

# EFFECTS OF THE DISTRIBUTION OF SYNAPTIC STRENGTH ON THE SPIKING DYNAMICS IN A NETWORK OF SPIKING NEURONS

Yeung Chun Yat (1155110319)

Supervisor: Prof. Ching Shuk Chi Emily

April, 2021

## Abstract

Networks are prevalent in sciences. This research project studies neuronal networks in particular, which have spike counts observed to be highly skewed and long-tailed in distribution, showing a significant deviation from Gaussian. In our model, the network dynamics respects a system of stochastic dynamical differential equations that governs the evolution of the nodal states, coupled through the network structure. In biology literatures, the synaptic strengths, like the spike counts, are also found to have a skewed and long-tailed distribution, and whether the two are associated, either statistically or causally, is the central question to answer. This project continues from its first part conducted in the first semester. In this project, different useful measures of a network are first introduced. Built upon a network structure reconstructed from experimental cortical electrical signals (the DIV25 dataset), we investigate the conductance-based synaptic spiking model, established upon E. M. Izhikevich's base model. It is found that among all previous attempted models, the synaptic spiking model has the best performance, succeeding in generating realistic neuronal spikes which recover some empirical statistical features that real neuronal spikes have. Finally, the effects of the distribution of synaptic strength on the spiking dynamics are investigated, and the analysis indicates that the long-tailed incoming strengths are causing the long-tailed spike counts.

# Contents

<b>1 Project Overview</b>	<b>3</b>
<b>2 Introduction</b>	<b>4</b>
2.1 Measures of a Network . . . . .	5
2.2 Summary of Network Measures . . . . .	6
2.3 Empirical Network Data . . . . .	7
2.4 ( <i>Summary</i> ) Logistic and FHN Network Model . . . . .	10
2.4.1 Network Models . . . . .	10
2.4.2 Key Results . . . . .	11
<b>3 Synaptic Spiking Model</b>	<b>14</b>
3.1 The Conductance-based Model . . . . .	15
3.2 Network Dynamics . . . . .	17
3.3 Comparison with the FHN Network Model . . . . .	20
<b>4 Effects of the Distribution of Synaptic Strength</b>	<b>22</b>
4.1 Reference Network Analysis . . . . .	23
4.2 Drivers of the Long-tailed Spike Count Distribution . . . . .	24
<b>5 Conclusion</b>	<b>26</b>
<b>6 Datasets and Codes</b>	<b>27</b>
<b>7 Acknowledgments</b>	<b>28</b>
<b>8 References</b>	<b>28</b>
<b>9 Appendix</b>	<b>29</b>

# 1 Project Overview

The project continues from its first part, conducted in the first semester. While the first part illustrates the effects of synaptic strength distribution on neuronal network spiking dynamics using the *logistic* and *FitzHugh-Nagumo* network model as case study, reaching the conclusion that the long-tailed incoming strengths are causing the long-tailed spike counts, this project seeks to reinforce the analysis by studying the *conductance-based synaptic spiking model* extended from E. M. Izhikevich's base model.

The report consists of the following three parts.

- (1) (*Summary*) In the first part of the project in the first semester, two models, each obeying a different intrinsic dynamics, are tested. One obeys *logistic* and the other obeys *FitzHugh-Nagumo*. The natures of the model time series generated are reviewed, and it has been concluded that the former *fails* at generating realistic neuronal spikes while the latter *succeeds* in giving realistic results while recovering *some* of the empirical statistical features. Yet, there is still room for improvement as the spike count tail is still far from achieving the long-tailedness experimental data have.
- (2) (*Model Implementation*) The *conductance-based synaptic spiking model* is discussed and implemented. Under this model framework, we attempt to understand the mechanism of a neuronal spiking cycle and compare the model dynamics with the experimental results. It is found that in comparison with our previous logistic and FHN network models, *the synaptic spiking model is superior at producing long-tailed spike counts with spike data covering a wide range*. The drawbacks of the model are also examined, by looking into the raster plot that illustrates the spike timestamps.
- (3) (*Analysis*) It is observed that the experimental neuronal spike counts are highly skewed and long-tailed in distribution, and so are the synaptic strengths. Researchers have long conjectured that there exists some association between the two, and that the distribution of the synaptic strengths affects the distribution of spike counts – *Can the very large spike counts be explained by the very large synaptic strengths?* It inevitably requires a model to establish the connection between the network structure and the spiking dynamics. Built upon the synaptic spiking model, the effects of different network features, including degrees and strengths, on the spiking dynamics are investigated, through the *reference network analysis* which preserve some features while varying some other. This follows tightly the methodology proposed in the first part.

**Objectives.** This research aims to address the following questions. They are in a broad sense, and the specific context under which they are to be understood will be supplied.

- (1) For an assumed network model, how does the distribution of synaptic strengths (and more generally, the “structure” of the network) affect the spiking dynamics?
- (2) A search for realistic models of neurons – How do different network models behave and perform? How do they compare in terms of resemblance to real, experimental neuronal time series?

## 2 Introduction

Networks [1] are prevalent in sciences, and are useful for modeling systems consisting of mutually interacting individual components. For example, connection of individuals that may be friends of each other in a social networking site and connection of websites that lead to each other on the internet may both be modeled (or abstracted) in the form of a network. Here, we study a biological example – **neuronal network**. In a network, individual components, known as **nodes**, connect to one another via **links** (or **edges**). Each node is identified with a label  $i$ , where  $i = 1, 2, \dots, N$  with  $N$  being the total number of nodes. Interaction between two linking nodes is quantified by the **coupling strength** that informs how strong the connection is. Often a node can affect the dynamics of another node but its own dynamics is not affected by the latter. (The meaning of “dynamics” will be made precise later.) Such behavior is captured in a **directed network** with directional links. In an **undirected network**, in contrast, the interaction between two nodes is mutual and symmetric with a two-way link. The structure of a network is a crucial piece of information for understanding the system it represents, e.g. how the individuals are interconnected, and the presence of “central bodies”, known as **hubs**, that make many connections to other individuals.

In a neuronal network, each node carries a value symbolizing the electrical signal generated by the neuron at a point in time. The measured signals vary in time, sometimes with steady fluctuations and sometimes with significant jumps, which are known as **neuronal spikes** (or **peaks**). Over time, we collect the values into a time series, and the time series of all the nodes in the network collectively are called the **network dynamics**. More generally, it refers to the time-dependence or the evolution of certain network characteristics. There are many possible ways to characterize the nature of a dynamics. For a neuronal dynamics, for example, we may count the number of spikes, compute the standard deviation of the fluctuation in the time series, or measure the correlation between the time series. The characterizations are specific to the problem we want to address, and there is no characterization that is universally good. It is worth noting that each characterization has its own flaws. For example, to count spikes, we use a peak detection algorithm but it inevitably misses some spikes that are not “obvious” enough; also, standard deviation may not be a representative measure of the fluctuation in time series when the values are long-tailed in distribution.

The coupling strengths of the edges in a network are the primitive and leading factor to drive the dynamics. A node impacts its connecting nodes through the edges, with a positive coupling strength effecting an **excitatory** impact and a negative coupling strength effecting an **inhibitory** impact.

## 2.1 Measures of a Network

There are useful measures to aid us in understanding a network. The most important two are the **degrees** and **strengths**. The number of nodes that connect to a specific node  $i$  is known as the incoming degrees of the node  $i$ , or **in-degrees**. Likewise, the number of nodes that a specific node  $i$  connects to is known as the outgoing degrees of the node  $i$ , or **out-degrees**. Strengths are an average measure of the coupling strengths. The incoming strengths, or **in-strengths**, are the average coupling strength of edges that link to a node. The outgoing strengths, or **out-strengths**, are the average coupling strength of edges that come out of a node to other nodes.

Mathematically, we label the coupling strength of an edge that links from node  $j$  to node  $i$  by  $g_{ij}$ . As discussed,  $g_{ij}$  can be positive (meaning an excitatory edge) or negative (meaning an inhibitory edge). If node  $i$  and  $j$  are unconnected,  $g_{ij} = g_{ji} = 0$ , and one's dynamics does not affect the other. For a directed network, generally,  $g_{ij} \neq g_{ji}$ , for node  $j$  may connect to node  $i$  but not the otherwise. The coupling strengths  $g_{ij}$ , with  $i, j = 1, 2, \dots, N$ , can be contained in an  $N \times N$  **coupling strength matrix**  $\mathbf{G}$ , which, generally, is asymmetric and sparse. In most cases, we assume the nodes are not self-connecting, so the diagonal entries of  $\mathbf{G}$  are all zero. There are in total  $N(N - 1)$  possible directed edges in a network, and practically, most *real* networks have a number of edges that is only a tiny fraction of the number, and the fraction is known as the **sparsity**. In the neuronal network used in this project, for example, the sparsity is approximately 1.4%. *The coupling strength matrix  $\mathbf{G}$  is the most crucial piece of information about a network*, as it specifies (1) the connectivity (i.e., where  $g_{ij} \neq 0$ ) and (2) the nodal interactions (i.e., signs and magnitudes of  $g_{ij}$ ). Other measures of the network, e.g. degrees and strengths, may be instantly derived once we have  $\mathbf{G}$ .  $\mathbf{G}$  has high theoretical importance, but *in practice*, it is difficult to extract and there has been ongoing research on the reconstruction of  $\mathbf{G}$ , in the field of **network reconstruction**.

With  $\mathbf{G}$ , we instantly know the degrees and strengths. The in-degree and out-degree of node  $i$  are respectively given by

$$k_{\text{in}}(i) = \sum_{j=1}^N \mathbb{1}(g_{ij} \neq 0) \text{ and } k_{\text{out}}(i) = \sum_{j=1}^N \mathbb{1}(g_{ji} \neq 0), \quad (1)$$

where  $\mathbb{1}(\cdot)$  is the indicator function of a condition. The in-strength and out-strength of node  $i$  are respectively given by

$$s_{\text{in}}(i) = \frac{1}{k_{\text{in}}(i)} \sum_{j=1}^N g_{ij} \text{ and } s_{\text{out}}(i) = \frac{1}{k_{\text{out}}(i)} \sum_{j=1}^N g_{ji}. \quad (2)$$

Computationally, regarding the coupling strength matrix  $\mathbf{G}$ ,  $k_{\text{in}}(i)$  is the number of non-zero entries in row  $i$  and  $k_{\text{out}}(i)$  is the number of non-zero entries in column  $i$ . Similarly,  $s_{\text{in}}(i)$  averages the non-zero entries in row  $i$  and  $s_{\text{out}}(i)$  averages the non-zero entries in column  $i$ . There are other finer measures that could be derived. For example, if we are interested in the positive entries, we may define the positive in/out-degree and the positive in/out-strength by

$$k_{\text{in}}^+(i) = \sum_{j=1}^N \mathbb{1}(g_{ij} > 0) \text{ and } k_{\text{out}}^+(i) = \sum_{j=1}^N \mathbb{1}(g_{ji} > 0), \quad (3)$$

and

$$s_{\text{in}}^+(i) = \frac{1}{k_{\text{in}}^+(i)} \sum_{j=1}^N g_{ij} \mathbb{1}(g_{ij} > 0) \text{ and } s_{\text{out}}^+(i) = \frac{1}{k_{\text{out}}^+(i)} \sum_{j=1}^N g_{ji} \mathbb{1}(g_{ji} > 0). \quad (4)$$

This is generalizable to other desired conditions by replacing the arguments in the indicator functions. Note that for a directed network, the in- and out-measures are generally different, while for an undirected (aka. bidirectional) network, they are equivalent because of the symmetric  $\mathbf{G}$ .

It is also a common practice to define the adjacency matrix  $\mathbf{A}$ , with entry  $A_{ij} = \mathbb{1}(g_{ij} \neq 0)$ , i.e., if node  $j$  links to node  $i$ ,  $A_{ij} = 1$ , 0 otherwise. The adjacency matrix is all we need to understand the connectivity in a network, but it contains less information than the coupling strength matrix as it does not tell us the coupling strengths associated with the connections. Therefore, in this project,  $\mathbf{A}$  is less useful than  $\mathbf{G}$ . Physically,  $k_{\text{in}}^+(i)$  counts the number of incoming excitatory links into node  $i$  while  $s_{\text{in}}^+(i)$  averages the strengths of the incoming excitatory links, and similar interpretations apply to  $k_{\text{out}}^+(i)$  and  $s_{\text{out}}^+(i)$  for outgoing links.

To network scientists, the distribution of degrees and strengths are of particular interest as they are two high-level summaries of the connectivity of a network. Besides the basic summary statistics, e.g. mean, median and quantiles, the **probability distribution** of a measure is frequently looked at. Interestingly, for *real* networks such as a neuronal network, world wide web, human social network and marine food web, it is argued that the degree distribution follows a heavy- or long-tailed distribution; notably, for the social network, the distribution approximates a power law  $P(k) \sim k^{-\gamma}$  [2]. This is an example of characteristics that happen to exist across different categories of networks. Is this a coincidence? This is an intriguing research topic in network science, and is nonetheless not the central subject to discuss here.

