EFFECTS OF SPIKE-DRIVEN FEEDBACK ON NEURAL GAIN AND PAIRWISE CORRELATION

by

John D. Bartels

B.S. Computer Science, Rensselaer Polytechnic Institute, 1996

Submitted to the Graduate Faculty of

the Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Sciences

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This thesis was presented

by

John D. Bartels

It was defended on

March 30 2010

and approved by

Brent Doiron, Ph.D., Assistant Professor, Department of Mathematics

G. Bard Ermentrout, Ph.D., University Professor, Department of Mathematics

William C. Troy, Ph.D., Professor, Department of Mathematics

Thesis Advisor: Brent Doiron, Ph.D., Assistant Professor, Department of Mathematics

EFFECTS OF SPIKE-DRIVEN FEEDBACK ON NEURAL GAIN AND PAIRWISE CORRELATION

John D. Bartels, M.S.

University of Pittsburgh, 2010

Both single neuron and neural population spiking statistics, such as firing rate or temporal patterning, are critical aspects of many neural codes. Tremendous experimental and theoretical effort has been devoted to understanding how nonlinear membrane dynamics and ambient synaptic activity determine the gain of single neuron firing rate responses. Furthermore, there is increasing experimental evidence that the same manipulations that affect firing rate gain also modulate the pairwise correlation between neurons. However, there is little understanding of the mechanistic links between rate and correlation modulation. In this thesis, we explore how spike-driven intrinsic feedback co-modulates firing rate gain and spike train correlation. Throughout our study, we focus on excitable LIF neurons subject to Gaussian white noise fluctuations. We first review prior work which develops linear response theory for studying spectral properties of LIF neurons. This theory is used to capture the influence of weak spike driven feedback in single neuron responses. We introduce a concept of "dynamic spike count gain" and study how this property is affected by intrinsic feedback, comparing theoretical results to simulations of stochastic ODE models. We then expand our scope to a pair of such neurons receiving weakly correlated noisy inputs. Extending previous work, we study the correlation between the spike trains of these neurons, comparing theoretical and simulation results. We observe that firing rate gain modulation from feedback is largely time-scale invariant, while correlation modulation exhibits marked temporal dependence. To discern whether these effects can be solely attributed to firing rate changes, we perform a perturbative analysis to derive conditions for correlation modulation over small

time scales beyond that expected from rate modulation. We find that correlation is not purely a function of firing rate change; rather it is also influenced by sufficiently fast feedback inputs. These results offer a glimpse into the connections between gain and correlation, indicating that attempts to manipulate either property via firing rates will affect both, and that achievability of modulation targets is constrained by the time scale of spike feedback.

TABLE OF CONTENTS

1.0 INTRODUCTION 1

2.0 THE SINGLE SELF-COUPLED NEURON 5

2.1 DERIVATION OF THE MODEL 5

2.2 SPIKE TRAIN STATISTICS 8

2.2.1 FIRING RATE ESTIMATION 9

2.2.2 LINEAR RESPONSE THEORY 11

2.2.3 DYNAMIC GAIN 12

2.2.3.1 COMPUTATIONAL APPROACH 13

2.2.3.2 THEORETICAL APPROACH 14

3.0 NETWORKS OF PAIRED SELF-COUPLED NEURONS 17

3.1 DERIVATION OF THE MODEL 17

3.2 SPIKE TRAIN STATISTICS 18

3.2.1 FIRING RATE ESTIMATION 18

3.2.2 SPIKE TRAIN CORRELATIONS AND LINEAR RESPONSE THEORY 19

3.3 CORRELATION SUSCEPTIBILITY 21

3.4 ANALYSIS OF RESULTS 24

4.0 CONCLUSIONS 37

APPENDIX A. DERIVATION OF SPECTRA AND TRANSFER FUNCTION 41

A.0.1 POPULATION DENSITY AND FOKKER-PLANCK TREATMENT 41

A.0.2 STATIONARY FIRING RATE 43

A.0.3 POWER SPECTRA 44
A.0.4 TRANSFER FUNCTION 45
APPENDIX B. ANALYSIS OF FIRING RATE MATCHING WITH θ . . 47
BIBLIOGRAPHY 51

LIST OF FIGURES

1 **Effects of ADP's and AHP's.** (A) Chemical inhibition of AHP reveals hyperpolarization present in control. (B) Gain (slope of FI curves) changes induced by amplifying an AHP with dynamic clamp. (C) Chemical inhibition of ADP reveals depolarizing effect present in control. (D) Gain changes induced by suppressing ADP. Panels A and B modified from figures donated by M. Higgs and W. Spain; C and D modified from figures donated by H. Mehaffey. 4

