EXAMINING HETEROGENEOUS WEIGHT PERTURBATIONS IN NEURAL NETWORKS WITH SPIKE-TIMING-DEPENDENT PLASTICITY

by

Colin Bredenberg

University of Pittsburgh, 2017

Submitted to the Graduate Faculty of

the Department of Mathematics in partial fulfillment

of the requirements for the degree of

Bachelor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

MATHEMATICS DEPARTMENT

This thesis was presented

by

Colin Bredenberg

It was defended on

April 14, 2017

and approved by

Brent Doiron, Department of Mathematics

Bard Ermentrout, Department of Mathematics

Jonathan Rubin, Department of Mathematics

Ashok Litwin-Kumar, Columbia University

Thesis Advisor: Brent Doiron, Department of Mathematics

EXAMINING HETEROGENEOUS WEIGHT PERTURBATIONS IN NEURAL NETWORKS WITH SPIKE-TIMING-DEPENDENT PLASTICITY

Colin Bredenberg, BPhil

University of Pittsburgh, 2017

Large-scale cortical networks employing homeostatic mechanisms and synaptic plasticity rules have been shown to differentiate into neural ensembles when common stimuli are applied in tandem to selected subsets of neurons. These ensembles were found to be stable in response to small perturbations to synaptic strengths[13]—such ensemble stability is a critical feature for network-based memory. Previous studies applied relatively simple perturbations to probe the stability of the network—all synapses within a given population were lowered by a uniform percentage[13]. The goal of this work has been to analyze whether more complex perturbations can reveal more information about network stability. Towards this aim, we constructed a reduced stochastic Wilson-Cowan model[2], which captures the same perturbation phenomenon observed in spiking simulations, but which is analytically much simpler. We found that when the mean self-excitatory synaptic weight for a population was preserved, perturbations that were distributed more evenly among synapses would lead to a more stable response than focused perturbations, and that this was caused by quantization of neural activity levels within a population.

TABLE OF CONTENTS

LIST OF FIGURES

1.0 INTRODUCTION

Neuronal ensembles have the potential to serve as a critical mechanism for memory, stimulus encoding, and computation in the brain[23],[12]. There is significant experimental evidence of the existence of these ensembles[17],[16],[15], but critical aspects of their function are as of yet unexplored. Specifically, the dynamics cortical ensemble dissolution has not been examined in response to stimuli, either simulated or experimental, designed to remove or weaken connections select subsets of neurons from within an ensemble.

Spiking network simulations of assembly formation may be constructed through spike-timing dependent plasticity (STDP), coupled with spatial structure[16], lateral inhibition[12], or inhibitory STDP [13], in addition to heterosynaptic competitive mechanisms[18]. These simulations show ensembles active during both stimulus phases and quiescent states, reproducing experimental data [19],[14],[1]. STDP is often mediated by NMDA receptors, which have been linked to sequential stimulus representation in the visual cortex[23]. Further, the presence of STDP has been linked to the spontaneous formation of clusters in theoretical simulations[8], which gain excitatory interconnective strength when the constituent neurons are co-stimulated[11]. These clusters are co-activated with predictable temporal sequences, and are stable on the order of at least 24 hours[11]. Inhibitory interneurons prevent the coactivation of more than a few subsets at a time [13]. This leads to a paradigm where in response to a stimulus, the ensemble corresponding to this stimulus will fire, suppressing the activity of all other neurons, causing a marked reduction in the populations Fano factor[13],[6], as predicted by experiments[4]. Thus, ensemble formation is a demonstrated property of cortical neuron populations, and these ensembles are able to explain electrophysiological phenomena.

Theoretical constructs for analyzing assembly formation and dynamics are also growing, but remain deficient. Bressloff provides a tractable mechanism for the analysis of a single

population of neurons [2], and [13] integrates this approach into a scheme for the group interactions of a number of homogenous populations. Further, [13] is able to analyze perturbations to the average synaptic strength of an ensemble. This study did not, however, analyze perturbations to subsets of an ensemble.

To do this, one must first make a number of approximations for the sake of efficiency and tractability. Mean-field approximations are one such popular method for reducing the computation necessary for simulating a large number of interacting components. This approximation seeks to express the time-evolution of the state of the system in terms of only average values. In the context of a collection of neurons, this mean-field approximation would reduce the enormous number of synaptic connections to just a few mean valuesone self-excitatory weight for each assembly, and a weight expressing input strengths from each ensemble to each other ensemble.

Ensemble stability is an expression of whether or not the neurons of an ensemble have strong excitatory synaptic connections between one another that are maintained in the absence of the stimulus input. This is an important feature for any neural memory structure. However, memory structures also need to be able to adapt and rearrange. Some perturbations to the synaptic architecture, when sufficiently strong, can cause the dissolution of an ensemble[13], allowing for this loss of memory and reformatting.

This study asks: is a mean-field approximation valid for predicting weight dynamics in response to perturbations to the connective architecture of neuronal ensembles? More specifically, if we perturb the mean self-excitatory weight for an ensemble by a fixed amount, across a large number of independent trials, can we cause the system to exhibit different stability dynamics by perturbing a greater or lesser number of neurons? In order to answer this question, we develop a model which combines the computational tractability of [2] and [13] with the resolution and accuracy of brute force methods. We use this method to analyze the impact that perturbations on different numbers of neurons have on the stability of the perturbed ensemble, when we preserve the mean self-excitatory weight of that ensemble.

2.0 MODEL CONSTRUCTION

2.1 SPIKING NETWORK SIMULATIONS

A spiking network with 20 distinct populations was constructed in keeping with [13]. The populations had strong excitatory self-coupling, and inhibitory interconnections. Below, we discuss the basic structure of these neurons and explore the effects of two different types of perturbations that can be applied to a specific population. For more implementation details, please refer to [13].

2.1.1 Exponential Integrate-and-Fire Neurons

The excitatory group was comprised of exponential integrate-and-fire neurons with an adaptation current and an adaptive threshold. The voltage dynamics for the ith neuron was as follows, with all relevant constants in Table 1:

$$\frac{d}{dt}V_i^X(t) = \frac{1}{\tau^X}\left(E_L^X - V_i^X(t) + \Delta_T^X \exp\left(\frac{V_i^X(t) - V_{T,i}^X(t)}{\Delta_T^X}\right)\right) + \frac{g_i^{XE}(t)}{C}(E^E - V_i^X(t)) + \frac{g_i^{XI}(t)}{C}(E^I - V_i^X(t)) - \frac{w_i^X(t)}{C}. \tag{2.1}$$

$\frac{d}{dt}V_i^X(t)$ is written in terms of $V_{T,i}^X(t)$, the threshold for firing an action potential for population X; g_i^{XY}, the conductance from population X to population Y; and $w_i^X(t)$, the adaptation current. $V_{T,i}^E(t)$ increases with each successive spike, and relaxes back to a preset value. Its dynamics are illustrated with the following equation:

$$\frac{d}{dt}V_{T,i}^E(t) = \frac{1}{\tau_T}(V_T - V_{T,i}^E(t)) \tag{2.2}$$

The conductances due to neuronal input amount to weighted spikes passed through a synaptic kernel. They are described by the general formula:

$$g_i^{XY}(t) = F^Y(t) * (J_{ext}^{XY} s_{i,ext}^{XY} + \sum_j J_{ij}^{XY} s_j^Y(t)). \tag{2.3}$$

Here, the synaptic kernel $F^Y(t)$ is:

$$F^Y(t) = \frac{1}{\tau_d^Y - \tau_T^Y}(e^{-t/\tau_d^Y} - e^{-t/\tau_T^Y}). \tag{2.4}$$

Lastly, the adaptation current is as follows:

$$\frac{d}{dt} w_i^E(t) = \frac{1}{\tau_w}(a_w(V_i^E(t) - E_L^E) - w_i^E(t)). \tag{2.5}$$

These dynamics were selected to be in keeping with [5], in order to best match the voltage-based spike timing-dependent plasticity (STDP) rule they describe.

