Journal of Theoretical Biology 501 (2020) 110310

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

# Entropy, mutual information, and systematic measures of structured spiking neural networks

Wenjie Li [a], Yao Li [b,1,*]

[a] Department of Mathematics and Statistics, Washington University, St. Louis, MO 63130, USA
[b] Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA 01002, USA

ARTICLE INFO

*Article history:*
Received 7 October 2019
Revised 23 April 2020
Accepted 27 April 2020
Available online 19 May 2020

*Keywords:*
Neural field models
Entropy
Mutual information
Degeneracy
Complexity

ABSTRACT

The aim of this paper is to investigate various information-theoretic measures, including entropy, mutual information, and some systematic measures that are based on mutual information, for a class of structured spiking neuronal networks. In order to analyze and compute these information-theoretic measures for large networks, we coarse-grained the data by ignoring the order of spikes that fall into the same small time bin. The resultant coarse-grained entropy mainly captures the information contained in the rhythm produced by a local population of the network. We first show that these information theoretical measures are well-defined and computable by proving stochastic stability and the law of large numbers. Then we use three neuronal network examples, from simple to complex, to investigate these information-theoretic measures. Several analytical and computational results about properties of these information-theoretic measures are given.



## 1. Introduction

There has been a long history of researchers using information theoretic measures, such as entropy and mutual information, to study activities of neurons (Strong et al., 1998; Nemenman et al., 2004; Borst et al., 1999). It is important to understand how neuronal networks, including our brains, encode and decode information. It is well known that neurons transmit information by time series of spike trains. A common approach to estimate neuronal entropy is to divide the time series of spike trains into a collection of binary "words". More precisely, the time axis is divided into many time windows with $m$ "sub-windows". A sub-window takes value 1 if a spike is recorded in it and 0 otherwise. This gives a binary "word" with $m$ "letters". Entropy is then estimated through frequencies of those "words".

However, estimating entropy becomes difficult for larger neuronal networks. If one considers the time series generated by each neuron separately, then one needs to consider all possible values of a large vector of "words", which grows exponentially with the network size. If one considers the time series of all spikes produced by the neuronal network, the time window has to be extremely small to avoid two spikes falling into the same bin. Either approach makes a practical sample size much smaller than the number of possible configurations. This makes estimating entropy very difficult, in spite of many results on estimations in the undersampled regime (Nemenman et al., 2004; Strong et al., 1998; Strong et al., 1998).

The first aim of this paper is to study information theoretic measures of a structured neural network model introduced in Li et al. (2019) and Li and Hui (2019). Neurons in this network have integrate-and-fire dynamics. Both the configuration of neurons and the rule of interactions among neurons are simplified to make the model mathematically and computationally tractable. It was shown in Li et al. (2019) and Li and Hui (2019) that this model is still able to produce a rich dynamics of spiking patterns. In particular, this model can produce multiple firing events (MFEs), which are partially synchronized spiking activities that have been observed in other more realistic models (Rangan and Young, 2013; Aaditya and Young, 2013; Chariker and Young, 2015; Chariker et al., 2016). The only difference is that postsynaptic neurons in this paper are given by a *fixed* connection graph rather than decided on-the-fly.

To study information theoretical measures for large networks, we need to consider the coarse-grained entropy instead. The idea is still to divide a time series of spikes into many time windows and construct "words". But different from traditional approaches, here we do not distinguish the order of spikes that fall into the same time window. This method has some similarities to the multiscale entropy analysis used in many applications (Costa et al., 2005). In

* Corresponding author.
*E-mail addresses:* li.wenjie@wustl.edu (W. Li), yaoli@math.umass.edu (Y. Li).
[1] Yao Li is partially supported by NSF DMS-1813246.

https://doi.org/10.1016/j.jtbi.2020.110310
0022-5193/

a large neuronal network, the spike count in a time window can be large. Hence we use a partition function to further reduce the total number of possible "words". In addition, we prove a law of large numbers of spike counts, which says that the coarse-grained entropy defined in this paper is both well-defined and computable.

The biological motivation of our definition is that MFEs produced by a neuronal network can have fairly diverse spiking patterns (See Fig. 5 and Fig. 6). In addition, it has been argued that some information in the brain is indeed communicated through resonance or synchronization of the Gamma oscillation (Hahn et al., 2018; Hahn et al., 2014), which is believed to be modeled by MFEs in neuronal network models (Aaditya and Young, 2013; Chariker and Young, 2015). This prompts us to study information contained in those spiking volleys. By collecting spike counts in time windows, we are able to obtain the uncertainty of a spiking pattern. Heuristically, if a spiking pattern is completely homogeneous, it contains little information from a coarse-grained sense, as the spike count in a time window has little variation. Same thing happens if the spiking activity is completely synchronized, at which we have "all-or-none" spike count in a time window. In contrast, the spiking pattern contains more information if it consists of MFEs, which are only partially synchronized and have high variation in the degree of synchronizations. This is confirmed by our numerical study.

The definition of coarse-grained entropy can be extended to multiple local populations. This gives the concept of mutual information, which measures the amount of information shared by two local populations of a neuronal network. To numerically study the mutual information, we introduce three cortex models, from simple to complicated. The first model only has two interconnected hypercolumns, with no geometry structure. The second model aims to describe two layers of a piece of the cortex, each of which consists of many hypercolumns. We are interested in the effect of feedforward and feedback connections expressed in information theoretical measures. The third model is about layer 4 and layer 6 of the primary visual cortex. In addition to hypercolumns, there are orientation columns in both layers and long range connections in layer 6. Control parameters are magnitudes of feedforward connections, feedback connections, and long range connections. Our simulation shows that suitable feedforward and feedback connections can enhance mutual information between the two layers.

The second aim of this paper is to quantify some systematic measures, such as degeneracy and complexity, for spiking neuronal networks. These systematic measures are proposed in the study of systems biology (Gerald and Gally, 2001; Rangan and Young, 1999; Whitacre, 2010) and quantified for ODE-modeled networks in our earlier work (Li et al., 2012; Li and Yi, 2016). Both degeneracy and complexity can be measured by a linear combination of mutual information between components of a network. The introduction of coarse-grained entropy makes these systematic measures both well-defined and computable.

Biologically speaking, the degeneracy measures the ability of structurally different components of a neuronal network to perform similar functions on a certain target. And the (structural) complexity measures how different components in a neuronal network functionally depend on each other. In this paper, degeneracy and complexity are defined using the coarse-grained entropy. We also prove that a neuronal network with high degeneracy must be (structurally) complex. Finally, the dependency of degeneracy and complexity on certain network parameters is studied for our cortex models.

The organization of this paper is as follows: Section 2 defines our structured spiking neuronal network and three cortex models that are used in later numerical studies. Section 3 defines the coarse-grained entropy and proves the law of large numbers of spike counts, which implies that the coarse-grained entropy is well-defined and computable. Section 4 and 5 study mutual information and systematic measures, respectively. Section 6 is the conclusion.

## 2. Structured spiking neural network model

It is well-known that the cerebral cortex has many substructures. In particular, a functional organization called cortical column or hypercolumn is believed to be the "functional unit of information processing". Neurons in the same hypercolumn usually have similar receptive fields. In the visual cortex, a hypercolumn can be further divided into many orientation columns. Each orientation column only responds to stimulations with a certain orientation. This motivates us to propose a structured spiking neural network model that has two scales at the level of individual neurons and hypercolumns, respectively.

### 2.1. Network description

We consider a large population of neurons that is divided into many local structures (hypercolumns or orientational hypercolumns), called local populations. The following assumptions are made in order to describe the neuronal activity of this population by a mathematically tractable Markov process. Note that the first three assumptions are identical to those in our earlier papers (Li et al., 2019; Li and Hui, 2019).

- The membrane potential of a neuron takes finitely many discrete values.
- External synaptic input to each neuron is modeled by an independent Poisson process. Rates of Poisson processes are identical in the same local population.
- A neuron spikes when its membrane potential reaches a certain threshold. A post-spike neuron stays in a refractory state for an exponentially distributed random time.
- The set of postsynaptic neurons is given by a graph $G$.

The detailed description of this model is divided into the following aspects.

**Neuron indices.** Consider a neuronal network model with $K$ local populations, denoted by $L_1, \cdots, L_K$. Each local population has $N_E$ excitatory neurons and $N_I$ inhibitory neurons. A type $Q$ neuron in local population $L_k$ is denoted as neuron $(k, n, Q)$, where $k \in \{1, \cdots, K\}, Q \in \{E, I\}$, and $n \in \{1, \cdots, N_Q\}$. The triplet $(k, n, Q)$ is called the *label* of a neuron. We further assign an integer-valued *index*, denoted by $\mathrm{id}_{(k,n,Q)}$, to neuron $(k, n, Q)$, such that

$$\mathrm{id}_{(k,n,Q)} = (k-1)(N_E + N_I) + n + N_E \mathbf{1}_{\{Q=E\}}.$$

In other words, id is a function from the set of labels, denoted by $\mathcal{L} = \{(k, n, Q) \,|\, 1 \leqslant k \leqslant K, 1 \leqslant n \leqslant N_Q, Q \in \{E, I\}\}$, to the set of indices, denoted by $\mathcal{ID} = \{1, \cdots, K(N_E + N_I)\}$. We call a neuron with index $i$ "neuron $i$" when it does not lead to confusions. For each index $i \in \mathcal{ID}$, we denote $\mathrm{Label}_i$ as the label of the corresponding neuron. Obviously $\mathrm{Label}_i$ is the inverse function of id from the set of indices to the set of labels. A connection graph $G$ is said to be *static* if the edge set $E$ is fixed, and *random* if $E$ is updated after each spike.

**Connection graph.** The connection graph of the neural network is a directed graph $G = (V, \mathcal{E})$ with $V = \mathcal{ID}$. The edge set $\mathcal{E}$ is a collection of ordered pairs of neurons such that

$$\mathcal{E} \subset \{(pre, post) \,|\, pre \in V, post \in V, pre \neq post\},$$

where *pre* and *post* are indices of presynaptic and postsynaptic neurons, respectively. For each $(pre, post) \in E$, when neuron *pre* fires a spike, its postsynaptic neuron *post* receives the spike and changes

its membrane potential. The collection of postsynaptic neurons (resp. presynaptic neurons) of a neuron with index $i$ is denoted by $\mathrm{Po}(i)$ (resp. $\mathrm{Pr}(i)$).

**Single neuron state.** We denote the membrane potential of a neuron with label $i \in \mathcal{ID}$ by

$$V_i \in \Gamma := \{-M_r, -M_r+1, \dots c, -1, 0, 1, 2, \dots c, M\} \cup \{\mathcal{R}\},$$

where $M, M_r \in \mathbb{N}_+$ denote the threshold for spiking and the reversal potential, respectively. $\mathcal{R}$ represents the refractory state. When $V_i$ reaches $M$, a neuron fires a spike, and instantaneously moves to the refractory state $\mathcal{R}$. At the refractory state, a neuron stays inactive for an exponentially distributed amount of time with mean $\tau_{\mathcal{R}} > 0$. After that, $V_i$ is reset to 0.

**External drive.** Sources of stimuli that a neuron receives can be divided into external drives and postsynaptic kicks from in-network neurons. The external drive comes from outside of the neuronal network in this model, either from sensory input or from other parts of the brain. We assume that neurons of the same type in a local population receive external drive with the same rate, and we model external drives to excitatory and inhibitory neurons in local population $L_l$ by Poisson kicks with rates $\lambda_E^l, \lambda_I^l > 0$ respectively, for $l \in \{4, 6\}$. More precisely, assume the label of neuron $i \in \mathcal{ID}$ is $\mathrm{Label}_i = (k, n, Q)$. Then the time that neuron $i$ receives external kicks is given by a Poisson process with rate $\lambda_Q^k$. If $V_i \neq \mathcal{R}$ when neuron $i$ receives an external kick, $V_i$ immediately increases by 1.

**Neuron spikes and postsynaptic kicks.**

When neuron $i$ fires a spike immediately after $V_i$ reaches the threshold, all postsynaptic neurons in $\mathrm{Po}(i)$ receive a postsynaptic kick. The effect of a postsynaptic kick is delayed by an i.i.d. exponentially distributed random time with mean $\tau_E, \tau_I > 0$, for E, I kicks, respectively. When the delay is over, the kick takes effect instantaneously to neuron $j \in \mathrm{Po}(i)$ if $V_j \neq \mathcal{R}$. After the postsynaptic kick, $V_j$ jumps by $S_{QQ\prime}, Q, Q\prime \in \{E, I\}$, where $Q$ and $Q\prime$ represent the neuron types of $j$ and $i$ respectively. $S_{QQ\prime}$ is the strength of a postsynaptic kick, which is positive if $Q\prime = E$, and negative if $Q\prime = I$. If $Q\prime = E$ and $V_j$ jumps to $\geqslant M$, neuron $j$ jumps to $\mathcal{R}$ instead and fires a spike. If $Q\prime = I$ and $V_j$ jumps to $< -M_r$, it takes value $V_j = -M_r$.