## 2.2 Summary of Network Measures

Network Measure	Formula	Interpretation
In-degree	$k_{\text{in}}(i) = \sum_{j=1}^N \mathbb{1}(g_{ij} \neq 0)$	No. of incoming links
Out-degree	$k_{\text{out}}(i) = \sum_{j=1}^N \mathbb{1}(g_{ji} \neq 0)$	No. of outgoing links
In-strength	$s_{\text{in}}(i) = \frac{1}{k_{\text{in}}(i)} \sum_{j=1}^N g_{ij}$	Avg. strength of incoming links
Out-strength	$s_{\text{out}}(i) = \frac{1}{k_{\text{out}}(i)} \sum_{j=1}^N g_{ji}$	Avg. strength of outgoing links
Positive in-degree	$k_{\text{in}}^+(i) = \sum_{j=1}^N \mathbb{1}(g_{ij} > 0)$	No. of exc. incoming links
Positive out-degree	$k_{\text{out}}^+(i) = \sum_{j=1}^N \mathbb{1}(g_{ji} > 0)$	No. of exc. outgoing links
Positive in-strength	$s_{\text{in}}^+(i) = \frac{1}{k_{\text{in}}^+(i)} \sum_{j=1}^N g_{ij} \mathbb{1}(g_{ij} > 0)$	Avg. strength of exc. incoming links
Positive out-strength	$s_{\text{out}}^+(i) = \frac{1}{k_{\text{out}}^+(i)} \sum_{j=1}^N g_{ji} \mathbb{1}(g_{ji} > 0)$	Avg. strength of exc. outgoing links

**Table 1:** A summary of different network measures: formulas and interpretations.

## 2.3 Empirical Network Data

A **network reconstruction** method [3], based on a noise-induced relation, has been applied to the empirical neuronal electrical signals of cultures of rat embryonic cortices after 25 days in vitro to estimate the coupling strengths  $g_{ij}$ , with the estimated links known as the **effective connectivity** [4]. The following dataset of estimated  $g_{ij}$  is provided by my supervisor and forms the foundation of later simulations in this project. The dataset takes the following format, where unstated node pairs are implied to be unconnected. Note again that  $g_{ij}$  is the coupling strength associated with the directed edge linking from node  $j$  to node  $i$ . We call the following dataset of  $g_{ij}$  “**DIV25**”.

Node $i$	Node $j$	$g_{ij}$
1	196	0.0208720006
1	266	0.0156720001
1	267	0.0218959991
:	:	:
2	1	-0.0234200004
2	21	-0.00388139999
2	23	-0.00472760014
:	:	:
4095	4094	0.0089673996

**Table 2:** The dataset of the reconstructed coupling strengths  $g_{ij}$  based on the empirical neuronal time series data.

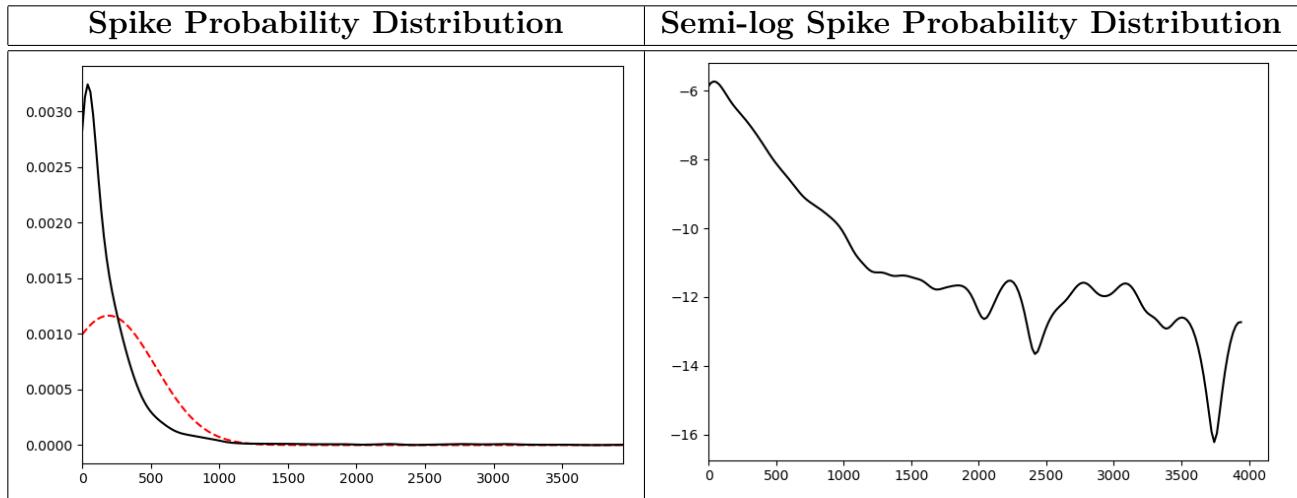
There are collectively  $64 \times 64 - 1 = 4095$  electrodes to measure electrical signals of the rat embryonic neurons, hence the neuronal network contains 4095 nodes. Precisely speaking, the nodes are not actually the neurons, but the electrodes serving as proxy for the “underlying” neurons. This approach is used because, experimentally, it is difficult to pinpoint a particular neuron. The network is sparse – only a tiny fraction (1.4%) of all possible node pairs are linked. Coupling strengths of the excitatory and inhibitory edges, i.e.,  $\{g_{ij} : g_{ij} > 0\}$  and  $\{|g_{ij}| : g_{ij} < 0\}$ , are discovered to exhibit a long-tailed distribution. This echoes with literatures. For example, it is well known to network scientists that biological parameters, e.g. neuron firing rate and synaptic weights, have long, heavy tails and are highly skewed [5, 6]. This is in contrast with the common belief in the past that brain parameters obey bell-shaped distributions and, therefore, biologists often rely on *mean* and *standard deviation* to understand the data. This is inadequate for describing data that have *a long tail* because the degree of dispersion is not comprehensively summarized by the standard deviation and the mean is also subject to heavy influence of the extreme outliers. In this case, higher-order moments and quantiles are relatively useful. This will be further discussed.

Neuronal spiking dynamics are *highly skewed* and *long-tailed*. Here, we analyze the spike distribution based on the following (experimentally obtained) dataset of rat embryonic neuronal spike counts after 25 days in vitro, provided by my supervisor.

Node $i$	Spike Count
1	521
2	9
3	6
:	:
4095	11

**Table 3:** The dataset of the rat embryonic neuronal spike counts after 25 days in vitro.

The spike counts form the following probability distribution plots.



**Figure 1:** In the spike probability distribution plot, the curve is seen to have a long right tail. The Gaussian (red dashed line) with the same mean and standard deviation is shown as comparison. In the semi-log plot, a linearity is seen for small spike counts while there are fluctuations for large spike counts.

Distribution	Mean	Standard Deviation	Skewness	Excess Kurtosis
Gaussian	192.32	342.84	0	0
Spike	192.32	342.84	5.39	39.08

**Table 4:** The first four moments of the spike counts and the Gaussian with the same mean and standard deviation.

The histogram of the spike counts is normalized and smoothed using kernel techniques (with the `scipy.stats.gaussian_kde` in the SciPy library of Python) to give the spike probability distribution curve. The probability distribution is compared to a Gaussian (red dashed line) that has the same mean (192.32) and standard deviation (342.84). It can be seen from the plot that the probability distribution has a peak value three times as high as that of the Gaussian and a tail extending far to the right. The  $x$ -axis stops at the largest value of spike count (3941).

The table shows the first four moments of the Gaussian and the spike probability distribution, where the skewness and the excess kurtosis are respectively given by

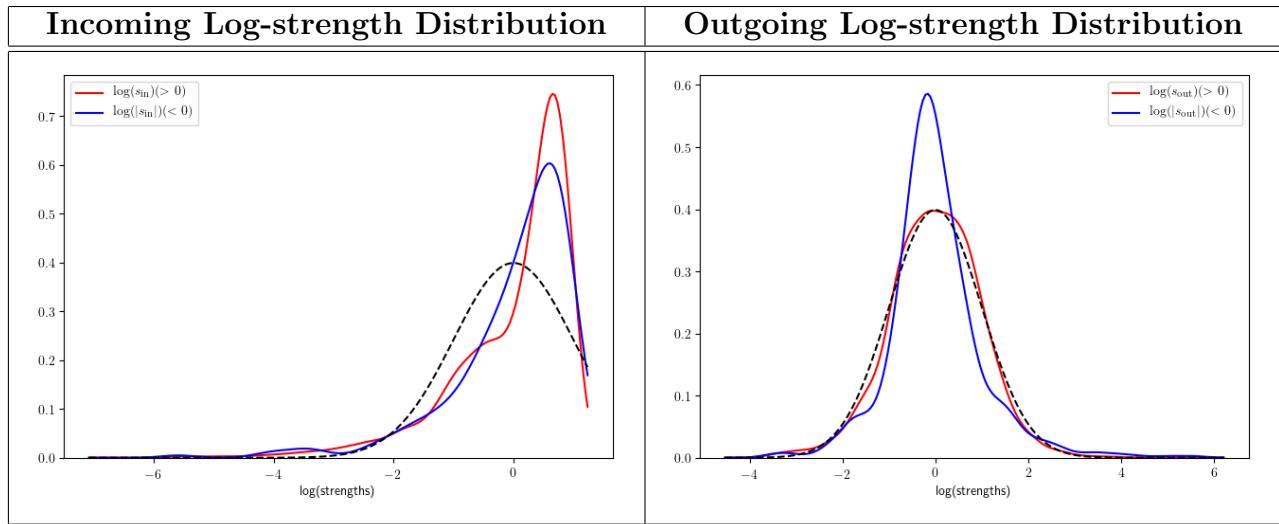
$$\text{Skew.} = \frac{\overline{(X - \mu)^3}}{\sigma^3} \text{ and Excess Kurt.} = \frac{\overline{(X - \mu)^4}}{\sigma^4} - 3, \quad (5)$$

with  $\mu$  being the mean and  $\sigma$  being the standard deviation of the data. For a Gaussian, the skewness and the excess kurtosis are both zero. Unlike standard deviation, skewness and kurtosis are scale-free measures that can be compared across different datasets. The positive skewness implies that the distribution is *right-skewed* with a mean greater than the median and the large kurtosis implies that it has a *significantly longer tail* than the Gaussian.

While a large majority of nodes have spike counts fewer than 100, there are 14 nodes with spike counts larger than 3000, which are 8 *standard deviations* from the mean. This would be practically impossible if spike counts were to obey a normal distribution because, at 8 standard deviations, the probability density would take an order of  $10^{-15}$ , requiring on average  $10^{15}$  nodes to have one such occurrence. Nevertheless, in our sample that comprises only 4095 nodes, there are already 14 occurrences! The extreme outliers again indicate that the distribution has a long, heavy tail.

The *exponential distribution* is one candidate to model a long/heavy-tailed distribution. In the semi-log plot, we take logarithm on the probability density to examine how well it may be approximated by an exponential distribution, which theoretically has the form  $P(x) = \lambda e^{-\lambda x}$  and logarithm gives  $\log P(x) = \log \lambda - \lambda x$  (linear). We can see that for the nodes with spike counts fewer than 1000, there is a clear linearity but there are a lot of fluctuations for nodes with greater spike counts. Although nodes with large spike counts do exist, the number is significantly smaller, subjecting the line fit to statistical noise. It is hard to draw a conclusion on the nature of the distribution for the large-spike nodes. *But for the low-spike nodes, we can conclude that the distribution is well approximated by an exponential distribution.*

To summarize, the empirical rat embryonic spike counts have the following characteristics: *highly right-skewed, long-tailed and have very extreme outliers*. A good neuronal network model should be able to generate time series that are successful in recovering these spike characteristics.



**Figure 2:** The probability distribution of the standardized incoming and outgoing log-strengths. Both are seen to have a long tail extending to 6 standard deviations from the mean.

Like the neuronal spikes, the incoming and outgoing synaptic strengths  $s_{in/out}(i)$  derived from the coupling strengths  $g_{ij}$  (extracted using the network reconstruction method) are also long-tailed. The plots show the probability distributions of the standardized incoming and outgoing log-strengths compared to a standard Gaussian (black dashed line). Note that to *standardize* a dataset  $\{X_i\}$ , the data are transformed according to  $X_i \rightarrow (X_i - \mu)/\sigma$ , where  $\mu$  and  $\sigma$  are respectively the mean and standard deviation.

The in/out-strengths are both grouped into two classes according to the signs. For both the in- and out-strengths, the tails are long and, to convert the values to a more reasonable scale, logarithms are taken. For the incoming log-strengths, we can see that the probability distribution is left-skewed with a long left tail. For the outgoing log-strengths, the probability distribution is relatively symmetric, and the positive strength class is well approximated by a Gaussian while the negative strength class has a thinner but taller peak. For both in- and out-strengths, there are outliers at over 6 standard deviations from the mean.