2 **Aftercurrents and afterpotentials for a self-coupled neuron.** (A) Our LIF neuron is driven by constant current μ, stochastic noise ξ, and a delayed feedback current $gx(t)$ after each spike. Aftercurrents are shown for g=1 (red) and g=-1 (blue). (B) Voltage deflections induced by afterpotentials in our stochastic ODE model 2.9 with $\theta = 1$ (i.e. μ=0.861, σ=0.61) with varying feedback strengths: g=-2 (blue), g=0 (black), and g=1 (red). Our model's normalized voltage outputs have been scaled by a factor of 15 for comparison to realistic neural voltages. 8

3 **Schematic of the single neuron scenario**. Gaussian white noise $\xi(t)$ drives the neuron, eliciting changes in voltage. The spike train y(t) records points at which action potentials are fired. Y(t) records the cumulative spike counts over time, while x(t) shows the level of feedback current to the neuron due to its prior spiking. Here g=1 and θ=1 (i.e. μ=0.861, σ=0.61). Our model's normalized voltage outputs have been scaled by a factor of 15 for comparison to realistic neural voltages. 9

4 **FI curves: theory (lines) vs simulation (circles).** Here g=-2 (blue), g=0 (black), g=1 (red). The x-axis shows $\mu(\theta)$ as θ ranges over [0,1] 11

5 **Numerical approximation of dynamic gain for θ=1** (i.e. μ=0.861, σ=0.61). The shaded region illustrates the window of integration when T=5 ms. The constant current μ is delivered up to $t^* = 0$, at which point the current is increased by ϵ=0.1. Firing rates are plotted for g=-2 (blue), g=0 (black), and g=1 (red), and were computed over bins of 0.5 ms width, averaged over 10 million noise realizations. Note that firing rates are seen to rise slightly before t=0: this is an artifact of plot interpolation for continuous values of t between the discrete bins at t=-0.5 ms and t=0 ms. 14

6 **Temporal changes in dynamic gain in theory and simulation for θ=1** (i.e. μ=0.861, σ=0.61). Results are expressed as ratios relative to the g=0 case, for g=-2 (blue) and g=1 (red) for both theory (lines) and simulation (circles). 16

7 **Schematic of paired self-coupled neurons** 19

8 **Spectral estimates.** Theoretical (solid) and numerical estimates (circles) of power-spectra (S1) and cross-spectra (S12) of spike trains from paired self-coupled neurons at different values of θ, for g=-2 (blue), g=0 (black), g=1 (red). 22

9 **Temporal changes in covariance, variance, and susceptibility ratios due to feedback at θ=1** (i.e. μ=0.861, σ=0.61). Theory(solid) and simulation statistics (circles) at g=-2 (blue) and g=1 (red) are shown here relative to results for the baseline g=0. 24

10 **Susceptibility changes due to feedback over our full range of θ, for short (T=3 ms), medium (T=10 ms), and very large time windows (T=256 ms).** Theory (solid) and simulation (circles) are shown at g=-2 (blue) and g=1 (red). 25

11 **Susceptibility *ratio* changes due to feedback over our full range of θ, for short (T=3 ms), medium (T=10 ms), and very large time windows (T=256 ms).** Theory (solid) and simulation (circles) at g=-2 (blue) and g=1 (red) are shown relative to baseline g=0. 26

12 **Cross-correlation and autocorrelation for θ=1 (i.e. μ=0.861, σ=0.61), where firing rate matching has been performed with μ.** These statistics are computed from simulations, and are shown for g=-2 (blue), g=0 (black), and g=1 (red). 27

13 **Changes in susceptibility ratios with firing rate matching for θ=1**(i.e. μ=0.861, σ=0.61). The baseline θ=1 produces 50Hz firing for g=0. The g=-2 (blue) and g=1 (red) show susceptibilities relative to susceptibility at g=0. 33

14 **Changes in susceptibility ratios with μ-based firing rate matching for θ=1 (i.e. μ=0.861, σ=0.61) at α=0.1 ($\tau_s = 10$).** 34

15 **Range of correlation modulation as a function of τ_s.** Here we use μ-based firing rate matching for θ=1 (i.e. μ=0.861, σ=0.61). 35

16 **Extinction of theoretical condition for correlation modulation as feedback time scale slows.** 36

17 **Progression from (A) uncoupled neurons, to (B) self-coupled neurons, to (C) cross-coupled neurons.** 40

