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

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

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

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

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

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

5 Center for Human Nature, Artificial Intelligence, and Neuroscience (CHAIN), Hokkaido University, Japan

* Corresponding author: safura@ipm.ir

Abstract

Using observed neuronal activity, we try to unveil hidden microcircuits. A key requirement is the knowledge of statistical input-output relation of single neurons in vivo. We use a recent exact solution of spike-timing for leaky integrate-and-fire neurons under noisy inputs balanced near threshold, and construct a framework that links synaptic type/strength, and spiking nonlinearity, with statistics of neuronal activity. The framework explains structured higher-order interactions of neurons receiving common inputs under different architectures. Comparing model's prediction with an empirical dataset of monkey V1 neurons, we find that excitatory inputs to pairs explain the observed sparse activity characterized by negative triple-wise interactions, ruling out the intuitive shared inhibition. We show that the strong interactions are in general the signature of excitatory rather than inhibitory inputs whenever spontaneous activity is low. Finally, we present a guide map that can be used to reveal the hidden motifs underlying observed interactions found in empirical data.

Introduction

One interest in neuroscience is to *reveal in vivo neural circuitries* using recorded neuronal activities. The recent technological advances in *Connectome projects* do reveal complete wiring diagrams of certain animals (Markram et al., 2015; Oh et al., 2014; Xu et al., 2020), nonetheless, to address what computation the neural circuitry performs, it is still important to identify network architecture from *in vivo* recordings of multiple neurons. The *simultaneous intracellular recordings* are the most reliable way to identify physical connections *in vivo* (Allen et al., 2018; Arroyo et al., 2018; Gentet et al., 2010; Poulet et al., 2019; Poulet and Petersen, 2008); yet, using patch-clamp technique, one should record from neurons and all their presynaptic inputs simultaneously, to successfully find the influential synapses which form the underlying circuits, but this reliable method is limited to the small number of neurons. Instead, one could use neuronal spiking activity simultaneously recorded from a large number of neurons (Stringer et al., 2019). The *cross-correlograms* (Kobayashi et al., 2019; Perkel et al., 1967) or constructing point-process network models are *classical* approaches to infer the connectivity from spiking data (Pillow et al., 2008; Truccolo et al., 2005; Volgushev et al., 2015). However, these approaches aim at discovering connections among the recorded neurons, despite that majority of the synaptic inputs to them come from unobserved neurons. Therefore, it remains a challenge to successfully reveal the hidden neuronal circuitries, using the *activity statistics* of a limited number of neurons, *in vivo*.

The hallmark of cortical spiking activity in vivo is its *variability* (Shadlen and Newsome, 1998; Softky and Koch, 1993). It was suggested that the variability of spiking activity is the result of *balanced inputs* from excitatory and inhibitory neurons fluctuating *near spiking threshold* (Shadlen and Newsome, 1998; Vreeswijk and Sompolinsky, 1996, 1998). Such balanced inputs were confirmed by intracellular recordings of in vivo neurons (Okun and Lampl, 2008), during stimulus presentation (Tan et al., 2014). In this condition, even a moderate synaptic input can result in the spiking of the postsynaptic neuron. This, however, does not mean we can safely ignore *strong synaptic* inputs. Such *strong synaptic connections* are indeed observed in cortical and hippocampal neurons, where a *log-normal distribution* for the synaptic strength, *i.e.*, a few strong synapses on top of a large number of weak synapses, is reported (Buzsáki and Mizuseki, 2014; Cossell et al., 2015; Lefort et al., 2009; Song et al., 2005). These influential synapses seem to act as the backbone of microcircuits. Therefore, we need to find out the architecture of these strong synapses to reveal the basic motifs of microcircuits. Previous models that link architecture to statistics of neural activity, assume *linear responses* to the synaptic input (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) (but see Curto and Morrison 2019; Ocker et al. 2017b). However, to identify the influential inputs, the nonlinearity of input-output relation does not let us use the *linear response* methods, exclusively designed for weak synapses. There is a recent analytical solution for the Leaky Integrate and Fire (LIF) neuron, which provides the dependency of output spikes on *arbitrary synaptic input* of interest, while the effect of many weak synapses accumulates as noisy background inputs, balanced near the spiking threshold. It certifies that a strong synaptic input results in a very different and nontrivial response, compared to the weak/moderate inputs (Shomali et al., 2018).

In this study, using the aforementioned analytic solution (Shomali et al., 2018), we investi-

gate the problem of network identification from observed correlations among spiking activity of neurons. We look at the simplest scenario: The experimentalist records spiking activity of three neurons in vivo (e.g., (Ohiorhenuan et al., 2010)), while s/he cannot directly reveal any synaptic 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 them all? Are shared inputs excitatory or inhibitory? And finally, does either of the three observed neurons make any direct synaptic connection to another of them?

We obtain *pairwise* and *triple-wise* interactions (Amari, 2009a; Nakahara and Amari, 2002) in the simultaneous spiking activity of three LIF neurons, and compare them with experimentally observed results of Ohiorhenuan et al. in monkey V1. They found significant positive pairwise and negative *triple-wise interactions* for spatially close neurons (Ohiorhenuan et al., 2010; Ohiorhenuan and Victor, 2011). Negative triple-wise interactions, observed in cortical and hippocampal neurons (Ohiorhenuan et al., 2010; Shimazaki et al., 2015; Yu et al., 2011), indicate a significantly higher probability of simultaneous silence among three neurons than those expected from their rates and pairwise correlations. Intuitively, excess simultaneous silence can be induced by suppression of these neurons caused by common inhibitory inputs. However, the analytical input-output relation quantitatively reveals that a non-intuitive architecture of *common excitatory inputs, each shared by a pair of neurons* (excitatory-to-pairs), does explain the observed negative triple-wise interactions, and rules out shared inhibition. We investigate the robustness of our results to sub/suprathreshold regimes, directional/recurrent connections among observed neurons, and to adaptative neurons. We confirm many of these results, particularly the significance of excitatory inputs to pairs' motif, remain intact. Plotting triple-wise versus pairwise interactions, for all basic motifs, we analytically provide a 2D guide map; each motif occupies its own region there. This guide map can be used to identify the hidden motifs underlying observed interactions found in empirical data.

Results

Spike probability of a leaky integrate-and-fire neuron: Near threshold regime

First, we introduce the statistical properties of our cortical neuron model operating under in-vivo like conditions (Shomali et al., 2018). We evaluate the probability of spiking within a given time window; it becomes the building block to construct the population activity of such neurons. To this end, we begin with a Leaky Integrate and Fire (LIF) *postsynaptic neuron* with membrane's time constant of τ_{m}, and a resting potential of V_{r}:

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

The neuron spikes when its membrane potential, $V(t)$, hits the spiking threshold, V_θ; then $V(t)$ resets to V_r. The input current of $I(t)$, also, consists of two parts: (a) a transient signaling input which represents the input from the influential synapses with arbitrary strength, $\Delta I(t, A, \tau_b)$, and (b) the effect of all other independent presynaptic inputs accumulated as a *fluctuating background input*, $I_0(t)$:

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

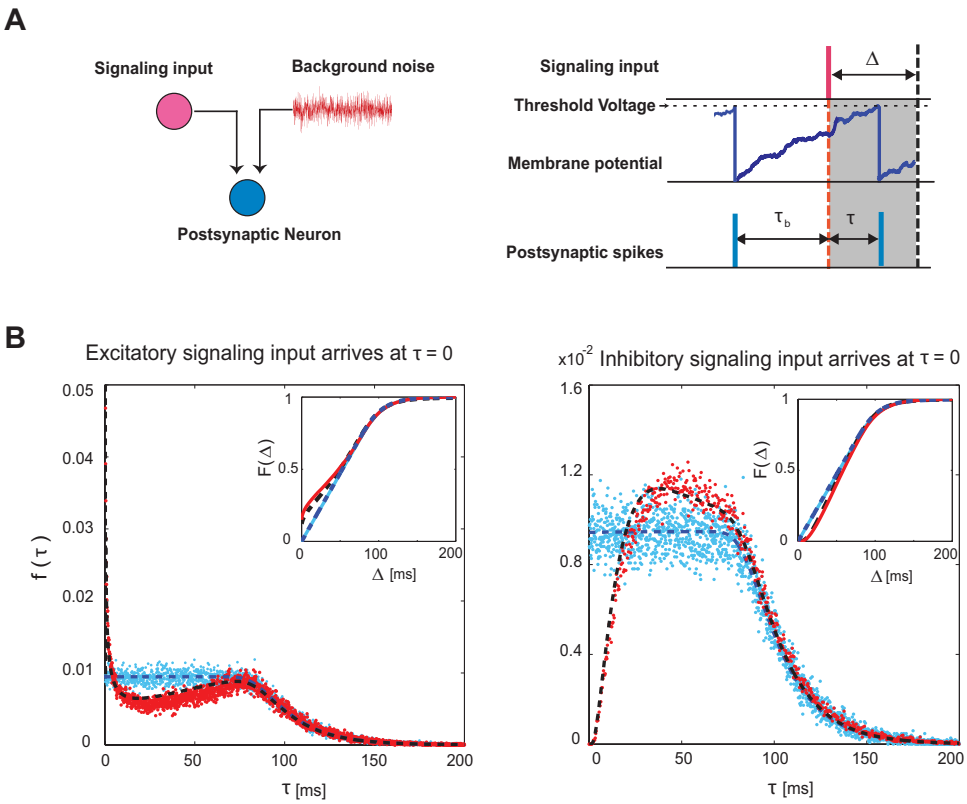

Fig 1. Analysis of a postsynaptic neuron receiving signaling input on top of background noise. (**A**) Left: A schematic model of a postsynaptic neuron driven by background Gaussian noise and a transient signaling input. Right: Timing of postsynaptic spikes and the arrival time of the signaling input. The postsynaptic neuron generates a spike (blue tick) then its membrane potential resets. While the potential rises, a signaling input (red tick) arrives at τ_b after the last spike, which changes the trajectory of the membrane potential. 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. 11) of excitatory (left) and inhibitory (right) signaling input. The simulation results (red dots) and the analytical solutions (dashed black line, Eq. 11) are well matched. The same results for a signaling input with zero amplitude are depicted. Inset: The probability of spiking within Δ after signaling input arrival, Eq. 12. The parameters are: $V_\theta = 20\,\mathrm{mV}$, $\tau_\mathrm{m} = 20\,\mathrm{ms}$, $A = 5\,\mathrm{mVms}$, and the diffusion coefficient is $D = 0.74\,(\mathrm{mV})^2\mathrm{ms}$. The values of the parameters are chosen from the physiologically plausible range (McCormick et al., 1985).

We model the fluctuating background input as Gaussian white noise, so it well replicates the synaptic inputs to V1 neurons when a visual stimulus is presented (Tan et al., 2014). Conclusively, $I_0(t)$ has a mean drive of $\bar{I}$ and a variance of $2D/\tau_\mathrm{m}$; here the diffusion coefficient of D

measures $I_0(t)$'s level of noise. The signaling input, $\Delta I(t, A, \tau_b)$, is characterized by its amplitude (or efficacy), A; and its arrival time, τ_b.

The fluctuating $I_0(t)$, is one important source of *variability*; its *stochastic* nature, however, makes it impossible to solve Eq.(1) and find the exact spike-time *deterministically*. Thus, people try to address the *probability of spiking* (Brunel and Hakim, 1999; Gerstner et al., 2014a; Shomali et al., 2018). Their essential mathematical tool is the *Fokker-Planck* (or diffusion) equations (Risken and Eberly, 1985), which addresses the probability density that postsynaptic neuron spikes at time t, given it had a known value of membrane potential at the initial time of t_0. However, even in the absence of any signaling input, the corresponding Fokker-Planck equation has not been solved yet. There exists an analytical solution but for a very specific case $\bar{I} = V_\theta$ (Bulsara et al., 1996; Wang and Uhlenbeck, 1945), which is known as the threshold regime representing a physiologically plausible situation for in vivo neurons. Recently, Shomali et al. were able to extend that analytic solution for spike density when signaling inputs arrive on top of background noise (Shomali et al., 2018). They considered a near threshold neuron, $\bar{I} \simeq V_\theta$, that receives a transient signaling input (*i.e.*, the synaptic time constant of $\tau_{\rm s}$ is smaller enough than the membrane time constant, $\tau_{\rm m}$). They solved the Fokker-Planck equation, and *analytically* found the probability density of spiking (known also as Inter-Spike Interval distribution, ISI) for arbitrary strength and shape of the signaling input.

Using that framework, we assess the effect of the signaling input on the activity of postsynaptic neurons (Fig. 1A, right). We ask two successive questions: First, what is the probability density of a spike occurrence at time τ after signal arrival, $f(\tau)$? And second, what is the probability of observing one or more spikes in a time window of Δ, after the occurrence of presynaptic signal with a strength A, $F_A(\Delta)$? Figure 1 depicts our results for these two questions (see Methods for the detailed derivation of $f(\tau)$ and $F_A(\Delta)$). Particularly, $F_A(\Delta)$ is the predictable quantity that relates the neuron model to the observed spiking activity of neurons.

Figure 1B shows the marginalized spiking density at time τ after signaling input arrival for the square shape input (Eq. 9, dashed black lines). Early spiking, i.e., small τ, after the arrival of excitatory (or inhibitory) input is much more (or less) probable, compared with no-signaling input case (dashed blue lines). However, the spiking densities with and 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. 1B insets) with and 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 a large time window.

Using $F_A(\Delta)$, one can find the probability of various spiking patterns for multiple neurons receiving common signaling inputs, under the assumption that the neurons are conditionally independent. This calculation is described in the next section.

Spike density of in-vivo LIF neurons can be used to model population activity driven by common inputs

We provide a framework for computing the statistical properties of the population activity of the LIF neurons that receive common excitatory or inhibitory inputs (Fig. 2). We investigate both pairwise and triple-wise interactions among neurons while they are receiving common inputs

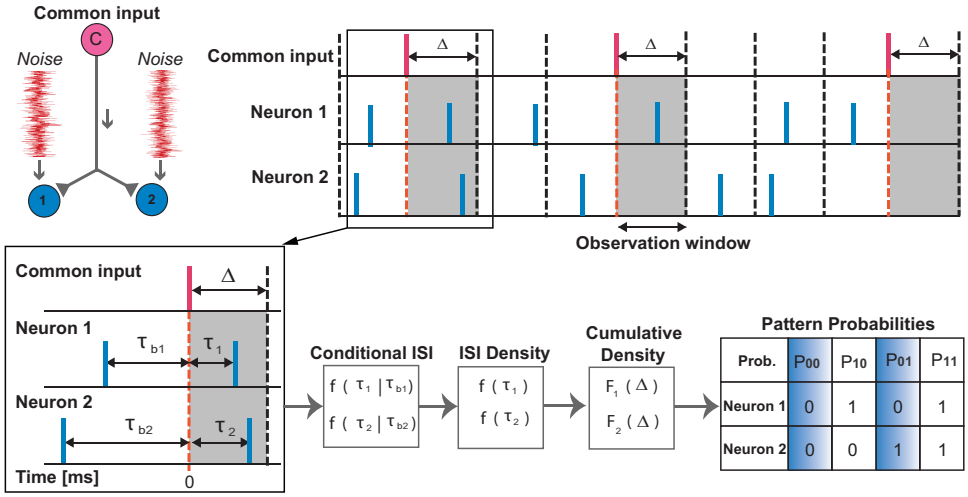

Fig 2. Analysis of two neurons receiving common signaling input on top of background noise. Top, left: A schematic model of two postsynaptic neurons (Neuron 1 and Neuron 2, blue circles) driven by independent noise and a common signaling input (pink circle). Top, middle and right: Spike trains of two postsynaptic neurons (blue spikes) receiving common signaling inputs (red spikes) with rate λ. The gray shaded area of width Δ indicates the 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). Bottom, middle: The conditional ISI density after input arrival is calculated by Eq. 10. Marginalizing over previous spike (τ_b), one obtains the probability of spiking after input arrival (ISI density, Eq. 11). The next step is to calculate the cumulative distribution function (Eq. 12) that is the probability of having one or more spikes within window Δ. Bottom, right: Based on the cumulative distribution function and the fact that neurons are conditionally 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 the occurrence of at least one spike within the Δ time window, whereas the '0' represents the silence of the neuron within this window.

on top of noise in basic possible motifs. Firstly, we consider only two neurons for simplicity, whereas the framework will be extended to three neurons in the next section. Suppose that the two postsynaptic neurons receive a common signaling input in addition to independent background noise. Figure 2 illustrates the timing of the postsynaptic spikes before and after the common input arrival. By binning their spike sequences with a Δ time window (Fig. 2, middle, and left), we compute the 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 $x_i = 1$ means that the ith neuron emitted one

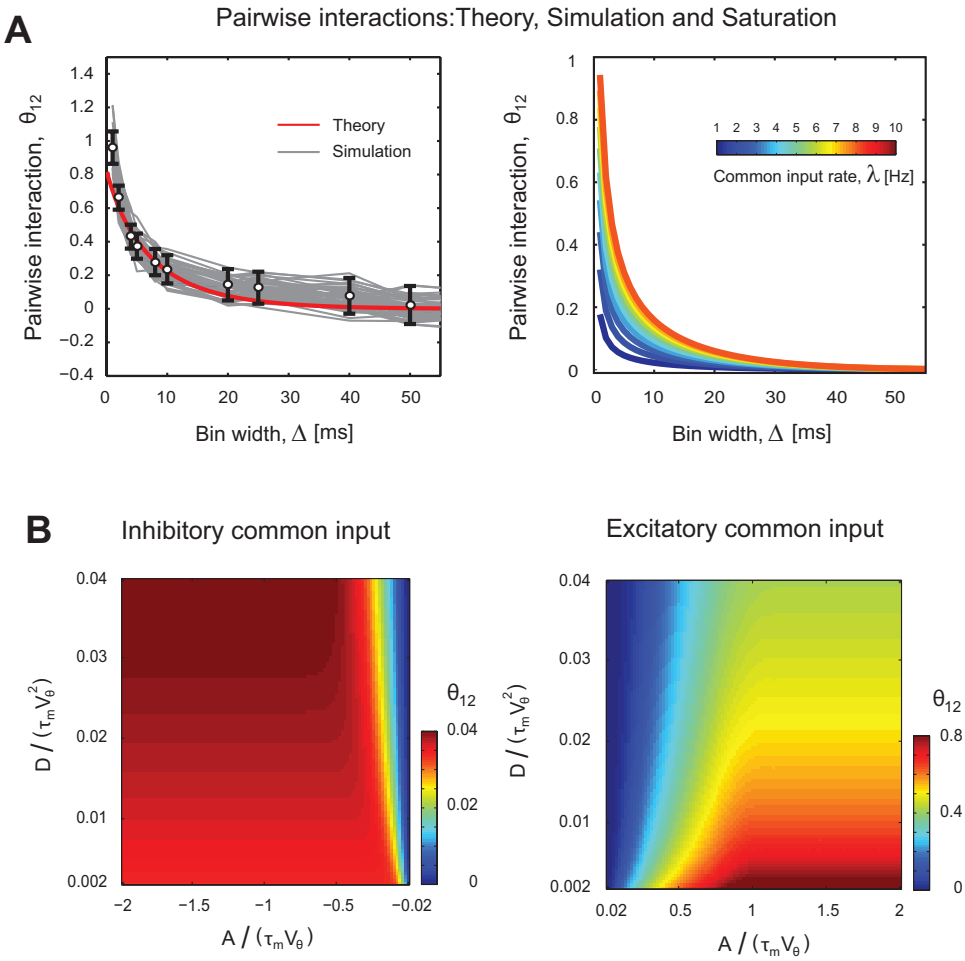

Fig 3. Analysis of pairwise interaction of two neurons receiving common signaling input on top of background noise: (**A**) The interaction (θ_{12}, Eq. 14) of two postsynaptic neurons as a function of bin size, Δ. We use a physiologically plausible range of parameters as: $V_\theta = 20\,\mathrm{mV}$, $\tau_\mathrm{m} = 20\,\mathrm{ms}$, $A = 5\,\mathrm{mVms}$, and the diffusion coefficient $D = 0.74\,(\mathrm{mV})^2\mathrm{ms}$ (McCormick et al., 1985). 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) compared with the analytic result of the mixture model (red line, Eq. 14) for common input rate $\lambda = 5\,\mathrm{Hz}$. Right: Analytical value of θ_{12} as a function of Δ for different common input rates, λ. (**B**) The pairwise interaction as a function of the scaled diffusion coefficient, $D/(\tau_\mathrm{m}V_\theta^2)$, and the shared signal strength, $A/(\tau_\mathrm{m}V_\theta)$, for excitatory (right) and inhibitory (left) common inputs, with $\lambda = 5\,\mathrm{Hz}$ and $\Delta = 5\,\mathrm{ms}$.

or more spikes in the bin, while $x_i = 0$ means that the neuron is silent.