2.1.2 Excitatory STDP

While the clustered neuronal population constructed in [13] must employ both excitatory STDP and inhibitory STDP (iSTDP), inhibitory plasticity is outside of the focus of the present work. For further information about the role of iSTDP in cluster formation, refer to [13]. The excitatory STDP rule for updating the synaptic weight J_{ij}^{EE} from excitatory neuron j to excitatory neuron i is described below:

$$\frac{d}{dt} J_{ij}^{EE}(t) = -A_{LTD} s_j^E(t) R(u_i^E(t) - \theta_{LTD}) + A_{LTP} x_j^E(t) R(V_i^E(t) - \theta_{LTP}) R(v_i^E(t) - \theta_{LTD}). \tag{2.6}$$

Here, x_j^E is a low-pass filtered version of the spike train s_j^E with time constant τ_X, A_{LTD} is the long-term depression (LTD) strength, A_{LTP} is the long-term potentiation strength (LTP), and u_i^E and v_i^E are the membrane voltage V_i^E low-pass filtered with the corresponding time constants τ_u and τ_v. In addition, synaptic normalization was employed in the following form:

$$J_{ij}^{EE}(t) \leftarrow J_{ij}^{EE}(t) - ((\sum_j J_{ij}^{EE}(t) - J_{ij}^E E(0))/N_i^E). \tag{2.7}$$

[5] derived a way to estimate the net weight update strictly based upon firing rate:

$$\langle J_{ij}^{EE} \rangle = \eta A_{LTP} v_j v_i \left(v_i - \frac{A_{LTD}}{\beta A_{LTP}} \right) \tag{2.8}$$

Here, v_j and v_i are the presynaptic and postsynaptic neurons' firing rates, respectively. In addition, β is the time-averaged integral of a single action potential's impact on the postsynaptic neuron's voltage.

3.0 FULL MODEL

Previous work has detailed the formation of distinct neuronal assemblies from a large, homogenous network[13]. Here we assume that such populations have already formed, and we model their behavior in terms of a stochastic Wilson-Cowan system of equations. Each neuron population was assigned a stochastic variable n_i, symbolizing the number of active neurons within the ith population. These n_i were given the following transition rates, with T_i^- as the rate of decrement and T_i^+ as the rate of increment:

$$T_i^- = n_i \tag{3.1}$$

$$T_i^+ = N_i F(\sum_j W_{ij} \frac{n_j}{N_T}) \tag{3.2}$$

where $F(x)$ is given by:

$$F(x) = \frac{1}{1 + \exp -\gamma(x - \theta)} \tag{3.3}$$

and where $N_T = 45$ is the total system size, N_i is the size of the specific population governed by n_i, W_{ij} is the strength of the synaptic connections from population j to i, and γ and θ are free parameters (4 and 0.85 respectively). In a network with two subpopulations, the bidirectional linkages between the two subpopulations are excitatory, while the linkages with the third population are inhibitory. Figure 5 provides a graphical illustration of the connective architecture.

Figure 6 show the fixed points of a prototypical coupled three-population network. Saddle points provide the boundaries between the different steady-states–as [2] shows, a process attempting to transition from one metastable state to another must travel through the

saddle point[2]. As a consequence, calculation of these fixed points is absolutely critical to classify a particular set of activity levels as belonging to a given stability well.

As will soon become evident, these classifications are valuable for reducing the computational complexity of the system. The weight dynamics obey the following equation:

$$\frac{d}{dt}W_{ij} = \eta y_j \cdot y_i(y_i - \theta_w). \tag{3.4}$$

Here, $y_i = n_i/N_i$, η is a learning rate parameter, and $\theta_w = 1.8$. Note the similarity to Eq. 2.8–the two are identical in form. We want to analyze the expected increase in W_{ij} over a long period of time, assuming that y_i equilibrates quickly relative to the evolution of W_{ij}. This allows for the following approximate separation of time scales[13]:

$$\langle \dot{W}_{ij} \rangle = \eta \int P((\mathbf{y}))y_j \cdot y_i(y_i - \theta_w) \approx \eta \sum_{\mathbf{y}^*} P(\mathbf{y}^*)y_j^* \cdot y_i^*(y_i^* - \theta_w) \tag{3.5}$$

In these equations, $\mathbf{y}^*$ is the vector of a stable steady state solution to the governing dynamical system, and $P(\mathbf{y}^*)$ is the corresponding probability that the system is in that state.

$P(\mathbf{y}^*)$ can be calculated by preparing a transition matrix according Eqs. 3.1-3.2, with a diagonal such that the rows sum to zero. From there, by the Perron-Frobenius theorem, the invariant density of the small-scale system may be calculated[2, 20, 13]. The right eigenvector corresponding to a zero eigenvalue must be calculated from the following transition matrix to obtain this invariant density:

$$Q = \begin{bmatrix} -\sum_{q=1}^{S} r_{1q} & r_{12} & \dots & r_{1(q-1)} & r_{1q} \\ r_{21} & \ddots & & & r_{2q} \\ \vdots & & \ddots & & \vdots \\ r_{(S-1)1} & & & \ddots & r_{(S-1)S} \\ r_{S1} & r_{S2} & \dots & r_{S(S-1)} & -\sum_{q=1}^{S} r_{Sq} \end{bmatrix}, \tag{3.6}$$

where r_{pq} is the transition rate from state q to state p, and S is the total number of states.

After classifying each position according to its bounding saddle points, the probability of a system-level state may be obtained by summing over all invariant density probabilities of micro-states belonging to that system-level state. This approach, though less general than

those suggested by other studies[2, 3, 7], has the advantage of simplicity. The fixed points and saddle points may be calculated from the following system:

$$0 = -x_i + \frac{N_i}{N_T} F(\sum_j W_{ij} x_j) \tag{3.7}$$

In this equation, $x_i = \frac{n_i}{N_T}$. Note that this relation implies that $y_i = N_T/N_i * x_i$. The reason for this change will be discussed below. Now, all components necessary for the calculation of the derivative of the weights with respect to time are available. All that remains to calculate the progression of weights through time is to systematically update the system in an Euler scheme using Eq. 3.5, and to recalculate all components at each time step.

Before the response of the system to perturbations can be analyzed, we must discuss some of the generalizations and modifications at play in the above equations. First, within the positive transition rate calculation, all activity levels must be proportional to the size of their corresponding populations. This way, the total activity of a population may be treated as the sum of its sub-populations, as shown below. In previous work [2], however, the fixed points were calculated as n_i/N, where N is the size of a population, homogenous across all populations[13]. In order for activities to sum properly, x_i had to be recast as n_i/N_T, resulting in the changes visible in Equation 3.7. Note that Equation 3.5 is still in terms of the original interpretation of activity (y_i). This is again necessary for consistency, as explained below section.

3.1 GILLESPIE SIMULATION

Gillespie simulations are intended to generate specific instantiations of the time course of a probabilistic system. Here, we have activity levels for each population, and these activity levels each increase or decrease with a predetermined rate. To analyze the efficacy of our theoretical reductions, we generate a large number of trial runs and average their weight evolutions throughout time by interpolating about select time points.

Gillespie simulations operate by first selecting a time τ at which the next transition occurs, and then selecting a specific transition, based on likelihood.

Figure 9 displays common state transitions for the Gillespie simulation constructed according to the aforementioned transition rates. The populations alternate between states of high activity and very low activity. The two partitioned subpopulations transition together, which is intuitively expected–an entire population must be activated or deactivated at the same time to reach a high activity fixed point.

