For this aim, we simulate all possible motifs that have directional recurrent connection among three neurons. At first view, the number of motifs for three neurons that can have directional connections would be 64 (each directed connection can be present or absent, so for 6 connections, it is 2^6). Since we have symmetry for three neurons, these 64 motifs are reduced to 16 motifs (Fig. 8). The question here is whether the inhibitory to trio with the help of recurrent activity can reach to the value of experimental data (red symbols in Fig. 7) or excitatory to pairs is the architecture behind it. Figure 8 shows the result of triple-wise interaction for each motif averaging over 50000 runs. We see four clusters of motifs, which are ordered as the number of inputs to pairs increases. The first cluster (blue) contains motifs that do not have any architecture like excitatory input to pairs in the recurrent motifs. For this cluster, the average of triple-wise interaction is a small value. Second cluster (red) has motifs that have one excitatory neuron as input to pairs of neurons in their recurrent architecture. Third and fourth clusters (green and black) contains architectures with two and three excitatory inputs to pairs of neurons. The message is simple; as the number of excitatory to pairs increases, the value of triple-wise interaction shifts to higher amounts (Fig. 7). The inset shows triple-wise versus pairwise interactions for these four clusters. This picture is consistent with our finding in Figure 7 that excitatory input to pairs can generate such a large negative triple-wise and positive pairwise interactions. The excitatory input to pairs which can exist either in common inputs or in recurrent connectivity can generate such activity and thus is the basic architecture behind the data presented here (Fig 7).

Then we answer quantitatively the question of why inhibitory input cannot generate strong pairwise and triple-wise interactions in Figure 7. We find out it is attributed to spontaneous probability of spiking for postsynaptic neurons in a small time-window Δ. It is shown in the Appendix, how the pairwise and triple-wise interactions depend on the firing rate of signaling input and the spontaneous probability of spiking for postsynaptic neurons. The result (Fig 9) shows that in a regime of strong signaling input, when the spontaneous spiking of postsynaptic neuron in time-window Δ is ($F_\Delta \ll 1$), the pairwise and triple-wise interactions imposed by excitatory inputs to pairs is much stronger than inhibitory to trio architecture. At the limit of high spiking probability for postsynaptic neurons ($F_0(\Delta) \simeq 1$), inhibitory to trio can generate strong interactions (Fig 9) while excitatory to pairs cannot. There is also intermediate regimes ($F_0(\Delta) \simeq 0.5$) where the strength of interaction induced by inhibitory to trio and excitatory to pairs are nearly the same(Fig 9). In Victor and colleagues experiment (Ohiorhenuan et al., 2010; Ohiorhenuan and Victor, 2011)(Figure 7), the range of firing rates neurons generate (Fig.4 in (Ohiorhenuan and Victor, 2011)) places between $10^{-2} < \mu < 70$ (Ohiorhenuan et al., 2010) and $10 < \mu < 70$ (Ohiorhenuan and Victor, 2011) which makes $F_0 < 0.5$. In this regime, excitatory to pairs architecture can induce large negative triple-wise and positive pairwise interactions while inhibitory to trio can't.

As a short summary, using our analytical solution, we clarified that: (i) Positive pairwise and negative triple-wise interactions are produced if excitatory signaling inputs are shared by two but not by more neurons, or inhibitory inputs are shared by at least three neurons. This scenario is consistent with the observed sparse connections among excitatory neurons (Holmgren et al., 2003; Ikegaya et al., 2013; Lefort et al., 2009) and also suggests dense inhibitory to excitatory