**Motivation.** We can see that both the neuronal spike distribution and the in/out-strength distribution are *skewed* and have *a long tail*. A natural question one may raise is: *Does the distribution of synaptic strength have an effect on the spiking dynamics? If it does, how does the spiking dynamics correlate with the synaptic strengths?* This is the central question this project aims to address.

## 2.4 (*Summary*) Logistic and FHN Network Model

In the first part of the project conducted in the first semester, two network models, namely the *logistic* and *FitzHugh-Nagumo* (FHN) network model, were attempted [7]. Here, they are briefly discussed, drawing on the core conclusions we reached, with the objective to contrast with the *conductance-based synaptic spiking model* that we later shall focus on (in the immediate next section). Below, we elaborate on the dynamical equations that govern the time-evolution of the network dynamics and the resulting time series.

### 2.4.1 Network Models

The two models are based on the following two sets of differential equations, with the former being one-dimensional, where each node carries a one-dimensional state  $x_i(t)$  symbolizing the electrical signal generated by a node at a point in time, and the latter being two-dimensional, where each node carries a two-dimensional state  $(x_i(t), y_i(t))$  with one variable symbolizing the electrical signal and the other being auxiliary. The nodes, as an abstraction of the neurons, are labeled by  $i = 1, 2, \dots, N$  with  $N$  being the total number of nodes.

The time-evolution is determined by the three terms on the right: (1) **intrinsic dynamics**, (2) **nodal interaction**, and (3) **noise**. The intrinsic dynamics term describes the evolution one node would exhibit if it were to be left out alone with connections with other nodes cut off and noise shut down. The nodal interaction term describes the interaction of one node with the other, with the coupling strength  $g_{ij}$  signifying the extent of the (directed) interaction. That is, the greater the magnitude of  $g_{ij}$  is, the stronger the influence node  $j$  imposes on node  $i$ . For our purpose, we substitute the reconstructed coupling strengths using the noise-induced relation proposed by Ching and Tam [3]. There are two types of coupling functions  $h(x, y)$  attempted: (1) **diffusive**  $h^{\text{diff}}(x, y) = y - x$ , and (2) **synaptic**  $h^{\text{syn}}(x, y) = 1/\beta_1\{1 + \tanh[\beta_2(y - y_0)]\}$ . Lastly, the noise term  $\eta_i$  summarizes external disturbances, modeled by a Gaussian white noise of zero mean and standard deviation  $\sigma_i$ .

Such a system of *stochastic* differential equations is then solved using the *Euler-Maruyama algorithm*, which is essentially the Euler algorithm that iteratively applies a linear update to  $x_i(t)$  by  $x_i \leftarrow x_i + (\text{RHS}) \times \Delta t$ , except that a noise term is added:  $x_i \leftarrow x_i + (\text{RHS}) \times \Delta t + \sigma_i \epsilon \times \sqrt{\Delta t}$  where  $\epsilon$  is a random number sampled from  $\mathcal{N}(0, 1)$ , the standard Gaussian. The resultant time series returned by the algorithm represent the network dynamics, and are subject to further analysis.

(1) **Logistic Network Model.** The free parameters are  $\{r_i\}$ , as well as those in  $h$ .

$$\frac{dx_i}{dt} = r_i x_i (1 - x_i) + \sum_{j \neq i} g_{ij} h(x_i, x_j) + \eta_i \quad \text{where } i = 1, 2, \dots, N \quad (\text{One-dimensional})$$

(2) **FitzHugh-Nagumo (FHN) Network Model.** The free parameters are  $\{\epsilon, \alpha\}$ , as well as those in  $h$ .

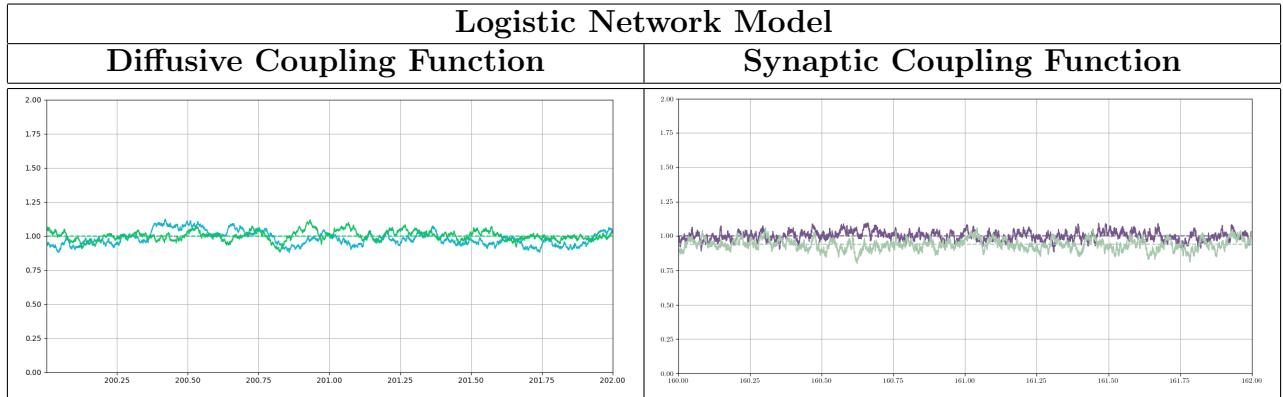
$$\frac{dx_i}{dt} = (x_i - x_i^3/3 - y_i)/\epsilon + \sum_{j \neq i} g_{ij} h(x_i, x_j) + \eta_i \quad (\text{Two-dimensional})$$

$$\frac{dy_i}{dt} = x_i + \alpha \quad \text{where } i = 1, 2, \dots, N$$

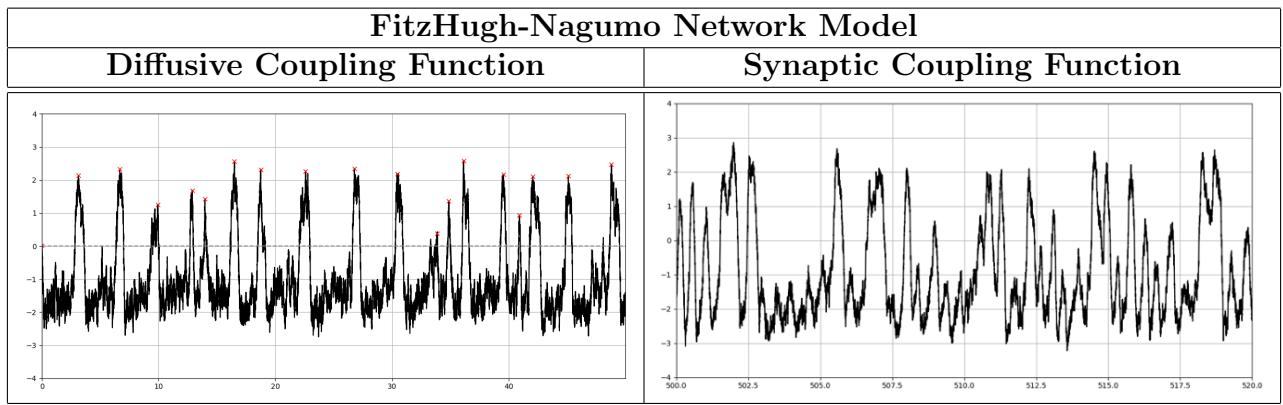
### 2.4.2 Key Results

We performed testing based on some chosen sets of parameters, but for our purpose here, the parameters and rationales of our choice will not be discussed. Here, we only seek to examine the key results and obtain a high-level summary, abstracting all technical details. Interested readers may refer to my previous report [7].

For the **logistic** network model, built upon the reconstructed coupling strength matrix  $\mathbf{G}$ , the model neuronal time series are generated by numerically solving the set of dynamical differential equations under the assumption of both a *diffusive* and *synaptic* nodal interaction  $h$ . *This, however, is a vain attempt to represent the neurons.* Below, we show the typical time series given by two sampled nodes amid a finite time interval in the whole dynamics. It is seen that no active spiking, let alone spike train, can be observed, and this in no way resemble the dynamics of real neurons. Rather, in my view, it is more appropriate to recognize them as *fluctuations* instead of *spikes*. *It is therefore concluded that the logistic network model combined with either a diffusive or synaptic nodal dynamics does not have a good performance in producing neuronal spikes.*



**Figure 3:** Typical time series of two sampled nodes in the logistic network model are seen *not* to produce realistic neuronal spikes, but rather, just fluctuations.



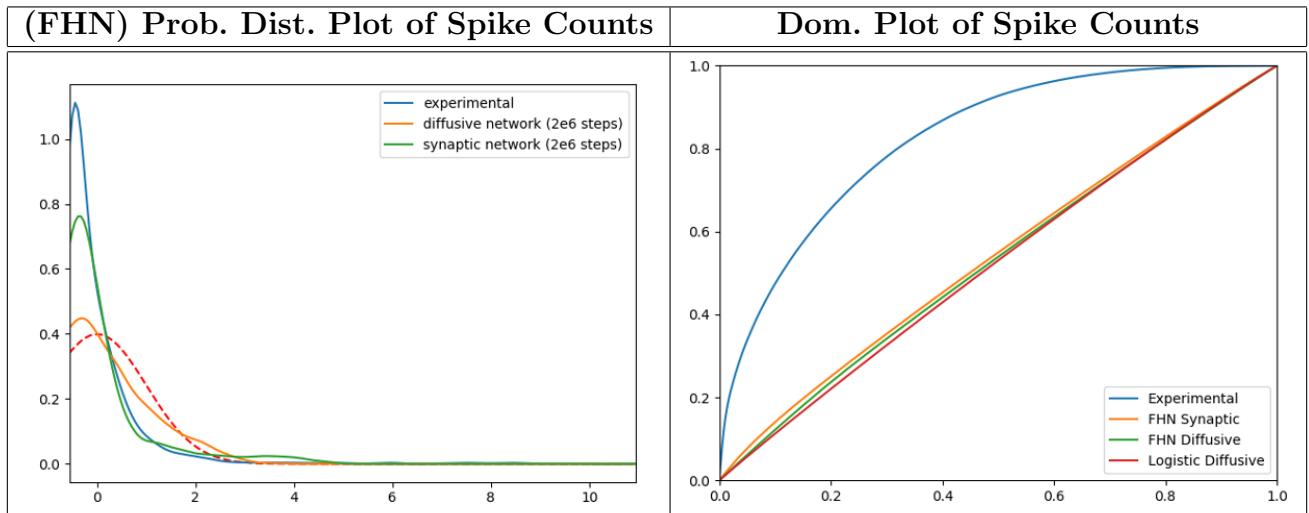
**Figure 4:** Typical time series of two sampled nodes in the FitzHugh-Nagumo network model are seen to produce realistic neuronal spikes.

For the **FitzHugh-Nagumo** network model, similarly, time series are obtained under the assumption of both a *diffusive* and *synaptic* nodal interaction  $h$ . In the FHN model, the neurons take two-dimensional states  $(x_i(t), y_i(t))$ , and with this model, *the generated time series are significantly more realistic and the spiking dynamics exhibit certain features that are empirically observed.* We show the typical time series given by two sampled nodes amid a finite time interval in the whole dynamics. It

is seen that the nodes do actively spike and spiking behavior persists throughout the whole dynamics. The peaks are detected using the peak detection algorithm (`scipy.signal.find_peaks`) in the SciPy library of Python, and labeled by red crosses in the left panel. Algorithmically the total spike counts of each node are collected and passed to obtain the spike count probability distribution plot.

To examine how realistic a network model is, we compare the model and experimental spike count distribution. A successful model has to replicate the *long-tailed* distribution that is empirically observed. Recall that in experimental observation, neurons are found to exhibit spike counts of a very wide range, with some neurons giving no spike at all and some having a spike count extending 8 standard deviations from the mean. In our project, to compare the model and experimental spike counts, we use (1) probability distribution plot (with data standardized), and (2) dominance plot of the spike counts.

While a *probability distribution plot* shows how data are located in values, a *dominance plot* illustrates the fraction of the total explained by the largest  $n$  data, and can be constructed by (1) ordering the data from the largest to the smallest and (2) connecting data points  $(n/N, \sum_{i=1}^n X_i/S)$  for  $n = 1, 2, \dots, N$ , with  $N$  being the number of data and  $S$  being the sum of all data. The dominance plot is commonly used for analyzing the behavior of the data lying in the tail. The more curvature the plot shows, *the longer the tail is and the further away from the mean the extreme outliers are*. Below, we compare the spike counts across the different attempted models with the experimentally observed one.



**Figure 5:** A comparison of model spike counts with the experimentally observed one. It is seen that among the attempted models, the FHN network model combined with the synaptic coupling function produces the most realistic result, thus regarded as the most successful. Nonetheless, there is still a significant room for improvement as, in the dom. plot, the extreme curvature exhibited by experimental neurons are still far from reach.