3.1.1 Model Description

For the purposes of Gillespie Simulation, the only necessary state variables are the weights W_{ij} and the current activity level $\mathbf{n}$. This is all that is necessary to generate a single trial run of the system's evolution.

First, we select the time interval until the next event, τ, as follows: Begin by drawing two uniformly distributed random numbers r_1, r_2. Then,

$$\tau = \frac{1}{T_0} \ln \frac{1}{r_1}. \tag{3.8}$$

Here, $T_0 = \sum_{i,+/-} T_i^{+/-}$ is the sum over all possible transitions rates out of the current state. Now we select which reaction will occur. We linearize the transition rate indices, so that they are denoted by T_i for $i \in 1, ..., 2m$, where $2m$ is the total number of possible transitions. Then, we select as follows:

$$i = \text{the smallest integer s.t.} \sum_{j=1}^{i} T_j(\mathbf{n}) > r_2 T_0. \tag{3.9}$$

This process is repeated at every time step. Prior to perturbation, a lengthy transient period is imposed to allow the system to formally settle into its steady-state distribution. At the end of this period, as will be described in Eq. 3.11, the perturbation is applied, and the weight dynamics are activated. The weight update equation is, in keeping with Eq. 3.4, as follows:

$$\Delta W_{ij} = \tau \eta y_j \cdot y_i (y_i - \theta_w). \tag{3.10}$$

3.1.2 Perturbation

Next, we will explore the effect of perturbations to weight strength in each of our models–this is the core of the model, and of the thesis. In the spiking network, the first perturbation affects all within-population excitatory synapses, and the second randomly selects 1/3 of these, and has three times the strength. From the perspective of a mean-field model, the average self-coupling strength of the population is preserved between the two cases, but there are striking differences in the stability dynamics between the two. Figure 3 illustrates these effects. It would appear that, as intuition would suggest, a weaker but more diffuse perturbation is less likely to dissolve a population when compared to a stronger but more focused synaptic loss.

It will be useful to establish a formal expression for our perturbation that illustrates the conservation in play. In what follows, A will refer to the subpopulation that is not perturbed, and B will refer to the subpopulation that is. The conservation is as follows:

$$W_p \cdot N_p = K. \tag{3.11}$$

Here, W_p stands for the strength of the perturbation to weights W_{AB} and W_{BB}, and N_p stands for size parameter of the perturbed population, N_B. By keeping this factor constant, the effective "work" performed on the system is constant. The system was tested along "K-curves", where K was fixed and W_p and N_p were varied. This allowed for easy comparison of perturbations with fixed net synaptic loss across varying strengths and sizes. This concludes the discussion of the basic model. Now we explore simplification methods that give analytic power and increased efficiency.

3.2 HALF-REDUCED DIMENSIONALITY SYSTEM

We want to increase the efficiency and analytical workability of our model. To do this, we capitalize on a previous framework[2] to provide macroscopic transition rates for external populations, and we use the microscopic Perron-Frobenius framework to obtain the probability distribution within the two coupled subpopulations.

3.2.1 Full Reduced Dimensionality System

The first order of business is to obtain a probability distribution for the macroscopic activity levels of the unperturbed states. We want to create a single population that, for the purposes of interactions with other populations, behave exactly the same as the two coupled subpopulations would. We want to reduce the dimensionality of our system so that it is no longer necessary to calculate the Perron-Frobenius eigenvector of an intractably large matrix. In what follows, we express the two subpopulations, A and B (unperturbed and perturbed, respectively) as an ostensibly equivalent total population, which allows us to reduce to a 3-dimensional continuous time Markov chain[2, 13].

Our desire is to find a W_{11} s.t. $(N_A + N_B)F(W_{11}x_1) = N_A F(W_{AA}x_A + W_{AB}x_B) + N_B F(W_{BB}x_B + W_{BA}x_A)$, where $x_1 = x_A + x_B$. The easiest way for this to work is if $W_{11}x_1 = W_{AA}x_A + W_{AB}x_B = W_{BB}x_B + W_{BA}x_A$. Such a W_{11} does not exist under normal circumstances, however, it will work under the constraints of our system. The relations that make our reduction possible follow.

x_A and x_B always occupy the same stable state, therefore:

$$x_1 N_A/(N_A + N_B) = x_A \tag{3.12}$$

and also,

$$x_1 N_B/(N_A + N_B) = x_B \tag{3.13}$$

Our perturbation targets W_{AB} and W_{BB}, therefore: $W_{AB} = W_{BB}$, and also $W_{AA} = W_{BA}$. This is true throughout time, due to the weight update rule (in the continuous case), because $y_A = y_B$. The following is a consequence of the equality in the weight matrix:

$$\begin{bmatrix} W_{AA} \\ W_{BA} \end{bmatrix} = \alpha_t \begin{bmatrix} W_{AB} \\ W_{BB} \end{bmatrix} \tag{3.14}$$

Here, α_t is a constant at any given time step, because W_{AB} and W_{BB} may update at a different rate from W_{AA} and W_{BA}, but that update will always be equivalent within each pair. This constraint becomes weaker for finite N_p, but W_{12} and W_{22} still track relatively closely.

Combining all of these simple relationships, we reformulate our goal:

$$W_{11}x_1 = W_{BB}(\frac{N_B}{N_A+N_B})x_1 + \alpha_t W_{BB}(\frac{N_A}{N_A+N_B})x_1 \tag{3.15}$$

$$W_{11}x_1 = \alpha_t W_{AB}(\frac{N_A}{N_A+N_B})x_1 + W_{AB}(\frac{N_B}{N_A+N_B})x_1 \tag{3.16}$$

These equations are identical, because $W_{AB} = W_{BB}$. Therefore, we need only keep the first equation. Canceling x_1 and using the definition of α_t from Eq. 3.14, we get:

$$W_{11} = W_{AA}(\frac{N_A}{N_A+N_B}) + W_{AB}(\frac{N_B}{N_A+N_B}) \tag{3.17}$$

Recall that W_{AB} has been perturbed away from W_{AA} by a value $W_p = K/N_B$, so $W_{AB} = W_{AA} - K/N_B$. Plugging this into our previous equation, we get:

$$W_{11} = W_{AA}(\frac{N_A+N_B}{N_A+N_B}) - W_p(\frac{N_B}{N_A+N_B}) = W_{AA} - \frac{K}{N_A+N_B} \tag{3.18}$$

With this equation, we have effectively reduced the system to 3-dimensions with constant N across populations. This system has already been studied[13], and we may capitalize upon previous work to obtain our steady-state probability distribution for any set of weights. Rates r_i^+ and r_i^-, for transitions out of the upper and lower stability wells for the ith population, respectively, can be calculated with the following equation [2, 13]

$$r_i^{+/-} = \frac{\Omega_i^+(\mathbf{x}_-^*)}{2\pi}\sqrt{|S_i''(\mathbf{x}_0^*)|S_i''(\mathbf{x}_{+/-}^*)}e^{-N_T[S_i(\mathbf{x}_0^*)-S_i(\mathbf{x}_{+/-}^*)]}, \tag{3.19}$$

where x_0 is a saddle point, x_+ is the upper stable point, and x_- is the lower stable point. Below, we define the subcomponents of this equation:

$$S_i(\mathbf{x}) = \int_0^{x_i} \ln\frac{\Omega_i^-(\mathbf{y})}{\Omega_i^+(\mathbf{y})}dy_i, \tag{3.20}$$

$$\Omega_i^-(\mathbf{x}) = x_i, \text{ and} \tag{3.21}$$

$$\Omega_i^+(\mathbf{x}) = \frac{N_i}{N_T}F(\sum_j W_{ij}x_j). \tag{3.22}$$

These allow us to calculate a macroscopic reduced invariant density without having to worry about the co-activation of populations A and B.