To formally investigate the neuronal correlation, we use the information-geometric measure of interaction, θ_{12} (Amari, 2009b; Martignon et al., 2000; Nakahara and Amari, 2002;

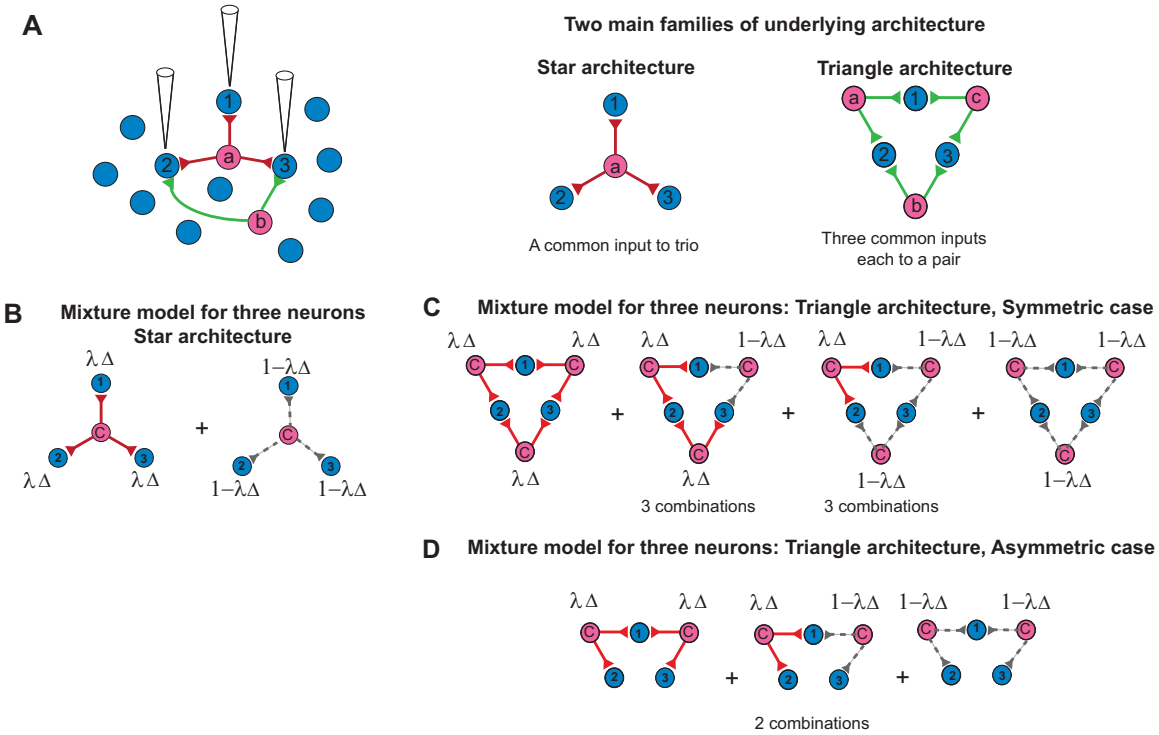

Fig 4. (**A**) A schematic of simultaneous recording of three postsynaptic neurons. The neurons (blue circles) operate independently in the absence of any *common input*. A common input (pink circle) can be an input to three of them (red connections), or at least two of them (green connections). Right: Assuming symmetric architectures, we have two main families: One common input to trio (star architecture, left), or three common inputs each to a pair (triangle architecture, right). (**B**) A mixture model for three neurons in the star architecture is composed of two conditions in which the neurons receive the common input (red lines) with probability $\lambda\Delta$ and they do not receive it (gray dashed lines) with probability $1 - \lambda\Delta$. (**C**) A mixture model for three neurons with the symmetric triangle architecture is composed of four distinct conditions that arise by the combination of the presence or absence of each common input (see main text). (**D**) For asymmetric triangle architecture, the number of possible cases is reduced to three.

Tatsuno and Okada, 2004) (Eq. 14 in Methods). This information-geometric measure of the correlation (SVI) is recommended over the classical covariance or correlation coefficient (Amari, 2009a; Martignon et al., 2000; Nakahara and Amari, 2002; Tatsuno and Okada, 2004) because this measure is not affected by the estimated firing rates of neurons while other measures are influenced by firing rates (i.e, this method is invariant or orthogonal to firing rates in terms of the Fisher metric, see SVI). For two neurons, the probabilities of the activity patterns can be constructed using the probability of spike occurrence in a bin Δ (Eq. 12, Fig. 2). Since two neurons receive common inputs, they are conditionally independent. Therefore, the probabilities

of activity patterns in a bin are given by:

$$P_A(x_1, x_2) = P_A(x_1)P_A(x_2) = \prod_{i=1}^{2} F_A(\Delta)^{x_i}(1 - F_A(\Delta))^{1-x_i}, \tag{3}$$

where A stands for the amplitude of the common input current ($A = 0$ represents the absence of common input). The rate of common input (λ) 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: Neurons either receive ($A \neq 0$) or do not receive ($A = 0$) common input. Hence the probabilities of the activity patterns are given by

$$P(x_1, x_2) = (\lambda\Delta)P_A(x_1, x_2) + (1 - \lambda\Delta)P_0(x_1, x_2). \tag{4}$$

From this probability mass function, Eq. 4, one can compute the neuron's pairwise interaction, denoted as θ_{12} (Eq. 14). This mixture model provides approximated probabilities for the activity patterns of the LIF neurons. This is an approximation of the actual dynamics because, if a neuron does not spike within Δ [ms] after the common input, the effect of the augmented membrane potential is carried over to the next bin, and then the binary activities are no longer a simple mixture of the two conditions. Such situations should happen often if the bin size is small compared to the mean postsynaptic inter-spike interval. So we test if this approximation predicts the interaction in the parallel sequences of the two LIF neurons, and examine reasonable bin sizes. We compare the pairwise interaction predicted by the mixture model with the simulated spike sequences in Figure 3A. It 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). The result also shows θ_{12} increases with the rate of common input (Fig. 3A, Right). However, if we increase the bin size, the probability of having one or more spikes within Δ increases and saturates to 1 (Fig. 1B, Inset) regardless of the presence or absence of the signaling input. This means $F_A(\Delta)/F_0(\Delta) \to 1$, which results in vanishing pairwise interaction as the bin size increases. Therefore, we can hardly discriminate between the presence and absence of common input if we use a large bin size.

We examine the pairwise interactions by changing two independent parameters, the scaled amplitude of the signaling input $A/(\tau_{\mathrm{m}}V_\theta)$ and the scaled variability of the noisy background input $D/(\tau_{\mathrm{m}}V_\theta^2)$ (Fig. 3B). As expected, the pairwise interactions are positive for both common excitatory and inhibitory inputs. However, the interactions are significantly weaker in the inhibitory case. This indicates that it is difficult to observe the effect of common inhibitory input for this range of postsynaptic firing rates, and that the strong pairwise interactions are the indicator of having common excitatory inputs.

For each value of normalized diffusion coefficient (level of inputs' noise), $D/(\tau_{\mathrm{m}}V_\theta^2)$, there exists a critical normalized amplitude for common *excitatory* input, $A/(\tau_{\mathrm{m}}V_\theta) \sim 1$. Upward this critical value, the postsynaptic neuron's spiking density, and consequently pairwise interaction, does not change anymore (Fig. 3B, right). The saturation value of pairwise interaction is inversely correlated with the normalized diffusion coefficient; since higher normalized diffusion coefficient (level of the noise) disperses the voltage of the membrane, the probability of spiking after common input arrival decreases. Similar behavior for θ_{12} is observed for inhibitory

input (Fig. 3B, left). Nevertheless, in contrast to the common excitatory input case, pairwise interaction and $D/(\tau_{\mathrm{m}} V_\theta^2)$ are directly correlated.

Higher-order interaction of three neurons depends on types of common inputs and network architecture

We now extend the analysis of neural interactions to three neurons. The motivation to investigate the interactions among three neurons comes from experimental studies (Ohiorhenuan et al., 2010; Ohiorhenuan and Victor, 2011) that investigated the activities of three neurons simultaneously (Fig. 4A). For two neurons, there is one possible shared input's architecture: A common input to both of them. For three neurons to induce correlation among them, however, it can be either (i) a shared input among three of them (red connections in Fig. 4A), or (ii) one or more shared inputs to each pair among them (green connections). Assuming symmetry, the former one leads to a *star architecture* or a "common input to trio" (Fig. 4A, middle) while the latter one makes a *triangle architecture* or "common inputs to pairs" (Fig. 4A, right).

For investigation the neuronal correlation among three neurons, we use the information-geometric measure of a triple-wise interaction, θ_{123} (Amari, 2009b; Martignon et al., 2000; Nakahara and Amari, 2002) (Eq. 16 in Methods). As the pairwise interaction extracted a pure interaction of two neurons, this triple-wise interaction measure is also not affected by firing rates and joint firing rates of two neurons, and extracts a pure triple-wise effect that can not be inferred from the first and second-order statistics of the population (SVI). There are two basic motifs that can induce triple-wise interaction among neurons (Fig. 4A, middle and right) as described below.

I. Common input is given to three neurons: Star architecture

In star architecture, three neurons receive a single common signaling input simultaneously (Fig. 4B). The conditional probability of activity patterns when a common input generates a spike given to all three neurons with probability $\lambda\Delta$ (red lines) is $P_A(\mathbf{x}) = \prod_{i=1}^{3} F_A(\Delta)^{x_i}(1 - F_A(\Delta))^{1-x_i}$, where $\mathbf{x} = (x_1, x_2, x_3)$ is the spiking activity for three neurons. Similarly, the probability mass function for three neurons receiving no common input with probability $1-\lambda\Delta$ (gray dashed lines), is obtained by $P_0(\mathbf{x}) = \prod_{i=1}^{3} F_0(\Delta)^{x_i}(1 - F_0(\Delta))^{1-x_i}$. Thus, we model spike occurrence as a mixture of the two conditions in which neurons receive and do not receive common input:

$$P(\mathbf{x}) = \lambda\Delta P_A(\mathbf{x}) + (1 - \lambda\Delta)P_0(\mathbf{x}). \tag{5}$$

From this probability mass function, we can compute the triple-wise interaction of three neurons according to Eq. 16.

II. Common inputs are given to pairs of three neurons: Triangle architecture

In triangular architecture, we assume that each pair of three neurons receives common signaling input from an independent presynaptic neuron with frequency λ (Fig. 4C). The first, second, and third common inputs project to neurons 1 and 2, neurons 2 and 3, and neurons 1 and 3, respectively (symmetric case). The three common inputs are independent, and occur with equal frequency, λ. By taking into account the occurrence probabilities, one could obtain the mixture model. The resulting mixture models are given in Methods, which also include the asymmetric common input architecture in which there are only two common inputs out of three (asymmetric case) (Fig. 4D).

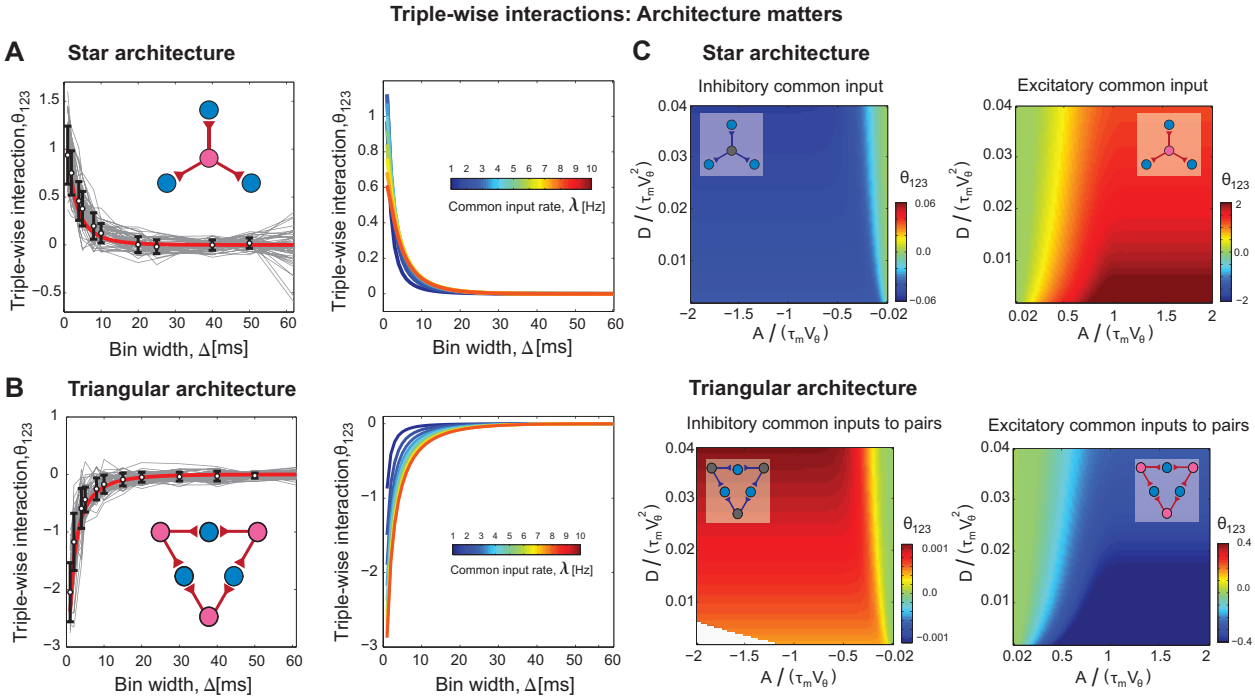

Fig 5. Comparison of the *triple-wise interaction* of 3 LIF neurons, θ_{123}, in the two leading architectures: (**A**) The interaction when three postsynaptic neurons, blue ones, are in a *star architecture* receiving a common excitatory signal $A = 5\,\mathrm{mV}$ ms, simultaneously. The parameters 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) compared with the *analytic* result of the mixture model (red line, Eq. 16) for $\lambda = 5\,\mathrm{Hz}$. **Right**: Analytical value of θ_{123} as a function of the bin size, 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 shares an independent common excitatory input. All parameters are as in (A). (**C**) Triple-wise interactions of 3 neurons receiving common excitatory or inhibitory inputs under star (top panel) and triangular (bottom panel) architectures. θ_{123} is represented as a function of *scaled diffusion coefficient*, $D/(\tau_\mathrm{m}V_\theta^2)$, and *scaled shared signal strength*, $A/(\tau_\mathrm{m}V_\theta)$; the other parameters are: $\Delta = 5\,\mathrm{ms}$ and $\lambda = 5\,\mathrm{Hz}$. The white region in the bottom left panel shows the numerically indeterminate region due to a very small diffusion coefficient (level of noise) and strong inhibition. The right-hand side figures are for excitatory and the left-hand side ones are for common inhibitory inputs.

The triple-wise interaction parameters computed from the simulated spike sequences of postsynaptic neurons are compared with the theoretical predictions, using the mixture models (Fig. 5A and B, left). The activities of neurons that receive a simultaneous common excitatory input (star architecture) are characterized by positive triple-wise interactions (Fig. 5A, left)

whereas the 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 that triple-wise interaction decreases as the bin size increases for the same reason given for pairwise interaction (Fig. 3A, right). The dependency of triple-wise interaction on the common input rate is also shown in the right panels in Figure 5A and Figure 5B.

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 (level of input noise) $D/(\tau_{\mathrm{m}} V_\theta^2)$ and scaled amplitude $A/(\tau_{\mathrm{m}} V_\theta)$. A single common excitatory input, in the 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 θ_{123} in Eq. 16. In the triangular architecture, with common excitatory input, however, each common input causes postsynaptic spikes for two neurons and does not drive the other one. This primarily increases $P(1,1,0)$ (or any of its permutations) and attenuates the fraction (Eq. 16) that makes θ_{123} negative. For common inhibitory input, the probability of the reversed pattern, $P(0,0,1)$ (or any permutations) increases; this results in a larger numerator in Eq. 16 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.

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

The 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 6A shows the first-order parameter, θ_1^{t}, in a star or triangular architecture receiving either common excitatory or inhibitory inputs. The θ_1^{t} strongly depends on $D/(\tau_{\mathrm{m}} V_\theta^2)$, which measures the level of the noise in background inputs, but shows a weak dependence of the signal's amplitude, $A/(\tau_{\mathrm{m}} V_\theta)$. More importantly, it does not show any conclusive dependence, neither on the choice of architectures nor type of common input. Thus, it is not possible to identify the underlying architecture nor the type of common input using the first-order parameter only.

Nonetheless, the 2D plane of θ_{123} versus θ_{12} does differentiate motifs, *c.f.* Fig. 6B; there, each motif clearly occupies its *distinct region*. Figure 6B shows $5 = 2 + 3$ motifs in the θ_{123} versus θ_{12} plane. Two inhibitory motifs (*triangle* and *star*) occupy tiny areas; thus are expressly shown in two panels on its left (top and bottom). The three excitatory motifs cover much wider areas. Here we have also considered the *asymmetric excitatory-to-pairs*; it is the only other architecture with shared excitatory inputs, which can produce non-zero θ_{123}. All five regions initiate from origin, i.e., $\theta_{12} = \theta_{123} = 0$; simply because both interactions vanish for zero signal's amplitude, i.e., $A/(\tau_{\mathrm{m}} V_\theta) = 0$. As we increase the signal's strength, both θ_{12} and θ_{123} deviate from zero.