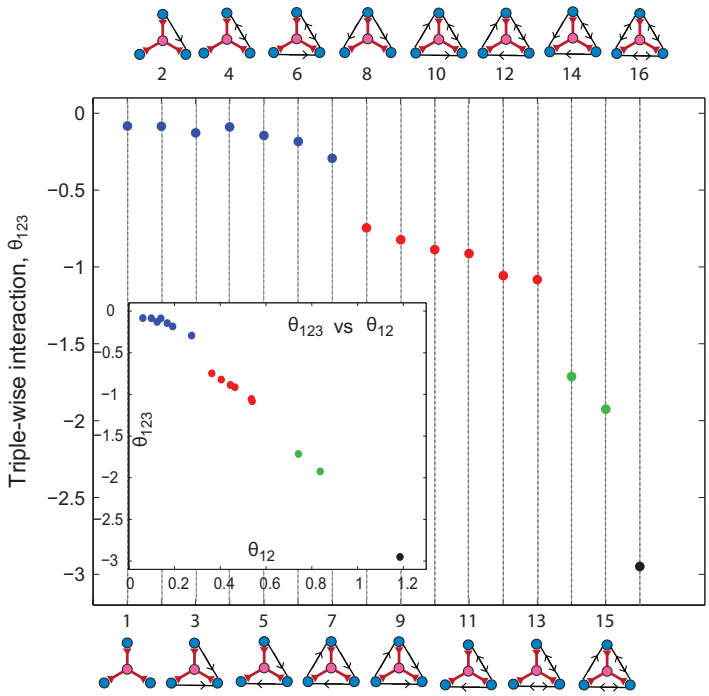

Fig 8. Triple-wise interactions for 16 motifs of directional recurrent connections among three neurons, when they receive independent noise and inhibitory common input to trio. The architectures are divided into four clusters based on the number of excitatory inputs to pairs in each motif. The first cluster (blue) contains recurrent connections that does not have any excitatory to pairs in their architectures. The second cluster (red) has one excitatory to pairs in recurrent motifs and the third and fourth (green and black) are related to two and three excitatory to pairs in motifs with directional recurrent connections. Inset shows the triple-wise interaction versus pairwise interaction for all 16 motifs. Each motif is simulated 50000 times and each time, 500 s of spike trains with time resolution of 0.05 ms are considered. The time window to calculate triple-wise and pairwise interactions is $\Delta = 5$ ms, and the shape of presynaptic input is a square function for both common input and recurrent input, the same as analytic calculation. The parameters are: the scaled diffusion $D/(\tau_m {V_\theta}^2) = 9 \times 10^{-5}$, the scaled amplitude $A/(\tau_m V_\theta) = 0.0125$, the common input and recurrent input's rate $\lambda = 5$ Hz, and $A'/A = 1$, where A' is the amplitude of the recurrent connections.

connections observed in some experimental studies (Fino and Yuste, 2011; Packer and Yuste, 2011). (ii) Interaction values are significantly weaker when all neurons receive common inhibitory input. Therefore, sufficiently strong negative triple-wise interactions observed in the experiments (Ohiorhenuan et al., 2010; Ohiorhenuan and Victor, 2011; Shimazaki et al., 2015; Yu et al., 2011) are not likely to be reproduced by common inhibitory inputs to trios. Shared

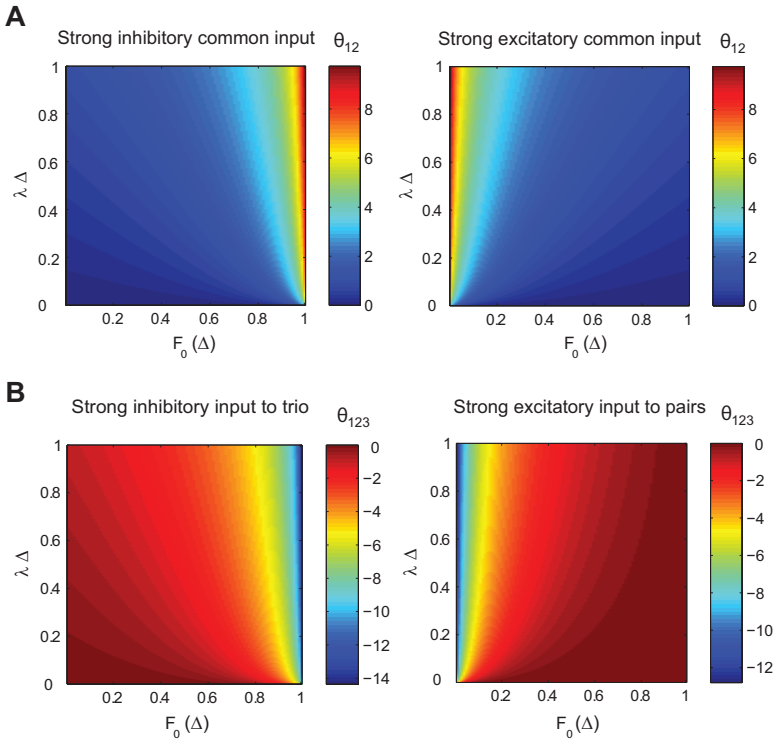

Fig 9. The pairwise and triple-wise interactions for strong excitatory and inhibitory inputs as a function of spontaneous spiking of postsynaptic neuron (F_0) and input rate (λ) in small time window Δ. (**A**) inhibitory inputs generate strong pairwise interaction in high spontaneous spiking of postsynaptic neuron, i.e., $F_0 \simeq 1$, while excitatory inputs generate such strong interactions in low spontaneous spiking regime, $F_0 \ll 1$. (**B**) The same story is true for triple-wise interaction independent of excitatory to pairs and inhibitory to trio architectures. Excitatory to pairs can generate strong triple-wise interactions in low spontaneous spiking but inhibitory to trio can't induce such strong interaction in this regime.