One problem remains. It should be clear that since $N_A + N_B$ is a constant for all N_B, varying N_B has no effect on W_{11} for constant K. The fact that it is initially invariant to changes in N_B does not directly imply that it is invariant throughout time, but below we will show that this is in fact the case.

We cannot immediately say that $\frac{dW_{11}}{dt} = \eta y_1 \cdot y_1(y_1 - \theta)$ where $y_1 = \frac{N_T}{N_1}x_1$, in keeping with Eq. 3.4, because W_{11} is just a dummy variable–it's evolution is strictly determined by W_{AA} and W_{AB}. Using Eq. 3.17, we arrive at:

$$\frac{d}{dt}W_{11} = \frac{dW_{AA}}{dt}(\frac{N_A}{N_A + N_B}) + \frac{dW_{AB}}{dt}(\frac{N_B}{N_A + N_B}) \tag{3.23}$$

Plugging in the definitions for $\frac{dW_{AA}}{dt}$ and $\frac{dW_{AB}}{dt}$, we get:

$$\begin{aligned}\frac{d}{dt}W_{11} = \eta\frac{N_T^2}{N_A^2}x_A^2(x_A\frac{N_T}{N_A} - \theta)\frac{N_A}{N_A + N_B} + \eta\frac{N_T^2}{N_A N_B}x_A x_B(x_A\frac{N_T}{N_A} - \theta)\frac{N_B}{N_A + N_B} \\ = \frac{\eta\frac{N_T^2}{N_A}(x_A\frac{N_T}{N_A} - \theta)[x_A^2 + x_A x_B]}{N_A + N_B}\end{aligned} \tag{3.24}$$

Now use $\frac{x_A}{N_A} = \frac{x_B}{N_B}$ and take $c = N_A + N_B$ (note that $1 + \frac{c - N_A}{N_A} = \frac{c}{N_A}$) to reduce to:

$$\frac{dW_{11}}{dt} = \eta(\frac{N_T}{N_A})^2 x_A^2(x_A\frac{N_T}{N_A} - \theta) \tag{3.25}$$

We want this in terms of x_1. Using $y_1 = \frac{N_T}{c}x_1 = \frac{N_T}{N_A}x_A$ we arrive at our final result:

$$\frac{dW_{11}}{dt} = \eta(\frac{N_T}{c})^2 x_1^2(\frac{N_T}{c}x_1 - \theta) = \eta y_1^2(y_1 - \theta) \tag{3.26}$$

All of the variables in this expression are independent of changes in N_A and N_B, because $N_A + N_B$ is a constant. Clearly we have a problem, because the simulations show a marked increase in weight updates between diffuse perturbations and focused perturbations, as seen in Fig. 8. Strictly employing this fully reduced model is not sufficient to capture the requisite dynamics. Instead, it will be necessary to use a method that is capable of capturing the microscopic dynamics of the subpopulations.

3.2.2 Model Description

Now, despite its deficiencies, the aforementioned fully reduced model *is* capable of generating a valid probability distribution, given a set of weights. We can condition our expected weight update equation (3.5) on the probability of existing in a given macroscopic state (stability well):

$$\langle \dot{W}_{ij} \rangle = \sum_{\mathbf{y}^*} P_s(\mathbf{n}^*) \sum_{n_A=1}^{\infty} \sum_{n_B=1}^{\infty} P_s(n_A, n_B|\mathbf{n}^*) y_j \cdot y_i(y_i - \theta_w). \tag{3.27}$$

$$y_{i/j} = \begin{cases} \frac{n^*_{i/j}}{N_{i/j}} & i/j \notin \{A, B\} \\ \frac{n_{A/B}}{N_{A/B}} & i/j \in \{A, B\} \end{cases} \tag{3.28}$$

Now, for $i/j \notin \{A, B\}$, we have simply fixed $x_{i/j}$ at its average within a stability well. The values are not dependent on n_A or n_B, so we may extract the y_i and y_j from the probability sums, and we are left with Eq. 3.5. If only one of the values is dependent on n_A or n_B, the double sum over the joint probability distribution reduces to a single sum over a marginal probability distribution. For the mutually excitatory subpopulations, we can find the joint probability distribution by using the Perron-Frobenius theorem with the following transition rates:

For a given n_A, n_B, $\mathbf{x}^*$, and $i \in \{A, B\}$,

$$T_i^- = n_i \tag{3.29}$$

and,

$$T_i^+(n_A, n_B|\mathbf{x}^*) = N_i F\Big(\sum_{j \in \{A,B\}} W_{ij} \frac{n_j}{N_T} + \sum_{j=2:m} W_{ij} x_j^* \Big). \tag{3.30}$$

Here, m is the total number of populations, and we assume that subpopulations A and B are the mutally excitatory subpopulations.

This reduces our system an incredible amount, making the Perron-Frobenius eigenvalue calculation tractable, and allowing for visualization of the resultant $P_s(n_i|\mathbf{n}^*)$ probability distributions in two dimensions (see Fig. 14).

4.0 MODEL ANALYSIS

Having fully examined the construction of our spiking network, Gillespie simulations, and reduced-dimensionality model, it is now time to turn towards properties of these systems and their responses to perturbations.

4.1 CONSISTENCY CRITERION

One fundamental criterion for our model, which is designed to allow for perturbations applied to a subset of the neurons in a given population, is that the activities of subpopulations must sum to the activity of the total population under certain weight criteria. Internal coupling weights may be perturbed from this position, and the model may deviate from the homogenous case, but if the trivial situation where the activities of two unperturbed subpopulations do not match the activity of the total population, the system is meaningless.

We wonder, specifically, if the following equality holds:

$$\frac{dx_1}{dt} = \frac{dx_A}{dt} + \frac{dx_B}{dt}. \tag{4.1}$$

Here, $\frac{dx_1}{dt}$ is the time derivative of the activity of the entire population, and $\frac{dx_A}{dt} + \frac{dx_B}{dt}$ is the addition of the activity derivatives of the two sub-populations. We want to know if there is a set of weights such that the sum of the sub-populations behaves equivalently to the entire population. In full detail, we have:

$$\frac{dx_A}{dt} + \frac{dx_B}{dt} = -x_A - x_B + \frac{N_A}{N_T}F(W_{AA}x_A + W_{AB}u_B + C) + \frac{N_B}{N_T}F(W_{BB}x_B + W_{BA}x_A + C), \tag{4.2}$$

where C is external input. Take $W_{AA} = W_{BB} = W_{AB} = W_{BA}$, which we refer to as W_{11}. Also take $N_A + N_B = N_1$ and $W_c = W_{11}$, where W_{11} is the weight associated with x_1. Simplifying Eq. 4.2, we arrive at:

$$\frac{dx_A}{dt} + \frac{dx_B}{dt} = -x_A - x_B + \frac{N_1}{N_T} F(W_{11}(x_A + x_B) + C) \tag{4.3}$$

This directly implies Eq. 4.1 so long as the weight updates are equivalent for each subpopulation. Now here, $x_A = \frac{n_A}{N_T}$. Obviously, x_A does not equal x_B. However, for our system to work, the weights must be updated equivalently. Thus, weight updates occur according to local population activity, rather than global population activity: $y_A = \frac{N_T}{N_A} x_A = \frac{n_A}{N_A}$.

Recall that the weight dynamics obey Eq. 3.4, so if $y_A = y_B$, then the weight updates will be equivalent between the two populations. So, does $y_A = y_B$? We know, by the equality of W_{AA} and W_{BB}:

$$\frac{dy_A}{dt} = \frac{N_T}{N_A}\frac{dx_A}{dt} = -y_A + F(W_{AA}x_A + W_{AB}x_B + C) = \frac{dy_B}{dt} \tag{4.4}$$

Thus, for identical initial conditions, $y_A(0) = y_B(0) = y_v(0)$, we see that $W_{11} = W_{12} = W_{22} = W_{21} = W_v$ throughout time. This gives us sufficient conditions such that the sum of sub-populations remains equivalent to the activity of the entire population.