The example of excitatory-to-trio is the most visible case (SI, for comparable trends in other architectures). Consider postsynaptic neurons with the fixed spontaneous rate of $\mu = 1$Hz (dashed-dotted purple curve); this curve shows how interactions change as we increase the shared

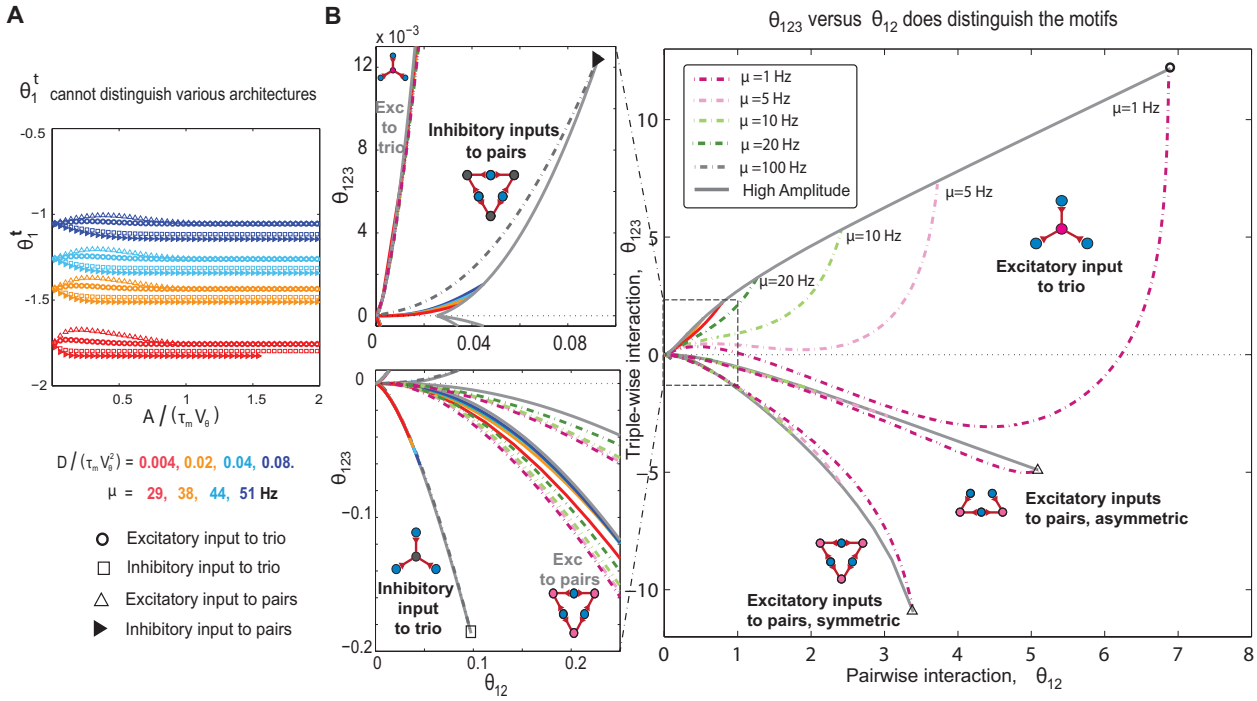

Fig 6. Population activity of three neurons caused by hidden common inputs occupies distinct regions in the plain of interactions, depending on the architecture and synaptic types. (**A**) The natural parameter for individual neurons, θ_1^t, in *star* and *triangular* architectures with common *excitatory* or *inhibitory* inputs (four cases, shown by four symbols) versus *scaled* shared signal strength $A/(\tau_{\mathrm{m}}V_\theta)$. The θ_1^{t} significantly varies with $D/(\tau_{\mathrm{m}}V_\theta^2)$, which measures *background spontaneous activity*, μ. However θ_1^{t} does not show any significant dependence on the shared signal strength, nor any conclusive dependence on the type of architecture. (**B**) In the plane of the triple-wise versus pairwise interactions, the star and triangular architectures with common excitatory or inhibitory inputs are distinguishable from each other. The regions associated with motifs of *excitatory common* input are expressly shown: From top to bottom, they attribute to *excitatory-to-trio*, *asymmetric excitatory-to-pairs*, and *symmetric excitatory-to-pairs* respectively. The regions associated with inhibitory common input are confined to around origin (i.e., small θ_{123} and θ_{12}), hence are shown with higher resolution in left panels. Each region is bounded by two analytic *boundary lines*; the gray solid lines attribute to the high signal amplitude, i.e., $|A|/(\tau_{\mathrm{m}}V_\theta) \gg 1$, see SI; while the purple (gray) dashed boundary lines for motifs with excitatory (inhibitory) common input attribute to low (high) diffusion limit. The low (high) diffusion limit corresponds to background spontaneous activity of $\mu = 1$ Hz ($\mu = 100$ Hz), see SII. We choose these limits to cover a wide range of postsynaptic spontaneous activity while maintaining the assumption of low activity rates: $\mu\Delta \simeq F_0(\Delta) \leq 0.5$, see SI. The colored solid lines are numerical results of the LIF neuron model for different scaled diffusions and signal amplitudes (Fig. S1). The fixed parameters are the bin size of $\Delta = 5\,\mathrm{ms}$ and the presynaptic rate of $\lambda = 5\,\mathrm{Hz}$.

signal's amplitude from zero to highest conceivable values, i.e., $A/(\tau_{\rm m}V_\theta) \gg 1$. The pairwise interaction monotonically increases with the signal's strength; but eventually saturates at its maximum of $\theta_{12} = 6.87$, see the hollow black circle. The triple-wise interaction, however, shows a *non-linear behaviour*: It initially increases to $\theta_{123} = +0.34$, then decreases to the negative value of -3.06, and finally increases to its saturation value of $\theta_{123} = +12.11$, again the hollow black circle. We analytically show that for any choice of the spontaneous firing rate μ, we reach its corresponding saturation point, at the high enough signal's strength (SI). The position of each saturation point (i.e. its θ_{12} & θ_{123}) is found to be independent of the *neuron model*, as well as the *near the threshold* assumption. Conclusively, the saturation points (thick gray curve) form a *universal upper boundary* in the θ_{123} versus θ_{12} plane; the corresponding point for any *excitatory-to-trio* motif is placed below it (Fig. 6B). To address the *lower boundary*, we have limited ourselves to spontaneous rates $\mu \geq 1$Hz. As seen in Fig. 6B, for any higher value of the background activity ($\mu > 1$Hz), the corresponding curve appears above the mentioned curve for $\mu = 1$Hz, yet below the upper boundary of saturation points. Thus, practically, the curve for $\mu = 1$Hz acts as the lower boundary.

A similar story holds for the other four motifs. Each corresponding region is composed of a bunch of curves. To obtain each curve, we assume a certain value of postsynaptic spontaneous rate, μ, then let the shared signal's amplitude vary from zero to very high values. This produces a curve that initiates from the origin and ends at its saturation point. Each region is the accumulation of all these curves, and has two boundaries: One boundary is composed of all saturation points (thick-gray boundary), while the other is the curve with the lowest firing rate of $\mu = 1$Hz (highest firing rate of $\mu = 100$Hz), for motifs with excitatory (inhibitory) shared inputs, see Fig. 6B.

One significant question is how much the obtained boundaries vary with change in the *neuronal model* or the *near threshold regime* assumption. Can one region entirely displace, or even two *distinct regions* overlap? Fortunately, the high amplitude boundaries (thick gray curves) are *analytically* verified to be independent of the neuron model, as well as near threshold assumption (see SI). The other boundaries of low (high) spontaneous rate, for the excitatory (inhibitory) shared inputs, however, have a non-trivial behaviour. For the *star* architectures, they remain independent of neuron model; while for the *triangle* architecture, they do depend on the choice of neuron model, and the near threshold assumption (see SII). This dependence is actually an important example of non-linearity of input-output relation: It wouldn't exist, if increasing the strength of presynaptic signal would linearly increase the probability of postsynaptic spike (technically, $F_A(\Delta) = F_0(\Delta) + cte \times A$, see SII). However, this is only correct for *week signals*. In general, the probability of postsynaptic spike *non-linearly* varies with signal's strength and saturates for very strong signals; the accurate description of this dependency requires a full knowledge of neuronal model (SII).

Figure 6 quantitatively answers the question we asked at the beginning of this part: It is possible to identify the underlying architecture and even the type of shared inputs (excitation or inhibition) for three homogeneous neurons, simply investigating their interaction parameters. As a practical example, we consider a careful study on V1 neurons of *anesthetized macaque monkey*; it investigated the relationship between the triple-wise interaction (Eq. 16) of three neurons (ordinate), and the average marginal pairwise interactions (Eq. 14) of neuron pairs in the group (abscissa) (Ohiorhenuan et al., 2010). They used extracellular recording of pyramidal

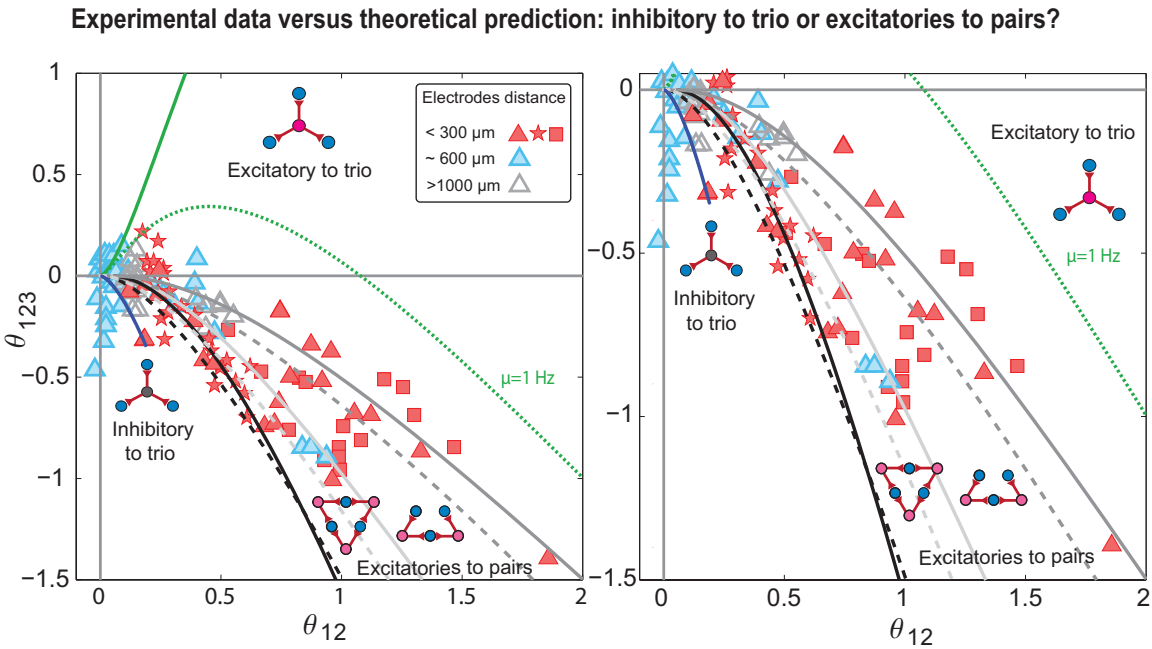

Fig 7. Comparison of theoretical predictions about neural interactions with experimental evidence extracted from (Ohiorhenuan and Victor 2011, Figure 4b). Ohiorhenuan et al. recorded the spike data from V1 neurons of macaque monkeys using tetrodes and analyzed the relation between triple-wise interaction (Eq. 16) of three neurons (ordinate) and an average marginal pairwise interaction (Eq. 14) of neuron pairs in the group (abscissa). Red and blue filled symbols represent the interactions of neurons within 300 and 600 μm vicinity respectively while unfilled grey symbol shows the interaction at > 1000 microns distance. The black lines show the region of interactions theoretically obtained by assuming triangular architecture with common excitatory inputs to each pair of neurons. The solid black line is the high amplitude limit for interactions while the dashed black line is the boundary for the low spontaneous firing limit ($\mu = 1$ Hz, low diffusion limit). Similarly, the dark gray solid and dashed lines determine the boundaries of region for triangular asymmetric architecture with common excitatory inputs given to two pairs among three neurons. The light gray lines are the same asymmetric excitatory inputs to pairs but all three pairwise interactions are taken into account for averaging while for the dark gray graphs, only nonzero pairwise interactions (for one pair, $\theta_{12} = 0$) are considered. The blue analytic lines represent a narrow region for star architecture with common inhibitory input (high common input's amplitude limit and high diffusion limit) whereas the orange lines show the wide region for excitatory inputs given to three neurons (dashed, low spontaneous activity limit, $\mu = 1$ Hz for scaled diffusion limit of $D/(\tau_m V_\theta^2) = 2 \times 10^{-19}$ and solid, high common input's amplitude limit). The fixed parameters are $\Delta = 10$ ms and common input's rate, $\lambda = 5$ Hz. The data reveals that nearby neurons (red filled symbols) receive pair of excitatory inputs in the triangular architecture rather than inhibitory or excitatory inputs in the star architecture.

neurons and found that many neurons, with mutual separations less than 300 μm, exhibited positive pairwise and strong negative triple-wise interactions (Fig. 7). The triple-wise interactions weaken as the electrodes' separations increase above 600μm. They attributed the observed *strong negative triple-wise* at near distances, to the hidden activity of small GABA-ergic inhibitory neurons, which presumably provide shared input to the larger excitatory pyramidal cells (Ohiorhenuan and Victor, 2011).

Fortunately, the spontaneous activity of V1 neurons are reported within 10-70Hz (Gur et al., 1997; Ohiorhenuan et al., 2010; Ohiorhenuan and Victor, 2011), higher enough than the 1Hz lower boundary which we considered for the excitatory-to-trio motif. Thus, we can safely compare our theoretical predictions with the empirical observations. Figure 7 clearly shows that empirical data coincides with regions associated with the motif of excitatory-to-pairs, neither excitatory-to-trio nor any of the inhibitory motifs. It clearly rules out, the initial intuitive picture that shared inhibition induces the observed strong negative triple-wise interactions.

Inevitably, the following question emerges: Why should the observed strong negative triple-wise interaction, be associated with excitatory common inputs, and the inhibitory shared inputs fail to produce any strong negative interactions? We will answer this question at the end of the results section. However, we firstly verify the robustness of our excitatory-to-pairs scenario, to the possible *complication of the motifs* due to recurrent interconnections, plausible change in *neuron model*, and even violation of *near threshold assumption*.

Excitatory directional/recurrent connections among three neurons can explain the observed negative triple-wise interaction

Although the previous section's analysis seems valid for common inputs' architecture among three postsynaptic neurons, the question arises whether considering interconnections among the three neurons can induce any change to our concluding result in Figure 7. We, therefore, simulate all possible motifs that have directional or reciprocal connections, among the three neurons. At first glance, the number of motifs for three neurons is $2^6 = 64$ (each directed connection can be present or absent; so for 6 possible interconnections, it yields 2^6). However, some of these motifs are structurally the same; they turn into each other, simply with changing (i.e. permuting) labels of three postsynaptic neurons. This means that 64 possible motifs could be further categorized in 16 main structures (Fig. 8). Regarding 16 main structures, we ask whether the inhibitory-to-trio with the help of either of them can reach the value of experimental data (red symbols in Fig. 7). Figure 8 shows the result of triple-wise interaction for each motif averaging over $50,000$ runs. We see four clusters of motifs, which are ordered as the number of inputs to pairs increases. The first cluster (blue, motifs 1 to 7) contains motifs with no simultaneous input from one excitatory neuron to two others (to pairs). For this cluster, the average triple-wise interaction is a small value. The second cluster (red, motifs 8 to 13) has motifs that have one excitatory neuron as the input to pairs of neurons in their architecture. The third and fourth clusters (green, motifs 14 and 15; and black, motif 16) contain architectures with respectively two and three excitatory inputs to pairs of neurons. Clearly, as the number of *excitatory-to-pairs* increases, the absolute value of triple-wise interaction boosts up, Fig. 8. The inset shows triple-wise versus pairwise interactions for these four clusters. This picture is consistent with our finding (shown in Fig. 7) that excitatory input to pairs induce large negative triple-wise

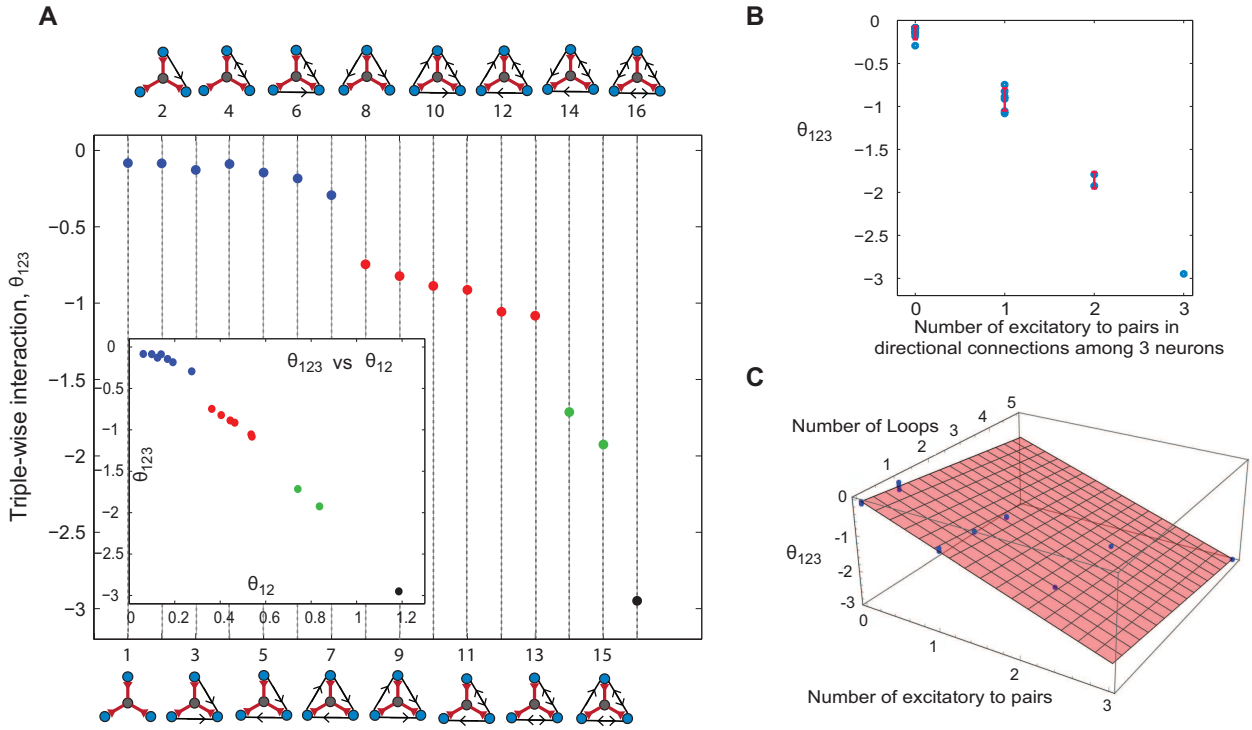

Fig 8. Triple-wise interactions for 16 motifs of directional and/or reciprocal connections among three neurons, when they receive independent noise and common inhibitory-to-trio. (**A**) The architectures are divided into four clusters based on the number of excitatory inputs to pairs in each motif. The first cluster (blue) contains directional connections that do not have any excitatory-to-pairs in their architectures. The second cluster (red) has one excitatory-to-pair in the motifs and the third and fourth (green and black) are related to two and three excitatory-to-pairs in their motifs. Inset shows the triple-wise interaction versus pairwise interaction for all 16 motifs. (**B**) The mean and error bar (standard deviation) of triple-wise interaction for each cluster as a number of excitatory-to-pairs. Clusters are separated from each other. (**C**) Triple-wise interaction for the 16 motifs is a linear function of both the number of excitatory-to-pairs motif and the number of loops. Each motif is simulated 50000 times, and each trial contains 500 seconds of spike trains with the time resolution of 0.05 ms. The time window to calculate the triple-wise and pairwise interactions is $\Delta = 5\,\mathrm{ms}$, and the shape of presynaptic input is a square function for both common input and directional connection's input, the same as analytic calculation. The parameters are: the scaled diffusion $D/(\tau_{\mathrm{m}} V_\theta{}^2) = 9{\times}10^{-5}$, the scaled amplitude $A/(\tau_{\mathrm{m}} V_\theta) = 0.0125$, the common presynaptic input and input rate of directional connectivity $\lambda = 5\,\mathrm{Hz}$, the time delay for directional interconnection $t' = 6\,\mathrm{ms} > \Delta$, and $A'/A = 1$, where A' is the amplitude of the directional connections.

and positive pairwise interactions. The excitatory input to pairs, either as common input or as directional connectivity, can generate such activity and thus is the basic architecture behind the data presented here.

Robustness: How does adaptation modulate the pairwise and triple-wise interactions?