We conclude that the FHN network class outperforms the logistic network class, and the use of synaptic coupling function generates a more realistic result. On the left panel, from the distribution plot, for the FHN synaptic model, we see a large concentration of nodes with spike counts near the mean, with some portion of nodes having spike counts extending far beyond the mean. The maximum spike count is at 5 standard deviations. This is closer to the behavior of neuronal spikes experimentally observed. In particular, for both the experimental and synaptic model curve, we can see an exponential-like decay. Meanwhile, on the right panel, the dominance plot offers a new perspective to understand the behavior of the extreme outliers, and we can see from the comparison plot that, assuming the FHN intrinsic dynamics, *the synaptic model is better at producing a curvature*

than the diffusive model. The logistic network model with the diffusive nodal interaction has the worst performance, with a curve that is almost entirely flat. We have seen from the prior analysis that the model even fails at generating realistic neuronal spikes, so it is an inferior model. While the FHN model has been a significant improvement compared to the logistic model, it is still far from achieving the extreme curvature that the empirical neuronal spikes exhibit.

This, in turn, motivates this project. Here we perform testing on an entirely new class of model – the **synaptic spiking model**, first proposed by E. M. Izhikevich. Modifications are made in order to incorporate **conductance**, so that the model more realistically replicates the dynamical response of a neuron upon receiving an electrical input.

### 3 Synaptic Spiking Model

E. M. Izhikevich's formulism [8] of the synaptic spiking model reproduces spiking and bursting behavior of known types of cortical neurons. The model combines the biological plausibility of Hodgkin–Huxley-type dynamics and the computational efficiency of integrate-and-fire neurons, allowing fast simulation of tens of thousands of spiking cortical neurons in real time (1 ms resolution) using a desktop PC.

In the model, every neuron is described by two variables: (1) **membrane potential** (or **voltage**)  $v(t)$  and (2) **membrane recovery variable**  $u(t)$ , which follow the coupled differential equations

$$\frac{dv}{dt} = \alpha v^2 + \beta v + \gamma - u + I(t) \quad (6)$$

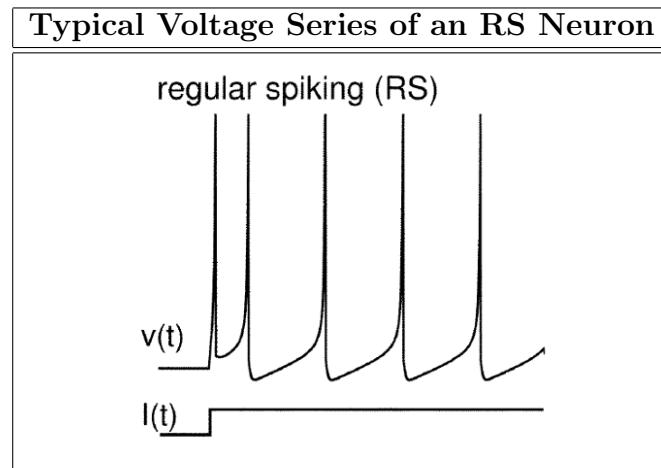
$$\frac{du}{dt} = a(bv - u), \quad (7)$$

with a *fire and reset rule*. Every time when  $v(t)$  assumes the *threshold value*  $v(t) = v_{\text{peak}}$ , both variables are instantaneously updated:

$$v(t) \leftarrow c \quad (8)$$

$$u(t) \leftarrow u(t) + d. \quad (9)$$

**Synaptic currents** or injected dc-currents are delivered via the variable  $I(t)$ . The part  $\alpha v^2 + \beta v + \gamma$  was obtained by fitting the spike initiation dynamics of a cortical neuron so that the membrane potential  $v$  has *mV* scale and time  $t$  has *ms* scale, and Izhikevich adopts the choice of  $(\alpha, \beta, \gamma) = (0.04, 5, 140)$  and a threshold value of  $v_{\text{peak}} = 30$ . The model has been successful in mimicking, by means of setting the appropriate values of parameters  $a, b, c, d$ , the behavior of neurons from different electrophysiological classes. *Regular spiking (RS) neurons* are the most typical neurons in the cortex. When presented with a prolonged stimulus (i.e., given a step-like injected  $I(t)$ ), the neurons fire a few spikes with short interspike period and then the period increases. In the model, this class of neuron corresponds to  $a = 0.02$ ,  $b = 0.2$ ,  $c = -65$  (deep voltage reset) and  $d = 8$  (large after-spike jump of  $u$ ). The typical voltage series generated by an RS neuron is shown below.



**Figure 6:** The spiking activity exhibited by a typical regular spiking neuron in Izhikevich's base model. The neuron is subject to an injected step-like dc-current. Adapted from [8].

### 3.1 The Conductance-based Model

R. F. O. Pena et al. [9] proposes the following extension, which includes conductance for a more realistic simulation of a neuron's dynamical response after receiving a current input signal, known as the **conductance-based synaptic spiking model**. The dynamical differential equations follow closely Izhikevich's formalism, and the synaptic current term in Eq. (7), for neuron  $i$ , reads:

$$I_i(t) = G_i^{\text{exc}}(t)(V_E - v_i(t)) + G_i^{\text{inh}}(t)(V_I - v_i(t)) + \eta_i, \quad (10)$$

where the current is controlled by **conductances**  $G_i^{\text{exc/inh}}$  and **reversal potentials**  $V_{E/I}$ , responsible for excitatory/inhibitory effects, and  $\eta_i$  is a **noise term**, whose common choices are Gaussian or uniform white noise. This project serves as a preliminary implementation of the model and here we choose to *drive* the dynamics using  $\eta_i$  uniformly distributed over a positive interval  $[0, \delta]$ . (Later after the project, Gaussian white noise will be attempted, which more closely mimics biological synapses.) Note that  $V_{E/I}$  are chosen to assume values larger/smaller than the typical of  $v_i$ , so that most of the time,  $v_i$  lies within the bound  $(V_I, V_E)$ . The time-evolution of the conductances is governed by

$$\frac{dG_i^{\text{exc}}}{dt} = -\frac{G_i^{\text{exc}}}{\tau_{\text{exc}}} + \beta \sum_{\{j: g_{ij} > 0\}} \left( g_{ij} \sum_k \delta(t - t_{j,k}) \right) \quad (11)$$

$$\frac{dG_i^{\text{inh}}}{dt} = -\frac{G_i^{\text{inh}}}{\tau_{\text{inh}}} + \beta \sum_{\{j: g_{ij} < 0\}} \left( |g_{ij}| \sum_k \delta(t - t_{j,k}) \right), \quad (12)$$

where  $\tau_{\text{exc/inh}}$  are the time scales over which conductances decay,  $\beta$  is a dimensionless scalar suitably chosen to set the interaction term to a reasonable scale and controls how strongly the neurons are interacting,  $g_{ij}$  follows our prior definitions and represents the coupling strength linking from node  $j$  to node  $i$ , conveniently contained inside the coupling strength matrix  $\mathbf{G}$ , and  $t_{j,k}$  is the  $k$ -th historical timestamp at which node  $j$  spikes, such that its voltage exceeds the threshold  $v_{\text{peak}} = 30$ . The index sets  $\{j : g_{ij} > 0\}$  and  $\{j : g_{ij} < 0\}$  respectively refer to the nodes that have excitatory/inhibitory links going into node  $i$ , and  $G_i^{\text{exc}}$  is subject only to the impacts of excitatory links while  $G_i^{\text{inh}}$  is subject only to the impacts of inhibitory links. The differential equations can be analytically solved to give

$$G_i^{\text{exc}} = \beta \sum_{\{j: g_{ij} > 0\}} \left( g_{ij} \sum_k \exp\left(-\frac{t - t_{j,k}}{\tau_{\text{exc}}}\right) \Theta(t - t_{j,k}) \right) \quad (13)$$

$$G_i^{\text{inh}} = \beta \sum_{\{j: g_{ij} < 0\}} \left( |g_{ij}| \sum_k \exp\left(-\frac{t - t_{j,k}}{\tau_{\text{inh}}}\right) \Theta(t - t_{j,k}) \right), \quad (14)$$

with  $\Theta(\cdot)$  being the Heaviside step function.

From the analytical solution, we see that the conductances  $G_i^{\text{exc/inh}}$  summarize the historical spike impacts of the connecting nodes and every spike in history drives the current conductances, although with an exponentially decaying magnitude. In theory, we require the whole history of spikes but one may choose to truncate spike data sufficiently far away from the current time  $t$  to speed up the calculation. For example, we can choose to discard timestamp  $t_{j,k}$  if  $t - t_{j,k} > 20 \times \max(\tau_{\text{exc}}, \tau_{\text{inh}})$  since  $\exp(-20) \sim 10^{-9}$  is smaller than the typical values of  $G_i$  by multiple orders. The excitatory conductance  $G_i^{\text{exc}}$  is driven by the connecting nodes with an excitatory link ( $g_{ij} > 0$ ) while the inhibitory conductance  $G_i^{\text{inh}}$  is driven by the connecting nodes with an inhibitory link ( $g_{ij} < 0$ ). The conductances, together with the noise term, then control the *current input* a node receives. The size

of the current is not only determined by the conductances but also by the reversal potentials  $V_{E/I}$ , which tend to drive the membrane potential towards either  $V_E$  or  $V_I$ . These two thresholds tend to *oppose* the directions that  $v_i$  moves because, for example, if  $v_i$  is close to  $V_E$  when it is going to spike,  $I_i$  gets dominated by the second negative term, causing a negative instantaneous rate of change, as seen from Eq. (6). Hence  $v_i$  is never in a state of settlement and instead, oscillates between  $(V_I, V_E)$ . Note that it is entirely possible for voltage  $v_i$  to shoot past the bound and usually this occurs when it spikes or gets reset immediately after a spike. This is clearly seen in the voltage time series in the subsequent section.

**Intuition.** Now, it is time to gain some intuition from the formulas. We can understand the *mechanism of spiking* under this framework. Suppose at some instant, nodes connecting to node  $i$  undergo simultaneous spiking. That is, within a short time interval prior to the current moment, a large collection of nodes spike. The conductances of node  $i$  immediately “detect” the spiking of its neighboring nodes as their spiking time stamps of the spikes that occur just now go into the summation over the whole spiking history in the conductance formulas. Now at this moment, the conductances take a jump. The current quickly becomes sensitive to the difference between the current membrane potential  $v_i$  and the two reversal potentials  $V_{E/I}$ , due to the large conductance multipliers. Assuming  $v_i$  is taking a negative voltage value close to  $V_I$  but far from  $V_E$ , in the synaptic current formula, the first excitatory term then dominates, imposing a tendency on node  $i$  for it to take a large instantaneous rate of change in  $v_i$ . This sequence of reactions causes  $v_i$  to rapidly grow in value as a result. But the voltage cannot grow without limit. As soon as it breaches the threshold  $v_{\text{peak}}$ , it immediately gets reset to value  $c$  by the fire and reset rule. This depicts the complete cycle of a spiking node. The spiking of node  $i$  also exercises a similar impact on other resting nodes, thus inducing a simultaneous spiking activity.

**Summary.** To fully specify a conductance-based synaptic spiking model, we require the following tabulated set of parameters. Then numerically, using the Euler algorithm, the dynamical differential equations Eq. (6) and (7) that govern the state variables  $v_i$  and  $u_i$ , together with the fire and reset rule Eq. (8) and (9), and synaptic current equation Eq. (10) driven by the conductance equations Eq. (13) and (14) are solved, generating a collection of  $N$  time series, one for each node.

Type	Parameters
Network	Coupling strength matrix $\mathbf{G}$
Dynamical	$\{(a_i, b_i, c_i, d_i)\}_{i=1:N}$
Current	$V_E, V_I$
Conductance	$\beta, \tau_{\text{exc}}, \tau_{\text{inh}}$

**Table 5:** Parameters required for specifying a conductance-based synaptic spiking model.

### 3.2 Network Dynamics

With the DIV25 reconstructed coupling strength matrix  $\mathbf{G}$  of 4095 nodes, the set of equations of the conductance-based synaptic spiking model is numerically solved. Recall that  $\mathbf{G}$  is constructed out of the experimental time series of electrical signals generated by real neurons. When  $\mathbf{G}$  is passed to the *logistic network model*, it fails to give realistic neuronal spikes. When  $\mathbf{G}$  is passed to the *FHN network model*, it gives active spiking activity and thus is a significant improvement compared to the logistic model, but on closer examination, its dominance plot of spike counts fails to reach a curvature that is as extreme as the experimental one. This time, we are interested in asking: *Does the synaptic spiking model produce a long-tailed spike count distribution comparable to the experimental one? How do different models compare in terms of the extent of tails?*

We substitute the DIV25 reconstructed coupling strength matrix  $\mathbf{G}$  along with the following tabulated choice of parameters. The neurons are classified into two types according to their excitatory/inhibitory nature: for node  $i$ ,

- (1) if its out-strength  $s_{\text{out}}(i) \geq 0$  (i.e., *excitatory*), it is identified as *regular spiking* (RS) neuron;
- (2) if its out-strength  $s_{\text{out}}(i) < 0$  (i.e., *inhibitory*), it is identified as *fast spiking* (FS) neuron.