This condition holds true in the case of continuous dynamics, but it is by no means necessarily correct for the discrete case, where y_A and y_B are restricted to meshes of differing resolution. Thus, for our purposes, this condition amounts to an approximation, rather than a proof.

4.2 COMPARISON TO GILLESPIE SIMULATION

Having obtained reasonably promising results in our comparison to the Perron-Frobenius method, the next step is to compare the theory with the exact stochastic simulations obtained by the Gillespie algorithm. Figures 12-13 display the average weight evolution for the two different methods, for a strong and a weak perturbation, and for $N_p = 12$ and $N_p = 3$. The error throughout time for these methods is consistently low. These comparisons are

intentionally far from either method's stability boundary, because slight deviations away from the stability boundary can have striking impacts on the time scale of recovery. Thus slightly different locations of the stability boundary can heavily confound the pointwise percent error between the two methods when the perturbation lands the system in this region. sectionStability Differences along a K-Curve To see the impact of such perturbations on the Gillespie simulations, Figure 8 shows the various differences between the weight updates for $N_p = 12$ compared to $N_p = 3$ for a perturbation where $K = 7.75$. We see that 46/60 of the simulations return to stability in the $N_p = 12$ case, whereas only 13/60 return to stability in the $N_p = 13$ case. We conclude that there is a statistically significant difference between the proportion of returned simulations in the $N_p = 12$ case, compared to the $N_p = 3$ case for $K = 7.75$ ($p < 0.00001$). This directly parallels spiking network simulations, where a 42% perturbation to 1/3 of the self-coupled excitatory weights within a population led to 25/120 simulation runs returning to stability, while a 14% perturbation to all self-coupled excitatory weights led to 81/120 simulations returning to stability. Again, we say that there is a statistically significant difference between these proportions ($p < 0.00001$).

Throughout time, Figure 8 reveals that the mean activity levels of stable $N_p = 12$ runs will return to pre-perturbation levels on a much faster time scale than stable $N_p = 3$ runs, and that unstable runs maintain a very low activity level. To understand why, it is constructive to look at Figure 14. In Figure 14, we see that the discretization of the $N_p = 3$ values leads to higher activity levels. In Figure 10, we see that higher activity levels that remain subthreshold will produce more negative weight updates. This simple depiction provides an intuitive reasoning for how the difference in stability may occur. Quantized size effects produced by forcing discrete activity levels, due to different population sizes, can have a direct impact on the activity distribution and subsequent weight update.

The goal now becomes to express this phenomenon in a more theoretically workable environment. The Perron-Frobenius method was too intractable–a dimensionality reduction is necessary to completely express the data.

Now that we have established that the reduced-dimensionality theory agrees reasonably well with the Gillespie simulations away from the stability point, it is necessary to look at agreement at the stability point itself. While the Gillespie simulations show a difference in

stability between $N_p = 12$ and $N_p = 3$ and $K \approx 7.75$, we see that same difference in the reduced theory at $K = 6.9$, giving a percent relative error of 11%. Figure 11 shows the stability difference in the reduced theory case, while Figure 8 shows the difference in the Gillespie simulations.

4.3 QUANTIZED SIZE EFFECTS

Lastly, we can more precisely view the impact of modifying size quantization on weight update by viewing directly the weight updates caused by specific activity levels. Figure 14 shows the differential effect on the weight update caused by a $K = 6.9$ perturbation. While the $N_p = 12$ case is able to have relatively strong positive weight updates relative to its negative updates, the vast majority of the weight update strength is negative and caused by low activity positions in the $N_p = 3$ case. This effect is likely due to quantization, because we have already shown that in the continuous case, weight updates are identical. The differences in weight update, must be due to relative differences in discretization, ie. the number of states that the population can hold with reasonable probability.

5.0 DISCUSSION

In the preceding work, we found that perturbations preserving the mean self-excitatory weight within a neuronal ensemble could lead to different weight updates if the number of targeted neurons was varied. More distributed perturbations with a weaker perturbation at each synapse were found to lead to a faster return to the stable fixed point in both spiking network and Gillespie simulations, and there were K values at which more focused perturbations would lead to a degradation of synaptic connectivity within the ensemble, but the distributed perturbations would lead to stable dynamics.

Further, we constructed a reduced-dimensionality model that leveraged both the simplified model in [13] and [2], and a brute force method employing the Perron-Frobenius theorem to calculate the average weight update for a stochastic Wilson-Cowan model of our system, allowing us to examine how quantized size effects within population activity wells (inactive or active) impacted the weight dynamics. We found that the reduced-dimensionality model matched the mean dynamics of the Gillespie simulations well, and that it exhibited all qualitative effects observed in the spiking network simulations, ie. there were K values at which distributed perturbations would lead to stability while focused perturbations would lead to a total loss of connective weight, and distributed perturbations always returned to stability faster.

We found in both the Gillespie simulations and the reduced-dimensionality model evidence that decreasing the number of neurons perturbed (by manipulating the N_p parameter) caused a significant increase in the negative weight update brought about by existing in the low activity state, caused by a greater spread of activity states. In a continuous model (ie. with populations of infinite size), this effect would disappear, as we have mathematically demonstrated.

Thus, though this effect is likely to disappear with extremely large numbers of neurons, it is still observable in spiking networks of the size we worked with, ie. on average 200 excitatory neurons per population. Conditioned on this prediction, then, we may call into question the validity of mean-field models specifically with regard to ensemble stability. These results do *not* indicate that a fault has yet been found in these models ability to predict the activity dynamics of a network of neuronal ensembles, given that they have already been formed and that their connective architecture remains within the stable regime. It is when within-ensemble heterogeneous perturbations are introduced that we encounter the insufficiencies of the mean-field paradigm. As of now, our newly developed reduced-dimensionality adaptation of existing models is capable of capturing the observed dynamics. Unlike the simplified homogenous systems existing in [13] and [2] that are capable of greatly reducing the computational workload of simulation, our model is, despite its gains in efficiency over a brute-force method, still quite computationally intensive. Thus, while our model is quite able to answer the questions regarding perturbations applied to subsets of neuronal ensembles addressed in this paper, its capability as a generalized model of neuronal dynamics is questionable. Future iterations attempting to capture these dynamics will likely have to totally obviate the need for any brute-force usage of the Perron-Frobenius theorem, as the calculation scales cubically with system size. Regardless, we can still conclude that the stability of neuronal ensembles in response to perturbations to synaptic weights is dependent on the number of synapses perturbed, even when the mean weight perturbed is kept constant.

5.1 ILLUSTRATIONS AND TABLES

Table 1 | Parameters for Spiking Network

Symbol	Description	Value
V_i^X	for neuron type X, cell i's membrane voltage	variable
τ^E	E neuron resting membrane time constant	20 ms
τ^I	I neuron resting membrane time constant	20 ms
E_L^E	E neuron resting potential	-70 mV
E_L^I	I neuron resting potential	-62 mV
Δ_T^E	E neuron EIF slope factor	2 mV
$V_{T,i}^X$	neuron type X voltage threshold for spike	variable
g_i^{XY}	conductance for neuron i from type Y to type X	variable
C	capacitance	300 pF
E^E	E reversal potential	0 mV
E^I	I reversal potential	-75 mV
w_i^X	adaptation current for neuron i	variable
τ_T	adaptive threshold time scale	30 ms
V_T	threshold potential	-52 mV
F^Y	synaptic kernel	variable
J_{ext}^{XY}	external synaptic weight from type Y to type X	variable
$s_{i,ext}^{XY}$	external spike train from type Y to type X	variable
J_{ij}^{XY}	synaptic weight from neuron i to j, from type X to Y	variable[13]
s_j^Y	spike train for neuron j of type Y	variable
A_{LTD}	Long-term depression (LTD) strength	0.0008 pA mV^{-1}
A_{LTP}	Long-term potentiation (LTP) strength	0.0014 pA mV^{-1}
θ_{LTD}	Threshold to recruit LTD	-70 mV
η	learning rate for weight dynamics	0.0001