It remains to discuss non-integer $S_{QQ\prime}$. We let $S_{QQ\prime} = p + u$ where $p = \lfloor S_{QQ\prime} \rfloor$ is the largest integer smaller than $S_{QQ\prime}$ and $u$ is a Bernoulli random variable taking value in $\{0, 1\}$ such that $\mathbb{P}[u = 1] = S_{QQ\prime} - p$.

**Markov process**

Because of the delay of postsynaptic kicks, the state of neuron $i$ is denoted by a triplet $(V_i, H_i^E, H_i^I)$. We use $H_i^E$ ($H_i^I$ respectively) to store the number of E (I respectively) kicks received from $\mathrm{Pr}(i)$ that have not taken effects. It is easy to see that the model described above generates a Markov process, denoted by $X_t$, on the state space

$$\mathbf{\Omega} := (\Gamma \times \mathbb{N}_+ \times \mathbb{N}_+)^{K(N_E+N_I)}.$$

A state $\mathbf{x} \in \mathbf{\Omega}$ has the form

$$\mathbf{x} = \{(V_i, H_i^E, H_i^I)\}_{i \in \mathcal{ID}}.$$

The transition probability of $X_t$ is denoted by

$$P^t(\mathbf{x}, \mathbf{y}) = \mathbb{P}[X_t = \mathbf{y} | X_0 = \mathbf{x}].$$

A probability measure $\pi$ on $\mathbf{\Omega}$ is said to be invariant if $\pi = \pi P^t$, where $\pi P^t$ is given by the left operator of $P^t$

$$\pi P^t(\mathbf{x}) = \sum_{\mathbf{y} \in \mathbf{\Omega}} \pi(\mathbf{y}) P^t(\mathbf{y}, \mathbf{x}).$$

Throughout this paper, we denote the conditional probability $\mathbb{P}[\cdot \,|\, \mathrm{law}(X_0) = \mu]$ by $P_\mu[\cdot]$ for the sake of simplicity, where $\mu$ is a probability measure on $\mathbf{\Omega}$, and $\mathrm{law}(X_0) = \mu$ means the initial distribution of $X_t$ is $\mu$.

### 2.2. Visual cortex models I-III

We use the following three visual cortex models, from simple to complicated, to demonstrate numerical results in this paper. Model parameters are largely consistent with earlier modeling result (Li et al., 2019; Chariker et al., 2016). In particular, (Chariker et al., 2016) explains how model parameters are determined based on existing modeling work and experimental studies. The first model (**Model I**) only has two local populations, one feedforward layer and one feedback layer. No spatial factor is considered in this model. The second model (**Model II**) has one feedforward layer and one feedback layer. Each layer consists of 16 local populations (hypercolumns). Each neuron in **Model II** has a coordinate. And the connection graph $G$ is generated according to locations of neurons. **Model III**, the most complicated model, has the same layers, hypercolumns, and neuron coordinates as in **Model II**. Each hypercolumn of **Model III** has 4 orientational columns. The connection graph $G$ depends on both location and orientation. We use a simpler model to illustrate entropy and mutual information, and discuss the role of mutual information by showing numerical results for more complicated models. **Model III** is mainly used to demonstrate degeneracy and complexity, which are two systematic measures defined on complex biological networks. In all three models, we choose $M = 100, M_r = 66, N_E = 300, N_I = 100$, $\tau_{\mathcal{R}} = 2.5$ ms, $S_{EE} = 5.0, S_{IE} = 2.3, S_{EI} = -3.5$, and $S_{II} = -3.0$ unless further specified. Other parameters including the number of local populations $K$, the connection graph $G$, delay times $\tau_E$ and $\tau_I$, and external drive rates $\lambda_E^1, \lambda_I^1, \cdots, \lambda_E^K, \lambda_I^K$, will be prescribed when introducing each model.

**Model I.** In the first model we have $K = 2$. Two local populations represent the feedforward and feedback layer respectively. We set $\lambda_Q^i = 5000$ for $i = 1, 2$ and $Q = E, I$. The connection graph $G$ is a spatially homogeneous random graph. For each pair of $i, j \in \mathcal{ID}$, let $\mathrm{Label}_i = (k_i, n_i, Q_i)$ and $\mathrm{Label}_j = (k_j, n_j, Q_j)$. The connection probability is divided into three cases: (i) $P_{Q_jQ_i}$ for $(i,j) \in \mathcal{E}$ if $k_i = k_j$ (intra-layer connection), (ii) $P^f_{Q_iQ_j}$ for $(i,j) \in \mathcal{E}$ if $k_i = 1$, $k_j = 2$ (feedforward connection), and $P^b_{Q_jQ_i}$ for $(i,j) \in \mathcal{E}$ if $k_i = 2, k_j = 1$ (feedback connection). In this model, we choose parameters $P_{EE} = 0.15, P_{IE} = 0.5, P_{EI} = 0.5$, and $P_{II} = 0.4, P^f_{EI} = P^f_{II} = P^b_{EI} = P^b_{II} = 0$. In other words inhibitory neurons only connect to neurons in the same layer. We set $P^f_{EE} = \rho_f P_{EE}, P^f_{IE} = \rho_f P_{IE}, P^b_{EE} = \rho_b P_{EE}$, and $P^b_{IE} = \rho_b P_{IE}$. Parameters $\rho_f$ and $\rho_b$ represent the strength of feedforward and feedback connection respectively. Without further specification, synapse delay times are chosen to be $\tau_E = 2.0$ ms and $\tau_I = 4.5$ ms. $\rho_b$ and $\rho_f$ are two main control parameters in **Model I**.

**Model II.** In the second model we have $K = 32$, with the first 16 local populations as hypercolums at the feedforward layer and the rest are hypercolumns in the feedback layer. We set $\lambda_Q^i = 5000$ for $i = 1, 2, \cdots, 15, 16$ and $\lambda_Q^i = 4500$ for $i = 17, 18, \cdots, 31, 32, Q = E, I$. Each neuron has a location coordinate. We assume neurons in a local population form a $10 \times 10$ lattice. At each lattice point, there are three E neurons and one I neuron. We assume one local population represents neurons in one hypercolumn of the visual cortex with a size 0.25 mm. In other words the grid size of this lattice is 25 μm, and the boundary neurons are 12.5 μm away from their nearest local boundary. Local populations in each layer is a $4 \times 4$ array. The connection graph is based on the types and locations of each pair of neurons. See Fig. 1 for the layout of this model.

For each pair of $i,j \in \mathcal{ID}$, let $\text{Label}_i = (k_i, n_i, Q_i)$ and $\text{Label}_j = (k_j, n_j, Q_j)$. The connection radius belongs to one of the three cases: (i) $\{i,j\} \in \mathcal{E}$ with radius $L_{Q_iQ_j}$ if $k_i$and $k_j$ are either both $\leqslant 16$ or both $> 16$ (intra-layer connection), (ii) $\{i,j\} \in \mathcal{E}$ with radius $L^f_{Q_iQ_j}$ if $k_i \leqslant 16, k_j > 16$ (feedforward connection), and (iii) $\{i,j\} \in \mathcal{E}$ with radius $L^b_{Q_iQ_j}$ if $k_i > 16, k_j \leqslant 16$ (feedback connection). Connection radii are $L_{EE} = L_{IE} = 0.15$ mm, $L_{EI} = L_{II} = 0.10$ mm, $L^f_{EI} = L^f_{II} = L^b_{EI} = L^b_{II} = 0.10$ mm.

The actual connection probability is the product of a baseline connection probability $p_{Q_iQ_j}$ (resp. $p^f_{Q_iQ_j}, p^b_{Q_iQ_j}$) as defined in **Model I** and the probability density function of a normal distribution with a standard deviation $L_{Q_iQ_j}$ (resp. $L^f_{Q_iQ_j}, L^b_{Q_iQ_j}$). More precisely, the probability that neurons $i$ and $j$ at the same layer are connected is

$$p_{Q_iQ_j} \frac{\exp\left(-\frac{d^2}{2L^2_{Q_iQ_j}}\right)}{(2\pi L^2_{Q_iQ_j})}$$

where $d$ is the distance between the two neurons in question. Cases of feedforward and feedback connection probabilities are analogous. Baseline connection probabilities are $p_{EE} = 0.01$, $p_{IE} = 0.035, p_{EI} = 0.03, p_{II} = 0.03$. Again, I neurons only have intra-layer connections so $p^f_{EI} = p^f_{II} = p^b_{EI} = p^b_{II} = 0$. The rule of feedforward and feedback connection probability is analogous to those of Model I. We have $p^f_{QE} = \rho^f p_{QE}$ and $p^b_{QE} = \rho^b p_{QE}$ for $Q = I, E$. $\rho^f$ and $\rho^b$ are two control parameters.

**Model III.** Then we add orientation columns into the model in order to model two layers in the primary visual cortex. In this model, each hypercolumn in Model II is further divided into four orientation columns that resemble the pinwheel structure (Hubel, 1995; Kang et al., 2003; Kaschube et al., 2010). The layout of orientation columns is demonstrated in Fig. 2. We assume the visual stimulation is vertical. The external drive rates to orientational columns with preferences 0 deg, 45 deg, 90 deg, and 135 deg are multiplied by coefficient $1.0, 0.6, 0.2$, and $0.6$ respectively.

We also add long-range excitatory connections that hits neurons with the same orientation preference to the feedback layer. The presence of long-range connections is consistent with experimental studies (Stettler et al., 2002; Gilbert and Wiesel, 1989; Malach et al., 1993), which show that many connections are between neurons with the same orientation preference. The connection probability is given by a linear function. For a pair of neurons $i,j \in \mathcal{ID}$, the probability of having long range connection is

$$p_{Q_iQ_j} C_{long}(1 - 0.5d),$$

if $Q_i = E, d > 2L_{EE}$, and neuron $i,j$ have the same orientation preference, where the baseline connection probability $p_{Q_iQ_j}$ is the same as before, $d$ is the distance between the two neurons in question, and $C_{long}$ is a parameter that controls the strength of long range connections. Here we assume that the probability of having long range connections decreases linearly and becomes zero if $d$ is larger than 2 mm.

Other configurations of Model III are identical to those of Model II.

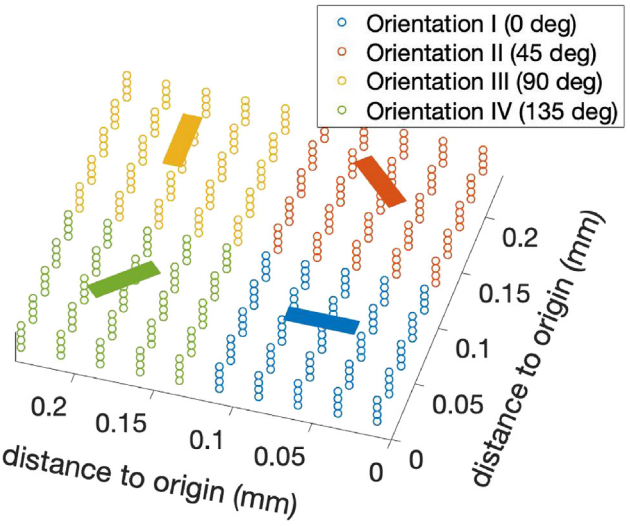
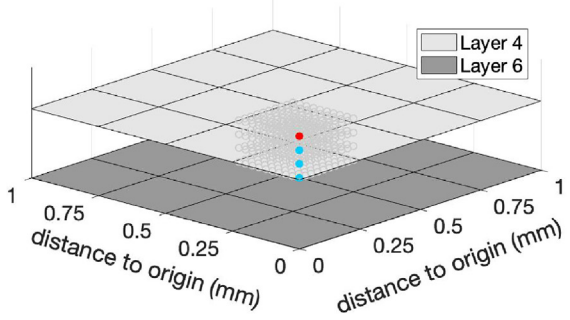



**Fig. 2.** The layout of neurons in one hypercolumn.

### 2.3. Firing rate and spiking pattern in models

In this subsection we will show some simulation results about our visual cortex models. The first result is about the mean firing rate. Fig. 3 gives mean firing rates of **Model I** versus the background drive rate $\lambda^i_E = \lambda^i_I = \lambda$ for $i = 1, 2$ when $\lambda$ increases from 1000 to 8000. We can see an increase of empirical firing rate with the background drive. This is consistent with our previous results in Li et al. (2019) and Li and Hui (2019). The heat map of firing rates in the most complicated model (Model III) is demonstrated in Fig. 4. We can clearly distinguish orientation columns in the heat map. The vertical-preferred orientation columns have the highest firing rate, while the horizontal-preferred orientation columns fire the slowest. See caption of Fig. 4 for the choice of control parameters.