Out of the 4095 neurons, 3425 are excitatory while the remaining 599 are inhibitory, therefore in our synaptic spiking model, regular spiking neurons occupy the majority, consistent with the fact that RS neurons are the most prevalent type of neuron in cortex. Collections of excitatory/inhibitory neurons respectively obey *separate* sets of dynamical parameters  $(a_i, b_i, c_i, d_i)$ . For an excitatory neuron, it is generated according to  $(0.02, 0.2, -65 + 15\epsilon_i^2, 8 - 6\epsilon_i^2)$ , whereas for an inhibitory neuron, it is generated according to  $(0.02 + 0.08\epsilon_i, 0.25 - 0.05\epsilon_i, -65, 2)$ , where  $\epsilon_i$  is a random variable uniformly distributed between  $[0, 1]$ . This follows the example network in [8]. The random variable here signifies the fact that each neuron is different, or *heterogeneous*, but respects the same distribution of parameters, hence belonging to the same class. The reversal potentials  $V_E = 0, V_I = -80$  and the conductance decay time scales  $\tau_{\text{exc}} = 5, \tau_{\text{inh}} = 6$  follow the standard choice suggested in [9].  $\beta = 1$  is such chosen that  $I(t)$  has a scale comparable to other terms in Eq. (6). If  $\beta$  is set too small, it causes weak correlations between voltage time series of different neurons; if  $\beta$  is set too large,  $v_i$  could easily “blow up” with very frequent breaching of  $v_{\text{peak}}$ . With these parameters, the model is run for 1000 time steps to give the voltage time series plots.

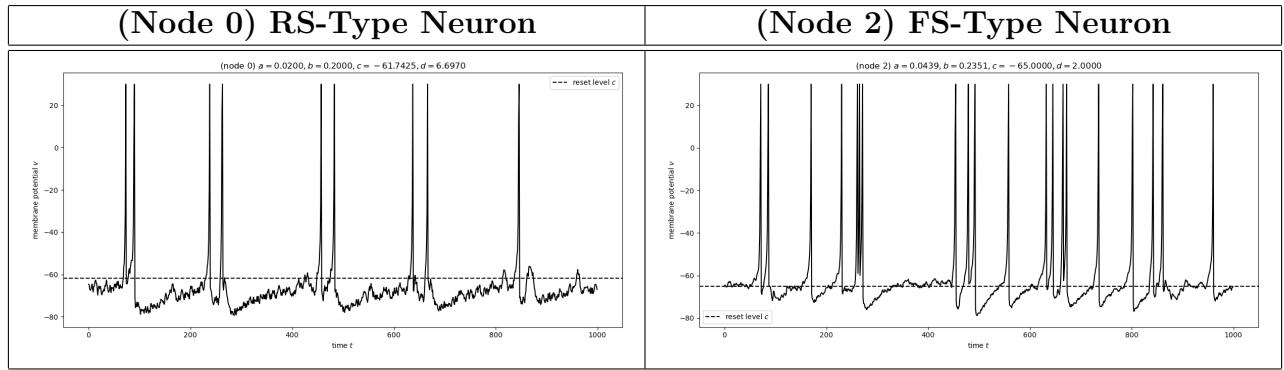
Type	Settings and Parameters
Network	The DIV25 reconstructed coupling strength matrix $\mathbf{G}$
Dynamical	$\{(a_i, b_i, c_i, d_i)\}_{i=1:N}$ generated according to: $(\text{Excitatory} \rightarrow \text{RS-type}) (0.02, 0.2, -65 + 15\epsilon_i^2, 8 - 6\epsilon_i^2)$ $(\text{Inhibitory} \rightarrow \text{FS-type}) (0.02 + 0.08\epsilon_i, 0.25 - 0.05\epsilon_i, -65, 2)$ where $\epsilon_i$ is uniformly distributed between $[0, 1]$
Current	$V_E = 0, V_I = -80$
Conductance	$\beta = 1, \tau_{\text{exc}} = 5, \tau_{\text{inh}} = 6$

**Table 6:** Parameters for the conductance-based synaptic spiking model based on DIV25 reconstructed coupling strength matrix.

Prior to performing the simulation, we can expect an RS-type neuron to fire a few spikes with short interspike period, then the period increases until another round of spike burst, and an FS-type neuron to fire periodic trains of spikes with high frequency. Inspection of the *voltage time series* verifies these.

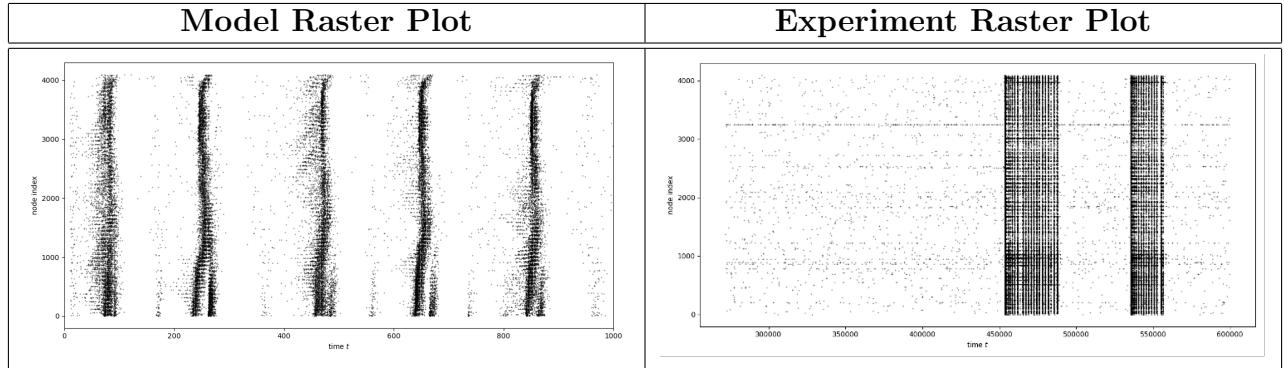
The model is successful in generating active spiking activity throughout the whole dynamics, and

features of regular spiking and fast spiking neurons can be clearly recognized by examining the voltage time series plots. Below the voltage time series of node 0, which belongs to RS since  $s_{\text{out}}(i = 0) \geq 0$ , and node 2, which belongs to FS since  $s_{\text{out}}(i = 2) < 0$ , are shown. It is seen that indeed the two types of neuron exhibit the expected kind of behavior. For RS-type, there exists a relatively long period of steady states after firing a few spikes, and the voltage gradually rises until the next round of spiking is triggered. For FS-type, throughout the whole time period, the neuron is persistently spiking with high frequency, and in particular, at time step 250, there is a dense concentration of spikes within a short time interval. Notice also that a spike cannot persist for any longer after the voltage breaches the  $v_{\text{peak}} = 30 \text{ mV}$ , because as soon as the voltage crosses 30 mV, in the next time step, it is immediately drawn to the reset level  $c_i$ , indicated by the black dashed line in the plots, by the fire and reset rule inherent in the model. Afterwards, the voltage is subject to random fluctuations and waits for the next spike burst.



**Figure 7:** Voltage time series plots of node 0, which is an RS-type neuron, and node 2, which is an FS-type neuron.

It is also informative to visualize the *timesteps* at which spikes happen, through the *spike raster plot*, which illustrates the spike timestamps for each node. In the model, the instant at which a spike occurs is well-defined: once the voltage breaches 30 mV, there is a spike. This allows spikes to be pinpointed accurately unlike our previously attempted models. For example, in the FHN model, we locate spikes through a peak detection algorithm and additional threshold criteria, and this inevitably involves the mis-counting of noise as spikes. There is no clear-cut difference between noise and spike, beyond the judgment of algorithm. The synaptic spiking model precludes the need of a peak detection algorithm and thus simplifies the spike counting process.



**Figure 8:** Raster plot generated from the synaptic spiking model in comparison with the experiment raster plot of real cortical neurons.

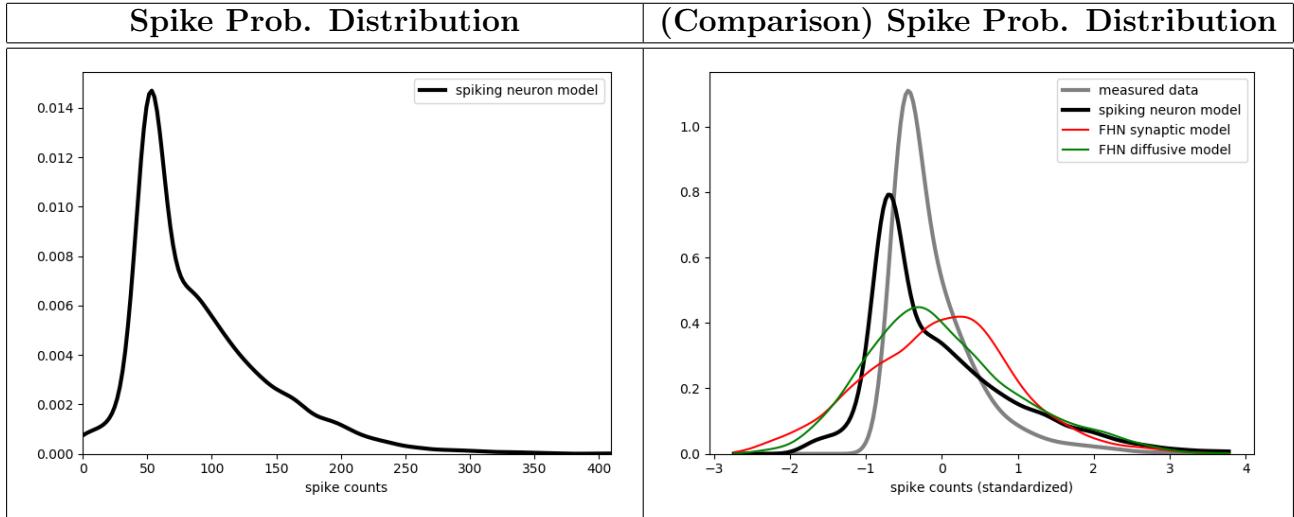
From the *model* raster plot, it is observed that there exists large-scale *periodic, simultaneous spiking activity* with a period that spans roughly 200 time steps. There is also sub-period spiking activity,

and there is a moderate number of nodes that spike midway between two large-scale bursts. Notice also that each spike burst is not contained a well-defined period, but rather, spread out. Prior to the consolidation of a spike train, a number of nodes are already spiking, and the number continuously builds up, eventually giving the solid concentration of spikes. Focusing on a spike train, behaviors of the neurons are not uniform, with some spiking more frequently and some spiking less. This gives rise to the “dispersion” of spike data points over a spike train.

In contrast with the model raster plot, the *experiment* raster plot that reflects the experimentally observed spiking activity shows the otherwise in some ways. The spike trains are relatively uniform for different neurons and there exists a well-defined period inside which a spike train is contained. On closer examination, one can see grid-like patterns within a spike train. The line of spike data points at a fixed time implies that a large number of neurons spike at the same exact moment, while the line of points for a fixed node implies that certain neurons spike more frequently than others, which is a clear sign of long-tailed spike counts. Such grid-like patterns are not observed in the model raster plot, and in that aspect, the synaptic spiking model is not accurate. Nonetheless, there is still room for improvement for the model. For example, the model is not calibrated and in this project, the parameters are chosen merely out of convenience. The dynamical parameters ( $a, b, c, d$ ), reversal potentials  $V_{E/I}$  and conductance decay time scales  $\tau_{\text{exc/inh}}$  have to be adjusted for a more realistic replication of the experimental scenario, and future efforts can be devoted to *parameter calibration*.

### 3.3 Comparison with the FHN Network Model

To further examine the network dynamics at a level beyond the visualization of time series, we look into the *distribution of spike counts*. To calculate the spike count of a node in a simulation, we count the number of timestamps at which the voltage  $v_i$  breaches the threshold  $v_{\text{peak}}$ , then from the collection of spike counts for each node, we generate the spike count distribution. Below we show (1) the spike count distribution of the synaptic spiking model, and (2) a comparison of spike count distribution of the measured data, FHN network models and synaptic spiking model.



**Figure 9:** The probability distribution plots of the spike counts measured experimentally, generated by synaptic spiking model and FHN models.

	Min.	Max.	Med.	Mean	Skew.	Ex. Kurt.	Tail Measure
Measured Data	0	3409	77	166.68	5.40	39.21	1.49
Spiking Neuron	0	409	74	90.46	1.43	2.66	1.01
FHN Synaptic	283	416	320	319.60	0.30	1.26	0.22
FHN Diffusive	308	374	330	331.29	0.59	0.27	0.48

**Table 7:** The quantiles, moments and tail measure of the spike counts measured experimentally, generated by synaptic spiking model and FHN models.