excitatory inputs to pairs can reproduce the observed strong negative triple-wise interactions (Fig 7). By considering recurrent connectivity among neurons, this picture is reconfirmed (Fig. 8). (iii) One can identify the underlying architecture and type of common input (excitation or inhibition) in the plane of pairwise and triple-wise interaction parameters (see Fig 7). Using our exact solution for the LIF neuron model, we showed that the excitatory to pairs and inhibitory to trio architectures have distinct signatures in the pairwise and triple-wise interaction plane; this makes identification of these two architectures possible. Investigating experimental data (Ohiorhenuan and Victor 2011, Fig 4b), our framework (Fig 7) suggests that nearby neurons within 300 μm receive excitatory common inputs to pairs, however, for far away neurons (≥ 600 μm) no concrete suggestion can be made. The weaker interactions of the distant neurons may be due to lack of common inputs, or it may also be attributed to the heterogeneity of these neurons that cancels out interactions among them.

Discussion

The aim of this study is to make a link between network architecture and statistics of neural activity through nonlinearity of the input-output relation of the neurons. Several studies of higher-order interactions among model neurons reveal key features of neural activity, and suggest mechanisms underlying the observed interactions in a neural pool (Cayco-Gajic et al., 2015; Montangie and Montani, 2017; Montani et al., 2013; Shimazaki et al., 2015; Zylberberg and Shea-Brown, 2015). But since strong synapses shape the backbone of network architecture, a model that accurately include nonlinear behavior of such synapses, is essential to reveal network architecture. We use the established conditional ISI density of an *LIF neuron* which is presumably in the near threshold regime, and receives a transient signaling input (Shomali et al., 2018). We use it to compute the probability density of spiking after shared signaling input's arrival, and hence the corresponding *cumulative density of spiking* within a given time window. It helps us to obtain the probabilities of binary activity of *two neurons* after arrival of strong shared input, a key that further connects network architecture with *population activity*.

The activity patterns in the binned spike data can be obtained as the probabilistic mixture of two possible events: The one that describes binary activity given a common input arrived upon stimulus onset, and the other that describes binary activity assuming two neurons received no common input. To assess the strength of neural interactions, we use the information geometric measure of correlations (Amari, 2001). Extending this scheme to three-neurons analysis, we derive triple-wise interactions of a set of postsynaptic neurons receiving common signaling input (excitatory or inhibitory) in different neural architectures (star or triangle). It let us accurately connect synaptic inputs, threshold nonlinearity, and network architecture to the *statistics of population activity*. Our results point to the possibility of inferring the underlying neuronal architectures and the type of common input, using pairwise and triple-wise neural interactions. Further, for a specific set of empirical observation (Ohiorhenuan et al., 2010) in comparison with our analytical result, we show that excitatory shared inputs to pairs of neurons rather than inhibition to trios explain the data. Consideration of recurrent connections among three neurons does not disturb this picture: Excitatory to pairs either as an independent common input or a recurrent connectivity is the architecture behind the observed *strong negative triple-wise* and positive pairwise interactions.

Previous studies showed that a simple model neuron known as the Dichotomized Gaussian (DG) model exhibits higher-order interactions (Amari et al., 2003; Macke et al., 2009, 2011). It is composed of threshold devices that receive inputs sampled from a correlated multivariate Gaussian distribution. The output binary patterns of this model exhibit characteristic higher-order interactions, namely positive pairwise and negative triple-wise interactions (Shimazaki et al., 2015; Yu et al., 2011), which results in the observed sparse activity (Shimazaki et al., 2015). However, it is speculated that common inhibitory inputs can also generate the same neural interactions (Ohiorhenuan and Victor, 2011; Shimazaki et al., 2015). The DG model based on the Gaussian approximation for common synaptic inputs (Amari et al., 2003; Leen and Shea-Brown, 2015; Macke et al., 2011; Montangie and Montani, 2015) as well as extended DG model for non-Gaussian inputs (Barreiro et al., 2014; Montangie and Montani, 2017, 2018) do not incorporate the dynamics of synapses and membrane potentials, hence are not suitable for investigating

types of synaptic inputs in different architectures. The aforementioned input-output relation for near threshold neurons (Shomali et al., 2018), on the other hand, does address dynamics of membrane potential using LIF neuron model; so it yields quantitative results with temporal accuracy, which let us quantitatively infer neural architecture and type of common input.