The spiking pattern is another important feature of spiking neuron models, as neurons pass information through spike trains. Similar to many pervious studies, all visual cortex models in our study exhibit partial synchronizations. Due to recurrent excitation, neurons tend to form a series of spike volleys, each of which involves neurons from a proportion of the total population. This



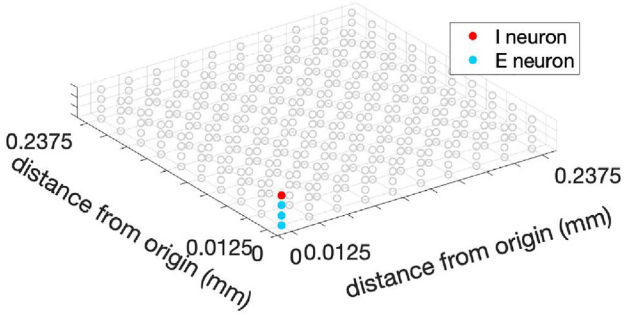




**Fig. 1.** Left: Layout of Model II. Each layer consists of $4 \times 4$ hypercolumns. Right: Geometry of each hypercolumn. Each lattice point is occupied by 3 excitatory neurons and 1 inhibitory neuron.

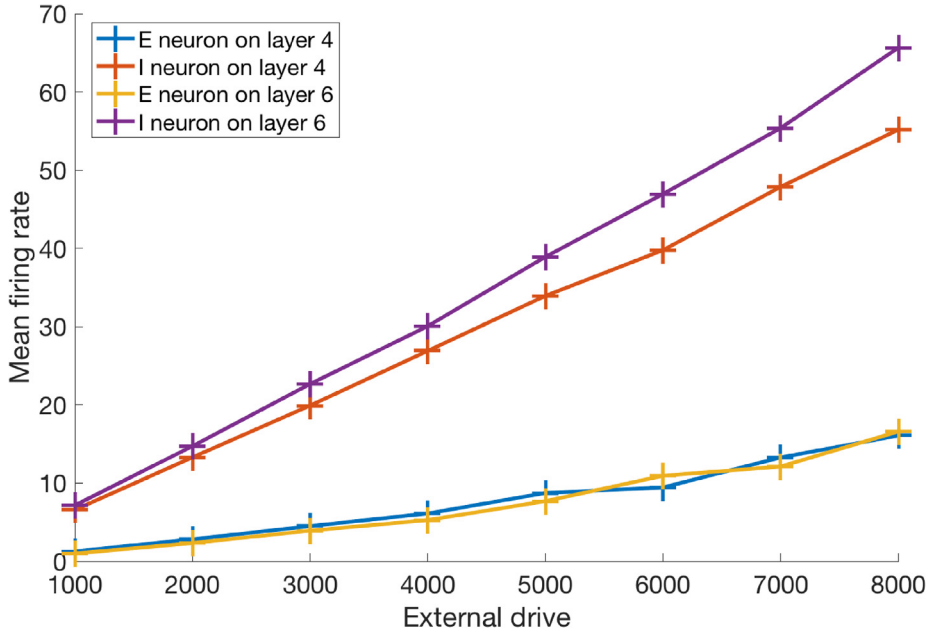



**Fig. 3.** E-population and I-population mean firing rates in layer 4 and 6 when $\lambda$ changes from 1000 to 8000. Parameters are $\rho_f = \rho_b = 0.6, \tau_E = 2.0, \tau_I = 4.5$.

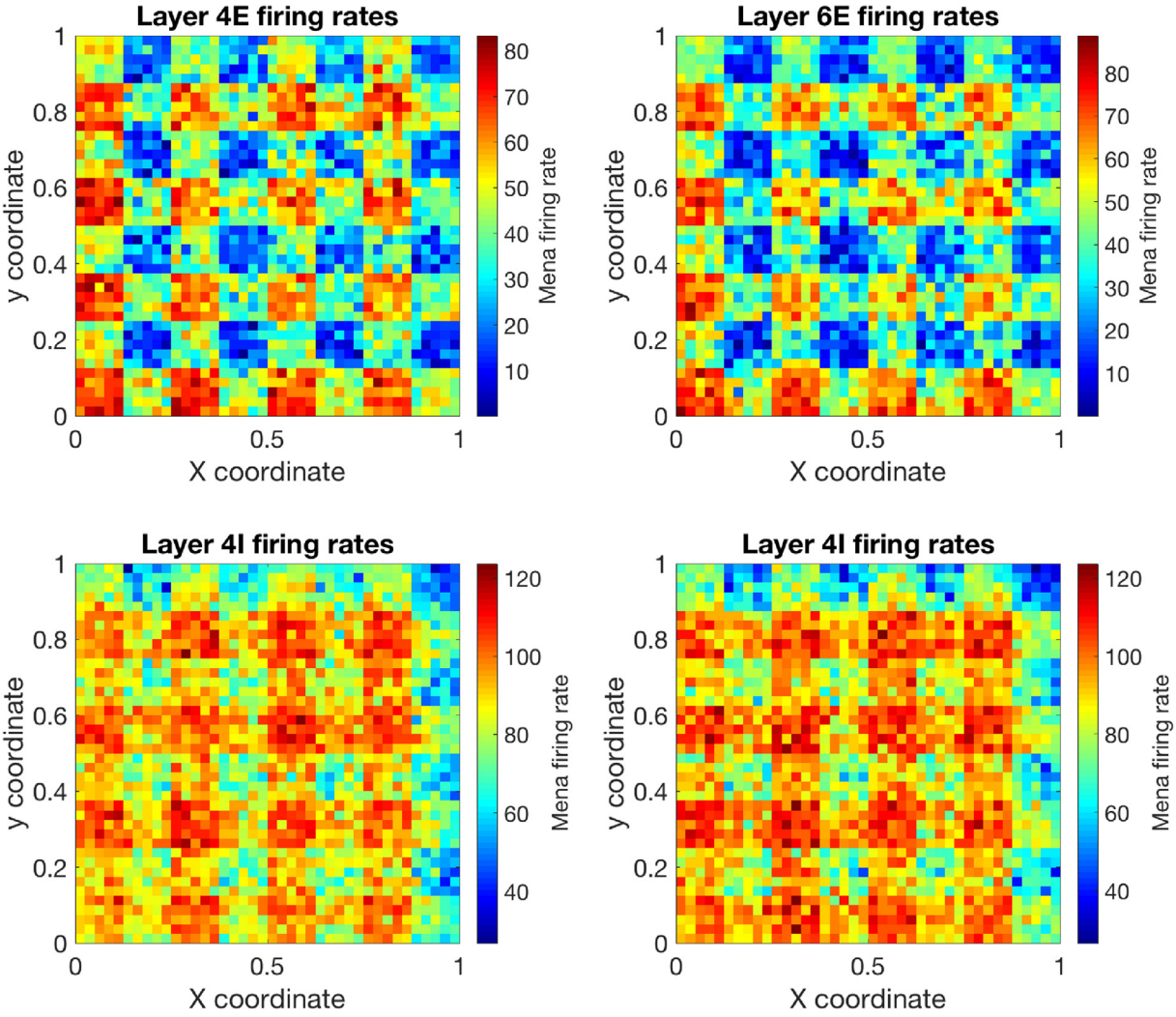



**Fig. 4.** Heat map for E and I firing rates of Model III. Parameters are $\rho_f = \rho_b = 0.6, \tau_E = 2.0, \tau_I = 4.5$.

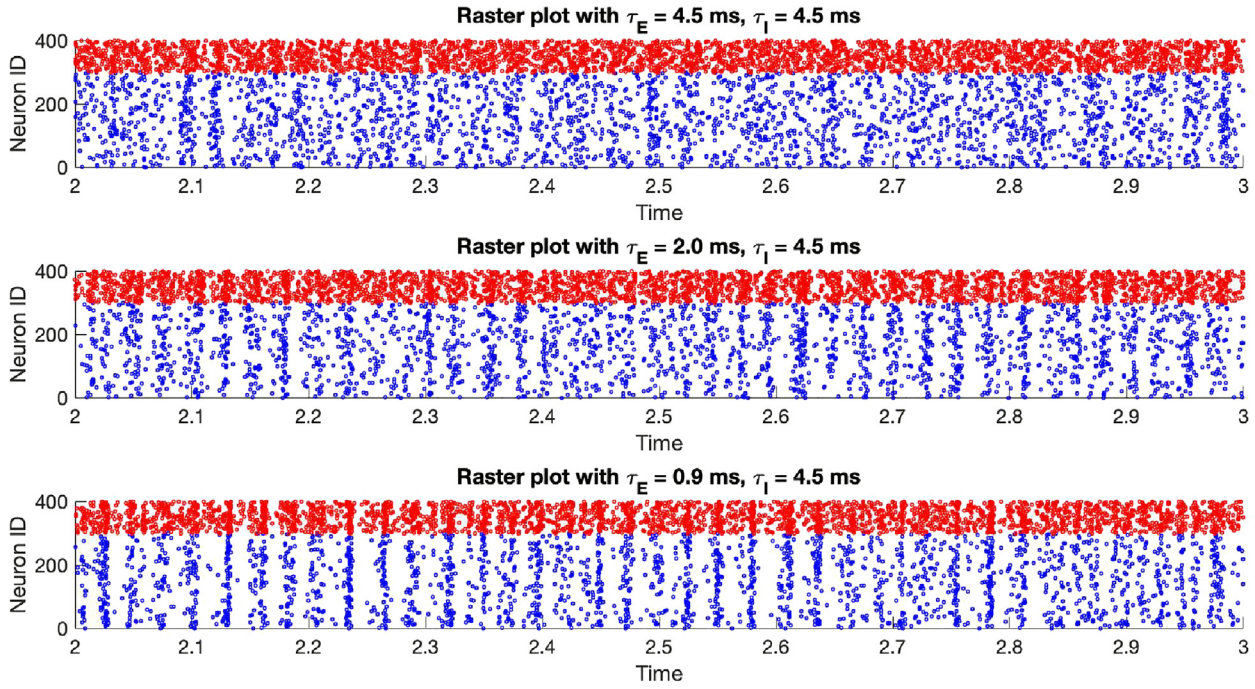



**Fig. 5.** Three raster plots for layer 4 of Model I without feedforward or feedback connection. Top: $\tau_E = 4.5$ ms, $\tau_I = 4.5$ ms. Middle: $\tau_E = 2.0$ ms, $\tau_I = 4.5$ ms. Bottom: $\tau_E = 0.9$ ms, $\tau_I = 4.5$ ms.

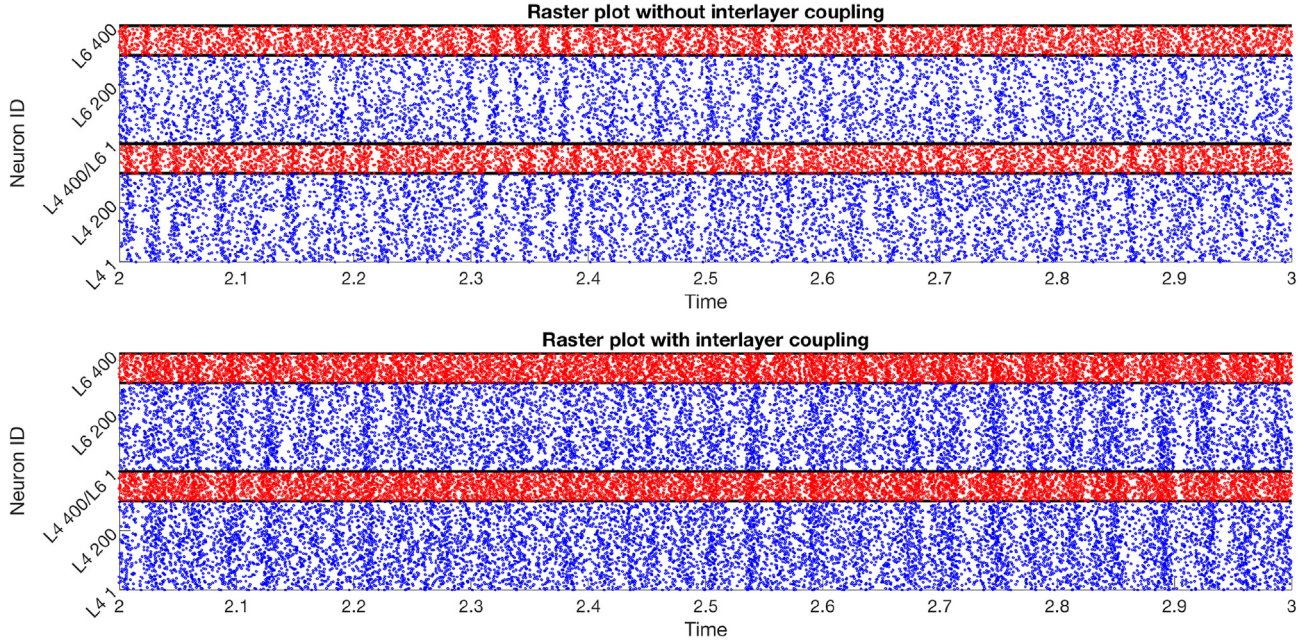



**Fig. 6.** Top. Two raster plots for the (2,2) hypercolumn in layer 4 (bottom) and the same hypercolumn in layer 6 (top). $\rho_b = \rho_f = 0$. $\tau_E = 2.0$ ms, $\tau_I = 4.5$ ms. Time span 2–3 s. Bottom. Same parameters but with intra-layer connections $\rho_b = \rho_f = 0.6$.

phenomenon is known as the multiple firing event (MFE), which is believed to be responsible for the Gamma rhythm (Rangan and Young, 2013; Aaditya and Young, 2013; Chariker and Young, 2015. Fig. 5) demonstrates raster plots produced by Model I. We can see from Fig. 5 that MFEs lie between homogeneous spiking activities and full synchronizations. In addition, the main control parameter of MFE sizes is the ratio $\tau_I/\tau_E$. Higher $\tau_I$-to-$\tau_E$ ratio is responsible for more synchronized spike activities. The mechanism of the dependence of MFE sizes on the synapse delay time is addressed in Li and Hui (2019). Longer $\tau_I$ means E-cascade can last longer time, which contributes to larger MFE sizes. In Fig. 6, we can see raster plots of $(2,2)$ hypercolumn in the feedforward and feedback layer produced by Model II. Top and bottom plots are the cases with and without synapse connections between these two layers, respectively. When the feedforward and feedback connection is turned on, one can see some correlations between MFEs. Later we will use mutual information to quantify such correlation.