It is a valid question, that how much our suggested framework (Fig.6), as well as its confrontation with the empirical data (Fig. 7), would change if we use a more physiologically plausible neuronal model? Regarding the predicted triple-wise and pairwise interactions, in Fig.6, we analytically prove that the curves for *high signal amplitude* are *independent* of the neuronal model, and the threshold assumption (see SI). This also holds for the curves with excitatory/inhibitory *to trio* architecture. Therefore, a good portion, not all, of the predictions would remain intact with or without a more plausible model. However, since a more physiologically plausible model can modify some predicted interactions for *excitatory-to-pairs* motifs; we have to study it, inevitably. The LIF neuron model is a good reduced model, that can reproduce in vivo spiking activity of neurons (Camera et al., 2004; Rauch et al., 2003). In comparison with data, however, it has some limitations and restrictions (Izhikevich, 2004; Jolivet et al., 2008; Ostojic and Brunel, 2011; Shinomoto et al., 1999). To have a more biologically plausible model, we add *adaptation term* to LIF (Brette and Gerstner, 2005; Gerstner et al., 2014a) and run simulation to see how the result of triple-wise versus pairwise interaction would change (more details about the model are in SIV). The simulation result shows adaptation reduces the firing rate of postsynaptic neuron (Fig. S4 and Fig. S5, SIV). As is explicitly shown at the end of the Results section, the lower background activity amplifies the effect of excitatory common inputs and diminishes that of inhibitory ones. This is why, the simulation results show that in the presence of adaptation, excitatory-to-pairs generates even stronger pairwise and triple-wise interactions, while inhibitory inputs induce weaker interactions (Fig. S6). Therefore, based on experimental evidence for low firing of V1 neurons (Ohiorhenuan et al., 2010), strong negative triple-wise interactions are induced by excitatory inputs to pairs motif, and *adaptation* simply strengthens this picture.

Robustness: How do pairwise and triple-wise interactions change if we go slightly away from the threshold regime?

Alongside the problem of the plausible neuronal model, one can ask if the particular assumption of *near threshold regime* does reduce the scope of validity of our results. Fortunately, as mentioned earlier, many of the boundaries we have found are *independent of neuronal model*, hence independent of the near threshold assumption. Moreover, the particular near threshold assumption is partly asserted by an empirical study (Tan et al., 2014) that shows during sensory stimulation, V1 neurons operate *near the threshold regime* while background noise is uncorrelated.

Nevertheless, we run simulation to see how deviation from the threshold regime, could modify our results (see SIII). Comparing to the threshold regime, the simulations show that in the subthreshold regime, excitatory common inputs produce stronger interactions while inhibitory common inputs produce weaker ones. This trend simply reverses in the suprathreshold

regime (Fig. S3, SIII). Now, considering the particular experimental study on macaque's V1 (Ohiorhenuan et al., 2010), we recognize other evidence that cortical neurons operate in the *subthreshold* (or near the threshold) regime (Shadlen and Newsome, 1998), depending on the state of the animal, and stimulus arrival (Tan et al., 2014). Assigning this fact to Fig.(7), it means that there is even a smaller region of inhibitory-to-trio which could be attributed to empirical data, due to the shift to subthreshold. While there is a larger portion of excitatory-to-pairs region which we can safely attribute to that data. This reaffirms our original conclusion that the observed strong negative triple-wise interactions are signature of excitatory-to-pairs, *exclusively*.

Finally, it is interesting to address why being in subthreshold (suprathreshold) results in increase (decrease) of the higher-order interactions induced by excitatory common inputs, and just do the reverse for interactions induced by inhibitory ones. Being in subthreshold (suprathreshold) mean lower (higher) spontaneous activity of postsynaptic neurons. As we show in the next part, it is the spontaneous activity that well explains all these observations.

Excitation versus inhibition: Which one can produce stronger triple-wise interactions?

So far, we have found that the empirical strong negative *triple-wise* combined with positive *pairwise* interactions, for V1 neurons, are signature of microcircuits with *excitatory common inputs* (Fig. 7). One crucial question is why should other microcircuits with *inhibitory common inputs* have failed to produce such strong negative triple-wise interactions? Can we always attribute strong higher-order interactions to excitatory common inputs, or does it depend on certain features which vary from experiment to experiment?

The measured pairwise and triple-wise interactions depend on various features of the postsynaptic neurons, as well as their possible shared inputs. For the analytically tractable regime of *strong signaling inputs*, however, we can reduce many factors to a few decisive ones. Then, analytical calculations show that, when the spontaneous rate of postsynaptic neurons in time-window Δ is low, i.e., $F_0(\Delta) \ll 1$, the excitatory common inputs can produce large pairwise and triple-wise interactions, while inhibitory common inputs can't (Fig. 9, SI). This picture simply reverses if the spontaneous firing rate of postsynaptic neurons happen to be high, i.e., $F_0(\Delta) \lesssim 1$. There is of course an intermediate regime, $F_0(\Delta) \simeq 0.5$, where the strength of interaction induced by *inhibitory input to trio* and *excitatory inputs to pairs* are nearly the same (Fig. 9).

Figure 10 illustrates how postsynaptic neurons' spontaneous activity, i.e. $F_0(\Delta)$, plays an essential role in relating the hidden underlying architecture with the observed interactions. If the regime of spontaneous rate is known, based on the statistics of neural data (pairwise and triple-wise interactions), one can predict the predominant architecture that induces the observed interactions. In a low spontaneous activity regime, motifs of excitatory inputs can induce strong triple-wise and pairwise interactions (regions in Fig. 10**A**); whereas, in high spontaneous activity, motifs with inhibitory inputs can generate strong interactions (Fig. 10**B**).

In the experiment by Victor and colleagues (Ohiorhenuan et al., 2010; Ohiorhenuan and Victor, 2011) the neuronal firing rates ranged within $10\text{Hz} \leq \mu \leq 70\text{Hz}$ (Fig.4 in (Ohiorhenuan and Victor, 2011)) while the time bin was $\Delta = 10\text{ms}$. The exact spontaneous firing rate of postsynaptic neurons is $F_0 = 1 - \exp(-\mu \times \Delta)$; this yields $0.1 \leq F_0 \leq 0.5$. For such

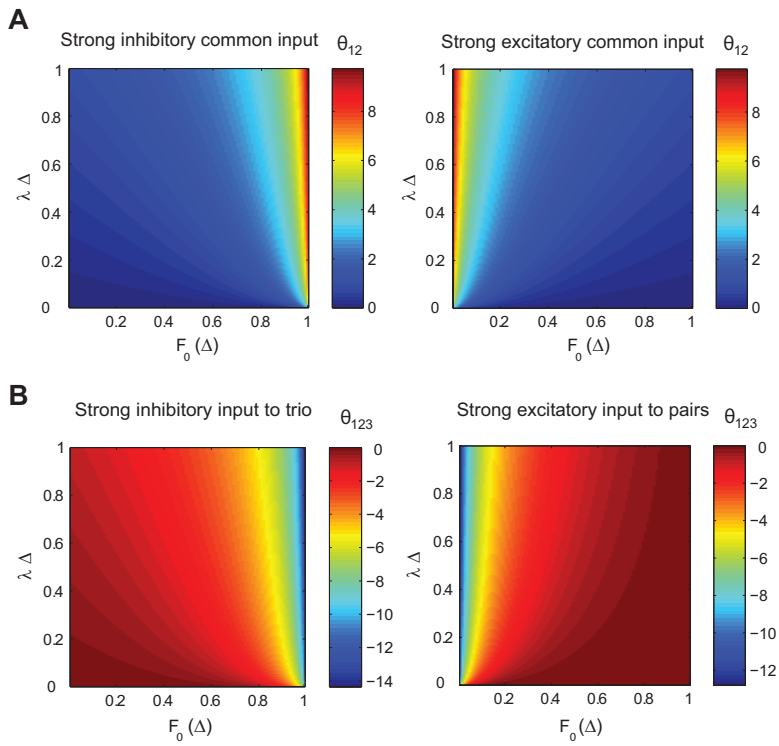

Fig 9. Pairwise and triple-wise interactions for strong excitatory and inhibitory inputs as a function of a spontaneous spike rate of a postsynaptic neuron ($F_0(\Delta)$) and an input rate (λ) in a small time window Δ. (**A**) inhibitory inputs generate strong pairwise interaction if the spontaneous rate of postsynaptic neuron is high (i.e., $F_0(\Delta) \simeq 1$) while excitatory inputs generate such strong interactions in a low spontaneous spiking regime, $F_0(\Delta) \ll 1$. (**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 cannot induce such strong interaction in this regime.

values of F_0, any observation of strong triple-wise interactions is an indication of excitatory common inputs as opposed to inhibitory ones. Furthermore, there is an 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 Fig.4c and 4a in (Ohiorhenuan and Victor, 2011), each set of data that has the smaller firing rate, produces larger pairwise interaction. This observed decrease of pairwise interaction with increasing the spontaneous activity (firing rate), is also revisited for *excitatory-to-pairs* architecture in figure 9-right, and not for the *inhibitory-to-trio*, see Fig. 9-left. This is another fact which reaffirms our original claim that the motif of excitatory inputs to pairs is the architecture behind this set of data.

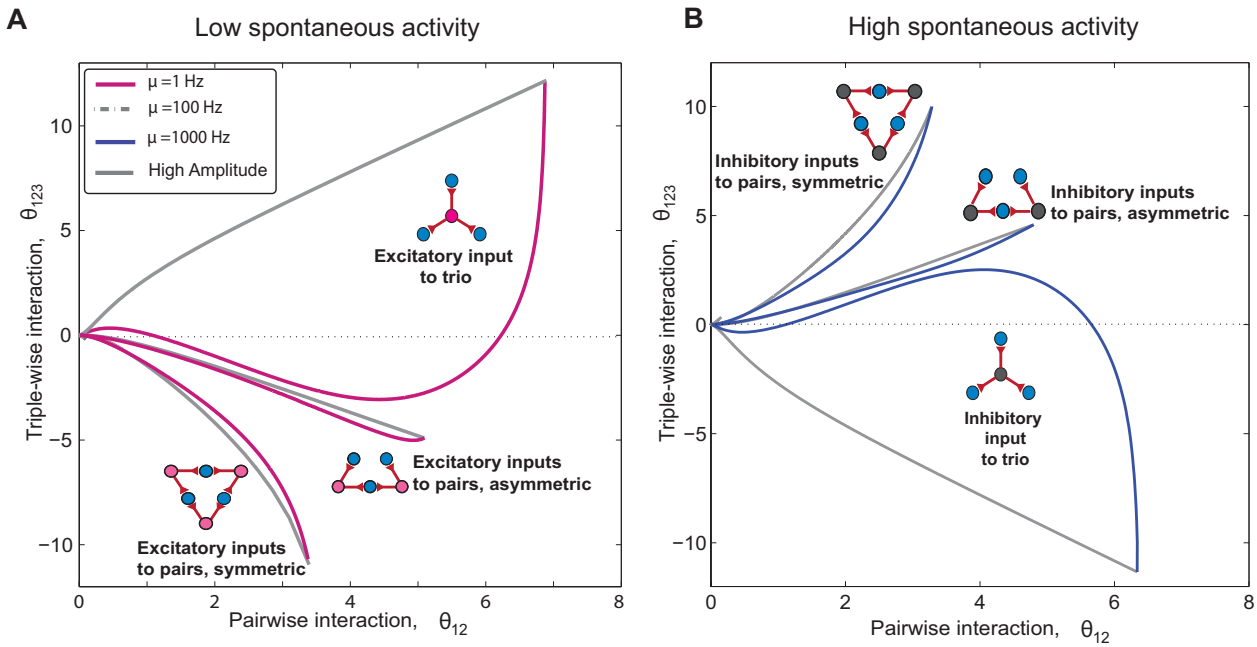

Fig 10. Uncovering the underlying architecture from observed higher-order interactions. **A** For spontaneous activity within the range $\mu = 1 - 100$ Hz, three regions for excitatory inputs' motifs are shown. The strong higher-order interactions regardless of their signs are the sole signature of common excitatory inputs when the spontaneous rate is low. This scenario reverses if the background spontaneous activity is high (for example, olfactory bulb (Burton and Urban, 2015)). **B** The regions for spontaneous activity within the range of $\mu = 100 - 1000$ Hz show that inhibitory inputs can induce strong interactions in a high spontaneous regime. The excitatory inputs' regions for high spontaneous activity (Fig. S2) shrink to a small size compared to the low spontaneous activity regime. The method for drawing the boundaries is similar to Fig. 6 and SII. The fixed parameters are bin size, $\Delta = 5$ ms, and input rate, $\lambda = 5$ Hz.

Discussion

Overview

Our results point to the possibility of revealing the underlying neuronal architectures and the type of common input by using pairwise and triple-wise neural interactions (Fig. 10). Furthermore, for a specific set of empirical observations (Ohiorhenuan et al., 2010) in comparison with our analytical result, we show that excitatory shared inputs to pairs rather than intuitive inhibitory inputs to trio explain the data. Considering directional connections among three neurons does not disturb this picture: Excitatory inputs to pairs either as hidden layer common input or directional connectivity input are sufficient (and necessary) to explain the observed *strong negative triple-wise* and positive pairwise interactions. We also investigate the robustness of our result when the neuron's voltage is *slightly away from threshold* (SIII) and when

adaptation is present in the model (SIV). We analytically investigate how the *extreme case* of strong common inputs influence pairwise and triple-wise interactions (see SI); the analysis is *independent of any neuron model*, as well as *the voltage's distance from the threshold*. It helps us to analytically find a *guide map*: Distinct regions with clear boundaries for each basic motifs in the triple-wise versus pairwise plane. This reveals that whenever spontaneous firing of neurons is *low*, motifs that have *excitatory inputs* can induce strong triple-wise interactions (Fig. 9); whereas when spontaneous firing rate is high, motifs with *inhibitory common input* can produce strong interactions. Likewise, a person among many others, if the majority are silent (low spontaneous activity), the one talkative person (excitatory input) is clearly noticed; whereas if the majority are talkative (high spontaneous activity), one silent person (inhibitory input) would be conspicuous.

Comparison with other approaches

A classical approach to infer synaptic connectivity from extracellular spiking activity is to construct cross-correlograms of simultaneous spike trains from pairs of neurons Perkel et al. (1967). However, this approach aims at discovering connections among recorded neurons. In fact, researchers made efforts to eliminate the effect of common drives from unobserved inputs on this measure to avoid erroneously reporting pseudo-connections (Brody, 1999; Kobayashi et al., 2019). Another approach is a model-based method that uses a stochastic model of neurons. Among them, the point process - generalized linear model (GLM) is a standard tool for analyzing the statistical connectivity of observed neurons (Pillow et al., 2008; Truccolo et al., 2005; Volgushev et al., 2015). However, these models describe neuronal activity from their past activities and/or known covariate signals such as stimulus and local field potential signals. Since recorded neurons are embedded in larger networks, we need to take into account the effects of inputs from unobserved neurons in order to accurately describe the population activity. Although there have been attempts to include common inputs from unobserved neurons into the GLM framework by treating them as hidden variables (Kulkarni and Paninski, 2007; Vidne et al., 2012), variations in the structure of hidden common inputs are limited. In addition, these statistical models are not directly constrained by physiologically plausible membrane dynamics and spiking threshold while the LIF neuron model is (Ladenbauer et al., 2019). Here, given knowledge about the balanced network, we introduce hidden inputs as background noise and additionally consider various architectures of arbitrarily strong hidden common inputs as shared signals.

Another approach for modeling the input-output relation of a neural population under in-vivo conditions is to use the dichotomized Gaussian (DG) model (Amari et al., 2003; Macke et al., 2009, 2011) and its extensions (Montangie and Montani, 2015, 2017, 2018; Montani et al., 2013). Previous studies have shown that this simple model exhibits positive pairwise and negative triple-wise interactions, which results in the observed sparse population activity (Shimazaki et al., 2015; Yu et al., 2011). The DG model is composed of threshold devices that receive inputs sampled from a correlated multivariate Gaussian distribution to model shared synaptic inputs (Amari et al., 2003; Leen and Shea-Brown, 2015; Macke et al., 2011). Limited by such a structure, one cannot test alternative hypotheses, e.g., if common inhibitory inputs can also generate the same neural interactions (Ohiorhenuan and Victor, 2011; Shimazaki et al., 2015).

In addition, the DG models do not incorporate the dynamics of synapses and membrane potentials. Conversely, the aforementioned input-output relation for near threshold neurons does address the dynamics of the membrane potential by using the LIF model neuron (Shomali et al., 2018); it thus yields quantitative results with temporal accuracy, enabling us to infer the types of common input under various architectures.

Assumptions, limitations, and justifications of the framework

The quantitative model we introduced here is based on two distinct network architectures (triangle and star) with either excitatory or inhibitory shared inputs. It is crucial to see how the directional connection among postsynaptic neurons 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, the combination of directional connections with shared inputs is observed: Excitatory inputs from layer 4 are shared to layer 2/3 connected pairs of excitatory pyramidal neurons in cortex (Yoshimura et al., 2005). Hence, we ran simulations for the directional connections among three neurons: The results reaffirm that excitatory-to-pairs either in recurrent or common input's motif, induces strong negative triple-wise and positive pairwise interactions in low spontaneous regimes (see Fig. 8). There is also another question, whether the *simultaneous existence* of common excitatory and inhibitory inputs in both triangle and star architectures damages this picture. We analyze the model in which two architectures are mixed, i.e., existing together and functioning simultaneously. The result shows the mixing of other motifs, while excluding excitatory-to-pairs, cannot induce strong negative triple-wise and positive pairwise interactions (see SV). It reiterates that the observed strong negative triple-wise interactions are the result of the *excitatory-to-pairs* motif. In other experimental results, of course, it is possible that the divergent common inhibition is mixed with the local common excitatory inputs. To extract evidence of the presence of such mixed inhibitory to trios from the data, one should carefully examine deviations from those observed interactions that are achieved solely by the excitatory inputs to pairs. As far as the spontaneous activities of neurons are low, however, we expect such deviations would be unfortunately small.

One of the assumptions for the analytical framework we introduced is that the firing rate of the signaling input is low (Wolfe et al., 2010) in comparison with postsynaptic neuron; 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 rates for signaling input (SIII in Shomali et al. 2018). However, as we consider a small time window, $\Delta = 10\mathrm{m}s$, the assumption of having *not more than one* signal arrival during such a short time window is practically acceptable. The other assumption is that the synaptic inputs set the voltage of the neuron near the threshold regime, which is reported to be the case when stimulus is presenting (Tan et al., 2014). We analytically calculate the *regions' boundaries* for each motif (see Fig. 6): High amplitude boundaries for all motifs as well as low (high) diffusion boundary for excitatory-to-trio (inhibitory-to-trio) motif, are shown to be independent of the neuron model, hence the near threshold assumption. The other boundaries of triangle architecture, however, do depend on neuronal model, and the near threshold assumption. Yet, we carried numerical simulations, plus many other verifications, to make sure (i) the particular observed data on macaque V1 is signature of excitatory-to-

pairs, and (ii) the suggested *guide map* practically remains reliable and intact, in more general situations. What does change the suggested *guide map*, and regions corresponding to motifs, are the spontaneous rate of neurons as well as bin size. Here, we compare the result for the infrequent and high spontaneous activity of postsynaptic neurons (Fig. 10) while assuming the low input rate, and small bin size. The bin size, on the other hand, cannot be too large, as it would diminish the effect of shared input (Fig. 5), hence the overall reliability of our formalism.