The quantitative model we introduced here is based on two distinct network architectures (triangle or star) with either excitatory or inhibitory shared inputs. It is crucial to see how the recurrent connection alters the view. The observed sparse connectivity of pyramidal neurons (Holmgren et al., 2003; Lefort et al., 2009; Markram et al., 1997; Mizusaki et al., 2016) shows that pyramidal neurons in visual cortex - in mature animals - are *not interconnected* (Jiang et al., 2015). However, some experimental studies show the combination of recurrent connections with shared inputs: Excitatory inputs from layer 4 are shared to layer 2/3 connected pairs of excitatory pyramidal neurons (Yoshimura et al., 2005). Hence, we ran simulations for the recurrent connections among three neurons: The simulations reaffirm that even with recurrent connections, the motif that does cause in strong negative triple-wise and positive pairwise interactions, is excitatory to pairs (see Fig. 8). Furthermore, it is questionable whether simultaneous existence of excitatory and inhibitory common inputs in both triangle and star architectures damages this picture. We thus analyzed the model in which two architectures are mixed i.e., existing together, at each time. The result reiterates that the excitatory triangle architecture is still the prevailing motif which generates the observed negative triple-wise interactions (data not shown here).

The *robustness* of the analytical tool, is another vital question. Firstly, the analytical tool is valid for neurons with (i) membrane potentials close to the threshold of spiking, which (ii) receive Gaussian background noise (Shomali et al., 2018). These assumptions are partly motivated by an empirical study (Tan et al., 2014) which shows V1 neurons operate near the threshold regime and the background noise would be uncorrelated during sensory stimulation. Nonetheless, we preform another analysis for *extreme case* of very strong common inputs (see Appendix); the analysis is *independent of model neuron*, and the fact of being near or far from the threshold. It reveals that for *sparse activity* of neurons, in general, the excitatory to pairs architecture is the sole structure behind strong negative triple-wise interactions. The other question of robustness involves our assumption of sparse (Wolfe et al., 2010) firing rate for the signaling input: We have implicitly assumed that the firing rate of the signaling input is low, so there is at most one signal arrival during two successive spikes of the postsynaptic neuron. It is possible to consider cases with higher firing rate for signaling input (see Appendix III in Shomali et al. 2018). However, as we consider small time window, $\Delta = 5\text{m}s$, the assumption of having *not more than one* signal arrival during such short time window, is easily acceptable. What happens for cases with very high firing rate? Our aforementioned *extreme case* analysis shows that we should expect a very different picture, where *inhibitory to trio* are the one architecture behind possible observation of negative triple-wise interaction.

Finally, we have attributed the observation of strong negative triple-wise interactions, in an already sparse network, to a simple motif: Excitatory to pairs. It is tempting to ask whether this structure has any specified computational advantage so has been boldly observed, or it is the overall setting of the particular experiment which has dictated such observation? On one hand, (a) there exist independent empirical evidences that excitatory to pairs is overexpressed - compared to random network - in rat's visual (Song et al., 2005) and somatosensory (Perin et al.,

2011) cortex. In a recent theoretical study this overexpression is explained as a consequence of maximizing capacity of the associative memory (Zhang et al., 2019); these suggest that the emergence of excitatory to pairs is not an accidental observation. On the other hand, (b) we have shown that whenever the background network is sparse, strong negative triple-wise is the signature of excitatory to pairs; and even simultaneous presence of excitation and inhibition does not disturb this conclusion. So, if we are to capture the other motif of inhibitory to trio, we should expect negative but small values for triple-wise interaction. Nevertheless, (c) there also exist other experimental evidences that, an inhibitory common input innervate multiple postsynaptic pyramidal neurons closer to each other than 100μm (Packer and Yuste, 2011); in longer inter-separations, the probability of inhibitory common inputs to two (and hence more) neurons annihilates (Fig 6B in Packer and Yuste 2011). This is attributed to the limited length of inhibitory neurons' axons and simply means that if electrodes' inter-separation is larger than 100μm, the chance of capturing a common inhibitory input (to pair, or to trio) has already diminished! So, we ask of curiosity, if one could increase the experimental precision, and reach smaller inter-distances for electrode's separation, was it possible to see some transition in the observation as s/he was decreasing electrodes' inter-distance?

In summary, we provided a theoretical tool to predict network architecture and types of input neurons (excitatory/inhibitory) from activity of neurons recorded in vivo, based on the dynamics of a standard neuron model. While it is quite challenging to perform in vivo patch clamp of common inputs and postsynaptic neurons at the same time, such an experiment that can directly identify input types and structure of the networks in living animals is helpful to improve the prediction of this method, and make it applicable to a variety of in vivo data.

Acknowledgments

S.R.S acknowledges T.Vogels' and C.Clopath's helpful comments; H.S. thanks S.Koyama and R.Kobayashi for valuable discussions; S.R.S and S.N.R acknowledge T.Fukai's kind supports and are grateful to him and H.Maboudi for valuable discussions.

Contributions

S.R.S, M.N.A, S.N.R and H.S designed research. S.R.S and H.S performed the research. S.RS and S.N.R analyzed the results. S.RS, M.N.A, S.N.R and H.S wrote the first draft of paper. S.R.S, S.N.R and H.S wrote and edited the paper.

Declaration of Interests

The authors declare no competing interests.