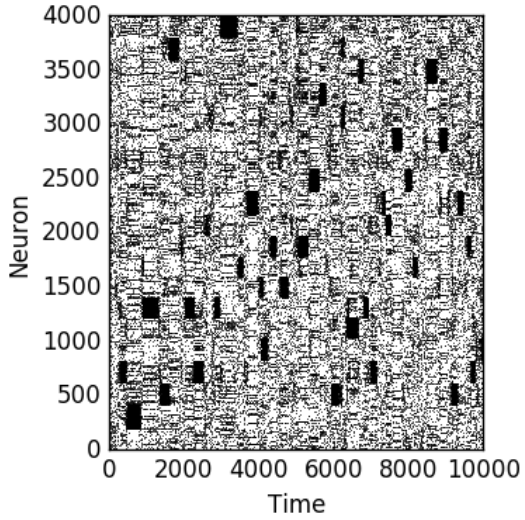

Figure 1: Spiking network raster of activity levels immediately after a perturbation to all neurons in one population. The population (bottom) shows no high activity states, while the other populations alternate stochastically between states of high and low activity. Reproduced from [13].

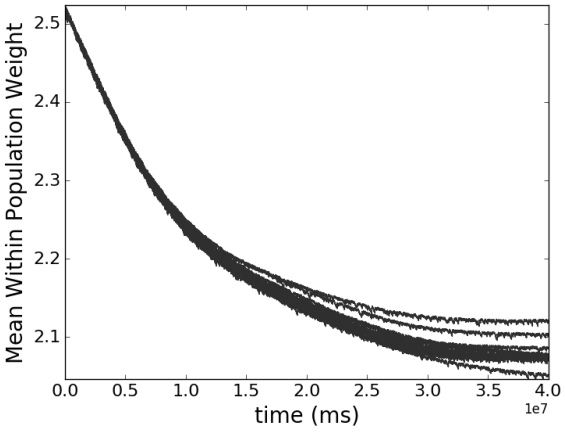

Figure 2: Control dynamics for the spiking network system. No perturbations or stimuli have been applied, and the mean within population weight is allowed to stabilize throughout the transient period. All subsequent figures will show the perturbation applied after this period of time, where the weight dynamics are stable in the absence of a perturbation.

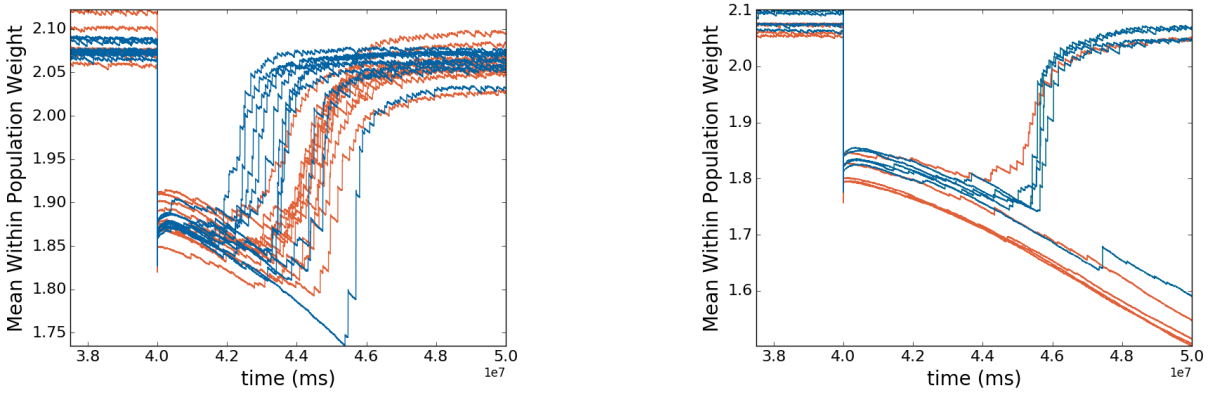

Figure 3: a) Example of a difference in the rate of recovery in response to a weak perturbation to synaptic strengths. Perturbations to all self-excitatory synapses in a population of magnitude 11.5% (blue) show faster recovery times when compared to perturbations to 1/3 of all self-excitatory synapses in a population with magnitude 34.5% (red). b) Demonstration of different impacts engendered by reducing self-coupling synaptic weights from a single neural assembly[13] by 14% (blue) and 42% (red), across thirty realizations. In blue, all weights were reduced, in red, only 1/3 of the weights were reduced. The constraint in Equation 3.11 is satisfied, yet the more diffuse perturbation exhibits gains in synaptic strength almost to the point that it matches the control, while the focused perturbation shows no such recovery.

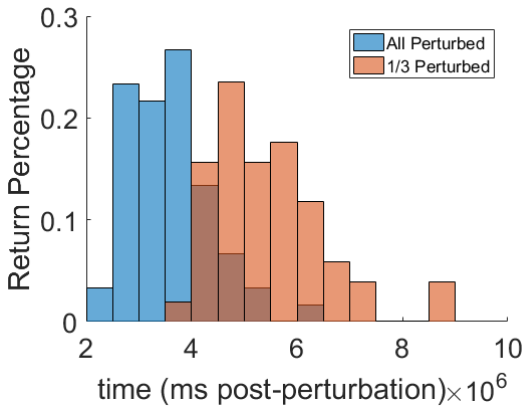

Figure 4: Histogram of the times at which 60 different simulation trials returned back to the fixed point in two cases: in blue, all self-excitatory synapses in the population were perturbed by 11.5%, and in red, 1/3 of the self-excitatory synapses in the population were perturbed by 34.5%. The histograms for the two cases were found to produce statistically significant differences in return time (Two-sample t-test, $p < 0.001$).

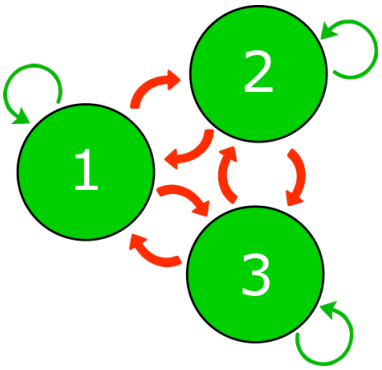

Figure 5: This depicts the weight structure. All self-coupling weights, and weights between the two subpopulations are excitatory, while weights between the two main populations are inhibitory.

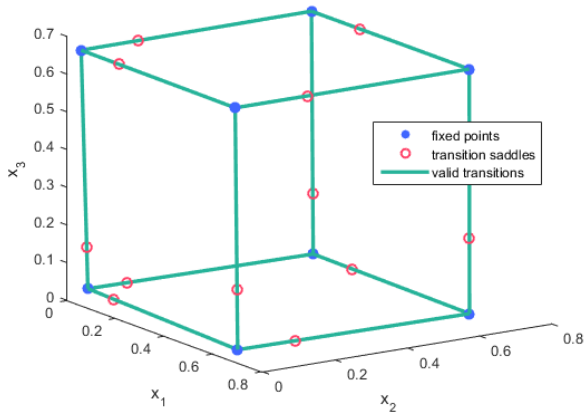

Figure 6: Graphic of the possible transitions within the reduced system. Two populations may never simultaneously co-activate here, though the two sub-populations must always activate and deactivate together. The stable points lie on a lattice containing each possible permutation of active and inactive states for each population.