## 3. Coarse-grained entropy and law of large numbers

### 3.1. Coarse-grained entropy and coarse-grained information theoretic measures.

Let $T > 0$ be a time window size that is associated to a coarse-grained information theoretic measure. Let

$$\mathcal{SP} = \mathbb{R}_+ \times \mathcal{ID}$$

be the configuration space of neuron spikes. A spike $s \in \mathcal{SP}$ has the form $s = (t, \xi)$, where $t$ is the time of a spike, and $\xi$ is the index of the spiking neuron. Let $S_{[0,T)} = \{s_1, \cdots, s_Z\}$ be set of spikes produced by $X_t$ within the time window $[0, T)$, where $s_i \in \mathcal{SP}$ denotes the $i$-th spike within this time window. (Note that we ignore the order of spikes with in $[0, T)$.) $Z$ is a random variable that represents the total spike count on $[0, T)$. Let

$$\mathcal{SPT} = E \cup \bigcup_{n \geqslant 1} \{\mathcal{SP}\}^n$$

be the configuration space of neural spike trains, where $E$ represents an empty spike train, and $\{\mathcal{SP}\}^n$ is the collection of unordered $n$-sets of elements in $\mathcal{SP}$. Finally, we let $\mathbf{D} = \{0, 1, \cdots, \Sigma\}$ be a dictionary set that consists of countably many distinct integers $0, 1, \cdots, \Sigma$. ($\Sigma$ could be infinity.) A function $\zeta : \mathcal{SPT} \to \mathbf{D}$ is a coarse-grained mapping that maps a spike train to an element in $\mathbf{D}$. For example, the simplest coarse-grained function is the number of spikes of $X_t$ on the interval $[0, T)$. In this case we have $\Sigma = \infty$.

For $i \in \mathbf{D}$, let

$$p_i = \mathbb{P}_\pi[\zeta(S_{[0,T)}) = i]$$

be the probability that the coarse-grained function maps a spike train $S_{[0,T)}$ to $i$ when starting from the invariant probability measure $\pi$. The *coarse-grained entropy* with respect to $T$ and $\zeta$ is

$$H_{T,\zeta} = -\sum_{i \in \mathbf{D}} p_i \log p_i.$$

If there are $q$ coarse-grained functions $\zeta_1, \cdots, \zeta_q$ and a function $F : \mathbb{R}^q \to \mathbb{R}$, an *information theoretical measure* $\mathfrak{M}$ with respect to $T$, $\zeta_1, \cdots, \zeta_q$, and $F$ is given by

$$\mathfrak{M} = F(H_{T,\zeta_1}, \cdots, H_{T,\zeta_q}).$$

**Remark 1.** The reason of defining the coarse-grained entropy is to make entropy and information theoretical measures computable. The classical definition of neural entropy can only allow at most one spike in each bin (time window). In a large neural network, neuronal activities are usually synchronized to some degree. As a result, the necessary time window size quickly becomes too small to be practical.

The main simplification we make is to ignore the order of spikes that are sufficiently close to each other. Needless to say this treatment loses some information. But it makes information theoretical measures more computable for large networks. In addition, we map a spike train within a time window to an integer to further reduce the state space. This is because the state space of naive spike counting is still huge. If we naively count the spikes in each hypercolumn in a model with $K$ local populations, there will be $(N_E + N_I)^K$ possible spike counting results even if we assume a neuron cannot spike twice in a time bin.

The definition of the coarse-grained entropy is very general. In practice, the function $\zeta$ can be given in the following ways to address different features of the neuronal network.

A. Spike counting for a certain local population. If we are interested in the information produced by a certain local population, say local population $k$, we can have

$$\zeta(S_{[0,T)}) = \zeta(\{(t_1, \xi_1) \cdots, (t_Z, \xi_Z)\}) = \sum_{j=1}^{Z} \mathbf{1}_{\{\text{Label}_{\xi_j}(1)=k\}},$$

where $\text{Label}_{\xi_j}(1)$ means the first entry of $\text{Label}_{\xi_j}$, which is the index of the local population of spike $j$. Here one can replace $k$ by either a set of local populations, or restrict it to certain type of neurons.

B. Coarse-grained spike counting. If the state space of spike counting is too big to estimate accurately, we can introduce a partition function $\Theta : \mathbb{Z}_{\geqslant 0} \to \{0, \cdots, d-1\}$, such that $\Theta(n) = i$ if $a_i \leqslant n < a_{i+1}$, where $0 = a_0 < a_1 < \cdots < a_d = \infty$ is a given sequence of numbers called a "dictionary". Then let

$$\zeta(S_{[0,T)}) = \Theta\left(\sum_{j=1}^{Z} \mathbf{1}_{\{\text{Label}_{\xi_j}(1)=k\}}\right).$$

C. Spike counting with delays. If one would like to address time lags of spike activities, the time window $[0, T)$ can be further evenly divided into many sub-windows $[0, T/m), [T/m, 2T/m), \cdots, [(m-1)T/m, T)$. Assume we still adopt the coarse grained spike counting in $B$, we have

$$\zeta(S_{[0,T)}) = \sum_{l=0}^{m-1} d^l \Theta\left(\sum_{j=1}^{Z} \mathbf{1}_{\{\text{Label}_{\xi_j}(1)=k, t_j \in [lT/m, (l+1)T/m)\}}\right).$$

In other words we take consideration of spike counts in each sub-windows. Integer $m$ is said to be the "word length".

### 3.2. Stochastic stability and law of large numbers

The aim of this section is to prove that the law of large numbers holds for entropy. In other words the entropy is computable by running Monte Carlo simulations. To do so, we first need to show the stochastic stability of $X_t$.

Let

$$U(\mathbf{x}) = 1 + \sum_{i \in \mathcal{ID}} (H_i^E + H_i^I)$$

be a function on $\mathbf{\Omega}$. For any signed distribution $\mu$ on the Borel $\sigma$-algebra of $\Omega$, let

$$\|\mu\|_U = \sum_{\mathbf{x} \in \mathbf{\Omega}} U(\mathbf{x}) |\mu(\mathbf{x})|$$

be the $U$-weighted total variation norm, and let

$$L_U(\Omega) = \{\mu \text{on} \mathbf{\Omega} \,|\, \|\mu\|_U < \infty\}$$

be the collection of probability distributions with finite $U$-norm. In addition, for any function $\eta(\mathbf{x})$ on $\Omega$, denote the $U$-weighted supremum norm by

$$\sup_{\mathbf{x} \in \Omega} \frac{|\eta(\mathbf{x})|}{U(\mathbf{x})}.$$

**Theorem 1.** *$X_t$ admits a unique invariant probability distribution $\pi \in L_U(\mathbf{X})$. In addition, there exists positive constants $c_1, c_2$ and $\gamma \in (0, 1)$ such that*

$$\|\mu_1 P^t - \mu_2 P^t\|_U \leqslant c_1 \gamma^t \|\mu_1 - \mu_2\|_U$$

*for any $\mu_1, \mu_2 \in L_U(\Omega)$ and*

$$\|P^t \eta - \pi(\eta)\|_U \leqslant c_2 \gamma^t \|\eta - \pi(\eta)\|_U$$

*for any test function $\eta$ with $\|\eta\|_U < \infty$.*

We skip the proof of Theorem 1 here because it is almost identical to that of Theorem 1 in Li and Hui (2019). The only difference is that postsynaptic neurons in the model of Li and Hui (2019) is chosen randomly *after* a spike, which does not affect the proof. To prove the existence of an invariant probability measure $\pi$ and the exponential convergence to it, we need to construct a Lyapunov function $U(\mathbf{x})$ such that

$$P^h U \leqslant \gamma U + K$$

for some $h > 0, \gamma \in (0,1)$ and $K < \infty$. In addition the "bottom" of $U$, denoted by $C = \{\mathbf{x} \in \Omega \,|\, U(\mathbf{x}) \leqslant R\}$ for some $R > 2K(1-\gamma)^{-1}$, must satisfy the minorization condition, which means there exists a constant $\alpha > 0$ and a probability distribution on $\Omega$, such that $P^h(\mathbf{x}, \cdot) \geqslant \alpha v(\cdot)$ uniformly for $\mathbf{x} \in C$. Then the existence and uniqueness of $\pi$ and the exponential speed of convergence to $\pi$ follows from Li and Hui (2019). We refer to the proof of Theorem 3.1 in Li and Hui (2019) for the full detail.

The next Theorem is about the computability of information-theoretical measures. The way of estimating $H_{T,\zeta}$ is called "plug-in estimate", which is known to be convergent if the spike train $S_{[0,T)}$ is i.i.d. sampled from a certain given probability distribution (Antos and Kontoyiannis, 2001) (also see Paninski (2003)). However, in our case $S_{[0,T)}$ is the spike count generated by a Markov chain, which depends on the sample path of $X_t$ on $[0,T]$. Therefore, one needs to sample $S_{[0,T)}$ by running $X_t$ over many time intervals $[0,T), [T,2T), \cdots$. Hence the result in Antos and Kontoyiannis (2001) does not apply directly. Instead, we need to use the concept of *sample path dependent observable* to show that $H_{T,\zeta}$ is a computable quantity.

**Theorem 2.** *Let $\mathcal{T} = NT$ be the length of a trajectory of $X_t$. For any coarse-grained mapping $\zeta$ and any $i \in \mathbf{D}$, denote the empirical probability of $\zeta = i$ by*

$$\hat{p}_i = \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{\{\zeta(S_{[(j-1)T,jT)})=i\}}(X_t).$$

*We have*

$$\lim_{N\to\infty} \hat{p}_i = \mathbb{P}_\pi[\zeta(S_{[0,T)}) = i] = p_i.$$

**Proof.** The proof of this theorem relies on the concept of *Markov sample path dependent observables*. Let $X_t$ be a Markov process. A function $Y$ is said to be a Markov sample-path dependent observable on an interval $[t_1, t_2)$ if

i. $Y$ is a real-valued function on $C_\Omega([t_1,t_2))$, where $C_\Omega([t_1,t_2))$ is the collection of cadlag paths from $t_1$ to $t_2$ on $\Omega$.
ii. The law of $Y$ only depends on the value of $X_{t_1}$.

Now let $Y_1, Y_2, \cdots, Y_n, \cdots$ be a sequence of sample-path dependent observables on $[0,T), [T,2T), \cdots, [(n-1)T, nT)$ respectively. Since the law of $X_t$ converges to $\pi$ as $t \to \infty$, by Theorem 3.3 in Li and Hui (2019), we have the law of large numbers for $\{Y_n\}$, i.e.,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} Y_n = \mathbb{E}_\pi[Y_1]$$

provided there exists a constant $M < \infty$, such that $\mathbb{E}[Y_n^2 \,|\, X_{(n-1)T} = \mathbf{x}] < M$ uniformly for all $\mathbf{x} \in \mathbf{\Omega}$.

Hence we only need to construct a sequence of sample-path dependent observables $\{Y_n\}$ whose expectations equal to $\hat{p}_i$. Let

$$Y_n = \mathbf{1}_{\{\zeta(S_{[(n-1)T,nT)})=i\}}.$$

It is easy to see that $Y_n$ is Markov because $X_t$ is a Markov process. In addition, $\mathbb{E}[Y_n^2] \leqslant 1$ uniformly because $Y_n$ is an indicator function. Therefore, $\{Y_n\}_{n=1}^{\infty}$ satisfies the law of large numbers. This completes the proof. □

### 3.3. Numerical result and discussion

We use the following two numerical simulations to demonstrate the role of coarse-grained entropy in our network models. Consider model I without feedforward or feedback connections. Without loss of generality we only compute the entropy of the feedforward layer. The synapse delay times are chosen to be $\tau_E = 2$ ms and $\tau_I = 4.5$ ms.