Appendix

The interaction induced by inhibitory and excitatory inputs

One main message of this paper is that the observed strong negative triple-wise interactions are signature of excitatory to pairs (see figure 7). The reasonable question is then, why should this happen? Is there any fundamental difference between excitatory to pairs, with inhibitory to trio, that the later one can not produce strong negative triple wise interactions? We approach this question, for the extreme case of very strong signaling input. We firstly consider pairwise interactions, and then extend the formalism to triple-wise. Combining Eq 11, Eq 12, and Eq 13, the pairwise interaction reads:

$$\theta_{12}^{\mathrm{p}} = \log \frac{(bF_A(\Delta)^2 + (1-b)F_0(\Delta)^2)(b(1-F_A(\Delta))^2 + (1-b)(1-F_0(\Delta))^2)}{(bF_A(\Delta)(1-F_A(\Delta)) + (1-b)F_0(\Delta)(1-F_0(\Delta)))^2}, \tag{18}$$

where $b = \lambda\Delta$. If the postsynaptic neuron receives *an extremely strong inhibitory signals*, the chance of its spike in a short time window of Δ immediately after signal arrival diminishes: $F_A(\Delta) \simeq 0$. This reduces Eq.(18 to:

$$\theta_{12}^{\mathrm{p}} = \log(1 + \frac{a}{\{1-F_0(\Delta)\}^2}), \tag{19}$$

where $a = b/(1-b) = \lambda\Delta/(1-\lambda\Delta)$. Conversely, for *an extremely strong excitatory signals*, the chance of spiking in the time window of Δ, immediately after signal arrival is almost 1; hence $F_A(\Delta) \simeq 1$. The pairwise interaction then simplifies to:

$$\theta_{12}^{\mathrm{p}} = \log(1 + \frac{a}{F_0(\Delta)^2}). \tag{20}$$

Comparing Eq 19 and Eq 20, we see that $F_0(\Delta)$ is simply replaced by $1 - F_0(\Delta)$. In Eq.(20), for excitatory signaling input and with $a > 0$, we obtain large θ_{12}^{p} when $F_0(\Delta) \ll 1$. In fact, θ_{12}^{p} indefinitely increases as $F_0(\Delta) \to 0^+$. On the other hand, for strong inhibitory signals in Eq.(19), this criteria changes to $1 - F_0(\Delta) \ll 1$; which means we need $F_0(\Delta) \lesssim 1$ to obtain high θ_{12}^{p}. Consequently, in regime of sparse activity i.e., $F_0(\Delta) \ll 1$, strong excitatory signals produce large θ_{12}^{p}; whereas in high activity regime i.e., $F_0(\Delta) \lesssim 1$, strong inhibitory inputs produce large pairwise interactions. The only approximation in above reasoning is our assumption of low firing rate: $\lambda\Delta \ll 1$. More accurately, the Poissonian probability of signal arrival in a time window of Δ, is $b = 1 - \exp(\lambda\Delta)$, which simplifies to $b \simeq \lambda\Delta$ in the low firing rate limit. Accordingly, the prefactor of a is in general $a = \{1 - \exp(-\lambda\Delta)\}/\exp(-\lambda\Delta)$ which simplifies to $a \simeq \lambda\Delta/(1-\lambda\Delta)$ in the low firing rate limit. However, this correction does not disturb the overall result of our analysis, at all.

Figure 9**A** shows how the pairwise interaction depends on spontaneous spiking rate and also the firing rate of signaling input in small time window of Δ for both inhibitory (left panel) and excitatory inputs (right panel). As the firing rate of signaling input increases, the amount of interaction induced by both excitatory and inhibitory inputs, increases as well. However, by increasing the spontaneous activity rate of $F_0(\Delta)$, interaction induced by excitatory inputs to pairs decreases while that of inhibitory to pairs increases.

We extend this analysis to triple-wise interaction. Firstly, we consider the case of *strong inhibitory to trio*. Combining Eq 14, Eq 15 and Eq 16 we have:

$$\theta^{\mathrm{t}}_{123} = \log \frac{P_{111}P_{100}^3}{P_{000}P_{110}^3}, \tag{21}$$

where

$$P_{111} = b \times F_A(\Delta)^3 + (1-b) \times F_0(\Delta)^3, \tag{22a}$$
$$P_{000} = b \times \{1 - F_A(\Delta)\}^3 + (1-b) \times \{1 - F_0(\Delta)\}^3, \tag{22b}$$
$$P_{100} = b \times F_A(\Delta)\{1 - F_A(\Delta)\}^2 + (1-b) \times F_0(\Delta)\{1 - F_0(\Delta)\}^2, \tag{22c}$$
$$P_{110} = b \times F_A(\Delta)^2\{1 - F_A(\Delta)\} + (1-b) \times F_0(\Delta)^2\{1 - F_0(\Delta)\}; \tag{22d}$$

and $b = 1 - \exp(\lambda\Delta)$, which simplifies to $b \simeq \lambda\Delta$ with low firing rate. Here, we have considered the aforementioned symmetry in the inhibitory to trio architecture, which enforces that $P_{011} = P_{101} = P_{110}$, etc. In the limit of strong inhibition, for small time window of $\Delta = 5\,\mathrm{ms}$, $F_A(\Delta) \simeq 0$. It drastically simplifies $\theta^{\mathrm{t}}_{123}$ to:

$$\theta^{\mathrm{t}}_{123} = -\log(1 + \frac{a}{(1 - F_0(\Delta))^3}). \tag{23}$$