First, we focus on the synaptic spiking model. From the distribution plot, a *long tail* is observed with the spike data spanning a wide interval, with the  $x$ -axis stopping at the maximum datum. Some nodes give no spike at all while some generate spikes at a significant higher frequency, with 409 being the maximum. This is precisely the phenomenon observed in experiment, where we have a large concentration of neurons with relatively few spike counts and in contrast, a minority exhibiting a very large number of spike counts. To *quantify* the distribution, we look at the quantiles, moments and the tail measure  $S(95\%)$ . In general, for a long-tailed distribution, we expect to observe (1) the mean being greater than the median, (2) the maximum being significantly greater than the median, (3) high skewness (if the tail is one-sided) and high excess kurtosis. *These hold to some extent for the synaptic spiking model so we identify it as being able to generate long-tailed spike counts.*

Then, in comparison with our previous models, namely the FHN diffusive and FHN synaptic model, from the distribution comparison plot which *standardizes* the spike counts, it is seen that the distribution curve of the synaptic spiking model more closely resembles the experimental one. Both have a sharp peak locating between  $(-1, 0)$ , indicating a concentration of neurons with small spike

counts. In terms of the statistics, the skewness and excess kurtosis of the synaptic spiking model are the highest among the other models, although still far away from that of the measured data.

In addition, we define for a dataset  $\{X_i\}$  a *tail measure*  $S(q)$  to quantify the extent of the tail, given by

$$S(q) = \frac{E[X|X > X_q] - X_{50\%}}{\sigma(X)} - S_0(q) \quad (15)$$

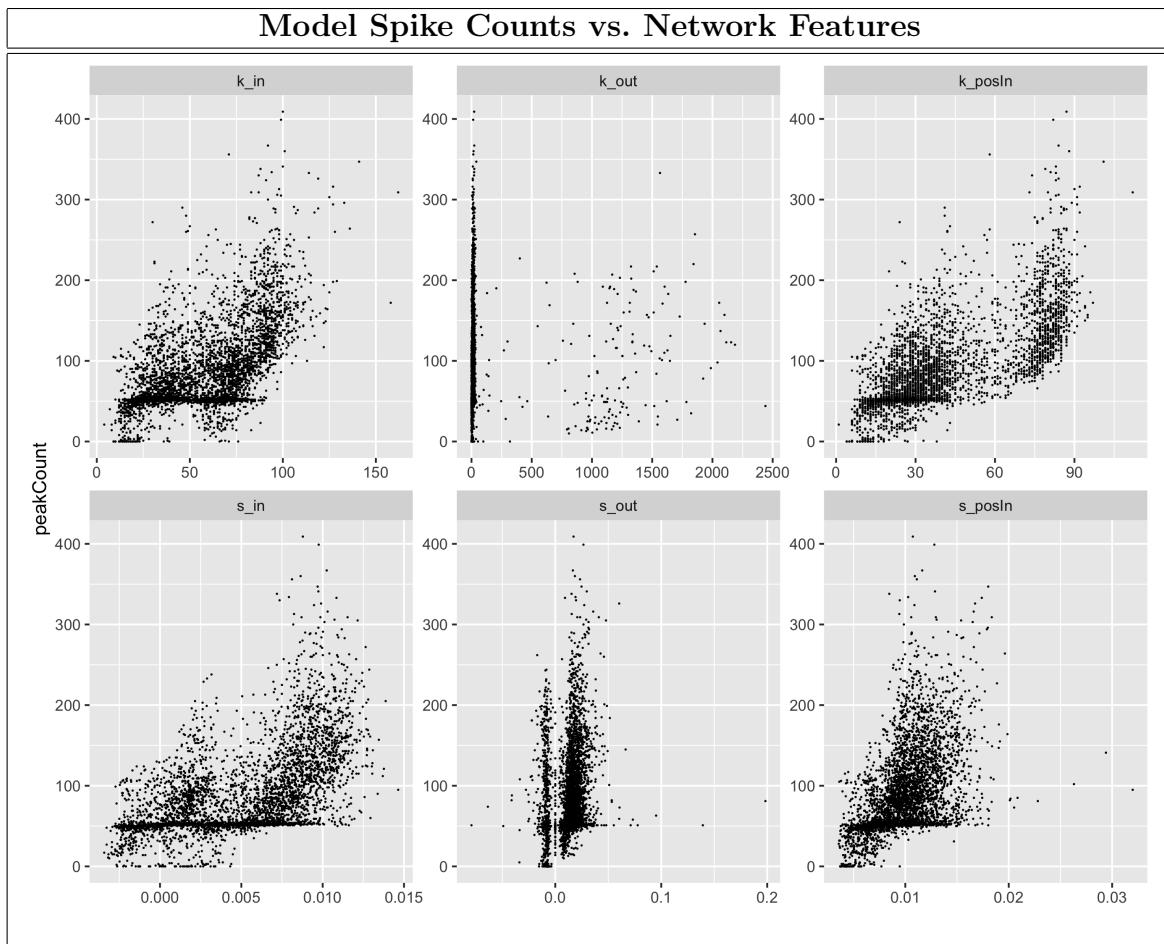
where  $E[X|X > X_q]$  is the conditional mean for data that are greater than the  $q$ -quantile,  $X_{50\%}$  is the median (i.e., 50%-quantile),  $\sigma(X)$  is the standard deviation for the whole dataset, and  $S_0(q) = E[Z|Z > Z_q]$  with  $Z$  being a random variable obeying the standard Gaussian  $\mathcal{N}(0, 1)$ . Fixing  $q$ , the tail measure first computes the mean of the data beyond the  $q$ -quantile, recognizing that as the “tail”, then standardizes the conditional mean but instead of the sample mean, the median is subtracted as it is less influenced by outliers for long-tailed data, and finally compares with a reference  $S_0(q)$ , which is the conditional mean applied to a standard Gaussian. The measure is zero if the data obey a Gaussian distribution. For a distribution with a tail “longer” than that of Gaussian, the measure is positive.

From the table, we see that the synaptic spiking model is more comparable to the measured data in terms of the tail measure than the FHN models. This, together with moment statistics, serve as evidence that *the synaptic spiking model is more successful in producing long-tailed spike counts than our previous FHN models*. But this does not end the journey for the search of better models yet. Recall that the parameters in our synaptic spiking model are still uncalibrated and we expect improvements in the spike tail upon further efforts in calibration.

## 4 Effects of the Distribution of Synaptic Strength

The conductance-based synaptic spiking model is so far the best-performing model among all the attempted ones – it succeeds in producing neuronal spikes and the spike counts have a long tail. Although the tail is not as extreme as the empirical one, we are one step closer. A question one may raise is: *Are the spike counts associated in any way with the network features, including degrees and strengths? Further, does the distribution of synaptic strengths affect the distribution of spike counts?* In this part, we attempt to answer the question *in the context of the synaptic spiking model*, so results here are model-specific and do not necessarily generalize to models that obey other kinds of dynamics.

As the first step, an *exploratory data analysis* is done to examine how the spike counts depend on the network features. Here, a narrow, but the most plausible, subset of network features is selected, and the spike counts are plotted against each of the network features, which include in-degrees ( $k_{in}$ ), out-degrees ( $k_{out}$ ), positive in-degrees ( $k_{posIn}$ ), in-strengths ( $s_{in}$ ), out-strengths ( $s_{out}$ ) and positive out-strengths ( $s_{posOut}$ ), the definitions of which can be found in the introduction. Some non-linear relations can be seen in all except the  $k_{out}$  plot. *It can be concluded that, in general, large spike counts are associated with large degrees and large strengths.* Nodes with moderately large in-degrees and in-strengths can have small spike counts, though. Interestingly, these results coincide with the analysis on the FHN network models, in the first part of the project, suggesting the generality of the conclusion.



**Figure 10:** An exploratory analysis on the relations between the spike counts of the synaptic spiking model and different network features.

## 4.1 Reference Network Analysis

The exploratory data analysis visually pinpoints the network features that are *associated with* the spike counts. Going further, another question one may raise is: *What network features cause the spike counts to obey the distribution that they do?* Here, we investigate the causality relation between the network features and the spike counts. We rely on the use of *reference networks*, which are networks derived out of the original network of interest but have certain network features varied and the rest fixed. *Ideally*, one particular feature is varied and its effect on the network dynamics can be understood. This, however, is practically impossible given that the network features have strong dependence on each other – they are all derived out of the coupling strength matrix  $\mathbf{G}$ . When we manipulate a certain feature, some other features are inevitably affected, so no feature can be *marginally* manipulated. In this part, we rely on a *shuffling technique* to derive reference networks out of the original network with the reconstructed  $\mathbf{G}$ , in an attempt to maximally preserve the other features while manipulating certain features of interest. This was inspired by the bootstrap technique in non-parametric statistics, which allows the generation of new samples that approximately obey the distribution that the original dataset respects through sampling with replacement, with the modification that, instead of sampling, *shuffling* is used.

The following five reference networks are designed. Based on the original network with DIV25 reconstructed coupling strengths  $g_{ij}$ ,

- (*Reference Network 1*) The non-zero  $g_{ij}$  are replaced with samples from  $\mathcal{N}(\mu, \sigma)$ , where  $\mu$  and  $\sigma$  are respectively the mean and the standard deviation of the non-zero  $g_{ij}$ .
- (*Reference Network 2*) For each  $i$ ,  $g_{ij}$  are shuffled. Effectively, each row in  $\mathbf{G}$  is replaced by a permuted vector of all entries (zeros included) in the row.
- (*Reference Network 3*) For each  $j$ ,  $g_{ij}$  are shuffled. Effectively, each column in  $\mathbf{G}$  is replaced by a permuted vector of all entries (zeros included) in the column.
- (*Reference Network 4*) An independently generated DWR network using the same connection probability  $p = \sum \mathbb{1}(g_{ij} \neq 0)/N(N - 1)$  with coupling strengths sampled from  $\mathcal{N}(\mu, \sigma)$ .
- (*Reference Network 5*) The non-zero  $g_{ij}$  are shuffled.

The computational construction and the intention of each reference network are summarized.

Ref. Net.	Computational Construction	Intention
1	Non-zero $g_{ij} \rightarrow \mathcal{N}(\mu, \sigma)$	Keep $\{k_{\text{in}}(i), k_{\text{out}}(i)\}$ but vary $\{s_{\text{in}}(i), s_{\text{out}}(i)\}$
2	Shuffle rows in $\mathbf{G}$	Keep $\{k_{\text{in}}(i), s_{\text{in}}(i)\}$ but vary $\{k_{\text{out}}(i), s_{\text{out}}(i)\}$
3	Shuffle columns in $\mathbf{G}$	Keep $\{k_{\text{out}}(i), s_{\text{out}}(i)\}$ but vary $\{k_{\text{in}}(i), s_{\text{in}}(i)\}$
4	Random network	All network features are structure-free
5	Shuffle non-zero entries in $\mathbf{G}$	Keep $\{k_{\text{in}}(i), k_{\text{out}}(i)\}$ but vary $\{s_{\text{in}}(i), s_{\text{out}}(i)\}$

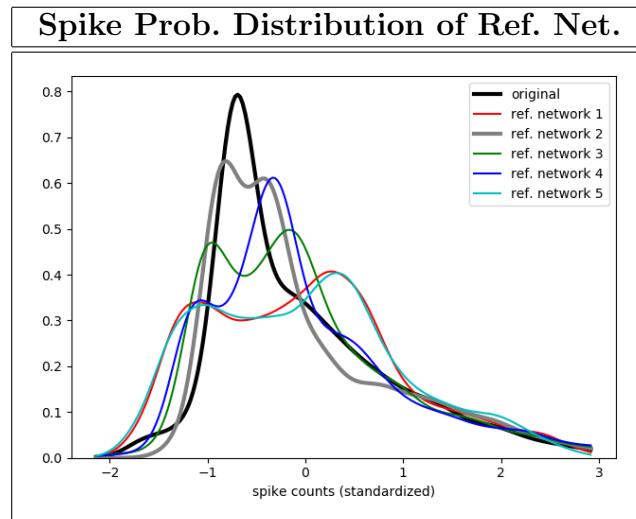
**Table 8:** The computational construction and the intention of the five reference networks used for understanding the effects of the varied network features.

## 4.2 Drivers of the Long-tailed Spike Count Distribution

For the reference networks, *the effects of the varied network features* are what we seek to investigate. For each, the conductance-based synaptic spiking model, *assuming the constructed reference coupling strength matrix  $\mathbf{G}$* , is run and the model time series are returned. If, for example, using ref. net. 2, the spiking dynamics turns out to be similar to what the original network produces, then we can conclude that network features  $k_{\text{out}}(i)$  and  $s_{\text{out}}(i)$  are not important driving factors for the spiking dynamics while  $k_{\text{in}}(i)$  and  $s_{\text{in}}(i)$  *probably* are. (As we shall see, this is indeed the case.) Note again that the features being *kept* have distributions *identical* to that of the original network, which are long-tailed and contain its “unique structure”, while the features being *varied* have distributions *close to Gaussian*. Interested readers may refer to the appendix for the precise distributions of the in/out-strengths for the reference networks.

The probability distributions of the standardized spike counts of the five ref. net. are shown. Notice that the distributions are *all* deviating from Gaussian with a long right tail. In particular, for ref. net. 4, although all its network features are structure-free and obey Gaussian distribution, its spike count distribution does not. This is unlike our previous result, where under the FHN dynamics, we observe that *Gaussian* network features lead to *Gaussian* spike counts. From this observation, *it is concluded that the network dynamics is an important factor of spike count distribution, and structure-free features do not necessarily lead to structure-free spike counts*.