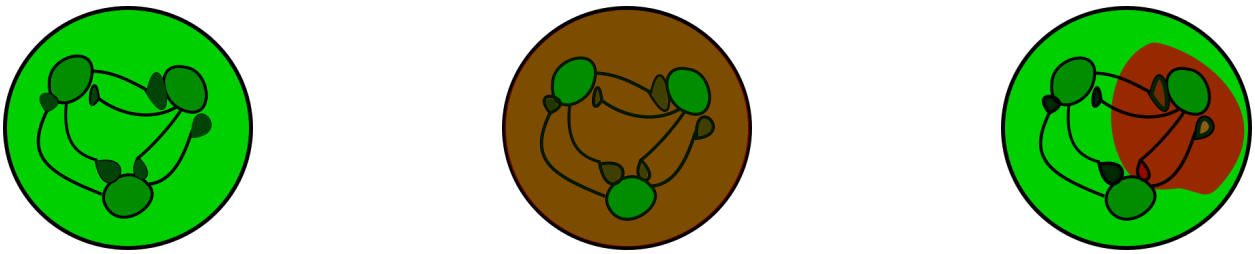

Figure 7: a) Depiction of an unperturbed network of neurons. b) The first type of perturbation, targeting self-excitatory weights within a cluster of neurons. c) The second type of perturbation, selective for a smaller portion of the self-excitatory synapses within a cluster of neurons. Here brown signifies a weak perturbation, and red signifies a stronger perturbation.

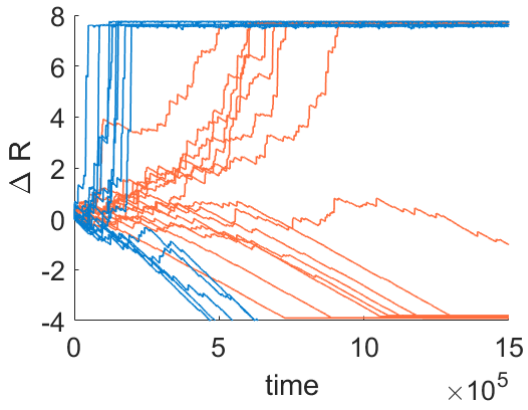

Figure 8: Weight dynamics for $N_p = 12$ (blue) and $N_p = 3$ (red). 46 simulations returned to their starting point for $N_p = 12$, while 13 simulations returned for $N_p = 3$. Here, $\Delta R = N_p * (W_{22}(t) - W_{22}(0))$.

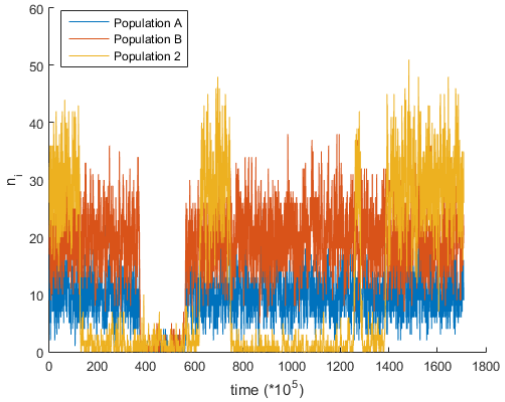

Figure 9: Activity dynamics for an $N_p = 10$, $K = 0$ partition of the subpopulations. The two subpopulations have smaller n_i than the external population, but add to be equal. Note that the two subpopulations transition in concert.

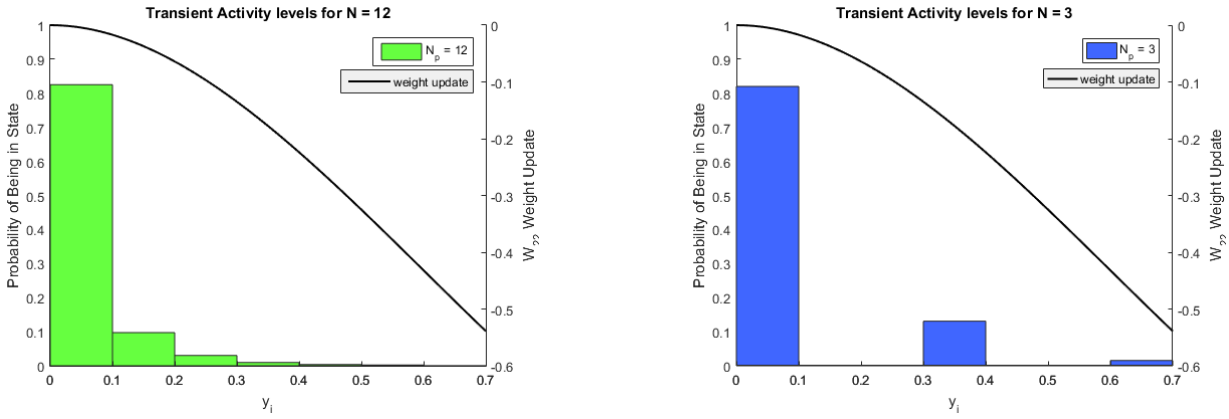

Figure 10: Frequency histogram of the average time spent at a given activity level immediately after a $K = 7.75$ perturbation in the Gillespie simulation. a) $N_p = 12$, and activity levels are confined to a tight space near zero. b) $N_p = 3$, and activity levels are considerably more spread, because there are fewer occupiable states.

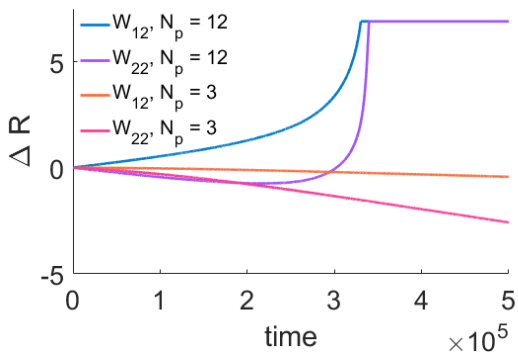

Figure 11: Response of the Half-reduced model to a $K = 6.9$ perturbation. The $N_p = 12$ (green) case exhibits a return to stability, while the $N_p = 3$ case shows continued weight loss.

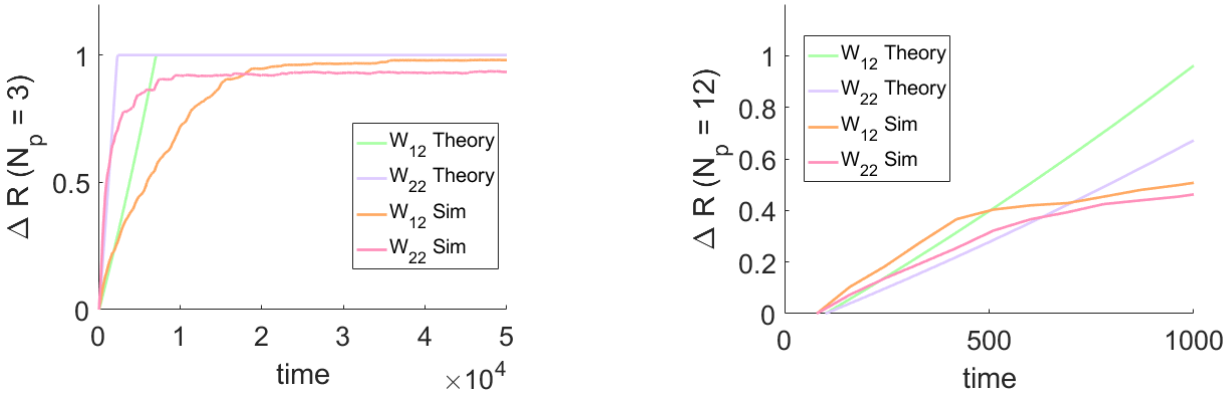

Figure 12: Comparison between an average of 40 Gillespie simulations to the Half-reduced model for $K = 1$. The early divergence of the theory from the simulation is due to the fact that faster-returning simulations hit the W_{max} boundary and ceased to provide their positive bias. a) $N_p = 12$. b) $N_p = 3$.