Implication of the results and future challenges

Finally, we have attributed the observation of strong negative triple-wise interactions to a simple motif: Excitatory inputs to pairs. It is tempting to ask whether this microcircuit has any specified *computational advantage* so has been boldly observed, or whether it is the overall setting of a specific experiment that has resulted in this observation. On the one hand, there exist independent empirical evidence that the motif of excitatory-to-pairs is overexpressed - compared with a random network - in rat 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 observations suggest that the emergence of excitatory inputs to pairs is not an accidental observation. On the other hand, common inhibitory input has a clear computational advantage as a well-known winner-take-all network for sparse coding (de Almeida et al., 2009). There also exist experimental evidence that a common inhibitory input innervate multiple postsynaptic pyramidal neurons closer to each other than 100μm (Packer and Yuste, 2011); at greater distances, the probability of common inhibitory inputs to two (and hence more) neurons decreases (Fig.6B in Packer and Yuste 2011). This is attributed to the limited length of inhibitory neurons' axons and simply means that, if electrodes' separation is greater than 100μm, the chance of capturing a common inhibitory input (to pair, or to trio) has already diminished. For the particular experiment of Ohiorhenuan *et al.*, the closest possible separation of recorded neurons is known to be less than 300 microns, *i.e.,* recorded neurons are expected to be gathered in a circle of radius $r \sim 150\mu$m (Ohiorhenuan and Victor, 2011). Then, the probability of having neurons closer to each other than 100μm is 32%, and that of having 3 neurons each closer than 100μm to two others would be very low, less than 7% (see SVII). Thus, the probability of finding an inhibitory presynaptic neuron, which innervates synapses to three reocorded postsynaptic neurons was already less than 7%. Consequently, Ohiorhenuan *et al.* observation does not rule out the presence of inhibitory-to-trio architecture in a more local microcircuitry less than 100μm, and cannot be used as empirical proof for the excitatory-to-pairs as an exclusive computational motif in microcircuits.

There need to be more precise experiments with higher control on the separation of electrode tips. If so, plotting how the observed triple-wise interaction varies with distances among neurons, would lead to a clear conclusion. Such a dependency for pairwise interactions varied with neurons' distance is observed in retina ganglion cells (Ganmor et al., 2011). If the chance of observing a strong negative triple-wise interaction, for neurons closer than 100μm, reduces, it indicates the absence of excitatory-to-pairs architecture in the local network less than 100μm; therefore Ohiorhenuan *et al.*'s observation was a specific result of the experimental setting. However, if prevailing of strong negative triple-wise interactions persists even for neurons closer than 100μm, it means that the excitatory-to-pairs are prevailing architecture in the microcircuit ($\leq 300\mu$m),

and would be another evidence to support for the *computational advantage* of *excitatory-to-pairs* microcircuit. On the contrary, it would be difficult to find evidence that inhibitory-to-trio exists or coexists with excitatory-to-pairs as computational units in the local microcircuits from activities of the three neurons as long as the postsynaptic firing rate is low, because of small negative values for triple-wise interactions induced by common inhibitory inputs (SI). Furthermore, although it is quite challenging to perform in vivo patch-clamp of common inputs and postsynaptic neurons at the same time, an experiment that can directly identify input types and the network's structure in living animals is helpful to improve the prediction of this method.

In summary, we have provided a theoretical tool based on the dynamics of a standard neuron model, to predict network architecture and types of hidden input neurons (excitatory/inhibitory) from the activity of neurons recorded in vivo. We define analytic regions for each motif, with boundaries mostly independent of neuron model, to show the basic motifs can be distinguishable from the statistical data. Our *guide map* helps to uncover hidden network motifs from neural interactions observed in a variety of in vivo data.

Methods

Leaky integrate-and-fire neuron at the threshold regime

According to (Shomali et al., 2018), the first-passage time density (inter-spike interval density) for the LIF neuron (Eq. 1) when neuron receives signaling input at time τ_b on top of noisy background input is given as

$$J(t) = \frac{\sqrt{\kappa\omega}}{\pi\tau_{\rm m}}\{\exp(-\frac{\varphi_+^2}{2})[1+\sqrt{\frac{\pi}{2\kappa}}\varphi_+\exp(\frac{\varphi_+^2}{2\kappa})\times(1+{\rm Erf}(\frac{\varphi_+}{\sqrt{2\kappa}}))] - \exp(-\frac{\varphi_-^2}{2})[1+\sqrt{\frac{\pi}{2\kappa}}\varphi_-\exp(\frac{\varphi_-^2}{2\kappa})\times(1+{\rm Erf}(\frac{\varphi_-}{\sqrt{2\kappa}}))]\}, \quad (6)$$

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

$$\begin{aligned} \kappa(t,\tau_b) &= (1-r(t)^2)/(1-r(t-\tau_b)^2), \quad (7)\\ \omega(t,\tau_b) &= \frac{r(t-\tau_b)^2(1-r(\tau_b)^2)}{(1-r(t)^2)^3},\\ \varphi_\pm(t,\tau_b) &= \sqrt{\frac{\tau_{\rm m}V_\theta^2}{D(1-r(\tau_b)^2)}}\{\pm r(\tau_b) - \int_{\tau_b}^t \frac{ds}{\tau_{\rm m}}\frac{\Delta I(s,A,\tau_b)}{V_\theta}\}, \end{aligned}$$

using $r(t) = \exp(-t/\tau_{\rm m})$. Before the occurrence of signal, i.e., $t < \tau_b$, the ISI density reduces to the known formula (Tuckwell, 1988; Wang and Uhlenbeck, 1945):

$$J_0(t) = \frac{1}{\tau_{\rm m}}\sqrt{\frac{2}{\pi}\frac{\tau_{\rm m}V_\theta^2}{D}\frac{r^2(t)}{(1-r^2(t))^3}}\exp[-\frac{\tau_{\rm m}{V_\theta}^2}{2D}\frac{r^2(t)}{1-r^2(t)}]. \quad (8)$$

In this article, we use Eq. 6 with a square shape of signaling input given by:

$$\Delta I(t, A, \tau_b) = A \times \begin{cases} 0 & t < \tau_b, \\ 1/\Delta t & \tau_b \leq t \leq \tau_b + \Delta t, \\ 0 & \tau_b + \Delta t < t, \end{cases} \tag{9}$$

where Δt measures the signal's lasting time i.e., $\Delta t \sim \tau_{\mathrm{s}}$, which is much smaller than τ_{m}.

Spiking density of LIF neuron after signaling input arrival

We derive the probability density of postsynaptic spike after the arrival of a signaling input. For this goal, we reset the *time origin* to signal arrival's time. Following the aforementioned formalism in Eq. 6, the last postsynaptic spike would have happened at τ_b before the new origin. The conditional probability that the next postsynaptic spike happens at τ after signal arrival is calculated easily (Shomali et al., 2018):

$$f(\tau|\tau_b) = \frac{J(\tau + \tau_b)}{1 - \int_0^{\tau_b} J(s)ds}, \tag{10}$$

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 that the postsynaptic neuron has spiked at τ_b before signal arrival, but has not spiked since then, $p_{\mathrm{back}}(\tau_b)$. It comes as the probability of backward recurrence time, following renewal point process theory (Cox, 1962): $p_{\mathrm{back}}(\tau_b) = \mu[1 - \int_0^{\tau_b} J_0(V_\theta, s)\, ds]$, where $\mu = (\int_0^\infty sJ_0(V_\theta, s)\, ds)^{-1}$ is the mean firing rate of the postsynaptic neuron when there is no signaling input. By marginalizing Eq. 10 with respect to τ_b using $p_{\mathrm{back}}(\tau_b)$, we obtain (Shomali et al., 2018):

$$\begin{aligned} f(\tau) &= \int_0^\infty f(\tau|\tau_b) \times p_{\mathrm{back}}(\tau_b)\, d\tau_b \\ &= \mu \int_0^\infty J(\tau + \tau_b)\, d\tau_b. \end{aligned} \tag{11}$$

Note that when the amplitude of the signaling input, A, is reduced 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 the probability of having one or more spikes in a *specific time window* of Δ, after stimulus onset. It is given as the cumulative density function of $f(\tau)$:

$$F_A(\Delta) = \int_0^\Delta f(\tau)d\tau, \tag{12}$$

where subscript A indicates that $F_A(\Delta)$ is a function of the amplitude of the signaling input. Thus, in the absence of the signaling input (i.e., A=0) we have $F_0(\Delta) = \int_0^\Delta f_0(\tau)d\tau$.

Pairwise and triple-wise interactions of neural populations

Using a binary representation of spiking activity for each postsynaptic neuron in a time window of Δ (schematically illustrated in Fig. 2), one can represent the population activity of the

postsynaptic neurons as a binary pattern. From the probabilities of the occurrence of all possible patterns, one can assess pairwise or higher-order interactions of the neural population. For example, let us consider two neurons. Let $x_i = \{0, 1\}$ $(i = 1, 2)$ be a binary variable, where $x_i = 1$ means that the ith neuron emitted one or more spikes in the bin while $x_i = 0$ means 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:

$$P(x_1, x_2) = \exp\left(\theta_1 x_1 + \theta_2 x_2 + \theta_{12} x_1 x_2 - \psi\right), \tag{13}$$

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

$$\theta_{12} = \log \frac{P(1, 1)P(0, 0)}{P(1, 0)P(0, 1)}. \tag{14}$$

If the binary activities of two neurons are independent, $\theta_{12} = 0$.

The same treatment is applied to three neurons. In an exponential form, the probability mass function for three neurons is written as

$$P(x_1, x_2, x_3) = \exp\left(\sum_{i=1}^{3} \theta_i^{\mathrm{t}} x_1 + \sum_{i<j} \theta_{ij}^{\mathrm{t}} x_i x_j + \theta_{123} x_1 x_2 x_3 - \psi\right). \tag{15}$$

If θ_{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 (less) often than the chance coincidence expected from the event rates of individual neurons and their pairwise correlations. From this equation, the triple-wise interaction among three neurons for the exponential family of probability mass function is calculated using (Amari, 2009a; Nakahara and Amari, 2002):

$$\theta_{123} = \log \frac{P(1, 1, 1)P(1, 0, 0)P(0, 1, 0)P(0, 0, 1)}{P(0, 0, 0)P(0, 1, 1)P(1, 0, 1)P(1, 1, 0)}. \tag{16}$$

Mixture model of three neurons receiving common inputs to their pairs (triangle architecture)

Here we explain the mixture model of three neurons whose pairs receive independent common inputs (a triangle architecture). As shown in Fig. 4D, there are 8 possible patterns to occur for the 3 independent common inputs. When the first common input is active (and the other two common inputs are silent), the pattern probabilities of three postsynaptic neurons are

given by $P_A^1(\mathbf{x}) = [\prod_{i=1}^2 F_A(\Delta)^{x_i}(1-F_A(\Delta))^{1-x_i}] \times [F_0(\Delta)^{x_3}(1-F_0(\Delta))^{1-x_3}]$. This situation happens in $(\lambda\Delta)(1-\lambda\Delta)^2 \times 100\%$ of the bins. The probabilities of activity patterns in which neurons receive the second (third) common input, $P_A^2(\mathbf{x})$ ($P_A^3(\mathbf{x})$), are given similarly to this equation. The common inputs may be simultaneously applied to the same bin due to 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 common inputs 2 and 3, and common inputs 1 and 3 coincide, respectively. Finally, all common inputs coincide at $(\lambda\Delta)^3 \times 100\%$ of the bins, for which the pattern probability is given by $P_A^{123}(\mathbf{x}) = \prod_{i=1,2,3} F_{2A}(\Delta)^{x_i}(1-F_{2A}(\Delta))^{1-x_i}$. The parallel spike sequences are modeled as a mixture of these probability mass functions,

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

For the asymmetric case, when two common inputs are shared among three neurons (Fig. 4E), the mixture model simplifies to:

$$P(\mathbf{x}) = (1-\lambda\Delta)^2 P_0(\mathbf{x}) + \sum_{i=1}^{2}(\lambda\Delta)(1-\lambda\Delta)P_A^i(\mathbf{x}) + (\lambda\Delta)^2 P_A^{12}(\mathbf{x}). \tag{18}$$

Acknowledgments

S.R.S acknowledges J. Victor's, and C. Clopath's helpful comments; H.S. thanks S. Koyama and R. Kobayashi for the valuable discussions; S.R.S and S.N.R acknowledge T. Fukai's kind support and are grateful to him and H. Maboudi for the valuable discussions. HS was supported by the Cooperative Intelligence Joint Research between Kyoto University and Honda Research Institute Japan, MEXT/JSPS KAKENHI Grant Number JP 20K11709, and the grant of Joint Research by the National Institutes of Natural Sciences (NINS Program No. 01112005).

Declaration of Interests

The authors declare no competing interests.

Supplementary materials

SI: The pairwise and triple-wise interactions induced by strong inhibitory and common excitatory inputs

One main message of this paper is that the observed strong negative triple-wise interactions are signatures of excitatory-to-pairs (Fig. 7). The reasonable question is then why this should happen. Is there any fundamental difference between excitatory-to-pairs and inhibitory-to-trio such that the later one cannot produce strong negative triple-wise interactions? We approach this question by considering the extreme case of very strong signaling input. We firstly consider pairwise interactions and then extend the formalism to triple-wise interaction. Combining Eq. 14, Eq. 3, and Eq. 4, the pairwise interaction reads:

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

where $b = \lambda\Delta$. If the postsynaptic neuron receives *an extremely strong inhibitory signal*, 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. S.1 to:

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

where $a = b/(1-b) = \lambda\Delta/(1-\lambda\Delta)$. Conversely, for *an extremely strong excitatory signal*, 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:

$$\theta_{12} = \log(1 + \frac{a}{F_0(\Delta)^2}). \tag{S.3}$$

Comparing Eq. S.2 and Eq. S.3, we see that $F_0(\Delta)$ is simply replaced by $1 - F_0(\Delta)$. In Eq. S.3, for excitatory signaling input and with $a > 0$, we obtain large θ_{12} when $F_0(\Delta) \ll 1$. In fact, θ_{12} indefinitely increases as $F_0(\Delta) \to 0^+$. On the other hand, for strong inhibitory signals in Eq. S.2, this criteria changes to $1 - F_0(\Delta) \ll 1$; which means we need $F_0(\Delta) \lesssim 1$ to obtain high θ_{12}. Consequently, in a regime of low activity i.e., $F_0(\Delta) \ll 1$, strong excitatory signals produce large θ_{12}; whereas in high activity regime i.e., $F_0(\Delta) \lesssim 1$, strong inhibitory inputs produce large pairwise interactions. The only approximation in the 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 a small time window 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. 15, Eq. 16, and Eq. 5 we have:

$$\theta_{123} = \log \frac{P_{111} P_{100}^3}{P_{000} P_{110}^3}, \tag{S.4}$$

where

$$P_{111} = b \times F_A(\Delta)^3 + (1-b) \times F_0(\Delta)^3, \tag{S.5a}$$
$$P_{000} = b \times \{1 - F_A(\Delta)\}^3 + (1-b) \times \{1 - F_0(\Delta)\}^3, \tag{S.5b}$$
$$P_{100} = b \times F_A(\Delta)\{1 - F_A(\Delta)\}^2 + (1-b) \times F_0(\Delta)\{1 - F_0(\Delta)\}^2, \tag{S.5c}$$
$$P_{110} = b \times F_A(\Delta)^2\{1 - F_A(\Delta)\} + (1-b) \times F_0(\Delta)^2\{1 - F_0(\Delta)\}, \tag{S.5d}$$

and $b = 1 - \exp(\lambda\Delta)$, which simplifies to $b \simeq \lambda\Delta$ for low firing rates. 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 $\Delta = 5 - 10\,\mathrm{ms}$, $F_A(\Delta) \simeq 0$. It drastically simplifies θ_{123} to:

$$\theta_{123} = -\log(1 + \frac{a}{(1 - F_0(\Delta))^3}). \tag{S.6}$$

Here $a = \{1 - \exp(-\lambda\Delta)\}/\exp(-\lambda\Delta)$, which is approximated to $a = \lambda\Delta/(1 - \lambda\Delta)$ whenever $\lambda\Delta \ll 1$. For the case of *strong excitatory-to-trio*, triple-wise interaction becomes:

$$\theta_{123} = \log(1 + \frac{a}{F_0(\Delta)^3}). \tag{S.7}$$

Similarly, we can consider the triple-wise interaction for excitatory-to-pairs architectures. It is a bit complicated compared to the inhibitory-to-trio case. Considering Eq. 15, Eq. 16 and Eq. 17, the triple-wise interaction is:

$$\theta_{123} = \log \frac{P_{111} P_{100}^3}{P_{000} P_{110}^3}, \tag{S.8}$$

where

$$\begin{aligned} P_{111} = {} & (1-b)^3 \times F_0(\Delta)^3 + 3b(1-b)^2 \times F_A(\Delta)^2 F_0(\Delta) \\ & + 3b^2(1-b) \times F_A(\Delta)^2 F_{2A}(\Delta) + b^3 \times F_{2A}(\Delta)^3, \end{aligned} \tag{S.9a}$$

$$\begin{aligned} P_{000} = {} & (1-b)^3 \times \{1 - F_0(\Delta)\}^3 + 3b(1-b)^2 \times \{1 - F_A(\Delta)\}^2\{1 - F_0(\Delta)\} \\ & + 3b^2(1-b) \times \{1 - F_A(\Delta)\}^2\{1 - F_{2A}(\Delta)\} + b^3 \times \{1 - F_{2A}(\Delta)\}^3, \end{aligned} \tag{S.9b}$$

$$\begin{aligned} P_{100} = {} & (1-b)^3 \times F_0(\Delta)\{1 - F_0(\Delta)\}^2 + b^3 \times F_{2A}(\Delta)\{1 - F_{2A}(\Delta)\}^2 \\ & + b(1-b)^2 \times \big\{2F_A(\Delta)\{1 - F_A(\Delta)\}\{1 - F_0(\Delta)\} + F_0(\Delta)\{1 - F_A(\Delta)\}^2\big\} \\ & + b^2(1-b) \times \big\{2F_A(\Delta)\{1 - F_A(\Delta)\}\{1 - F_{2A}(\Delta)\} + F_{2A}(\Delta)\{1 - F_A(\Delta)\}^2\big\}, \end{aligned} \tag{S.9c}$$

$$\begin{aligned} P_{110} = {} & (1-b)^3 \times F_0(\Delta)^2\{1 - F_0(\Delta)\} + b^3 \times F_{2A}(\Delta)^2\{1 - F_{2A}(\Delta)\} \\ & + b(1-b)^2 \times \big\{2F_0(\Delta)\, F_A(\Delta)\{1 - F_A(\Delta)\} + F_A(\Delta)^2\{1 - F_0(\Delta)\}\big\} \\ & + b^2(1-b) \times \big\{2F_A(\Delta)F_{2A}(\Delta)\{1 - F_A(\Delta)\} + F_A(\Delta)^2\{1 - F_{2A}(\Delta)\}\big\}. \end{aligned} \tag{S.9d}$$

For extremely strong excitation to pairs, we have $F_A(\Delta) \simeq 1$ and $F_{2A}(\Delta) \simeq 1$; we thus have:

$$\theta_{123} = -\log\left((1+\frac{a}{F_0(\Delta)^2})^3 \times \frac{1}{1+3a/F_0(\Delta)^2+3a^2/F_0(\Delta)^3+a^3/F_0(\Delta)^3}\right), \tag{S.10}$$

where $a = \{1-\exp(-\lambda\Delta)\}/\exp(-\lambda\Delta)$, as before. We can show that for the case of strong inhibitory inputs to pairs out of three neurons ($F_A(\Delta) \simeq 0$), the triple-wise interaction is:

$$\theta_{123} = \log\left((1+\frac{a}{(1-F_0(\Delta))^2})^3 \times \frac{1}{1+3a/(1-F_0(\Delta))^2+3a^2/(1-F_0(\Delta))^3+a^3/(1-F_0(\Delta))^3}\right), \tag{S.11}$$