Here $a = \{1 - \exp(-\lambda\Delta)\}/\exp(-\lambda\Delta)$, which is approximated to $a = \lambda\Delta/(1 - \lambda\Delta)$ whenever $\lambda\Delta \ll 1$.

Similarly, we can consider the triple-wise interaction for excitatory to pairs architecture. It is a bit complicated in comparison with inhibitory to trio case. Considering Eq 14, Eq 15 and Eq 17, the triple-wise interaction is:

$$\theta^{\mathrm{t}}_{123} = \log \frac{P_{111}P_{100}^3}{P_{000}P_{110}^3}, \tag{24}$$

however,

$$\begin{aligned} P_{111} = {} & (1-b)^3 \times F_0(\Delta)^3 + 3b(1-b)^2 \times F_A(\Delta)^3 F_0(\Delta) \\ & + 3b^2(1-b) \times F_A(\Delta)^2 F_{2A}(\Delta) + b^3 \times F_{2A}(\Delta)^3, \end{aligned} \tag{25a}$$
$$\begin{aligned} P_{000} = {} & (1-b)^3 \times \{1 - F_0(\Delta)\}^3 + 3b(1-b)^2 \times \{1 - F_A(\Delta)\}^2\{1 - F_0(\Delta)\} \\ & + 3b^2(1-b) \times \{1 - F_A(\Delta)\}^2\{1 - F_{2A}(\Delta)\} + b^3 \times \{1 - F_{2A}(\Delta)\}^3, \end{aligned} \tag{25b}$$
$$\begin{aligned} P_{100} = {} & (1-b)^3 \times F_0(\Delta)\{1 - F_0(\Delta)\}^2 + b^3 \times F_{2A}(\Delta)\{1 - F_{2A}(\Delta)\}^2 \\ & + b(1-b)^2 \times \big\{2F_A(\Delta)\{1 - F_A(\Delta)\}\{1 - F_0(\Delta)\} + F_0(\Delta)\{1 - F_A(\Delta)\}^2\big\} \\ & + b^2(1-b) \times \big\{2F_A(\Delta)\{1 - F_A(\Delta)\}\{1 - F_{2A}(\Delta)\} + F_{2A}(\Delta)\{1 - F_A(\Delta)\}^2\big\}, \end{aligned} \tag{25c}$$
$$\begin{aligned} P_{110} = {} & (1-b)^3 \times F_0(\Delta)^2\{1 - F_0(\Delta)\} + b^3 \times F_{2A}(\Delta)^2\{1 - F_{2A}(\Delta)\} \\ & + b(1-b)^2 \times \big\{2F_0(\Delta)\,F_A(\Delta)\{1 - F_A(\Delta)\} + F_A(\Delta)^2\{1 - F_0(\Delta)\}\big\} \\ & + b^2(1-b) \times \big\{2F_A(\Delta)F_{2A}(\Delta)\{1 - F_A(\Delta)\} + F_A(\Delta)^2\{1 - F_{2A}(\Delta)\}\big\}. \end{aligned} \tag{25d}$$

For extremely strong excitation to pairs $F_A(\Delta) \simeq 1$; we thus have:

$$\theta^{\mathrm{t}}_{123} = -\log\left((1 + \frac{a}{F_0(\Delta)^2})^3 \times \frac{1}{1 + 3a/F_0(\Delta)^2 + 3a^2/F_0(\Delta)^3 + a^3/F_0(\Delta)^3}\right), \tag{26}$$

where $a = \{1 - \exp(-\lambda\Delta)\}/\exp(-\lambda\Delta)$, as before.

Figure 9**B** shows, for *sparse activity* of postsynaptic neuron, i.e., $F_0(\Delta) \ll 1$, excitatory to pairs does generate strong negative triple-wise interaction (right panel); whereas for *high spontaneous activity*, i.e., $F_0(\Delta) \simeq 1$, inhibitory to trio produces such strong results. There is one interesting fact in the data from Ohiorhenuan and Victor (Ohiorhenuan and Victor, 2011): The neurons that have the lower firing rate, generate stronger pairwise interactions; we particularly compare figures 4c and 4a in (Ohiorhenuan and Victor, 2011), each set of data with smaller firing rate has indeed produced larger pairwise interaction.