The first test is about the entropy rate per second with increasing word lengths. We use spike counting with delays to define the coarse-grained mapping $\zeta$ (case C in Section 3.1). The duration of each sub-window is $T/m = 5$ ms. The time window size goes through $T = 5$ms (m = 1) to $T = 60$ ms (m = 12). Then we define the partition function $\Theta : \mathbf{Z}_+ \to \{0,1,2,3\}$ with $\Theta(n) = 0$ for $n < 4, \Theta(n) = 1$ for $4 \leqslant n < 8$, $\Theta(n) = 2$ for $8 \leqslant n < 12$, and $\Theta(n) = 3$ for $n \geqslant 12$. Fig. 7 shows the entropy rate (entropy divided by the time window size) versus time window size. The curve is similar to that given in Strong et al. (1998). Starting from word length = 2, and until the word length being too long for estimating entropy, the entropy rate declines linearly. But when a "word" is too long such that the entropy estimation is badly undersampled, the estimated entropy rate declines faster-than-linear, as explained in Strong et al. (1998). We use extrapolation to estimate the entropy rate at the infinite window size limit, as shown in Fig. 3 of Strong et al. (1998).

The second test is on the entropy for different degrees of synchrony. Here we only consider word length = 1 with time window size $T = 15$ ms, with a more refined partition function $\Theta : \mathbb{Z}_+ \to \{0,1,\cdots,45\}$ such that $\Theta(n) = i$ for $3i \leqslant n < 3(i+1)$ if $i = 0, \cdots, 44$ and $\Theta(n) = 45$ if $n \geqslant 135$. The excitatory synapse delay time is set to be $\tau_I = 4.5$ ms. And $\tau_I$-to-$\tau_E$ ratio varies between 0.5 and 3.5. In addition we let $S_{EE} = 6$ to increase the degree of synchrony. The entropy rate versus $\tau_I/\tau_E$ is plotted in Fig. 8 Left. We can see that the entropy reaches a peak in the middle of the tuning curve for both E and I neurons. The peak locations are different for E and I neurons partially due to different population sizes. The heuristic reason is straightforward. When $\tau_I/\tau_E$ is small, the spiking pattern is homogeneous, and the distribution of spike counts in each time window concentrates at a few small numbers. Larger $\tau_I/\tau_E$ means larger MFE sizes, which makes spike

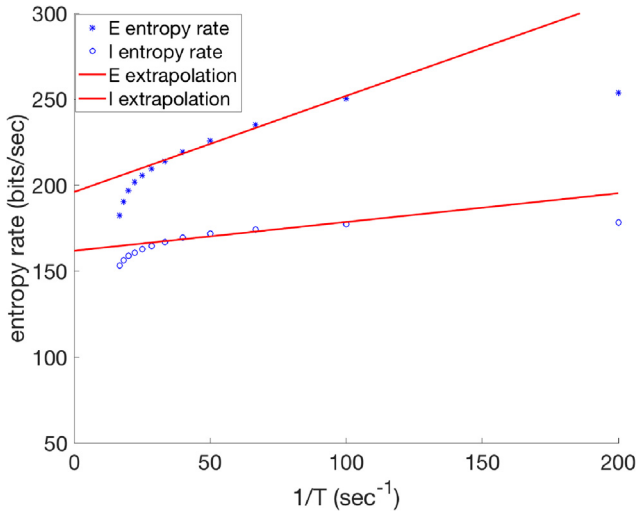



**Fig. 7.** Entropy rate versus time window length. Samples are collected from eight trajectories with length 10000. Linear extrapolation uses points with respect to $m = 2, 3, \cdots, 7$.

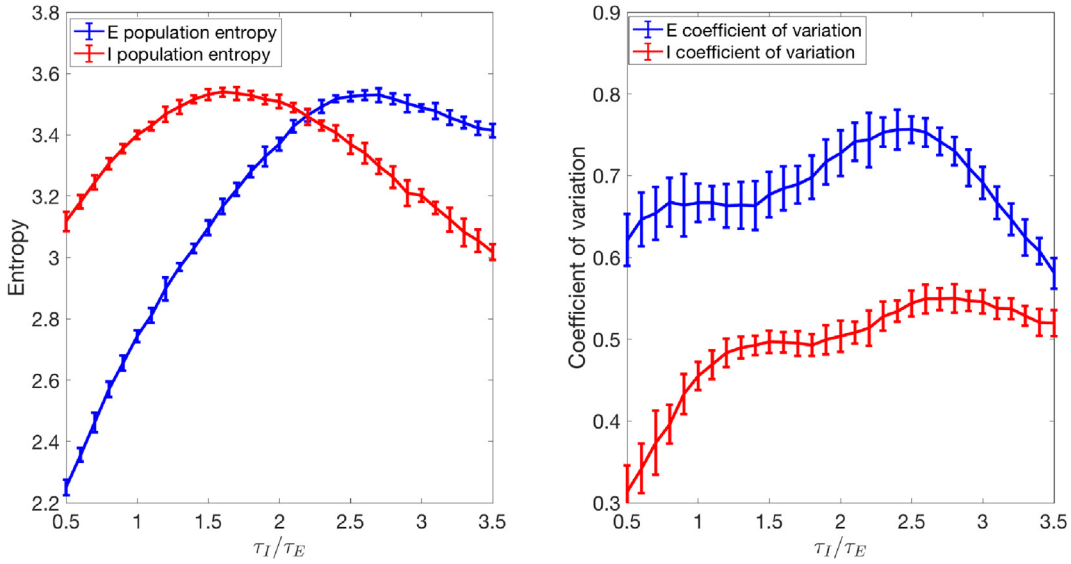



**Fig. 8.** Left: Entropy versus delay time for Model I. Right: Coefficient of variation of MFE sizes versus delay time for Model I. Each simulation runs up to $\mathcal{T} = 50$.

counts in each time window more diverse. But when $\tau_I/\tau_E$ is too large and such that all MFEs are large, such diversity shows slight decreases. This is because when MFEs are too large, many MFEs can use up all neurons in the network. Therefore, our simulation shows that partial synchronizations, rather than full synchronizations, helps a neuronal network to produce more information.

To further confirm the change of the diversity of MFE sizes with $\tau_I$-to-$\tau_E$ ratio, we compute the coefficient of variation of sizes of MFEs in Fig. 8 Right. The coefficient of variation is the ratio of the sample standard deviation to the sample mean, which measures the relative diversity of a sample. In the simulation, we count the number of spikes in each time bin with size 1 ms. If at least 10 spikes fall into a time bin, we say MFE occurs at this time bin. Then we merge consecutive time bins at which MFE occurs, and count sizes of MFEs. As seen in Fig. 8 Right, using the metric of coefficient of variation, we can also see an increase of the MFE diversity when $\tau_I/\tau_E$ increases, and a decrease of the MFE diversity when too big $\tau_I/\tau_E$ makes the network too synchronized. We remark that the coefficient of variation and the entropy are two different metrics of the diversity of MFEs, so curves in Fig. 8 Left and Right are not exactly that same. In particular, the entropy of I population decreases early (at $\tau_I/\tau_E = 1.5$) because I population has a much higher firing rate (Fig. 3). When the partition function $\Theta(n)$ maps many large spike counts ("letters") to 45, the coarse-grained entropy is reduced. Other three curves in Fig. 8 start to decrease when $\tau_I/\tau_E$ reaches about 2.5.

## 4. Mutual information

### 4.1. Mutual information and conditional mutual information

Based on the framework of spike train space defined in Section 3.1, one can also define the mutual information. Let $T > 0$ be a fixed time window size. A *coarse-grained function* $\mathcal{SPT} \to \mathbf{D} = \{0, 2, \cdots, d-1\}$ *with respect population set* $C \subset \{1, \cdots, K\}$ is denoted by $\zeta_C$ if it gives joint spike count distributions from populations that belong to $C$. Here $C = \{C_1, \cdots, C_{|C|}\}$ is a generic subset of $\{1, \cdots, K\}$. We have

$$\zeta_C(S_{[0,T]}) = \sum_{n=1}^{|C|} d^{n-1} \left( \Theta \left( \sum_{j=1}^{Z} \mathbf{1}_{\{\text{Label}_{\tilde{\varsigma}_j}(1) \in C_n\}} \right) \right), \tag{4.1}$$

where $\text{Label}_{\tilde{\varsigma}_j}(1)$ means the first entry (index of local population) of $j$-th spiked neuron, and $\Theta : \mathbb{Z}_{\geqslant 0} \to \{0, \cdots, d-1\}$ is a partition function that maps a spike count to an integer. Here the definition of $\zeta_C$ can also be extended to the case of a particular type of neurons, or the case of spike counts with delays.

Now consider two disjoint sets of local populations $A, B \subset \{1, \cdots, K\}$. The *mutual information* is an information theoretical measure that is given by

$$\text{MI}_{T,\zeta}(A : B) = H_{T,\zeta_A} + H_{T,\zeta_B} - H_{T,\zeta_{A\cup B}}.$$

In other words, the mutual information measures the information shared by populations in $A$ and populations in $B$ when the spike count is measured by the coarse-grained map $\xi$.

The coarse-grained mutual information $\text{MI}_{T,\zeta}(A : B)$ is the mutual information between two spike trains produced by local populations $A$ and $B$ that are processed by coarse-grained function $\zeta_A$ and $\zeta_B$ respectively. By the data processing inequality (Cover and Thomas, 2012), $\text{MI}_{T,\zeta}(A : B)$ is no greater than the actual mutual information between two unprocessed spike trains on $A$ and $B$ respectively. Therefore, the "true" mutual information is larger than our measurement by using coarse-graining functions.

### 4.2. Mutual information in visual cortex models

The following simulations aim to use mutual information to study the correlation in our visual cortex models. We present the following two numerical results.

**Simulation I: Role of feedforward and feedback.** In this study we first consider Model I, which is a feedforward-feedback network with only two local populations. We let the time window size be 5 ms with word length = 1. The partition function $\Theta$ maps $\mathbb{Z}_+$ to $\{0, 1, \cdots, 10\}$ such that $\Theta(n) = i$ for $5i \leqslant n < 5(i+1)$ if $i = 0, 1, \cdots, 9$ and $\Theta(n) = 10$ if $n \geqslant 50$. The mutual information between the feedforward layer and the feedback layer versus the coupling strength is plotted in Fig. 9 Left. We consider three cases: (i) $\rho_f = \rho_b = \rho$, (ii) $\rho_f = \rho, \rho_b = 0$, and (iii) $\rho_f = 0, \rho_b = \rho$. When $\rho$ is small, one can see a clear increase of mutual information between two layers with increasing coupling. And the presence of both feedforward and feedback connections can significantly increase the mutual information between two layers. This is somehow expected: communications between neuronal networks can increase the information shared between them. The same simulation is done in Model II, which has many hypercolumns and geometric structures. The mutual information is measured between $(2, 2)$ hypercolumn of layer 4 and $(2, 2)$ hypercolumn of layer 6. The corresponding parti-

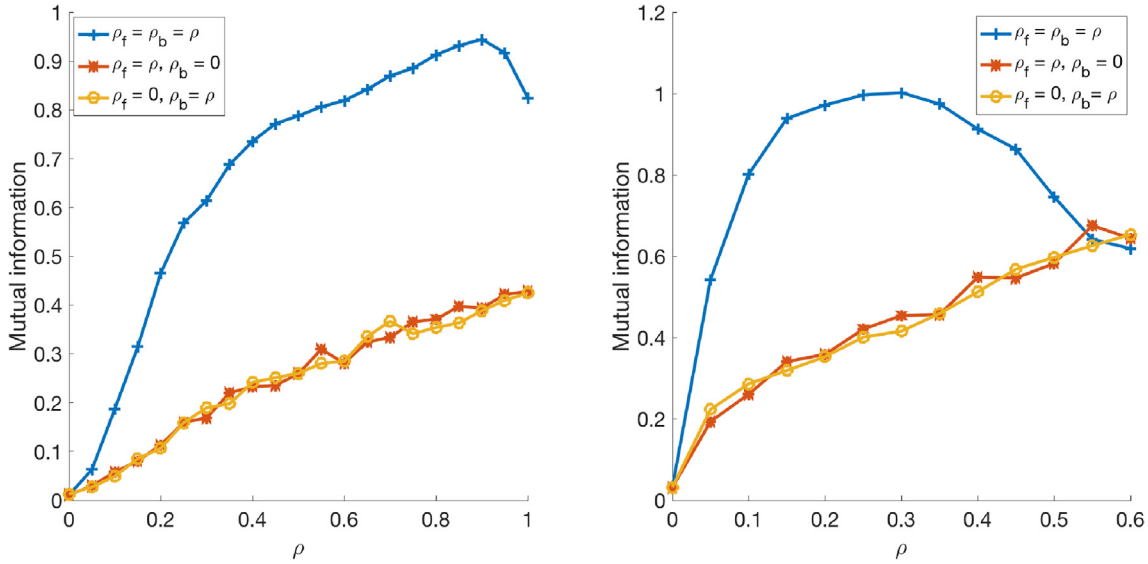



**Fig. 9.** Mutual information versus coupling strength for Model I (left) and Model II (right).

tion function for Model II is $\Theta(n) = i$ for $10i \leqslant n < 10(i+1)$ if $i = 0, 1, \cdots, 9$ and $\Theta(n) = 10$ if $n \geqslant 100$. All other parameters are same as those in Model I. The result is plotted in Fig. 9 Right. We can see the same pattern as in Fig. 9 Left. In the presence of both feedforward and feedback connections, when $\rho$ is large, the network becomes highly synchronized. The mutual information decreases for large $\rho$ in this regime because the entropy decreases in a highly synchronized network, as seen in Fig. 8.