For asymmetric excitatory-to-pairs architecture (two common inputs instead of three), the pattern probabilities S.9a changes as:

$$P_{111} = (1-b)^2 \times F_0(\Delta)^3 + 2b(1-b) \times F_A(\Delta)^2 F_0(\Delta) + b^2 \times F_A(\Delta)^2 F_{2A}(\Delta), \tag{S.12a}$$

$$\begin{aligned} P_{000} = {} & (1-b)^2 \times \{1-F_0(\Delta)\}^3 + 2b(1-b) \times \{1-F_A(\Delta)\}^2\{1-F_0(\Delta)\} \\ & + b^2 \times \{1-F_A(\Delta)\}^2\{1-F_{2A}(\Delta)\}, \end{aligned} \tag{S.12b}$$

$$\begin{aligned} P_{100} = {} & (1-b)^2 \times F_0(\Delta)\{1-F_0(\Delta)\}^2 + b^2 \times F_{2A}(\Delta)\{1-F_A(\Delta)\}^2 \\ & + 2b(1-b) \times F_A(\Delta)\{1-F_A(\Delta)\}\{1-F_0(\Delta)\} \\ P_{010} = P_{001} = {} & (1-b)^2 \times F_0(\Delta)\{1-F_0(\Delta)\}^2 + b^2 \times F_A(\Delta)\{1-F_A(\Delta)\}\{1-F_{2A}(\Delta)\} \\ & + b(1-b) \times \{F_A(\Delta)\{1-F_A(\Delta)\}\{1-F_0(\Delta)\} + F_0(\Delta)\{1-F_A(\Delta)\}^2\} \\ P_{110} = P_{101} = {} & (1-b)^2 \times F_0(\Delta)^2\{1-F_0(\Delta)\} + b^2 \times F_{2A}(\Delta)F_A(\Delta)\{1-F_A(\Delta)\} \\ & + b(1-b)\{F_A(\Delta)^2\{1-F_0(\Delta)\} + F_A(\Delta)F_0(\Delta)\{1-F_A(\Delta)\}\} \\ P_{011} = {} & (1-b)^2 \times F_0(\Delta)^2\{1-F_0(\Delta)\} + b^2 \times F_A(\Delta)^2\{1-F_{2A}(\Delta)\} \\ & + 2b(1-b)F_A(\Delta)F_0(\Delta)\{1-F_A(\Delta)\} \end{aligned} \tag{S.12c}$$

For a high amplitude limit of asymmetric excitatory inputs given to pairs of neurons ($F_A(\Delta) \simeq 1$), the triple-wise interaction becomes

$$\theta_{123} = -\log\left((1+\frac{a}{F_0(\Delta)^2})^2 \times \frac{1}{1+2a/F_0(\Delta)^2+a^2/F_0(\Delta)^3}\right), \tag{S.13}$$

The marginalized θ_{12} for symmetric excitatory-to-pairs is (for inhibitory-to-pairs replace $F_0(\Delta)$ with $1-F_0(\Delta)$):

$$\theta_{12} = \log\left((1+2a/F_0(\Delta)+a/F_0(\Delta)^2+3a^2/F_0(\Delta)^2+a^3/F_0(\Delta)^2)/(1+a/F_0(\Delta))^2\right) \tag{S.14}$$

and for asymmetric excitatory-to-pairs is :

$$\theta_{12} = \log\left((1+a/F_0(\Delta)+a/F_0(\Delta)^2+a^2/F_0(\Delta)^2)/(1+a/F_0(\Delta))\right) \tag{S.15}$$

Figure 9**B** shows, for the *low activity* of a postsynaptic neuron, i.e., $F_0(\Delta) \ll 1$, the motif of excitatory-to-pairs generates 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.

SI-a: The limits of high amplitude approximation

In the high amplitude limit for excitatory inputs given to trio or pairs, the relations obtained for triple-wise and pairwise interactions (Eq. S.3 and Eq. S.7) in the limit of $F_0(\Delta) \to 1$, diverge. For this regime, we take the limit of pairwise and triple-wise interactions (Eq. S.1, Eq. S.4 and Eq. S.5), considering $F_A(\Delta) \to 1$ and $F_0(\Delta) \to 1$ simultaneously. The pairwise equation, Eq. S.1, will be:

$$\theta_{12} = \log \frac{b(1 - F_A(\Delta))^2 + (1 - b)(1 - F_0(\Delta))^2}{(b(1 - F_A(\Delta)) + (1 - b)(1 - F_0(\Delta)))^2}, \tag{S.16}$$

where $b = \lambda\Delta$. Assuming $z = (1 - F_A(\Delta)/(1 - F_0(\Delta))$, it becomes

$$\theta_{12} = \log \frac{az^2 + 1}{(az + 1)^2}, \tag{S.17}$$

where $a = b/(1 - b)$ and z varies within the range of $[0, 1]$. The triple-wise interaction (Eq. S.4 and Eq. S.5) in the limit of $F_A(\Delta) \to 1$ and $F_0(\Delta) \to 1$ is:

$$\theta_{123} = \log \frac{(az^2 + 1)^3}{(az^3 + 1)(az + 1)^3}. \tag{S.18}$$

Similarly for inhibitory input given to three neurons, the limit is taken when $F_A(\Delta) \to 0$ and $F_0(\Delta) \to 0$ simultaneously. In this limit, we put $1 - F_A(\Delta) \to 1$ and $1 - F_0(\Delta) \to 1$. Then, pairwise interaction in Eq. S.1 reduces to:

$$\theta_{12} = \log \frac{ax^2 + 1}{(ax + 1)^2}, \tag{S.19}$$

where $x = F_A(\Delta)/F_0(\Delta)$. The triple-wise interactions (Eq. S.4 and Eq. S.5) will be:

$$\theta_{123} = \log \frac{(ax^3 + 1)(ax + 1)^3}{(ax^2 + 1)^3}. \tag{S.20}$$

SII: Defining the analytic regions for motifs of excitatory/inhibitory inputs to trios/pairs in the plane of θ_{123}/θ_{12}

Here we describe in brief how to achieve the boundaries for each motif in the plane of θ_{123}/θ_{12}. For all motifs, the first boundary arises from the limit of high amplitude ($F_A(\Delta) \to 1$ for excitatory signaling inputs and $F_A(\Delta) \to 0$ for inhibitory signaling inputs, see SI) when the spontaneous activity $F_0(\Delta)$ changes within $[0, 1]$. This boundary is shown in Figure 6 by solid gray lines.

For the star architecture with common excitatory input, the second boundary (dashed purple line, Fig.6) is defined by the lowest plausible spontaneous firing rate of a postsynaptic neuron when CDF ($F_A(\Delta)$) varies from zero (no common input) to one (high amplitude of common input). One can easily find how spontaneous postsynaptic neuron's firing rate relates to the level of background synaptic activities from a stationary distribution of voltage trajectory for

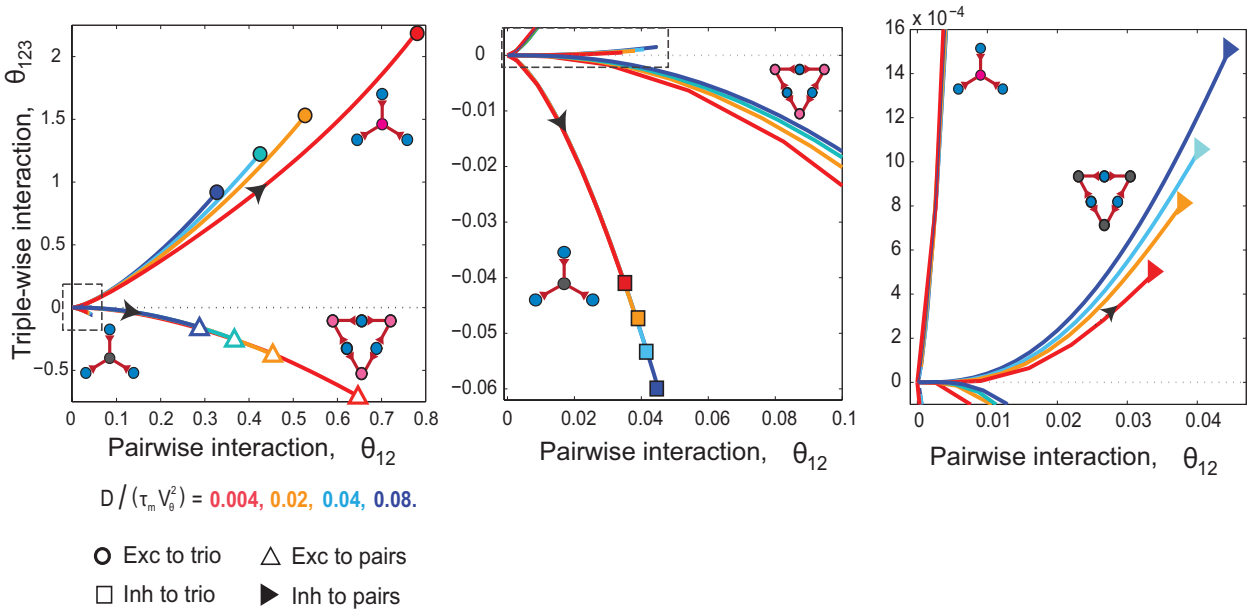

Fig S1. The triple-wise interaction versus pairwise interaction for star and triangular architectures with common excitatory and inhibitory inputs as a function of the scaled diffusion and scaled shared signal strength. The color codes are for four scaled diffusion coefficients and the symbols show the motifs. The arrows indicate increasing directions of the scaled amplitude parameter (i.e. $A/(\tau_{\mathrm{m}} V_\theta)$) and the symbols on graphs show the saturation points. The middle and right panels illustrate the interactions in the neighborhood of origin for negative and positive triple-wise interactions which correspond to inhibitory-to-trio and inhibitory-to-pairs motifs, respectively. The only fixed parameters are the bin size of $\Delta = 5\,\mathrm{ms}$ and presynaptic rate of $\lambda = 5\,\mathrm{Hz}$.

LIF model neuron (Brunel, 2000; Brunel and Hakim, 1999). We have (SV in (Shomali et al., 2018)):

$$P_s(V) = \frac{\mu \tau_{\mathrm{m}}^2}{D} \int_V^{\bar{I}} \Theta(V_s - V_r) \exp\left(\frac{-\tau_{\mathrm{m}}}{2D}(V^2 - V_s^2 - 2\bar{I}(V - V_s))\right) dV_s, \tag{S.21}$$

where $\Theta(V)$ is the Heaviside step function: $\Theta(V) = 1$ for $V > 0$ and otherwise $\Theta(V) = 0$. μ is the mean firing rate: $\mu = (\int_0^\infty s J_0(V_\theta, s)\, ds)^{-1}$ and V_r is the resting voltage that we assume to be zero here. $\bar{I}$ is the mean current which is equal to V_θ for the threshold regime. Since the stationary distribution is normalized ($\int_0^\infty P_s(V) dV_s = 1$), μ can be obtained explicitly (Brunel, 2000):

$$\mu^{-1} = \tau_{\mathrm{m}} \sqrt{\pi} \int_{(V_r - \bar{I})\sqrt{\frac{\tau_{\mathrm{m}}}{2D}}}^{(V_\theta - \bar{I})\sqrt{\frac{\tau_{\mathrm{m}}}{2D}}} e^{y^2} (1 + \mathrm{erf}(y)) dy, \tag{S.22}$$

where $\mathrm{erf}(x) = (2/\sqrt{\pi}) \int_0^x \exp(-t^2) dt$. For $V_r = 0$ and $\bar{I} = V_\theta$ (the threshold regime), the firing

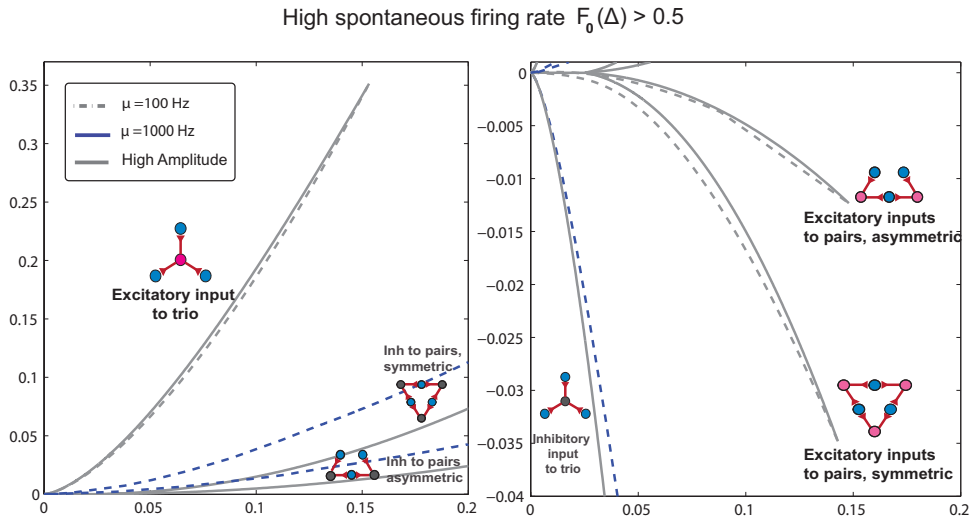

Fig S2. The regions of triple-wise interaction versus pairwise interaction for star and triangular architectures with common excitatory and inhibitory inputs for high spontaneous firing rates of postsynaptic neurons $\mu = 100 - 1000$ Hz (magnification of Fig. 10**B**). In this regime, the excitatory inputs' motifs induce small interactions and their regions shrink compared to the low spontaneous firing of the postsynaptic neuron (Fig. 6). The only fixed parameters are the bin size of $\Delta = 5\,\text{ms}$ and the presynaptic rate of $\lambda = 5\,\text{Hz}$.

rate becomes

$$\mu = (2\tau_{\text{m}} \int_{-\bar{I}\sqrt{\frac{\tau_{\text{m}}}{2D}}}^{0} \int_{-\infty}^{x} \exp\,(x^2 - t^2) dt\, dx)^{-1}. \tag{S.23}$$

Using Eq. S.23, each level of background noise (diffusion coefficient) gives us a specific spontaneous firing rate that we can use to calculate $F_0(\Delta)$. So by changing $F_A(\Delta)$ within the valid range of $[0, 1]$, we can calculate the whole range of triple-wise and pairwise interactions for excitatory inputs to trio (green to purple dashed lines in Fig.6). We observe that, as the background noise decreased, the region for excitatory-to-trio expands, so we use the lowest plausible spontaneous activity (here we use $\mu = 1Hz$ for $D = 5.5 \times 10^{-17} ms,\ mV^2$). This gives us the CDF of spontaneous activity, $F_0(\Delta)$. By changing $F_A(\Delta)$ within the range $[0, 1]$, we obtain the second boundary for the motif of excitatory input to trio (Fig.6, purple line).

For inhibitory inputs given to trios, higher spontaneous activity shows higher interactions. So the second boundary is the limit of high spontaneous activity. Since we are restricted to a low activity regime (i.e., $F_0(\Delta) \leq 0.5$), for this limit we choose the background noise for $F_0(\Delta) = 0.5$ that is ($D = 190ms,\ mV^2$ and $\mu = 100Hz$ for $\Delta = 5$ms). Here, like excitatory-to-trio case, $F_A(\Delta)$ varies within the range of $[0, 1]$, but this time for high spontaneous activity ($\mu = 100$ Hz). The two boundaries for inhibitory-to-trio, lie very close to each other and we get a narrow region in the plane of triple-wise/pairwise interactions.

For inhibitory inputs given to pairs of neurons, again the high spontaneous activity (here, ($D = 190ms,\ mV^2$ and $\mu = 100$Hz) defines the second boundary ($F_0(\Delta) = 0.5$ for $\Delta = 5$ms)

which is achieved by our analytic input-output method and depends on the neuron model (Eq. 12, from Eq. 14 and Eq. 16).

Finally, the second boundary for the motif of excitatory-to-pairs in the low activity regime comes from very low diffusion ($\mu = 1Hz$ for $D = 5.5 \times 10^{-17} ms,\ mV^2$). The solution of postsynaptic neuron's CDF after signal arrival (i.e. ($F_A(\Delta)$ and $F_{2A}(\Delta)$)) for very low diffusion regime is:

$$F_A = \int_{V_\theta - A/\tau_m}^{V_\theta} P_s(V) dV, \tag{S.24}$$

where $P_s(v)$ is the stationary distribution (Eq. S.21). When $\frac{A}{\tau_\mathrm{m}} \leq V_\theta$, the CDF is:

$$F_A = \mu \tau_\mathrm{m} \int_0^{A/\sqrt{D\tau_\mathrm{m}}} \frac{1}{u} (e^{-\frac{u^2}{2}} - e^{\frac{u^2}{2} - \frac{uA}{\sqrt{D\tau_\mathrm{m}}}}) du, \tag{S.25}$$

and for $\frac{A}{\tau_\mathrm{m}} > V_\theta$:

$$F_A = \mu \tau_\mathrm{m} \int_0^{\frac{A}{\sqrt{D\tau_\mathrm{m}}} - V_\theta \sqrt{\frac{\tau_m}{D}}} \frac{1}{u} (e^{-\frac{u^2}{2}} - e^{-\frac{u^2}{2} - uV_\theta \sqrt{\frac{\tau_\mathrm{m}}{D}}}) du + \int_{\frac{A}{\sqrt{D\tau_\mathrm{m}}} - V_\theta \sqrt{\frac{\tau_m}{D}}}^{\frac{A}{\sqrt{D\tau_\mathrm{m}}}} \frac{1}{u} (e^{-\frac{u^2}{2}} - e^{\frac{u^2}{2} - uA/\sqrt{\tau_\mathrm{m} D}}) du. \tag{S.26}$$

For a specific spontaneous firing rate of the postsynaptic neuron, from Eq. S.25 and Eq. S.26, pairwise and triple-wise interactions are calculated (Eq. 14 and Eq. 16) and drawn in lines in Fig.6. So the boundary is determined ($\mu = 1$ Hz) for both symmetric and asymmetric excitatory inputs given to pairs of neurons (purple dashed lines in regions of excitatory inputs to pairs, Fig.6). In high spontaneous activity (Fig. 10B), the boundary for lowest spontaneous activity ($\mu = 100$ Hz) of excitatory-to-pairs, is achieved by calculating the $F_A(\Delta)$ and $F_{2A}(\Delta)$ from input-output relation of a LIF neuron model in our method (Eq. 12) and then calculating interactions from Eq. 14 and Eq. 16. So just this boundary of excitatory-to-pairs like one boundary of inhibitory-to-pairs depends on the neuron model (here LIF).

SII-a: When the motif of excitatory inputs to trio has negative triple-wise interaction?