The distributions of the original network and ref. net. 2 are the closest in shape while the other ref. net. show obvious deviation from the original distribution. This is confirmed by the tabulated statistics, from which skewness, excess kurtosis and tail measure of ref. net. 2 are seen to be the most comparable to original. Recall that ref. net. 2 has the rows in the original reconstructed coupling strength matrix  $\mathbf{G}$  shuffled and, therefore,  $\{k_{\text{in}}(i), s_{\text{in}}(i)\}$  are preserved but  $\{k_{\text{out}}(i), s_{\text{out}}(i)\}$  are varied. That it has a spike probability distribution close to the original implies that  $k_{\text{out}}(i)$  and  $s_{\text{out}}(i)$  are not important driving factors of the spiking dynamics while  $k_{\text{in}}(i)$  and  $s_{\text{in}}(i)$  *could be*. For ref. net. 1 and 5,  $\{k_{\text{in}}(i), k_{\text{out}}(i)\}$  are preserved but  $\{s_{\text{in}}(i), s_{\text{out}}(i)\}$  are varied. The preservation of  $k_{\text{in}}(i)$  does not leave the spike probability distribution unchanged. Thus, we can further rule out  $k_{\text{in}}(i)$  as a driving factor of the spiking dynamics. *In summary, through understanding the effects of the varied network features of the ref. net., only  $s_{\text{in}}(i)$  is identified to be the factor that affects the spiking dynamics.*

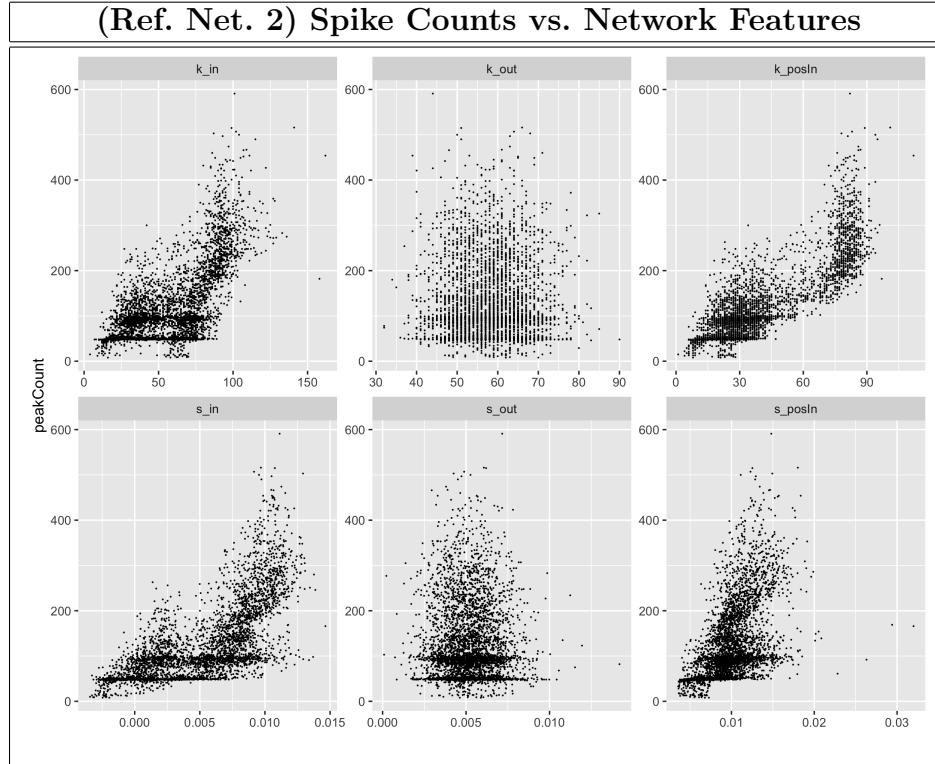


**Figure 11:** The probability distribution of the standardized spike counts of the five reference networks. That of ref. net. 2 is similar to the original while the rest show obvious deviation.

	Min.	Max.	Med.	Mean	Skew.	Ex. Kurt.	Tail Measure
<b>Original Net.</b>	0	409	74	90.46	1.43	2.66	1.01
<b>Ref. Net. 1</b>	0	693	207	210.76	0.53	-0.05	0.32
<b>Ref. Net. 2</b>	8	591	95	124.77	1.42	1.85	1.03
<b>Ref. Net. 3</b>	46	329	123	130.42	1.16	1.45	0.82
<b>Ref. Net. 4</b>	47	342	132	143.22	1.13	1.24	0.89
<b>Ref. Net. 5</b>	12	524	222	221.05	0.41	-0.44	0.15

**Table 9:** The quantiles, moments and tail measure of the spike counts generated by the original synaptic spiking model and its reference networks.

The correlations between the spike counts and the network features for the ref. net. are also looked at to reinforce the above argument. Here, only plots of ref. net. 2 is examined and for the rest, interested readers may refer to the appendix. For ref. net. 2, by construction, the in-measures are preserved, and the  $k_{in}$ ,  $k_{posIn}$ ,  $s_{in}$  and  $s_{posIn}$  plots are almost identical to that of the original network (see Figure 10). In other words, even though the distributions of  $k_{out}(i)$  and  $s_{out}(i)$  are disrupted by the row-shuffling during the construction of ref. net. 2, the correlation structures between the spike counts and the different in-measures, including  $k_{in}$  and  $s_{in}$ , are well preserved. *This suggests that the distributions of the out-measures do not have an important effect on the spiking dynamics while the distributions of the in-measures do.* We have seen that both the incoming strengths and the spike counts have a long-tailed distribution – this is also what motivates this research project. *Are they associated?* The plots suggest an affirmative answer, at least in the context of this synaptic spiking model with the diffusive nodal interaction. In the  $s_{in}$  plot, large  $s_{in}$  is associated with large spike counts – they respectively correspond to the tail in the incoming strength distribution and the spike distribution.



**Figure 12:** The relations between the spike counts of ref. net. 2 and different network features.

## 5 Conclusion

Continuing from its first part in the first semester and following tightly the methodology, this project studies the dynamics of the conductance-based synaptic spiking model, extended from its base model first proposed by E. M. Izhikevich, and investigates the effects of the distribution of synaptic strength on the spike count distribution through the *reference network analysis*.

There are many measures to aid us in understanding a network. In this project, we focus on degrees, including the in-degrees  $k_{\text{in}}(i)$  and the out-degrees  $k_{\text{out}}(i)$ , and strengths, including the in-strengths  $s_{\text{in}}(i)$  and the out-strengths  $s_{\text{out}}(i)$ . The *synaptic spiking model* is the core subject to examine. By numerically solving the dynamical equations with a fire and reset rule embedded, the model generates realistic neuronal spike dynamics, and certain features observed in experiments, including large-scale simultaneous, periodic spiking activity, can be recovered. In comparison with our previous models, namely the logistic and FHN network models, the synaptic spiking model is superior as it is able to produce a spike distribution curve with a higher skewness and a longer tail, with the spike data covering a wide range: a concentration of neurons with small spike counts and a minority exhibiting very large spike counts. Future efforts could be focused on the calibration of parameters because in our model, the parameters are simply chosen out of convenience and we expect improvements upon calibration, in particular, the dynamical parameters  $\{a, b, c, d\}$ .

To understand the relation between the long-tailed spike counts and different network features, we adopt the *reference network analysis*, which have artificially constructed  $\mathbf{G}$  to vary some features of interest while preserving the other. From the feature plots, we see some non-linear relations between the spike counts and different network in-measures. With the reference networks, the long-tailed *incoming strengths*  $s_{\text{in}}(i)$  are determined to be the driving factor for the long-tailed spike counts. The distribution of  $s_{\text{in}}(i)$  is found to have a concrete effect on the distribution of spike counts, as ref. net. 2, which preserves  $s_{\text{in}}(i)$ , generates heavy-tailed spike counts while the other ref. net. have  $s_{\text{in}}(i)$  that are approximately Gaussian and this causes their spike counts to deviate from that of the original network. Through comparing the ref. net., we can also rule out other network features as driving factors. Therefore, to answer the questions in the objectives,

- (1) For the conductance-based synaptic spiking model, the long-tailed distribution of *incoming strengths*  $s_{\text{in}}(i)$  leads to the long-tailed spike counts while other network features have no obvious effect on the spiking dynamics.
- (2) The *conductance-based synaptic spiking model* has so far the best performance in reproducing spike counts and is able to recover some of the empirical statistical features, such as the high skewness and the long tail, among all our previous attempts such as the logistic and FHN network models. However, the moment and tail measure statistics tell us that the model is still far from achieving the long-tailedness the experimental spike counts have, so there is still a lot of room for improvement, and the search for realistic models of neurons will still continue.

## 6 Datasets and Codes

All datasets and codes necessary for replicating the simulation results in this project are [here](#). The data format and their usage are discussed below.

There are two dataset files, namely `DIV25_PREmethod` and `DIV25_spks`.

- (*Reconstructed G*) `DIV25_PREmethod` contains the reconstructed coupling strengths  $g_{ij}$  in the format described in the introduction, and is provided by my supervisor. Each row takes the format `i j g_ij`, and `g_ij` is the coupling strength of the edge linking from node `j` to node `i`. Unstated node pairs are implied to be unconnected. The file is all the data necessary to construct a network.
- (*Spike Counts*) `DIV25_spks` contains the spike count data of collected by the 4095 electrodes, which measure the electrical signals generated by the rat embryonic neurons over 25 days in vitro. Each row takes the format `n t1 t2 ... tn`, where `n` is the total spike counts of a specific node measured over the period and `t1, t2, ..., tn` are the computational time steps at which a spike is detected to occur. In this project, only the first datum in the rows is used and the subsequent data, representing the spike timing, are not analyzed, which, however, constitute another important characterization of a neuronal dynamics. Looking at the first few rows of the data file immediately tells us that heavy-tailed nature of neuronal spikes – the first node has 521 spikes while the subsequent few have less than 10, which is a very sharp contrast.

For the codes, only an overview is provided here and the computational details are not discussed. The codes consist of a large class called `SpikingNeuronModel`. The most important variables and functions for the two classes are listed below. Functions are followed by a bracket () and the arguments, if any, are omitted.

- `N`: the number of nodes in the network
- `Coupling`: the coupling strength matrix `G`
- `Adjacency`: the adjacency matrix derived from `Coupling`, with entries  $A_{ij} = \mathbb{1}(g_{ij} \neq 0)$
- `a,b,c,d`: the dynamical parameters in the synaptic spiking model, and each is a vector with each entry assigned to a node
- `historicalDecayFactorSum()`: a vector of sums of the historical impacts of a collection of other nodes scaled by decay factors, used for the calculation of excitatory/inhibitory conductances
- `initDynamics()`: initialize the spiking neuron dynamics by passing in the total iterations that the model has to run. In particular, `voltage` and `recover` vectors are initialized inside, either with a new set of values or continued from previous simulation
- `runDynamics()`: iteratively update the nodal states using the Euler algorithm until the final time step is reached

The other functions in the `SpikingNeuronModel` class are for analysis purpose and are mainly used to generate plots. Some examples include `plotRaster()`, which generates a raster plot of the spiking timestamps for the whole set of neurons, and `plotTimeSeries()`, which plots the voltage time series for a user-defined set of neurons.