It is also interesting that this observed decrease of pairwise interaction with increasing of spontaneous activity, is also revisited for *excitatory to pairs* architecture in figure 9**B**-right, and not for the *inhibitory to trio*, see figure 9**B**-left. This is another fact which reaffirms our original claim that excitatory inputs to pairs is the architecture behind that set of data.

References

Amari, S.-i. (2001). Information geometry on hierarchy of probability distributions. IEEE transactions on information theory, 47, 1701–1711.

Amari, S.-i. (2009a). Information geometry and its applications: Convex function and dually flat manifold. In Emerging Trends in Visual Computing. (Springer), pp. 75–102.

Amari, S.-i. (2009b). Measure of correlation orthogonal to change in firing rate. Neural computation, 21, 960–972.

Amari, S.-i., Nakahara, H., Wu, S., and Sakai, Y. (2003). Synchronous firing and higher-order interactions in neuron pool. Neural computation, 15, 127–142.

Barreiro, A. K., Gjorgjieva, J., Rieke, F., and Shea-Brown, E. (2014). When do microcircuits produce beyond-pairwise correlations? Frontiers in computational neuroscience, 8, 10.

Cayco-Gajic, N. A., Zylberberg, J., and Shea-Brown, E. (2015). Triplet correlations among similarly tuned cells impact population coding. Frontiers in computational neuroscience, 9, 57.

Cover, T. M. and Thomas, J. A. (1991). Elements of information theory. (John Wiley & Sons).

Cox, D. R. (1962). Renewal theory, vol. 4. (Methuen London).

Fino, E. and Yuste, R. (2011). Dense inhibitory connectivity in neocortex. Neuron, 69, 1188–1203.

Holmgren, C., Harkany, T., Svennenfors, B., and Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. The Journal of physiology, 551, 139–153.

Hu, Y., Trousdale, J., Josic, K., and Shea-Brown, E. (2013). Motif statistics and spike correlations in neuronal networks. Journal of Statistical Mechanics: Theory and Experiment, 3, 03012.

Hu, Y., Trousdale, J., Josić, K., and Shea-Brown, E. (2014). Local paths to global coherence: cutting networks down to size. Physical Review E, 89, 032802.

Ikegaya, Y., Sasaki, T., Ishikawa, D., Honma, N., Tao, K., Takahashi, N., Minamisawa, G., Ujita, S., and Matsuki, N. (2013). Interpyramid spike transmission stabilizes the sparseness of recurrent network activity. Cerebral Cortex, 23, 293–304.

Jiang, X., Shen, S., Cadwell, C. R., Berens, P., Sinz, F., Ecker, A. S., Patel, S., and Tolias, A. S. (2015). Principles of connectivity among morphologically defined cell types in adult neocortex. Science, 350, aac9462.

Leen, D. A. and Shea-Brown, E. (2015). A simple mechanism for beyond-pairwise correlations in integrate-and-fire neurons. The Journal of Mathematical Neuroscience (JMN), 5, 17.

Lefort, S., Tomm, C., Sarria, J.-C. F., and Petersen, C. C. (2009). The excitatory neuronal network of the c2 barrel column in mouse primary somatosensory cortex. Neuron, 61, 301–316.

Macke, J. H., Berens, P., Ecker, A. S., Tolias, A. S., and Bethge, M. (2009). Generating spike trains with specified correlation coefficients. Neural computation, 21, 397–423.

Macke, J. H., Opper, M., and Bethge, M. (2011). Common input explains higher-order correlations and entropy in a simple model of neural population activity. Physical Review Letters, 106, 208102.

Markram, H., Lübke, J., Frotscher, M., Roth, A., and Sakmann, B. (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. The Journal of physiology, 500, 409.

Martignon, L., Deco, G., Laskey, K., Diamond, M., Freiwald, W., and Vaadia, E. (2000). Neural coding: higher-order temporal patterns in the neurostatistics of cell assemblies. Neural Computation, 12, 2621–2653.

McCormick, D. A., Connors, B. W., Lighthall, J. W., and Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. J Neurophysiol, 54, 782–806.

Mizusaki, B. E., Stepanyants, A., Chklovskii, D. B., and Sjöström, P. J. (2016). Neocortex: a lean mean storage machine. Nature neuroscience, 19, 643.

Montangie, L. and Montani, F. (2015). Quantifying higher-order correlations in a neuronal pool. Physica A: Statistical Mechanics and its Applications, 421, 388–400.

Montangie, L. and Montani, F. (2017). Higher-order correlations in common input shapes the output spiking activity of a neural population. Physica A: Statistical Mechanics and its Applications, 471, 845–861.