We observe that when the spontaneous activity of the postsynaptic neuron is less than $\mu = 5$ Hz, the triple-wise interaction in the motif of excitatory inputs given to three neurons, for some values of common input's amplitude, becomes negative (Fig. 6). We can rewrite the pattern probabilities and triple-wise interaction for excitatory inputs given to three neurons from Eq. S.4 and Eq. S.5:

$$P_{111} = F_0(\Delta)^3 (ax^3 + 1), \tag{S.27a}$$

$$P_{110} = F_0(\Delta)^2 (1 - F_0(\Delta))(ax^2 (\frac{1 - xF_0(\Delta)}{1 - F_0(\Delta)}) + 1), \tag{S.27b}$$

$$P_{100} = F_0(\Delta)(1 - F_0(\Delta))^2 (ax (\frac{1 - xF_0(\Delta)}{1 - F_0(\Delta)})^2 + 1), \tag{S.27c}$$

$$P_{000} = (1 - F_0(\Delta))^3 (a (\frac{1 - xF_0(\Delta)}{1 - F_0(\Delta)})^3 + 1), \tag{S.27d}$$

And triple-wise interaction is:

$$\theta_{123} = \log \varphi_1 \times \varphi_2^3, \tag{S.28}$$

that

$$\varphi_1 = \frac{ax^3 + 1}{a(\frac{1-xF_0(\Delta)}{1-F_0(\Delta)})^3 + 1}, \tag{S.29}$$

and

$$\varphi_2 = \frac{ax(\frac{1-xF_0(\Delta)}{1-F_0(\Delta)})^2 + 1}{ax^2\frac{1-xF_0(\Delta)}{1-F_0(\Delta)} + 1}, \tag{S.30}$$

where $a = b/(1-b)$ and $x = F_A(\Delta)/F_0(\Delta)$. For excitatory inputs to trio, x varies within the range $1 \leq x \leq 1/F_0(\Delta)$. The term that makes the triple-wise negative in some values of common input's amplitude, is φ_2, which is the fraction of pattern probability P_{100} over P_{110}. By investigation, we find that when $\eta < 0.25$ where $\eta = F_0(\Delta)(1 - F_0(\Delta))/a^{2/3}$, the triple-wise interaction becomes negative in motif of excitatory-to-trio. For this condition, negative triple-wise interaction occurs when the cumulative density function $F_A(\Delta)$ is in the range of $F_-(\Delta) \leq F_A(\Delta) \leq F_+(\Delta)$, where $F_\pm(\Delta) = 0.5 \pm \sqrt{0.25 - \eta}$. For example when the common input rate is $\lambda = 5$ Hz and time window is $\Delta = 5$ ms, postsynaptic spontaneous rate below $\mu = 4.4$ Hz, satisfy $\eta < 0.25$ and the triple-wise interaction becomes negative.

SIII: What happens to triple-wise and pairwise interactions when we go beyond threshold regime?

Here we investigate how the results change when we go beyond the threshold regime. We performed simulations when the postsynaptic neurons are in the subthreshold and suprathreshold regimes, and compare the results with simulations at the threshold regime. Figure S3 shows the interactions arisen from the simulations in the plane of pairwise versus triple-wise interaction parameters. The circles and rectangles marked in orange display the results when the postsynaptic neurons operate at the threshold regime. With the same parameters, those marked in yellow and pink are obtained when the mean inputs to postsynaptic neurons reach 2 and 10 percent below the threshold ($\delta I/\bar{I} = -0.02, -0.1$). The circles show the interactions obtained for the inhibitory-to-trio motifs, whereas the rectangles are for the excitatory-to-pairs motifs. The result is for three scaled amplitudes of the signaling inputs at $A/\tau_m V_\theta = 0.2, 0.4, 1$ shown with no line, one line, and two crossed lines inside the symbols respectively. These results show that, as the mean input goes away from the threshold in the subthreshold regime, the interactions for excitatory-to-pairs motifs get stronger while those for inhibitory-to-trio get weaker. These results are interpreted as follows. At the threshold regime, the mean input to a postsynaptic neuron is set very close to the threshold. Thus, even a small amount of noise induces spikes of the postsynaptic neuron. In subthreshold regime where the mean input given to the postsynaptic neuron is set below the threshold, the neuron generates a spike as far as the variance of the noise helps the voltage to reach the threshold. In this case, the postsynaptic neuron fires less frequently (smaller amount of $F_0(\Delta)$) compared to the case at the threshold regime. Since the neuron is in a regime of lower spontaneous activity, the interactions induced by the excitatory inputs increase while the interactions caused by inhibitory inputs decrease (see SI). Therefore,

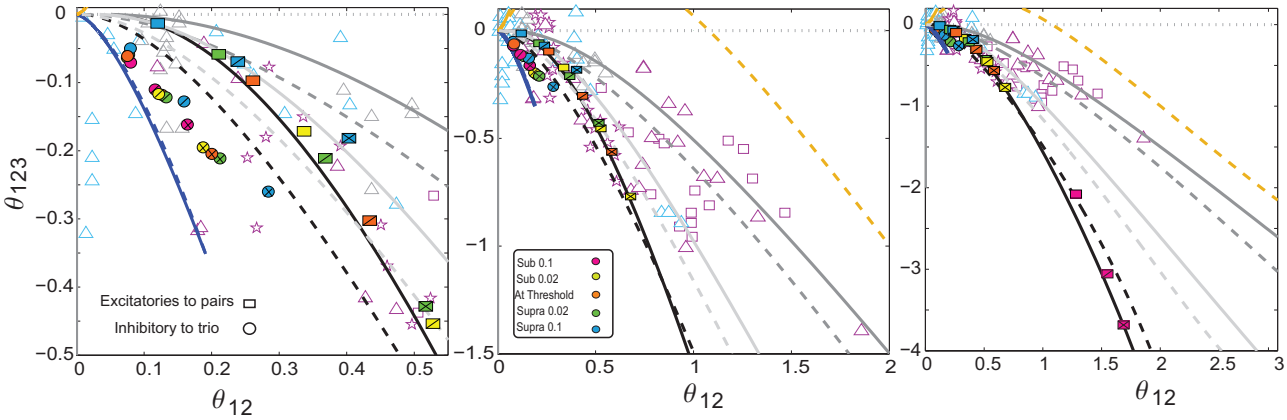

Fig S3. Modulation of interaction parameters in subthreshold and suprathreshold regimes, where mean synaptic input given to membrane potential is set below or above threshold. Left: Changes in interactions among inhibitory-to-trios are shown in subthreshold and suprathreshold regimes. Each color is for one deviation (see the color code) from the threshold regime at three scaled amplitudes of the signaling inputs at $A/\tau_m V_\theta = 0.2, 0.4, 1$ shown by no line, one line and two crossed lines inside symbols respectively. Middle and Right: The interactions are shown for the motif of excitatory inputs given to pairs of neurons. When a neuron is in the subthreshold regime, the interactions are reduced for inhibitory-to-trio of neurons (compare orange with pink and yellow circles) while they are increased for excitatory-to-pairs motifs (compare orange with pink and yellow rectangles). In the suprathreshold regime, the excitatory-to-pairs motif generates weaker interactions (green and blue rectangles) while the inhibitory-to-trio motif generates stronger interactions especially for stronger input (green and blue circles with crossed lines inside). Fixed parameters are $\Delta = 10$ ms, $D/\tau_m V_\theta^2 = 0.002$, and $\lambda = 5$ Hz. Each symbol results from at least 10^{10} steps of run to keep the mean squared error in order of $10^{-3} - 10^{-4}$.

the hypothesis that the excitatory-to-pairs motif is behind the empirically observed strong triple-wise and pairwise interactions is not only unchanged but also more strongly supported when the neuron operates in the subthreshold regime rather than the threshold regime.

We then move on to the suprathreshold regime, where the mean input is set above the threshold of postsynaptic neuron's voltage (Figure S3). Those marked in green and blue are obtained when the mean inputs to postsynaptic neurons reach 2 and 10 percent above the threshold ($\delta I/\bar{I} = 0.02, 0.1$). In the suprathreshold regime, the picture mentioned above reverses. Since the mean input is set above the threshold, a neuron generates regular spikes with high frequency, changing the regime of spontaneous activity from low to high firing rates $F_0(\Delta) > 0.5$. In this regime, the inhibitory signaling input can induce stronger interactions while excitatory signaling input cannot. We can reconfirm this interpretation in the simulation study (Fig. S3). For the range of scaling amplitudes of signaling inputs $A/\tau_m V_\theta = 0.2, 0.4, 1$, we observe the interactions induced by inhibitory inputs get stronger as the mean input increases to 2 and 10 percent above the threshold (compare orange circle with the green and blue circle for symbols

with crossed lines, for example).

SIV: Does adaptation render inhibitory-to-trio to induce strong negative triple-wise and positive pairwise interaction?

Here we want to use a more physiologically plausible model of the neuron and see how it affects the interaction parameters. For this aim, we add adaptation effect to the leaky integrate-and-fire neuron model (Brette and Gerstner, 2005; Gerstner et al., 2014b):

$$\tau_{\mathrm{m}} \frac{dV(t)}{dt} = -(V(t) - V_r) - w + I(t), \tag{S.31}$$

$$\tau_{\mathrm{w}} \frac{dw}{dt} = a(V(t) - V_r) - w + b\tau_{\mathrm{k}} \sum_{t_f} \delta(t - t^f), \tag{S.32}$$

where the adaptation current is w, the coupling term of voltage to adaptation is a, and τ_{w} is the time constant of the adaptation variable. When the neuron generates a spike, the voltage is reset to V_r and the adaptation variable w is increased by b which is a parameter for the spike-triggered adaptation. We use adaptation parameters consistent with Brette and Gerstner (Brette and Gerstner, 2005). Figure S4 shows the result with and without the adaptation at the threshold regime for the two motifs of excitatory-to-pairs (rectangles) and inhibitory-to-trio (circles) for three scaling amplitudes of signaling inputs $A/\tau_m V_\theta = 0.05, 0.1, 0.25$ shown with no line, one line and two crossed lines inside the symbols respectively. For adaptation, we consider two cases in that the coupling between voltage and adaptation current (a) is present or absent. The former (i.e. $a > 0$) considers subthreshold adaptation in addition to spike-triggered adaptation (b) while the latter (i.e. $a = 0$) assumes just the effect of spike-frequency adaptation.

Figure S4 shows how adaptation modulates the pairwise and triple-wise interactions for inhibitory-to-trio and excitatory-to-pairs motifs. We observe that the modulation by adaptation make the triple-wise (pairwise) interactions for the excitatory-to-pairs motif more negative (positive), while the triple-wise (pairwise) interactions in the inhibitory-to-trio motif get less negative (positive) for strong inhibitory signaling inputs. It makes a larger difference in triple-wise and pairwise interactions between these two motifs.

The result shows adding the effect of adaptation differentiates, even more, the triple-wise interactions between the excitatory-to-pairs and inhibitory-to-trio motifs in a way that the excitatory input to pairs can induce the strong negative triple-wise and positive pairwise interactions, while the strength of inhibitory-to-trio's interactions for strong common input decreases. This is attributed to decreasing the spontaneous activity of postsynaptic neurons in the presence of adaptation (Fig. S5). We observed that adding the effect of adaptation renders the firing rate of the postsynaptic neuron receiving common inputs on top of noisy background input, from 22 Hz (excitatory-to-pairs motif) and 19 Hz (inhibitory-to trio motif) to around 5 Hz and 3 Hz respectively (also Fig. S6, middle and right panels). When common input is off, adaptation also decreases the firing rate of postsynaptic neuron from 22 Hz to around 3 Hz (compare number 1 blue with red in Fig. S6, middle). Therefore by reducing the spontaneous activity, adaptation increases the strength of interactions in excitatory-to-pairs motifs but decreases the interactions' strength in inhibitory-to-trio motifs (Fig. S6, middle and right panels).

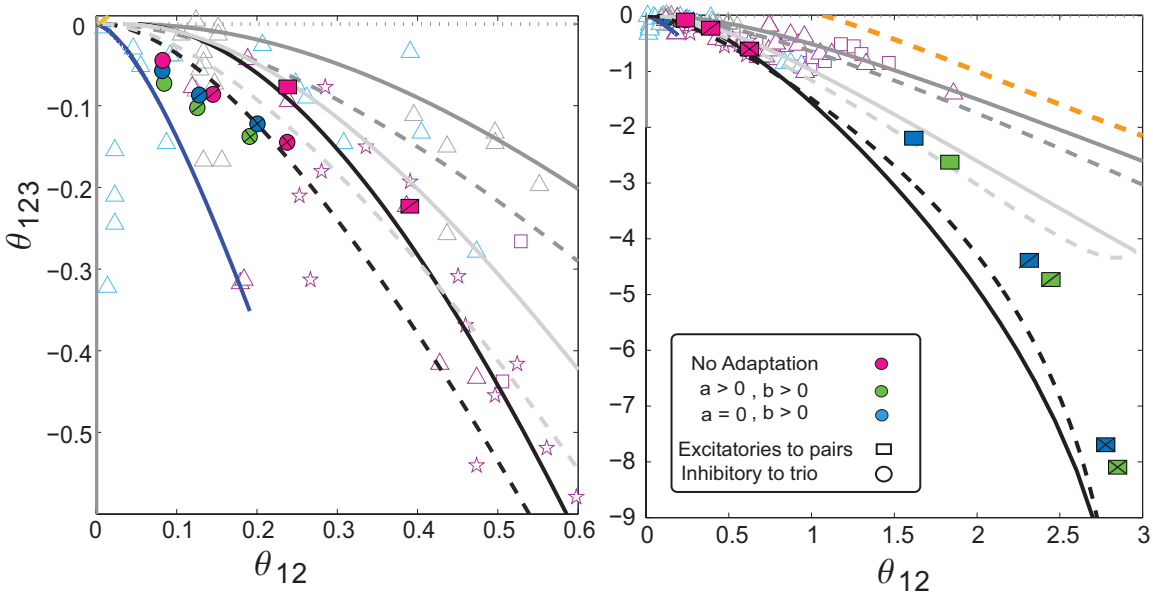

Fig S4. Modulation of interaction parameters in the presence of adaptation. The left panel shows the changes of interactions among the inhibitory inputs given to trios in the presence of adaptation (blue and green circles) compared with no adaptation case (purple circles). Each case is shown for three scaled amplitudes of the signaling inputs at $A/\tau_m V_\theta = 0.05, 0.1, 0.25$ shown with no line, one line and two crossed lines, inside the symbols respectively; the larger the amplitudes, the stronger the interactions. When the adaptation is involved, the interactions are weaker for strong inhibitory input given to three postsynaptic neurons (left: blue and green circles) while they are stronger for excitatory-to-pairs motifs (right: blue and green rectangles). The adaptation parameters are consistent with Brette and Gerstner paper (Brette and Gerstner, 2005). Fixed parameters are $a = 4\ nS$ (for $a > 0$ cases), $b = 0.0805\ nA$, $\tau_m = 10$ ms, $g_L = 30\ nS$, $C = 281\ pF$, $V_r = -70\ mV$, $V_\theta = -50\ mV$, $\tau_w = 144$ ms, $\Delta = 10$ ms, $D = 0.74\ msmV^2$, and $\lambda = 5$ Hz. Each symbol results from at least 10^{10} steps of run to keep the mean squared error in order of 10^{-4}.

The other interesting effect we observe is that excitatory inputs given to pairs of neurons with the help of adaptation can generate the observed CV (coefficient of variation) in V1 neurons (Gur et al., 1997), and by increasing the amplitude of signaling input, the CV reaches high values (Fig. S6 A) consistent with some experimental studies reporting high variability for neurons (Shadlen and Newsome, 1998). On the other side, the motif of inhibitory-to-trio cannot generate such a result of high CV (Fig. S6).

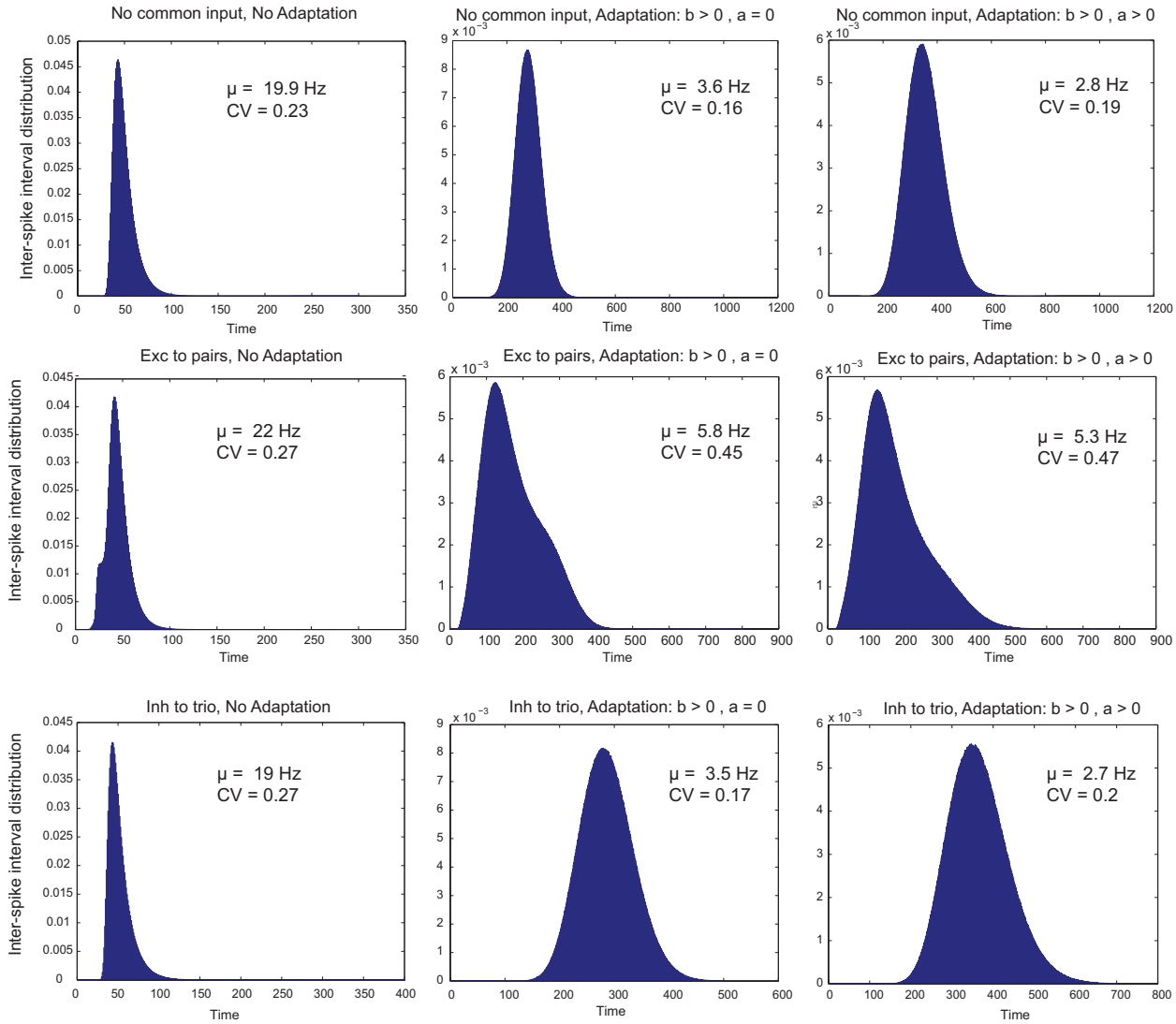

Fig S5. Inter-spike interval (ISI) distributions of postsynaptic neurons for motifs of excitatory-to-pairs and inhibitory-to-trio in the presence and absence of adaptation. The first row shows the ISI when there is no common (signaling) input and the postsynaptic neuron receives mean input which sets the voltage very near to the threshold in addition to a certain level of noise. In the presence of adaptation, when there is no common (signaling) input, the firing rate of postsynaptic neuron decreases, compared to no adaptation case. In the presence of common inputs under excitatory-to-pairs and inhibitory-to-trio motifs, the adaptation decreases the firing rate (second and third row). However, the adaptation increases the coefficient of variation (CV) when excitatory inputs are given to pairs, but decreases the CV in the inhibitory-to-trio case. Fixed parameters are $A/\tau_m V_\theta = 0.1$, $\Delta = 10$ ms, $D/\tau_m V_\theta^2 = 0.0002$, and $\lambda = 5$ Hz.