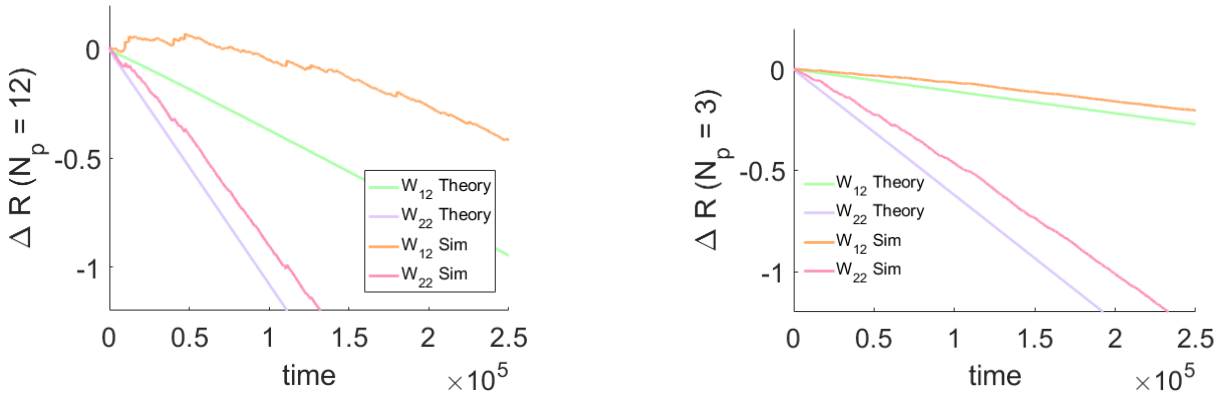

Figure 13: Comparison between an average of 40 Gillespie simulations to the Half-reduced model for $K = 10$. a) $N_p = 12$. b) $N_p = 3$.

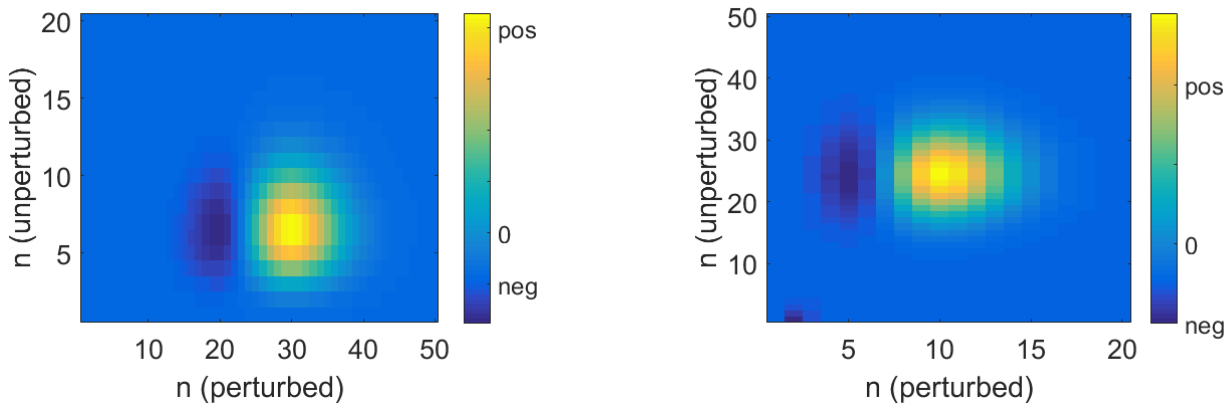

Figure 14: Illustration of quantized size effects in the Half-reduced model. Shown is the total weight update stratified by the contribution caused by each combination of activity levels between the two sub-populations. The contributions are weighted by the probability of the given activity level existing, summed across possible external population levels. The update values were scaled to all be greater than one, and were subsequently plotted on a log-scale for visibility purposes. a) The results for a $K = 6.9$, $N_p = 12$ perturbation. b) The results for a $K = 6.9$ $N_p = 3$ perturbation.

BIBLIOGRAPHY

[1] Berkes, P., et al. (2011). "Spontaneous Cortical Activity Reveals Hallmarks of an Optimal Internal Model of the Environment." Science 331: 83-87.

[2] Bressloff, P. C. (2010). "Metastable states and quasicycles in a stochastic Wilson-Cowan model of neuronal population dynamics." Physical Review E 82(5): 051903.

[3] Bressloff, P. C. and J. M. Newby (2013). "Metastability in a stochastic neural network modeled as a velocity jump Markov process." SIAM Journal on Applied Dynamical Systems 12(3): 1394-1435.

[4] Churchland, M. M., et al. (2010). "Stimulus onset quenches neural variability: a widespread cortical phenomenon." Nat Neurosci 13(3): 369-378.

[5] Clopath, C., et al. (2010). "Connectivity reflects coding: a model of voltage-based STDP with homeostasis." Nature neuroscience 13(3): 344-352.

[6] Deco, G. and E. Hugues (2012). "Neural Network Mechanisms Underlying Stimulus Driven Variability Reduction." PLOS Computational Biology 8(3): e1002395.

[7] Dellago, C., et al. (1998). "Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements." The Journal of chemical physics 108(22): 9236-9245.

[8] El Boustani, S., et al. (2012). "Stable Learning in Stochastic Network States." The Journal of Neuroscience 32: 194-214.

[9] Erban, R., et al. (2007). "A practical guide to stochastic simulations of reaction-diffusion processes." arXiv preprint arXiv:0704.1908.

[10] Gillespie, D. T. (2007). "Stochastic simulation of chemical kinetics." Annu. Rev. Phys. Chem. 58: 35-55.

[11] Izhikevich, E. M., et al. (2004). "Spike-timing Dynamics of Neuronal Groups." Cerebral Cortex 14(8): 933-944.

[12] Klampfl, S. and W. Maass (2013). "Emergence of Dynamic Memory Traces in Cortical Microcircuit Models through STDP." The Journal of Neuroscience 33: 11515-11529.

[13] Litwin-Kumar, A. and B. Doiron (2014). "Formation and maintenance of neuronal assemblies through synaptic plasticity." Nature Communications 5.

[14] Luczak, A., et al. "Spontaneous Events Outline the Realm of Possible Sensory Responses in Neocortical Populations." Neuron 62(3): 413-425.

[15] Montijn, J. S., Olcese, U., & Pennartz, C. M. A. (2016). Visual Stimulus Detection Correlates with the Consistency of Temporal Sequences within Stereotyped Events of V1 Neuronal Population Activity. The Journal of Neuroscience, 36(33), 8624.

[16] Ringach, D. L., et al. (2016). "Spatial clustering of tuning in mouse primary visual cortex." Nature Communications 7: 12270.

[17] Rothschild, G., et al. (2010). "Functional organization and population dynamics in the mouse primary auditory cortex." Nat Neurosci 13(3): 353-360.

[18] Royer, S. and D. Pare (2003). "Conservation of total synaptic weight through balanced synaptic depression and potentiation." Nature 422(6931): 518-522.

[19] Sutherland, G. R. and B. McNaughton (2000). "Memory trace reactivation in hippocampal and neocortical neuronal ensembles." Current Opinion in Neurobiology 10(2): 180-186.

[20] Van Kampen, N. G. (1992). Stochastic processes in physics and chemistry, Elsevier.

[21] Vogels, T., et al. (2011). "Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks." Science 334(6062): 1569-1573.

[22] Wilson, H. R. and J. D. Cowan (1972). "Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons." Biophysical Journal 12(1): 1-24.

[23] Xu, S., et al. (2012). "Activity recall in a visual cortical ensemble." Nat Neurosci 15(3): 449-455.