## 7 Acknowledgments

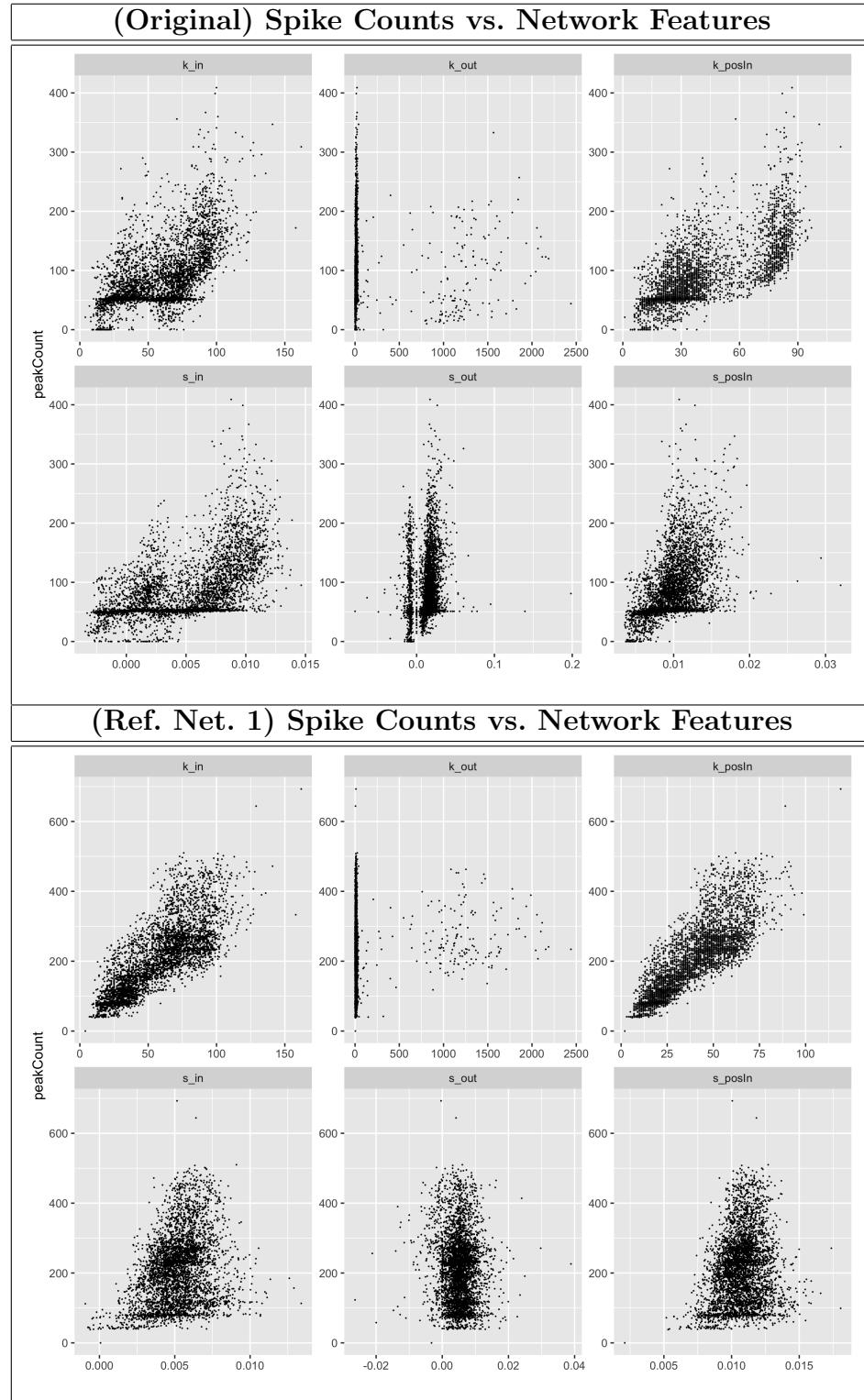
The author thanks Prof. Ching for her advice and guidance throughout the research project and on this report.

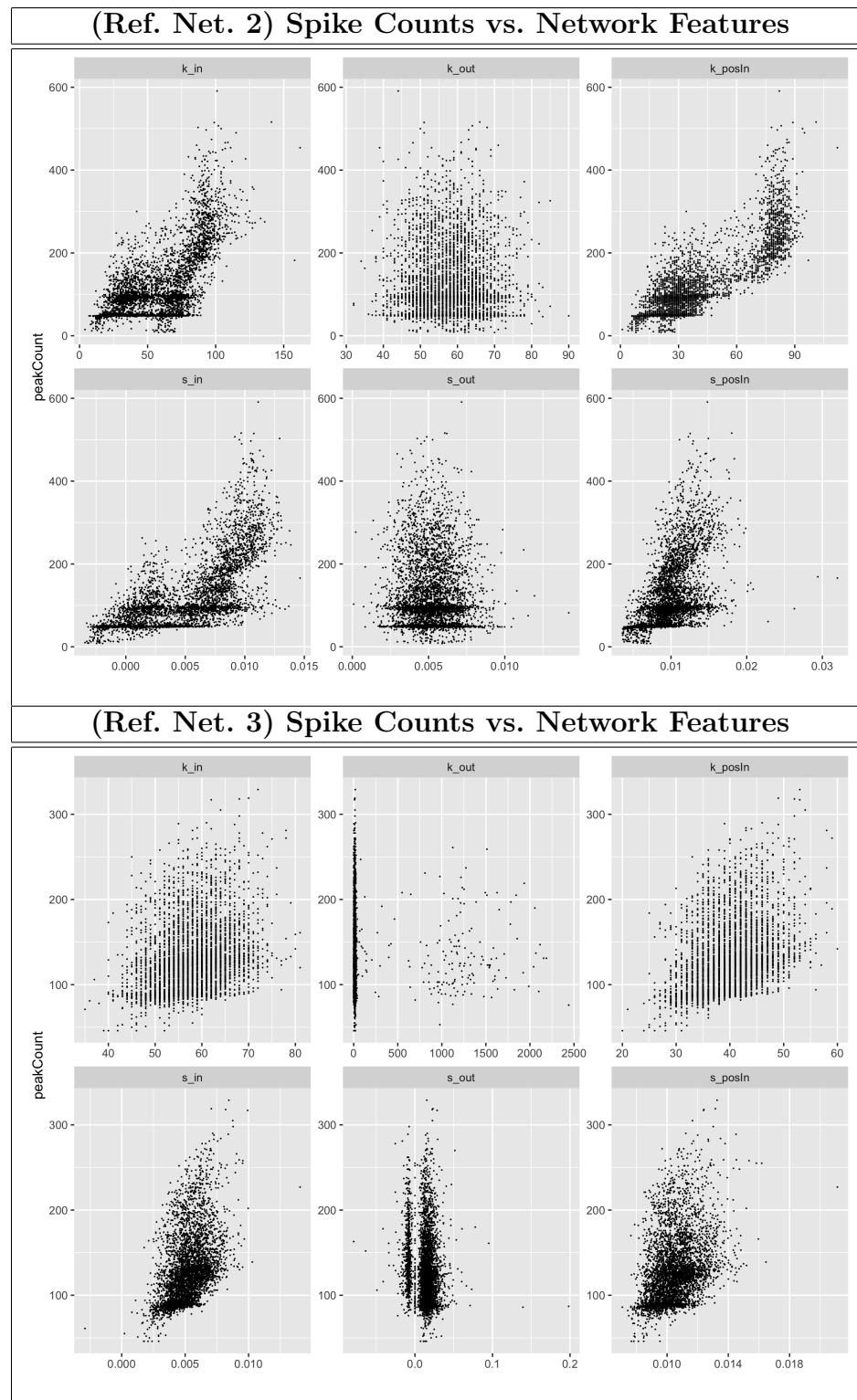
## 8 References

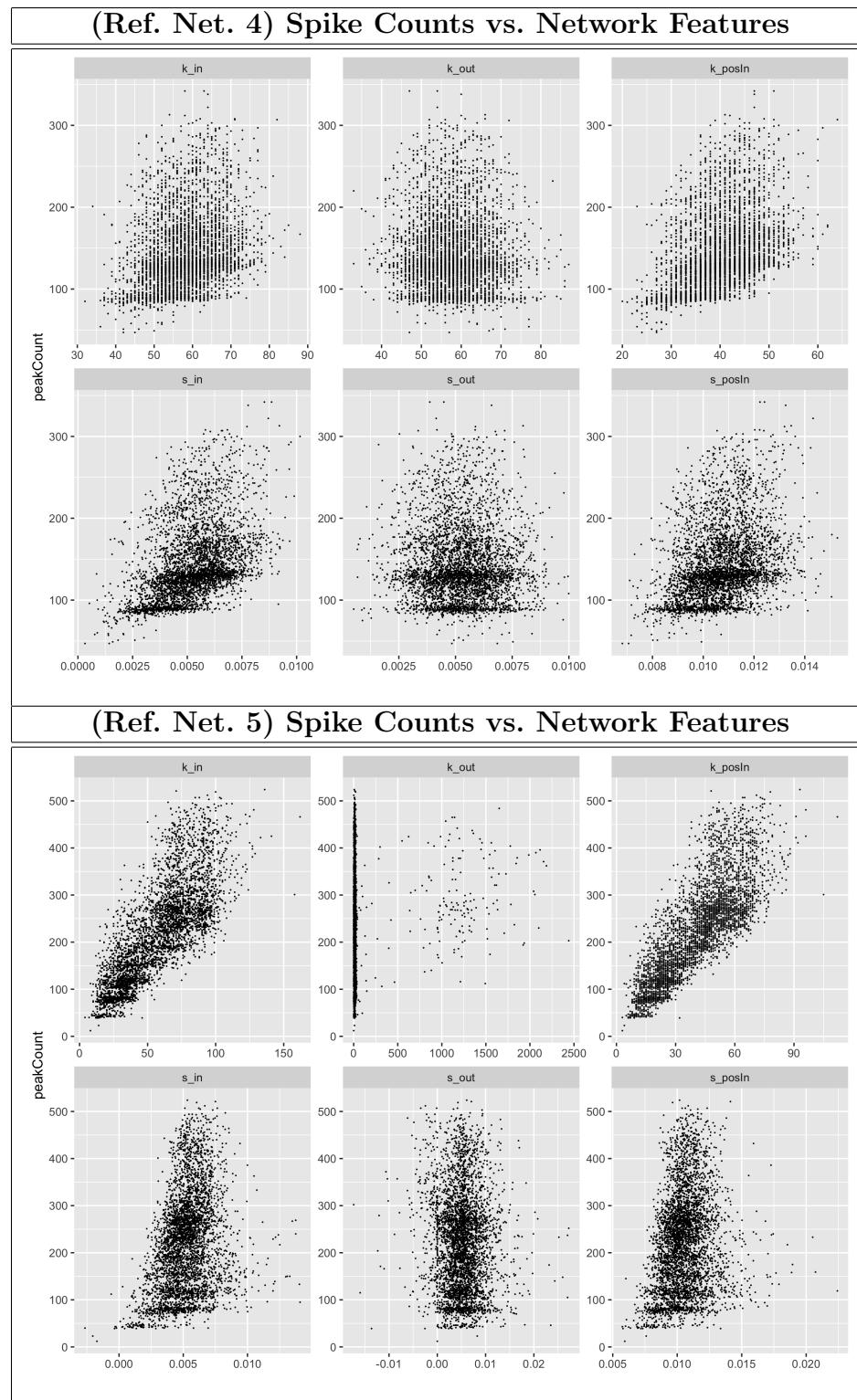
- [1] S. H. Strogatz, Exploring complex networks, *Nature* (London) 410, 268 (2001).
- [2] B. Albert-László and A. Réka, Emergence of scaling in random networks, *Science*. 286 (5439): 509–512.
- [3] E. S. C. Ching and H. C. Tam, Reconstructing links in directed networks from noisy dynamics, *Phys. Rev. E* 95, 010301(R) (2017).
- [4] C. Sun, K. C. Lin, Y. T. Huang, E. S. C. Ching, P. Y. Lai and C. K. Chan, Directed effective connectivity and synaptic weights of in vitro neuronal cultures revealed from high-density multielectrode array recordings, *BioRxiv* 2020.02.06.936781 [Preprint], available from <https://doi.org/10.1101/2020.02.06.936781>.
- [5] G. Buzsáki and K. Mizuseki, The log-dynamic brain: how skewed distributions affect network operations, *Nat. Rev. Neurosci.*, 2014;15(4):264–278.
- [6] L. Kuśmierz, S. Ogawa, and T. Toyoizumi, Edge of Chaos and Avalanches in Neural Networks with Heavy-Tailed Synaptic Weight Distribution, *Phys. Rev. Lett.* 125, 028101 (2020).
- [7] C. Y. Yeung, Effects of the Distribution of Synaptic Strength of Neuronal Network on its Spiking Dynamics (Final Year Project Report, 2020-21 Semester 1), available [here](#).
- [8] E. M. Izhikevich, Simple model of spiking neurons, *IEEE Transactions on Neural Networks*, 2003;14:1569–1572.
- [9] R. F. O. Pena, M. A. Zaks, A. C. Roque, Dynamics of spontaneous activity in random networks with multiple neuron subtypes and synaptic noise, *Journal of Computational Neuroscience*, 2018;45:1–28.

## 9 Appendix

Here, the feature plots of the original network and the five reference networks are shown. The original network obeys the reconstructed coupling strength matrix  $\mathbf{G}$  while the reference networks obey their respective constructed  $\mathbf{G}$ . They all respect the conductance-based synaptic spiking model. Compared to the original network, for ref. net. 2, the relations between the spike counts and different in-measures are basically preserved, while for the other ref. net., the spike counts no longer resemble the original distribution as the “structures” in the in-measures, in particular the incoming strength  $s_{in}$ , are disrupted through the shuffling construction.







Here, the distributions of the standardized strengths for the reference networks are shown. The ref. net. have certain features left untouched while some other varied. The network features being kept have distributions identical to that of the original network, while those being varied have bell-shaped distributions close to Gaussian. As an example, the coupling strength matrix  $\mathbf{G}$  of ref. net. 2 has the rows shuffled, keeping  $\{k_{\text{in}}(i), s_{\text{in}}(i)\}$  while varying  $\{k_{\text{out}}(i), s_{\text{out}}(i)\}$ . Therefore, we can see that the distribution of the in-strengths is preserved (identical to the original) but the distribution of the out-strengths becomes close to Gaussian (the long-tailed structure is destroyed through shuffling). The effects of the varied network features are studied through the *reference network analysis*.