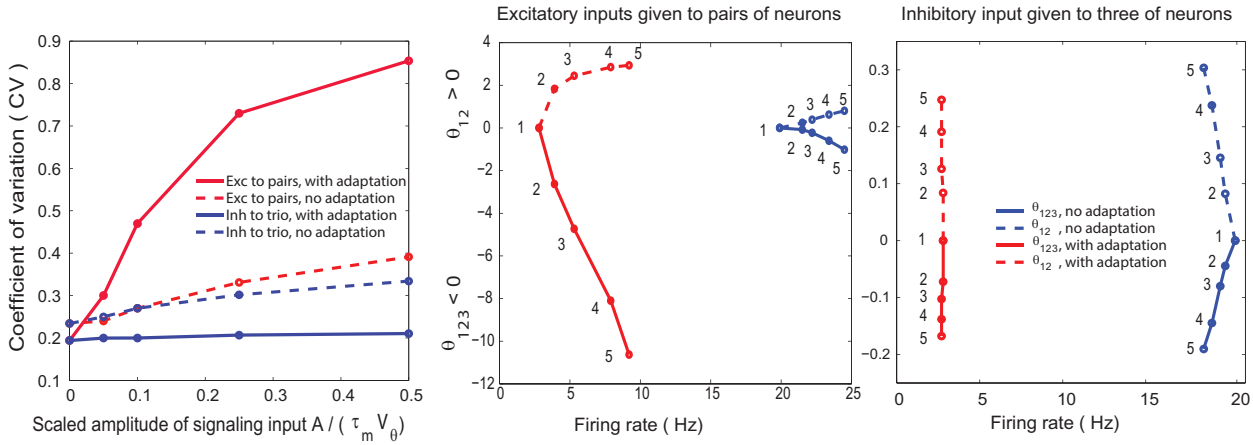

Fig S6. Adaptation can change the coefficient of variation (CV) and interaction parameters in both excitatory-to-pairs and inhibitory-to-trio motifs. (Left) By increasing the amplitude of excitatory inputs given to pairs of neurons, CV can reach high values with the adaptation (red solid line) compared to without the adaptation (left, red dashed line). The adaptation under the inhibitory-to-trio motif decreases the CV (blue lines). (Middle and right) The graph shows θ_{12} in the positive domain of the ordinate and θ_{123} in the negative domain, as a function of the neurons' firing rate for excitatory-to-pairs (middle) and inhibitory-to-trio (right) motifs. The red lines present the interactions in the presence of adaptation while the blue is for no adaptation. The numbers $1, 2, 3, 4, 5$ assigns to the amplitude of common inputs $A = 0, 10, 20, 50, 100$, respectively. The firing rate is decreased in the presence of adaptation for both excitatory-to-pairs and inhibitory-to-trio motifs (compare each number in blue and red graphs) and even when there is no common input (compare number 1 in blue with red graphs). Nevertheless, θ_{12} and θ_{123} of excitatory-to-pairs motif are increased for each amplitude of common inputs (middle, compare each number in blue and red graphs). The inhibitory-to-trio motif with adaptation induces smaller interactions, compared with the non-adaptation case (right, compare each number's interaction in blue with red). Fixed parameters are $\Delta = 10$ ms, $D/\tau_m V_\theta^2 = 0.0002$, $\tau_m = 10ms$, $V_\theta = 20$ and $\lambda = 5$ Hz.

SV: Does mixing motifs, excluding the excitatory-to-pairs motif, can generate strong negative triple-wise interactions?

We have four architectures, common inputs given to pairs or trio of postsynaptic neurons with either common excitatory or common inhibitory inputs. Based on our analysis in this paper about common input and also recurrent activity, we know that the motif of excitatory-to-pairs if exists, can induce strong interactions of $\theta_{12} > 0$ and $\theta_{123} < 0$. However, we can ask the following question: By mixing the other motifs (i.e, inhibitory-to-trio, excitatory-to-trio and inhibitory-to-pairs), can we reach the large magnitude of $\theta_{12}(> 0)$ and $\theta_{123}(< 0)$? For this purpose, we mix the three motifs, two by two to see whether such a case exists. We mix two motifs with firing rates of $\lambda_1 = 5Hz$ and $\lambda_2 = 5Hz$ for a range of common input's amplitude in each motif (i.e. from $A = 0$ to $A = 50$) (Fig. S7). The result for mixing excitatory input

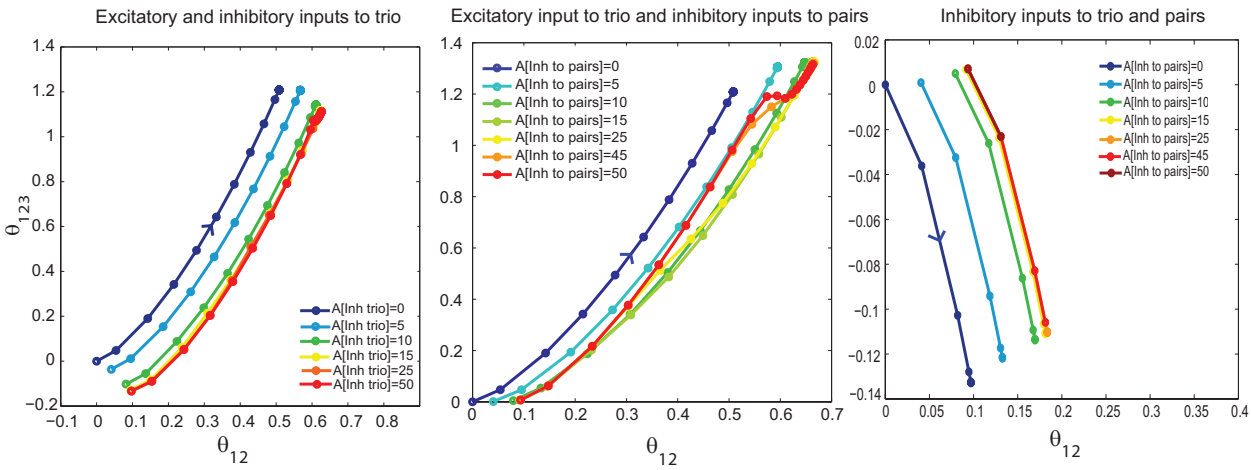

Fig S7. Mixing the motifs of excitatory-to-trio, inhibitory-to-trio, and inhibitory-to-pairs. The result of interactions are shown for the mixture of excitatory and inhibitory inputs each given to trio (left), excitatory-to-trio with inhibitory-to-pairs (middle), and inhibitory-to-trio with inhibitory-to-pairs (right). In the mixture, the amplitude of common input in each motif varied from 0 (i.e. the motif is inactive) to 50 (arrow shows increasing the amplitude in each panel), and the interactions (θ_{123} and θ_{12}) are calculated for different amplitudes of motifs. Along each colored line (following the arrow in each panel), one motif's amplitude is fixed while the other mixed motif's amplitude varied from 0 to 50. For example in the left panel, for each colored line, the amplitude of inhibitory-to-trio is fixed (see color box), while the amplitude of excitatory-to-trio has changed along the line from 0 to 50. The color box shows in each panel which motif has the fixed amplitude. Fixed parameters are $\Delta = 10$ ms, $D/\tau_m V_\theta^2 = 0.003$, $\tau_m = 10ms$, $V_\theta = 5$mV, $\lambda_1 = 5\,$Hz and $\lambda_2 = 5\,$Hz.

to trio with inhibitory input to trio is shown in the left panel. Although the inhibitory-to-trio architecture induces negative triple-wise interactions, we observed mostly positive triple-wise interactions when it is mixed with excitatory-to-trio. The only region that the mixing triple-wise interaction are negative is under weak or near zero amplitudes of excitatory inputs given to trio (Fig. S7, left). The mixture of excitatory-to-trio and inhibitory-to-pairs induce positive triple-wise interactions (Fig. S7, middle) as expected from its each individual motif. The result for mixing inhibitory input given to trio and inhibitory input given to pairs is shown in the right panel. Although the inhibitory-to-pairs induces positive triple-wise interactions, when it is mixed with inhibitory-to-trio, for most amplitudes of mixing, negative triple-wise interactions are observed (Fig. S7, right). The interactions induced in this mixing is in the range of inhibitory-to-trio's interaction which is not strong and does not interfere with excitatory-to-pairs' range of interactions. The triple-wise and pairwise interactions in each case, are saturated for a specific amplitude of common inputs.

SVI: Why the information-geometric measure is suitable for quantifying pairwise and higher-order interactions

In this study, we analyze the pairwise and higher-order (i.e., triple-wise) statistical dependency among the two and three neurons by using the information-geometric measure that is obtained from the neurons' interactions expressed in the form of exponential family distribution (Eqs. 13 and 15) (Amari, 2009b; Martignon et al., 2000; Nakahara and Amari, 2002). Here we explain why we employ this measure instead of the other alternatives.

For a system with two neurons (Eq. 13), the information-geometric measure θ_{12} of the pairwise interaction expresses the dependency of the two neurons, similarly to the Pearson's correlation coefficients computed for the binary patterns. Although the zero values of both measures indicate that the two neurons are uncorrelated, their mathematical expressions are different. The information-geometric measure of the pairwise interaction can take an unconstrained real-value whereas the Pearson's correlation coefficients are bounded in -1 and 1.

It is known that the information-geometric measures have a statistically suitable property to extract the pairwise and higher-order dependency of the binary representation of neural activity (Amari, 2009b; Martignon et al., 2000), compared to conventional measures such as Pearson's correlations, pairwise and triple-wise cross-correlations (Shlens et al., 2006), cumulants (Staude et al., 2010), and mutual information (Amari, 2009b; Martignon et al., 2000; Ohiorhenuan and Victor, 2011). For example, in the two neurons' system, the estimation of θ_{12} from the data is not affected by estimation of the activity rates of individual neurons. This property does not hold for the other measures such as the classical Pearson's correlation coefficient (Amari, 2009b). Namely, the simultaneous estimation of the firing rates and the neurons' dependency measured by the Pearson's correlation coefficients are correlated. On the contrary, the information-geometric measure of the correlation θ_{12} is independent of the estimated firing rates of neurons. In this sense, it quantifies the genuine interaction that we cannot obtain from activities of individual neurons, rather that can be inferred only if we simultaneously observe the two neurons. In information geometry, this property is known as the orthogonality of the interaction to the activity rates of neurons.

For a system with three neurons, the interaction among the three neurons (triple-wise interaction, θ_{123}) is obtained by Eq. 15. Similar to the information-geometric measure of the pairwise interaction, the triple-wise interactions estimated from the spike data are not correlated with the lower-order statistics, namely activity rates of individual neurons and joint activity rates of the two neurons. It quantifies a genuine triple-wise activity of the three neurons that can be inferred only if we observe the three neurons simultaneously.

The meaning of the non-zero triple-wise interaction becomes clear if we compare it with the neural activity that expresses no triple-wise interaction. Suppose the triple-wise interaction is fixed at zero. In that case, the model corresponds to the maximum entropy model that is derived as the most unstructured model when the individual firing rates and joint firing rates of pairs of neurons (equivalently, pairwise correlations) are given. Such a model produces the joint activities or in-activities of all three neurons (i.e., the pattern '111' or '000') with non-zero probabilities, but they occur as a chance coincidence given the neurons' firing rates and pairwise correlations. The non-zero information-geometric measure of the triple-wise interaction θ_{123} precisely measures the deviation of the joint activity or inactivity of the three neurons from this

expected level. If it is positive, the three neurons are jointly active more frequently than the chance level expected from their activity rates and pairwise correlations. If it is negative, the neurons are simultaneously silent more frequently than the chance level.

Comparison of the other methods with the information-geometric measure to calculate the triple-wise (or higher-order) dependency, such as covariance and mutual information (or Kullback-Leibler (KL) divergence), are also summarized in (Amari, 2009b). The former method (covariance) depends on firing rates, and the latter KL-divergence cannot show the sign of correlation. Ohiorhenuan and Victor compared the KL divergence and information-geometric measure, which pointed to the superiority of the information-geometric measure in practice (Ohiorhenuan and Victor, 2011): It can distinguish whether the interactions are less or more synchronous comparing with the chance level while the KL divergence can measure the magnitude only. The information-geometric measure is also easier to calculate the confidence bound, and it was reported that it is robust to spike-sorting errors (Ohiorhenuan and Victor, 2011). In summary, although information-geometric measure needs data to be binarized, its statistical and practical advantages make it suitable for measuring the dependency of neural population.

SVII: How geometrical constraints modify the probability of finding various microcircuits: A simple Monte Carlo approach

We wrote a simple Matlab code, using 10^8 times uniform random number generator. It put points on the XY-plane, in a $2r \times 2r$ square, with $r = 150\mu m$. Then, excluding any point further from the origin (i.e. at $x = 0$ and $y = 0$) than r, we simply have a pool of *random points* all sitting in a circle of radius r= $150\mu m$. Then, picking each pair (or trio), we verified if two (or three) picked points are closer to each other than the critical length of $l_{\mathrm{c}} = 100\mu m$. It simply produces two populations among *all possible choices* of two (or three) points. Out of 4.8×10^7 choices of pairs, we found that the probability of the two points being closer to each other than $l_{\mathrm{c}} = 100\mu m$ was 0.32; and the probability of finding three points, each closer to two others than $100\mu m$ was 0.07.

References

Allen, B. D., Moore-Kochlacs, C., Bernstein, J. G., Kinney, J. P., Scholvin, J., Seoane, L. F., Chronopoulos, C., Lamantia, C., Kodandaramaiah, S. B., Tegmark, M., et al. (2018). Automated in vivo patch-clamp evaluation of extracellular multielectrode array spike recording capability. Journal of neurophysiology, 120, 2182–2200.

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.

Arroyo, S., Bennett, C., and Hestrin, S. (2018). Correlation of synaptic inputs in the visual cortex of awake, behaving mice. Neuron, 99, 1289–1301.

Brette, R. and Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. Journal of neurophysiology, 94, 3637–3642.

Brody, C. D. (1999). Disambiguating different covariation types. Neural Computation, 11, 1527–1535.

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. Journal of computational neuroscience, 8, 183–208.

Brunel, N. and Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. Neural computation, 11, 1621.

Bulsara, A. R., Elston, T. C., Doering, C. R., Lowen, S. B., and Lindenberg, K. (1996). Cooperative behavior in periodically driven noisy integrate-fire models of neuronal dynamics. Physical Review E, 53, 3958.

Burton, S. D. and Urban, N. N. (2015). Rapid feedforward inhibition and asynchronous excitation regulate granule cell activity in the mammalian main olfactory bulb. Journal of Neuroscience, 35, 14103–14122.

Buzsáki, G. and Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. Nature Reviews Neuroscience, 15, 264–278.

Camera, G. L., Rauch, A., Lüscher, H.-R., Senn, W., and Fusi, S. (2004). Minimal models of adapted neuronal response to in vivo–like input currents. Neural computation, 16, 2101–2124.

Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., Hofer, S. B., and Mrsic-Flogel, T. D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. Nature, 518, 399–403.

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).

Curto, C. and Morrison, K. (2019). Relating network connectivity to dynamics: opportunities and challenges for theoretical neuroscience. Current opinion in neurobiology, 58, 11–20.

de Almeida, L., Idiart, M., and Lisman, J. E. (2009). The input–output transformation of the hippocampal granule cells: from grid cells to place fields. Journal of Neuroscience, 29, 7504–7512.

Ganmor, E., Segev, R., and Schneidman, E. (2011). The architecture of functional interaction networks in the retina. Journal of Neuroscience, 31, 3044–3054.

Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F., and Petersen, C. C. (2010). Membrane potential dynamics of gabaergic neurons in the barrel cortex of behaving mice. Neuron, 65, 422–435.

Gerstner, W., Kistler, W. M., Naud, R., and Paninski, L. (2014a). Neuronal dynamics: From single neurons to networks and models of cognition. (Cambridge University Press).

Gerstner, W., Werner, M. K., Naud, R., and Paninski, L. (2014b). Neuronal dynamics: From single neurons to networks and models of cognition. (Cambridge University Press).

Gur, M., Beylin, A., and Snodderly, D. M. (1997). Response variability of neurons in primary visual cortex (v1) of alert monkeys. Journal of Neuroscience, 17, 2914–2920.

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.

Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? IEEE transactions on neural networks, 15, 1063–1070.

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.

Jolivet, R., Kobayashi, R., Rauch, A., Naud, R., Shinomoto, S., and Gerstner, W. (2008). A benchmark test for a quantitative assessment of simple neuron models. Journal of neuroscience methods, 169, 417–424.

Kobayashi, R., Kurita, S., Kurth, A., Kitano, K., Mizuseki, K., Diesmann, M., Richmond, B. J., and Shinomoto, S. (2019). Reconstructing neuronal circuitry from parallel spike trains. Nature communications, 10, 1–13.

Kulkarni, J. E. and Paninski, L. (2007). Common-input models for multiple neural spike-train data. Network: Computation in Neural Systems, 18, 375–407.

Ladenbauer, J., McKenzie, S., English, D. F., Hagens, O., and Ostojic, S. (2019). Inferring and validating mechanistic models of neural microcircuits based on spike-train data. Nature communications, 10, 1–17.

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.

Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., et al. (2015). Reconstruction and simulation of neocortical microcircuitry. Cell, 163, 456–492.

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.

Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A. M., et al. (2014). A mesoscale connectome of the mouse brain. Nature, 508, 207–214.

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.

Okun, M. and Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. Nature neuroscience, 11, 535–537.

Ostojic, S. and Brunel, N. (2011). From spiking neuron models to linear-nonlinear models. PLoS Comput Biol, 7, e1001056.

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.

Perkel, D. H., Gerstein, G. L., and Moore, G. P. (1967). Neuronal spike trains and stochastic point processes: Ii. simultaneous spike trains. Biophysical journal, 7, 419–440.

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

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature, 454, 995–999.

Poulet, J. et al. (2019). Multiple two-photon targeted whole-cell patch-clamp recordings from monosynaptically connected neurons in vivo. Frontiers in synaptic neuroscience, 11, 15.

Poulet, J. F. and Petersen, C. C. (2008). Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. Nature, 454, 881–885.

Rauch, A., La Camera, G., Luscher, H.-R., Senn, W., and Fusi, S. (2003). Neocortical pyramidal cells respond as integrate-and-fire neurons to in vivo–like input currents. Journal of neurophysiology, 90, 1598–1612.

Risken, H. and Eberly, J. (1985). The fokker-planck equation, methods of solution and applications. Journal of the Optical Society of America B Optical Physics, 2, 508.

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

Shinomoto, S., Sakai, Y., and Funahashi, S. (1999). The ornstein-uhlenbeck process does not reproduce spiking statistics of neurons in prefrontal cortex. Neural Computation, 11, 935–951.

Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., Litke, A. M., and Chichilnisky, E. (2006). The structure of multi-neuron firing patterns in primate retina. Journal of Neuroscience, 26, 8254–8266.

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.

Staude, B., Rotter, S., and Grün, S. (2010). Cubic: cumulant based inference of higher-order correlations in massively parallel spike trains. Journal of computational neuroscience, 29, 327–350.

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. Nature, 571, 361–365.

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.

Tatsuno, M. and Okada, M. (2004). Investigation of possible neural architectures underlying information-geometric measures. Neural computation, 16, 737–765.

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.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., and Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. Journal of neurophysiology, 93, 1074–1089.

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

Vidne, M., Ahmadian, Y., Shlens, J., Pillow, J. W., Kulkarni, J., Litke, A. M., Chichilnisky, E., Simoncelli, E., and Paninski, L. (2012). Modeling the impact of common noise inputs on the network activity of retinal ganglion cells. Journal of computational neuroscience, 33, 97–121.

Volgushev, M., Ilin, V., and Stevenson, I. H. (2015). Identifying and tracking simulated synaptic inputs from neuronal firing: insights from in vitro experiments. PLoS computational biology, 11.

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.

Xu, C. S., Januszewski, M., Lu, Z., Takemura, S.-y., Hayworth, K., Huang, G., Shinomiya, K., Maitin-Shepard, J., Ackerman, D., Berg, S., et al. (2020). A connectome of the adult drosophila central brain. bioRxiv.

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.