**Simulation II: Mutual information versus distance.** The second simulation considers the dependence of mutual information on the distance between two hypercolumns. We fix the coupling strength as $\rho_f = \rho_b = 0.6$ and study mutual information between hypercolumns for model II and model III. In Fig. 10 Top, mutual information between $(1,1)$ hypercolumn and all other hypercolumns are demonstrated for model II. Model III has orientational columns. In Fig. 10 Bottom, we show the mutual information between the vertical-preferred orientation column in $(1,1)$ hypercolumn and all other orientation columns in model III. The logarithmic scale is used because the entropy at the bottom left corner is much larger than all mutual information. Numbers on grid boxes are the logarithm of mutual information. We can clearly see a decrease of mutual information with increasing distance. This is consistent with our results in Li and Hui (2019) that correlation of MFE sizes decays quickly with the distance. The fast decline of mutual information is also supported by experimental evidence. It is known that MFEs are responsible for the Gamma rhythm in the cortex, which is known to be local in many scenarios (Alex Goddard et al., 2012; Lee et al., 2003; Menon et al., 1996). In addition, in Fig. 10 Bottom, we can also see that orientation columns with the same preferred orientations share higher mutual information than those with orthogonal preferred orientations when two orientational columns are far apart. This phenomenon is more significant in layer 6 because only layer 6 has long-range connections.

## 5. Systematic measures: degeneracy and complexity

### 5.1. Definitions and rigorous results.

Systematic measures, including degeneracy, complexity, redundancy, and robustness, are first proposed in the study of systems biology. When a biological network is too large to be investigated in full detail, systematic measures are used to describe the global characteristics of the network. This idea was first proposed in the study of brains (Tononi et al., 1994). In Rangan and Young (1999), degeneracy and complexity are quantified as linear combinations of mutual information. A simple numerical study was also provided in Rangan and Young (1999), which is a small neural network whose activities are modeled by stationary Gaussian processes. This idea is then generalized by the corresponding author of this paper in Li et al. (2012) to ODE-modeled networks. Neither Gaussian processes nor ODE-modeled networks properly describe the activity of a spiking neuronal network. In this subsection, we use the idea of coarse-grained entropy to study two systematic measures, i.e., degeneracy and complexity, for more realistic spiking neuronal networks. The case of other systematic measures like the redundancy (Rangan and Young, 1999) can be investigated analogously.

Definitions of degeneracy and complexity relies on multivariate mutual information. Still consider a neuronal network with $K$ local populations. Let $\zeta_I$ be a coarse-grained function with respect to $I \subset \{1, \cdots, K\}$. For three sets $A, B, C \subset \{1, \cdots, K\}$, the *multivariate mutual information* $MI(A:B:C)$ is given as

$$\begin{aligned} MI_{T,\zeta}(A:B:C) &= MI_{T,\zeta}(A:C) + MI_{T,\zeta}(B:C) - MI_{T,\zeta}(A\cup B:C) \\ &= H_{T,\zeta_A} + H_{T,\zeta_B} + H_{T,\zeta_C} - H_{T,\zeta_{A\cup B}} - H_{T,\zeta_{B\cup C}} - H_{T,\zeta_{A\cup C}} + H_{T,\zeta_{A\cup B\cup C}}. \end{aligned} \quad (5.1)$$

The following proposition follows immediately from the definition of multivariate mutual information.

**Proposition 1.**

(a) If $\zeta_A(S_{[0,T)}), \zeta_B(S_{[0,T)})$, and $\zeta_C(S_{[0,T)})$ are identical random variables, then $MI_{T,\zeta}(A:B:C) = H_{T,\zeta_A}$.
(b) If $\zeta_A(S_{[0,T)}), \zeta_B(S_{[0,T)})$ are independent from $\zeta_C(S_{[0,T)})$, then $MI_{T,\zeta}(A:B:C) = 0$.

**Proof.** If all three random variables are equal, we have

$$H_{T,\zeta_{A\cup B}} = H_{T,\zeta_A}$$

because the histogram of $\zeta_{A\cup B}(S_{[0,T)})$ is supported by the diagonal set, which equals $\zeta_A(S_{[0,T)})$. Similar argument implies that

$$H_{T,\zeta_{B\cup C}} = H_{T,\zeta_{A\cup C}} = H_{T,\zeta_{A\cup B\cup C}} = H_{T,\zeta_A}.$$

This implies (a).

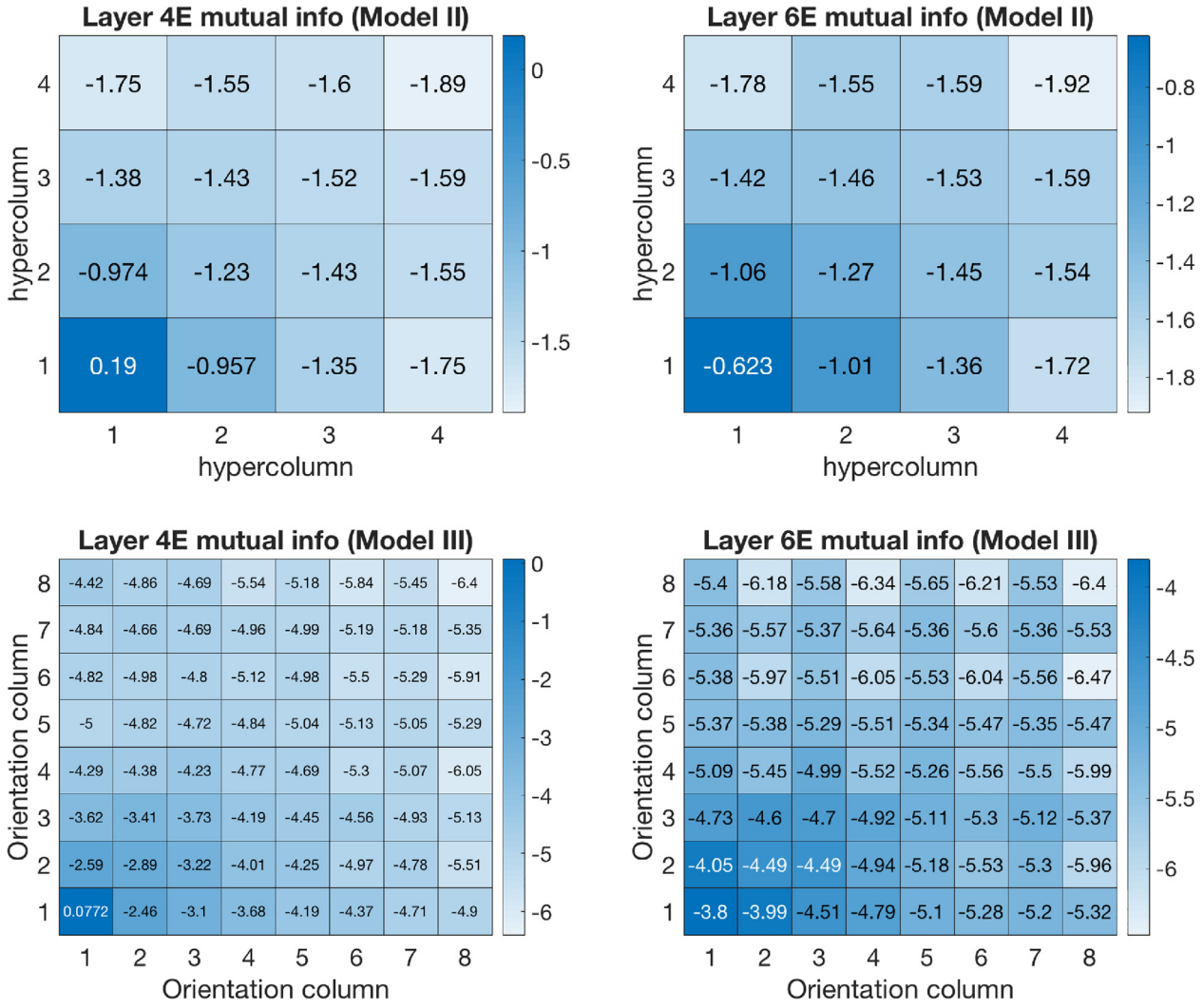



**Fig. 10.** Top: heat map of mutual information with $(1,1)$ hypercolumn in Model II. Top left: Layer 4. Top right: Layer 6. Bottom: heat map of mutual information with vertical-preferred orientation column in $(1,1)$ hypercolumn in Model III. Bottom left: Layer 4. Bottom right: Layer 6. Numbers on grid boxes are the logarithmic of mutual information.

If $\zeta_C(S_{[0,T)})$ is independent of $\zeta_A(S_{[0,T)}), \zeta_B(S_{[0,T)})$, we have

$$H_{T,\zeta_{B\cup C}} = H_{T,\zeta_B} + H_{T,\zeta_C}, \quad H_{T,\zeta_{A\cup C}} = H_{T,\zeta_A} + H_{T,\zeta_C}$$

and

$$H_{T,\zeta_{A\cup B\cup C}} = H_{T,\zeta_{A\cup B}} + H_{T,\zeta_C}.$$

It follows from Eq. (5.1) that

$$MI_{T,\zeta}(A:B:C) = MI_{T,\zeta}(A:B) - MI_{T,\zeta}(A:B) = 0.$$

Now consider a neuronal network with $K$ local populations. Let two disjoint sets $\mathcal{I}, \mathcal{O} \subset \{1,\cdots,K\}$ denote the *input set* and *output set* of this network. The *degeneracy* $\mathcal{D}_{T,\zeta}(\mathcal{I}:\mathcal{O})$ is given by a weighted sum of multivariate mutual informations:

$$\mathcal{D}_{T,\zeta}(\mathcal{I}:\mathcal{O}) = \sum_{0\leqslant k\leqslant |\mathcal{I}|}\sum_{\mathcal{I}_k\subset\mathcal{I}} \frac{1}{2\binom{|\mathcal{I}|}{k}} MI_{T,\zeta}(\mathcal{I}_k : \mathcal{I}_k^c : \mathcal{O}), \tag{5.2}$$

where the summation goes through all possible bipartitions of $\mathcal{I}$, and set $\mathcal{I}_k$ means a subset of $\mathcal{I}$ with $k$ local populations. Degeneracy measures how much more information different components of the input set share with the output set than expected if all components are independent. Degeneracy is high if many structurally different components in the input set can perform similar functions on a designated output set. A neuronal network is said to be *degenerate* if $\mathcal{D}_{T,\zeta}(\mathcal{I}:\mathcal{O}) > 0$ for some choice of $T, \zeta, \mathcal{I}$, and $\mathcal{O}$.

The (structural) *complexity* $\mathcal{C}_{T,\zeta}(\mathcal{I}:\mathcal{O})$ is given by the weighted sum of mutual information between components of the input set.

$$\mathcal{C}_{T,\zeta}(\mathcal{I}:\mathcal{O}) = \sum_{0\leqslant k\leqslant |\mathcal{I}|}\sum_{\mathcal{I}_k\subset\mathcal{I}} \frac{1}{2\binom{|\mathcal{I}|}{k}} MI_{T,\zeta}(\mathcal{I}_k : \mathcal{I}_k^c), \tag{5.3}$$

where the summation goes through all possible bipartitions of $\mathcal{I}$.

The complexity measures how much codependency in a network appears among different components of the input set. Again, a neuronal network is said to be (structurally) *complex* if $\mathcal{C}_{T,\zeta}(\mathcal{I}:\mathcal{O}) > 0$ for some choice of $T, \zeta, \mathcal{I}$, and $\mathcal{O}$.

The degeneracy at two limit cases can be given by Proposition 1 easily. When the input is independent of the output, we have zero degeneracy. When a neuronal network is fully synchronized, and $A, B, C$ are three local populations with $\mathcal{I} = \{A, B\}, \mathcal{O} = C$, then the degeneracy equals to $H_{\zeta,T}(A)$, which is positive. Our numerical simulation result will confirm this.

Finally, we have the following lemma regarding the connection between degeneracy and complexity.

**Lemma 1.** *For any choice of $T, \zeta, \mathcal{I}$, and $\mathcal{O}$ and any decomposition $\mathcal{I} = \mathcal{I}_k + \mathcal{I}_k^c$, we have*

$$MI_{T,\zeta}(\mathcal{I}_k : \mathcal{I}_k^c : \mathcal{O}) \leqslant \min\{MI_{T,\zeta}(\mathcal{I}_k : \mathcal{I}_k^c), MI_{T,\zeta}(\mathcal{I}_k : \mathcal{O}), MI_{T,\zeta}(\mathcal{I}_k^c : \mathcal{O})\}.$$

**Proof.** This lemma is a discrete version of Lemma 5.1 of Li and Yi (2016). We include it for the sake of completeness of this paper. It is sufficient to prove that for any three discrete random variables $X$, $Y$, and $Z$ with joint probability distribution function $p(x,y,z) = \mathbb{P}[X = x, Y = y, Z = z]$,

$$MI(X : Y : Z) \leqslant \min\{MI(X : Y), MI(X : Z), MI(Y : Z)\}.$$

It follows from the definition of the multivariate mutual information and some elementary calculations that

$$\begin{aligned} MI(X : Y : Z) &= H(X) + H(Y) + H(Z) - H(X,Y) - H(X,Z) - H(Y,Z) + H(X,Y,Z) \\ &= MI(X : Y) - (H(X,Z) + H(Y,Z) - H(Z) - H(X,Y,Z) \\ &= MI(X : Y) - MI(X : Y | Z), \end{aligned}$$

where the latter time is the conditional mutual information. The conditional mutual information is nonnegative, as a direct corollary of Kullback's inequality (Yeung, 2012). Hence

$$MI(X : Y : Z) \leqslant MI(X : Y).$$

Inequalities $MI(X : Y : Z) \leqslant MI(X : Z)$ and $MI(X : Y : Z) \leqslant MI(Y : Z)$ follow analogously. This completes the proof.