Montangie, L. and Montani, F. (2018). Common inputs in subthreshold membrane potential: The role of quiescent states in neuronal activity. Physical Review E, 97, 060302.

Montani, F., Phoka, E., Portesi, M., and Schultz, S. R. (2013). Statistical modelling of higher-order correlations in pools of neural activity. Physica A: Statistical Mechanics and its Applications, 392, 3066–3086.

Nakahara, H. and Amari, S.-i. (2002). Information-geometric measure for neural spikes. Neural Computation, 14, 2269–2316.

Ocker, G. K., Hu, Y., Buice, M. A., Doiron, B., Josić, K., Rosenbaum, R., and Shea-Brown, E. (2017a). From the statistics of connectivity to the statistics of spike times in neuronal networks. Current opinion in neurobiology, 46, 109–119.

Ocker, G. K., Josić, K., Shea-Brown, E., and Buice, M. A. (2017b). Linking structure and activity in nonlinear spiking networks. PLoS computational biology, 13, e1005583.

Ohiorhenuan, I. E., Mechler, F., Purpura, K. P., Schmid, A. M., Hu, Q., and Victor, J. D. (2010). Sparse coding and high-order correlations in fine-scale cortical networks. Nature, 466, 617–621.

Ohiorhenuan, I. E. and Victor, J. D. (2011). Information-geometric measure of 3-neuron firing patterns characterizes scale-dependence in cortical networks. Journal of computational neuroscience, 30, 125–141.

Ostojic, S., Brunel, N., and Hakim, V. (2009). How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. Journal of Neuroscience, 29, 10234–10253.

Packer, A. M. and Yuste, R. (2011). Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition? The Journal of Neuroscience, 31, 13260–13271.

Perin, R., Berger, T. K., and Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. Proceedings of the National Academy of Sciences, 108, 5419–5424.

Pernice, V., Staude, B., Cardanobile, S., and Rotter, S. (2011). How structure determines correlations in neuronal networks. PLoS computational biology, 7, e1002059.

Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E., and Doiron, B. (2017). The spatial structure of correlated neuronal variability. Nature neuroscience, 20, 107.

Shadlen, M. N. and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. The Journal of neuroscience, 18, 3870–3896.

Shimazaki, H., Amari, S.-i., Brown, E. N., and Grün, S. (2012). State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. PLoS computational biology, 8, e1002385.

Shimazaki, H., Sadeghi, K., Ishikawa, T., Ikegaya, Y., and Toyoizumi, T. (2015). Simultaneous silence organizes structured higher-order interactions in neural populations. Scientific Reports, 5, 9821.

Shomali, S. R., Ahmadabadi, M. N., Shimazaki, H., and Rasuli, S. N. (2018). How does transient signaling input affect the spike timing of postsynaptic neuron near the threshold regime: an analytical study. Journal of computational neuroscience, 44, 147–171.

Softky, W. R. and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. The Journal of Neuroscience, 13, 334–350.

Song, S., Sjöström, P. J., Reigl, M., Nelson, S., and Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. PLoS biology, 3, e68.

Tan, A. Y., Chen, Y., Scholl, B., Seidemann, E., and Priebe, N. J. (2014). Sensory stimulation shifts visual cortex from synchronous to asynchronous states. Nature, 509, 226–229.

Trousdale, J., Hu, Y., Shea-Brown, E., and Josić, K. (2012). Impact of network structure and cellular response on spike time correlations. PLoS computational biology, 8, e1002408.

Tuckwell, H. C. (1988). Introduction to Theoretical Neurobiology: Nonlinear and stochastic theories, vol. 2. (Cambridge University Press).

Vreeswijk, C. v. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science, 274, 1724–1726.

Vreeswijk, C. v. and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. Neural computation, 10, 1321–1371.

Wang, M. C. and Uhlenbeck, G. E. (1945). On the theory of the brownian motion ii. Reviews of Modern Physics, 17, 323.

Wolfe, J., Houweling, A. R., and Brecht, M. (2010). Sparse and powerful cortical spikes. Current opinion in neurobiology, 20, 306–312.

Yoshimura, Y., Dantzker, J. L., and Callaway, E. M. (2005). Excitatory cortical neurons form fine-scale functional networks. Nature, 433, 868.

Yu, S., Yang, H., Nakahara, H., Santos, G. S., Nikolić, D., and Plenz, D. (2011). Higher-order interactions characterized in cortical activity. The Journal of Neuroscience, 31, 17514–17526.

Zhang, D., Zhang, C., and Stepanyants, A. (2019). Robust associative learning is sufficient to explain the structural and dynamical properties of local cortical circuits. Journal of Neuroscience, pp. 3218–18.

Zylberberg, J. and Shea-Brown, E. (2015). Input nonlinearities can shape beyond-pairwise correlations and improve information transmission by neural populations. Physical Review E, 92, 062707.