The following theorem is a straightforward corollary of Lemma 1.

**Theorem 3.** *For any choice of $T, \zeta, \mathcal{I}$, and $\mathcal{O}$,*

$$\mathcal{D}_{T,\zeta}(\mathcal{I} : \mathcal{O}) \leqslant \mathcal{C}_{T,\zeta}(\mathcal{I} : \mathcal{O}).$$

Heuristically, Theorem 3 says that a neuronal network with high degeneracy must have high (structural) complexity as well.

### 5.2. Degeneracy and complexity in visual cortex models

We consider the following three numerical simulations to measure the degeneracy and complexity in model II and model III. Because the estimation of entropy is not satisfactory when the simulation is under-sampled, we limit the cardinality of $|\mathcal{I}|$ to 3 and only consider the case of $|\mathcal{O}| = 1$.

**Numerical simulation I.** In Model II, we consider the input set $\mathcal{I} = \{L_6, L_7, L_{10}\}$ and $\mathcal{O} = L_{22}$. In other words the input set consists of hypercolumns $(2,2), (2,3)$, and $(3,2)$ in layer 4 and the output set is local population $(2,2)$ in layer 6. The synapse delay times are chosen to be $\tau_E = 2$ ms and $\tau_I = 4.5$ ms. The time window size is $T = 5$ ms. The coarse-grained function $\zeta_C$ with respect to a generic set $C$ is given as in Eq. (4.1), with a partition function $\Theta$ maps $\mathbb{Z}_+$ to $\{0,1,2,3,4,5\}$ such that $\Theta(n) = i$ for $20i \leqslant n < 20(i+1)$ if $i = 0, 1, \cdots, 4$ and $\Theta(n) = 5$ if $n \geqslant 100$. Fig. 11 Left shows the dependence of degeneracy $\mathcal{D}_{\zeta,T}(\mathcal{I}, \mathcal{O})$ on the coupling strength between layers $\rho_f = \rho_b = \rho$. We can see that the degeneracy and complexity increase with $\rho$ in general. When $\rho$ is too large, the recurrent excitation cannot be tempered by inhibitions and the network fires at a very high rate that can not be captured by the partition map $\Theta$. As a result, both $\mathcal{D}$ and $\mathcal{C}$ drops to very small values. To further illustrate this, we provide raster plots of excitatory neurons in the four corresponding hypercolumns when the coupling strengths are 0.25, 0.5, and 0.75, respectively. We can see that in general stronger, more synchronized MFEs produces higher degeneracy and complexity.

**Numerical simulation II.** Now we take orientation columns into considerations. In Model III, we let the input set $\mathcal{I}$ consist of

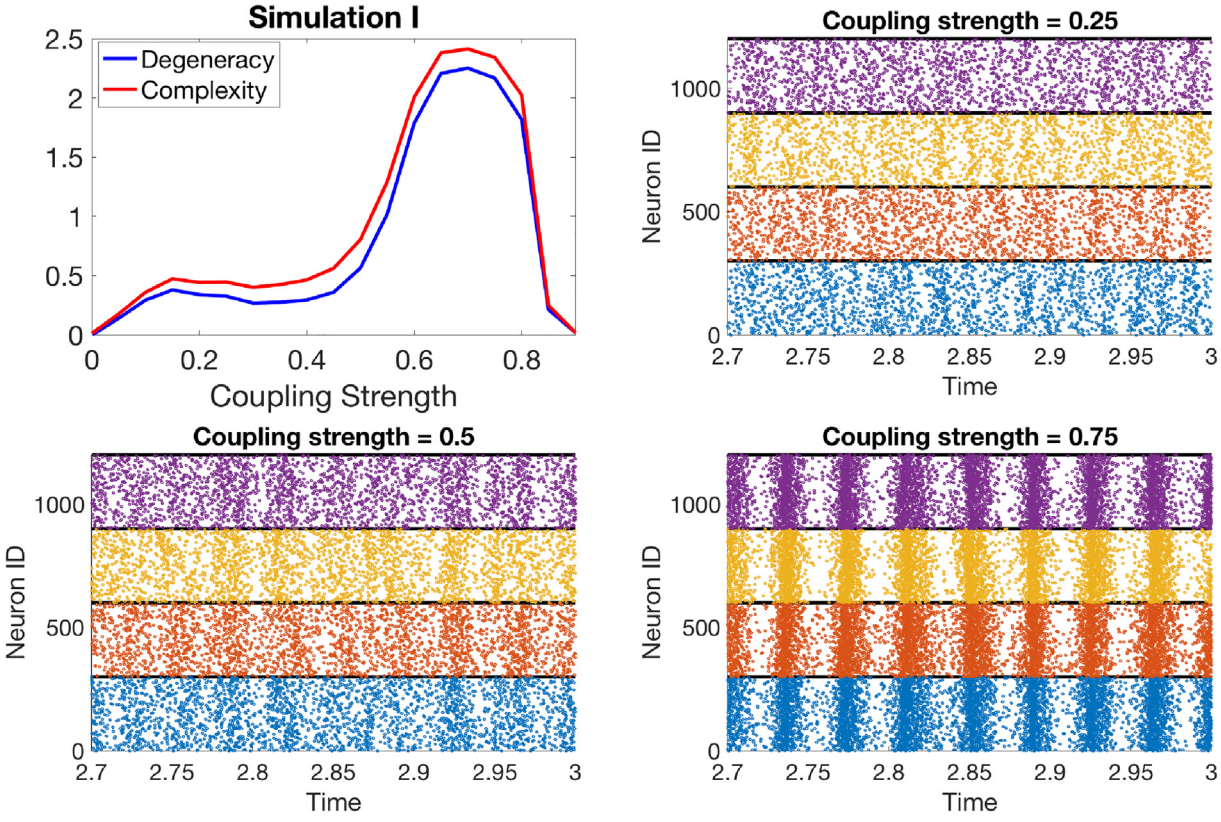



**Fig. 11.** Degeneracy and complexity in model II from **Numerical simulation I**. The simulation runs 16 independent trajectories with length 20 s each. Three raster plots of excitatory neurons with coupling strengths 0.25, 0.5, and 0.75 are provided. Neurons in the raster plot from bottom to top are from layer 4 hypercolumn $(2,2), (2,3), (3,2)$, and layer 6 hypercolumn $(2,2)$.

three orientation columns in hypercolumn $(2,2)$ of layer 4, with orientation preferences 0 deg, 45 deg, and 90 deg respectively. The output set is the orientation column in hypercolumn $(2,2)$ of layer 6, with an orientation preference 0 deg. The synapse delay times are chosen to be $\tau_E = 2$ ms and $\tau_I = 4.5$ ms. The time window size is $T = 5$ ms. The coarse-grained function $\zeta_C$ is similar to the one in **Numerical simulation I**, except the partition function takes value $\Theta(n) = i$ for $10i \leqslant n < 10(i+1)$ if $i = 0, 1, \cdots, 4$ and $\Theta(n) = 5$ if $n \geqslant 50$, because an orientation column contains less neurons than a hypercolumn. The strength of long range connection is given by $C_{long} = 0.5$. Fig. 11 Middle shows the dependence of degeneracy $\mathcal{D}_{\zeta,T}(\mathcal{I},\mathcal{O})$ on the coupling strength between layers $\rho_f = \rho_b = \rho$. Again, stronger coupling between layers leads to larger degeneracy, until the network firing rate is too high to be captured by the given partition function. The degeneracy and complexity vs. coupling strength is plotted in Fig. 12. Raster plots of excitatory neurons in the four corresponding orientation columns with coupling strength 0.25, 0.5, and 0.75, respectively, are also provided in Fig. 12.

**Numerical simulation III.** The third simulation studies the effect of long range connections. Still in Model III, we let the input set $\mathcal{I}$ consist of three orientation columns with orientation preference 0 deg in hypercolumns $(2,2), (2,3)$, and $(3,2)$ of layer 4. The output set is the orientation column in hypercolumn $(2,2)$ of layer 6, with an orientation preference 0 deg. Parameters such as the synapse delay times, the time window size, the coarse-grained function $\zeta_C$ are identical to those of **Numerical simulation II**. The feedforward and feedback strengths are $\rho_b = \rho_f = 0.6$. And the strength of long range connections is the main control parameter. Fig. 13 shows the dependence of degeneracy $\mathcal{D}_{\zeta,T}(\mathcal{I},\mathcal{O})$ on the strength of long range connections $C_{long}$. We can see that a stronger long range coupling between orientation columns also leads to a higher degeneracy. This is further verified by the three raster plots in Fig. 13 with long range coupling strength equals to 0.5, 1.0, and 1.5, respectively.

## 6. Conclusion

This paper investigates a few information-theoretic measures of a class of structured neuronal networks. Neurons in the network are of the integrate-and-fire type. Being consistent with our earlier papers (Li et al., 2019; Li and Hui, 2019), membrane potentials are set to be discrete to make the model mathematically and computationally simple. The network consists of many local populations, each of which has its own external drive rate. Biologically, one local population could be a hypercolumn in the cortex or an orientational column in the visual cortex. We provide several different network models for the purpose of examining information-theoretic measures. The most complicated model (Model III) aims to model two layers (layer 4 and layer 6) of the primary visual cortex, each layer has $4 \times 4$ hypercolumns, and each hypercolumn has 4 orientational columns.

Then we use the idea of coarse-graining to define the entropy. The motivation is that the naive definition of neuronal entropy works poorly for large neuronal networks. One needs unrealistically large sample to estimate the entropy in a trustable way. In

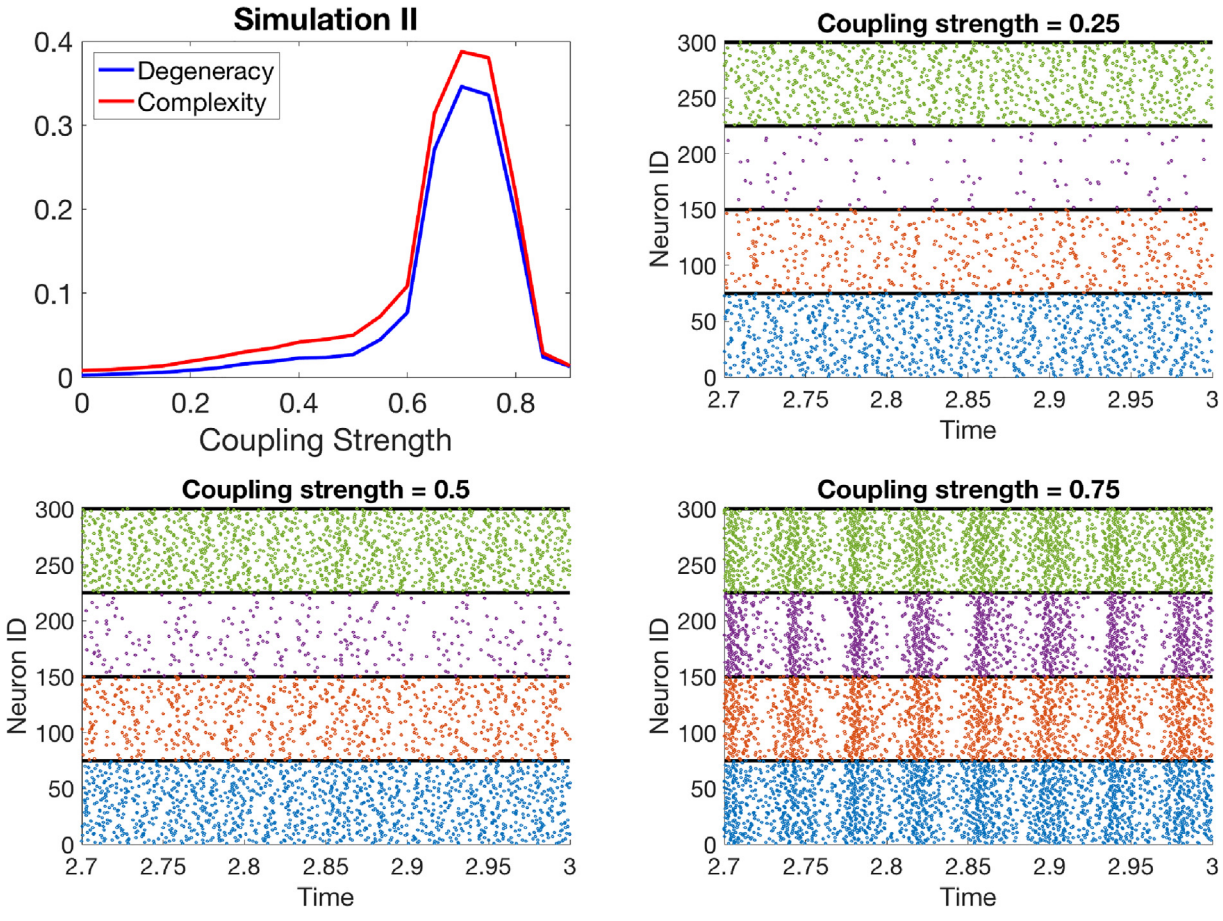



**Fig. 12.** Degeneracy and complexity in model III from **Numerical simulation II**. The simulation runs 16 independent trajectories with length 20 s each. Three raster plots of excitatory neuron with coupling strengths 0.25, 0.5, and 0.75 are provided. Neurons in the raster plot from bottom to top are from layer 4 hypercolumn $(2,2)$ with orientation preferences 0 deg, 45 deg, and 90 deg respectively, and layer 6 hypercolumn $(2,2)$ with orientation preference 0 deg.

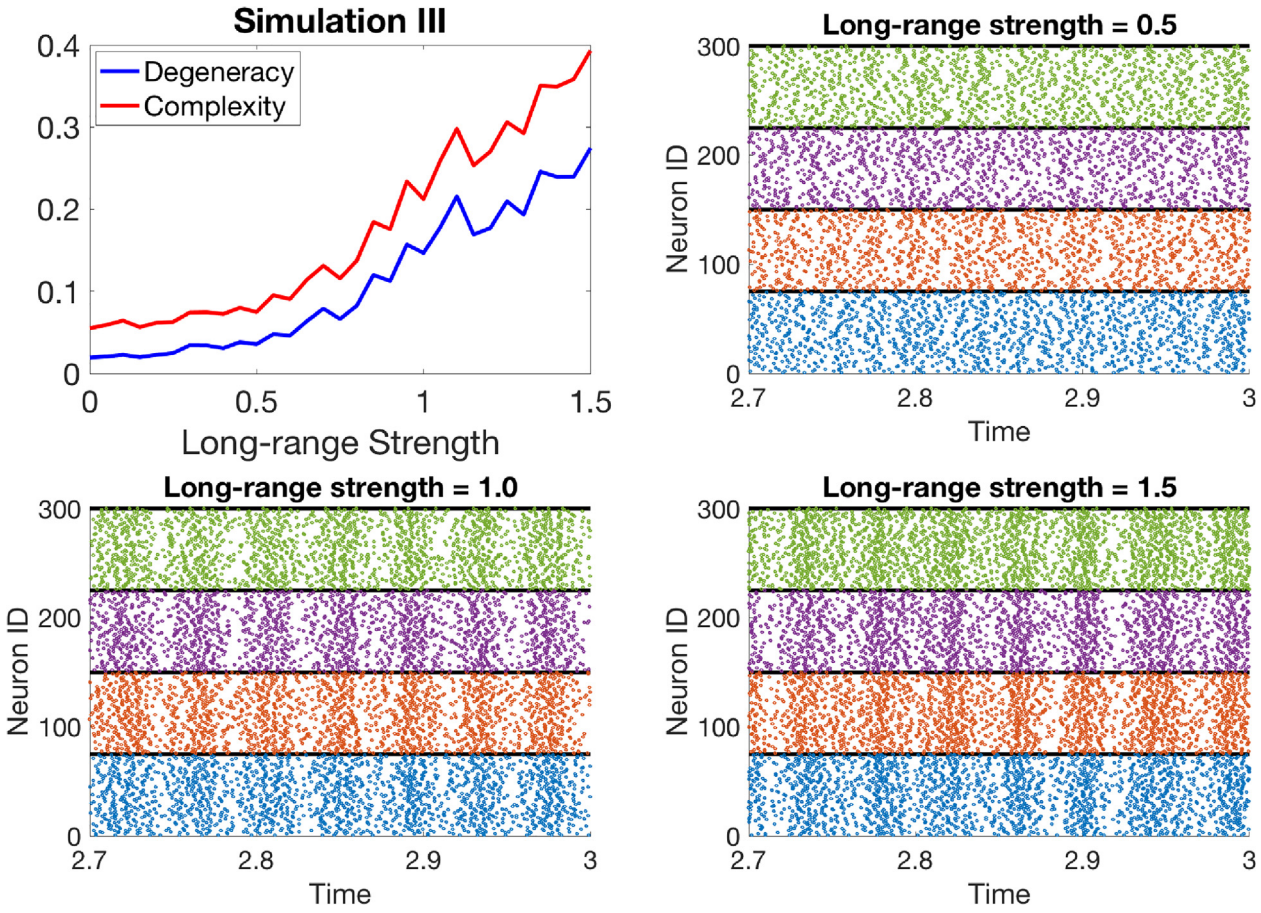



**Fig. 13.** Degeneracy and complexity in model III from **Numerical simulation III**. The simulation runs16 independent trajectories with length 20 s each. Three raster plots of excitatory neurons with long-range coupling strengths $0.5, 1.0$, and $1.5$ are provided. Neurons in the raster plot from bottom to top are from layer 4 hypercolumns $(2,2), (2,3), (3,2)$ with orientation preference 0 deg, and layer 6 hypercolumn $(2,2)$ with orientation preference 0 deg.

addition it is known that many neuronal network can generate Gamma-type rhythms. Hence when calculating the entropy, we choose to ignore the precise order of neuronal spikes from the same local population if they fall into the same small time bin. This entropy mainly captures the uncertainty in Gamma oscillations produced by the neuronal network. One can easily compute the entropy for large scale neuronal networks. The mutual information can be defined analogously.

The coarse-grained entropy and the mutual information are examined through our examples. We find that the coarse-grained entropy mainly capture the information contained in the rhythm produced by a local population. Under suitable setting of the partition function, the coarse-grained entropy reaches maximal value when the partial synchronization has the most diversity, and decreases when the spiking pattern is either homogeneous or fully synchronized. Furthermore, in our two-layer network model, we find that stronger connections between layers can produce higher mutual information between layers until the entropy starts to decrease due to too much synchronization. This is intuitive, as stronger coupling between layers makes the firing patterns from two layers more synchronized.

In the end, we attempt to quantify two systematic measures, namely degeneracy and complexity, for spiking neuronal network models. These systematic measures are originally proposed in the study of systems biology. They can be written as linear combinations of mutual information. Therefore, after defining coarse-grained entropy and mutual information, these systematic measures can be defined analogously. We find that the inequality proved in our earlier paper (Li and Yi, 2016) still holds, which says that the degeneracy is always smaller than the complexity, or a system with high degeneracy must be structurally complex. Finally, we numerically compute degeneracy and complexity for our two-layer cortex models. We find that at certain range of parameters, stronger coupling between layers, as well as stronger long-range connectivities, contribute to both higher degeneracy and higher complexity.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

Aaditya, Aaditya V., Young, Lai-Sang, 2013. Dynamics of spiking neurons: between homogeneity and synchrony. J. Comput. Neurosci. 34 (3), 433–460.
Aaditya, Aaditya V., Young, Lai-Sang, 2013. Emergent dynamics in a model of visual cortex. J. Comput. Neurosci. 35 (2), 155–167.
Alex Goddard, C., Sridharan, Devarajan, Huguenard, John R., Knudsen, Eric I., 2012. Gamma oscillations are generated locally in an attention-related midbrain network. Neuron 73 (3), 567–580.
Antos, András, Kontoyiannis, Ioannis, 2001. Convergence properties of functional estimates for discrete distributions. Random Struct. Algor. 19 (3–4), 163–193.
Borst, Alexander, Theunissen, Frédéric E., 1999. Information theory and neural coding. Nat. Neurosci. 11, 947.
Chariker, Logan, Young, Lai-Sang, 2015. Emergent spike patterns in neuronal populations. J. Computat. Neurosci. 38 (1), 203–220.
Chariker, Logan, Shapley, Robert, Young, Lai-Sang, 2016. Orientation selectivity from very sparse lgn inputs in a comprehensive model of macaque v1 cortex. J. Neurosci. 36 (49), 12368–12384.

Costa, Madalena, Goldberger, Ary L., Peng, C.-K., 2005. Multiscale entropy analysis of biological signals. Phys. Rev. E 71, (2) 021906.
Cover, Thomas M., Thomas, Joy A., 2012. Elements of information theory. John Wiley & Sons.
Edelman, Gerald M., Gally, Joseph A., 2001. Degeneracy and complexity in biological systems. Proc. Natl. Acad. Sci. 98 (24), 3763–13768.
Gilbert, Charles D., Wiesel, Torsten N., 1989. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. J. Neurosci. 9 (7), 2432–2442.
Hahn, Gerald, Bujan, Alejandro F., Frégnac, Yves, Aertsen, Ad., Kumar, Arvind, 2014. Communication through resonance in spiking neuronal networks. PLoS Computat. Biol. 10 (8), e1003811.
Hahn, Gerald, Ponce-Alvarez, Adrian, Deco, Gustavo, DecoAd., Ad., Kumar, Arvind, 2018. Portraits of communication in neuronal networks. Nat. Rev. Neurosci. 1.
Hubel, David H., 1995. Eye, brain, and vision. Sci. Am.Library/Sci. Am. Books.
Kang, Kukjin, Shelley, Michael, Sompolinsky, Haim, 2003. Mexican hats and pinwheels in visual cortex. Proc. Nat. Acad. Sci. 100 (5), 2848–2853.
Kaschube, Matthias, Schnabel, Michael, Löwel, Siegrid, Coppola, David M., White, Leonard E., Wolf, Fred, 2010. Universality in the evolution of orientation columns in the visual cortex. Science 330 (6007), 1113–1116.
Lee, Kwang-Hyuk, Williams, Leanne M., Breakspear, Michael, Gordon, Evian, 2003. Synchronous gamma activity: a review and contribution to an integrative neuroscience model of schizophrenia. Brain Res. Rev. 41 (1), 57–78.
Li, Yao, Hui, Xu., 2019. Stochastic neural field model: multiple firing events and correlations. J. Math. Biol., 1–36
Li, Yao, Yi, Yingfei, 2016. Systematic measures of biological networks ii: Degeneracy, complexity, and robustness. Commun. Pure Appl. Math. 69 (10), 1952–1983.
Li, Yao, Dwivedi, Gaurav, Huang, Wen, Kemp, Melissa L., Yi, Yingfei, 2012. Quantification of degeneracy in biological systems for characterization of functional interactions between modules. J. Theor. Biol. 302, 29–38.
Li, Yao, Chariker, Logan, Young, Lai-Sang, 2019. How well do reduced models capture the dynamics in models of interacting neurons? J. Math. Biol. 78 (1–2), 83–115.
Malach, R., Amir, Y., Harel, M., Grinvald, A., 1993. Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. Proc. Nat. Acad. Sci. 90 (22), 10469–10473.
Menon, V., Freeman, W.J., Cutillo, B.A., Desmond, J.E., Ward, M.F., Bressler, S.L., Laxer, K.D., Barbaro, N., Gevins, A.S., 1996. Spatio-temporal correlations in human gamma band electrocorticograms. Electroencephal. Clin. Neurophysiol. 98 (2), 89–102.
Nemenman, Ilya, Bialek, William, De Ruyter, Rob, Steveninck, Van, 2004. Entropy and information in neural spike trains: progress on the sampling problem. Phys. Rev. E 69, (5) 056111.
Paninski, Liam, 2003. Estimation of entropy and mutual information. Neural Comput. 15 (6), 1191–1253.
Rangan, Aaditya V., Young, Lai-Sang, 1999. Measures of degeneracy and redundancy in biological networks. Proc. Nat. Acad. Sci. 96 (6), 3257–3262.
Stettler, Dan D., Das, Aniruddha, Bennett, Jean, Gilbert, Charles D., 2002. Lateral connectivity and contextual interactions in macaque primary visual cortex. Neuron 36 (4), 739–750.
Strong, S.P., De Ruyter Van, R.R., Bialek, Steveninck William, Koberle, Roland, 1998. On the application of information theory to neural spike trains. Pac. Symp. Biocomput. 1998, 621–632.
Strong, Steven P., Koberle, Roland, De Ruyter, Rob R., Steveninck, Van, Bialek, William, 1998. Entropy and information in neural spike trains. Phys. Rev. Lett. 80 (1), 197.
Tononi, Giulio, Sporns, Olaf, Edelman, Gerald M., 1994. A measure for brain complexity: relating functional segregation and integration in the nervous system. Proc. Natl. Acad. Sci. 91 (11), 5033–5037.
Whitacre, James M., 2010. Degeneracy: a link between evolvability, robustness and complexity in biological systems. Theor. Biol. Medical Modell. 7 (1), 6.
Yeung, Raymond W., 2012. A first course in information theory. Springer Science & Business Media.