18 **Firing rate matching for θ=1.** We introduce a parameter "p" for traversing firing rate isolines in (θ, g) space. Panel (a) shows the curves in parameter space which produce identical firing rates. Panel (b) shows the firing rates corresponding to the isolines in panel (a). Panel (c) shows the gain at each point along the firing rate isolines. 48

19 **Changes in susceptibility ratios with firing rate matching for θ=1.** The baseline θ=1 produces 50Hz firing for g=0. The g=-2 (blue) and g=1 (red) show susceptibilities relative to susceptibility at g=0. Firing rate matching is achieved with θ_g=1.168 (g=-2) and θ_g=.915 (g=1). 49

1.0 INTRODUCTION

A central tenet of modern neuroscience is that communication between neurons is accomplished by encoding information in sequences of action potentials or "spike trains". Consequently, much research has been devoted to looking for patterns in spike trains recorded from animals engaged in prototypical tasks. Among the many spike-train statistics which have been studied, single neuron gain [34, 37] and pairwise correlation [2] have been shown to be particularly relevant to behavior and cognitive function. For instance, single neuron recordings from monkeys performing visual discrimination tasks in attentive and inattentive states show that attention to a stimulus increases the gain of a neuron's response to that stimulus [27]. Gain control mechanisms have also been implicated in the brain's mapping of visual spaces [39] and auditory spaces [41], and may produce orientation selectivity to contrast changes in visual stimuli [15]. The degree of spike train correlation is also influenced by behavioral state. Subjects who are resting quietly exhibit high neural correlation which is disrupted when the subject is roused to activity [21]. The impact of subject attention on pairwise correlation is complex, with several studies reporting an attention-related decrease in long time scale spike correlation [32, 8], while others report an increase in spike time synchrony with attention [36]. Gain and correlation have also been shown to be significant in sensory encoding. Changes in pairwise correlation between neurons in visual cortex have been shown to differ between exposure to contrast discrimination and orientation discrimination [20]. Experiments with electric fish have shown differences in correlation arising from communication signals versus predatory ones [7]. Recent modeling work has also bolstered arguments for the relevance of divisive gain control in response to time-varying stimuli [25]. Despite the overwhelming evidence that both spike rate gain and spike time correlation are affected by similar behavioral and neural manipulations, there is a lack of understanding of

the underlying mechanisms which relate both gain and correlation control.

An important mechanistic component of firing rate gain is the membrane currents that are recruited during single cell activity. Action potentials occur when a neuron's electrical potential reaches a critical threshold, causing the soma to release current through the dendrite. Following a spike, there is a rapid change in the distribution of open and closed ion channels as the neuron repolarizes. During this period, ion flux in and out of the neuron changes, creating an after-current which in turn affects the potential of the neuron. Depending on the net effect of this current on membrane potential, it is referred to as either a depolarizing afterpotential (ADP) or a hyperpolarizing afterpotential (AHP) [9]. Publications in recent years have shown self-coupling or feedback to affect many aspects of neural computation. AHP's have been implicated as an important mechanism behind the gain decreases observed with increasing stimulus noise [18], while ADPs have been shown to increase gain [28]. Fig 1 shows data collected in these studies, revealing both the voltage deflections induced by aftercurrents (panels A and C), as well as their associated changes in firing rate gain (panels B and D).

It is known from [11] that neural correlation increases with firing rate, and firing rate gain is, by definition, a function of firing rate. The experimental evidence cited above indicates that self-coupled feedback alters firing rates and gains, and so would be expected to affect correlation as well. It therefore seems reasonable to examine the effects of self-feedback on gain and correlation, looking for similarities and differences in the underlying neural mechanisms that affect gain and correlation control. To this end, our study investigates how feedback from spikes affects both gain and correlation in self-coupled neurons. This work builds off of previous theoretical results reported for leaky integrate-and-fire (LIF) neurons subject to correlated noise inputs, including [24, 38, 16, 5, 35, 30]. While these studies focused on a simple LIF model without internal feedback, other studies have developed tools for studying feedback. In particular, a linear response approach for computing power and cross-spectra of spike trains for neurons with spike-driven feedback was presented in [12, 24], and [37] discussed calculation of gain in the presence of feedback. Our study aims to integrate this work to better understand how gain and correlation are comodulated by feedback. We first introduce a concept of "dynamic gain", and observe how it is affected by feedback.

We then synthesize the work of [35] and [24] to explore how feedback shapes correlation at different time scales, and finally analyze the results to explain how these effects arise.

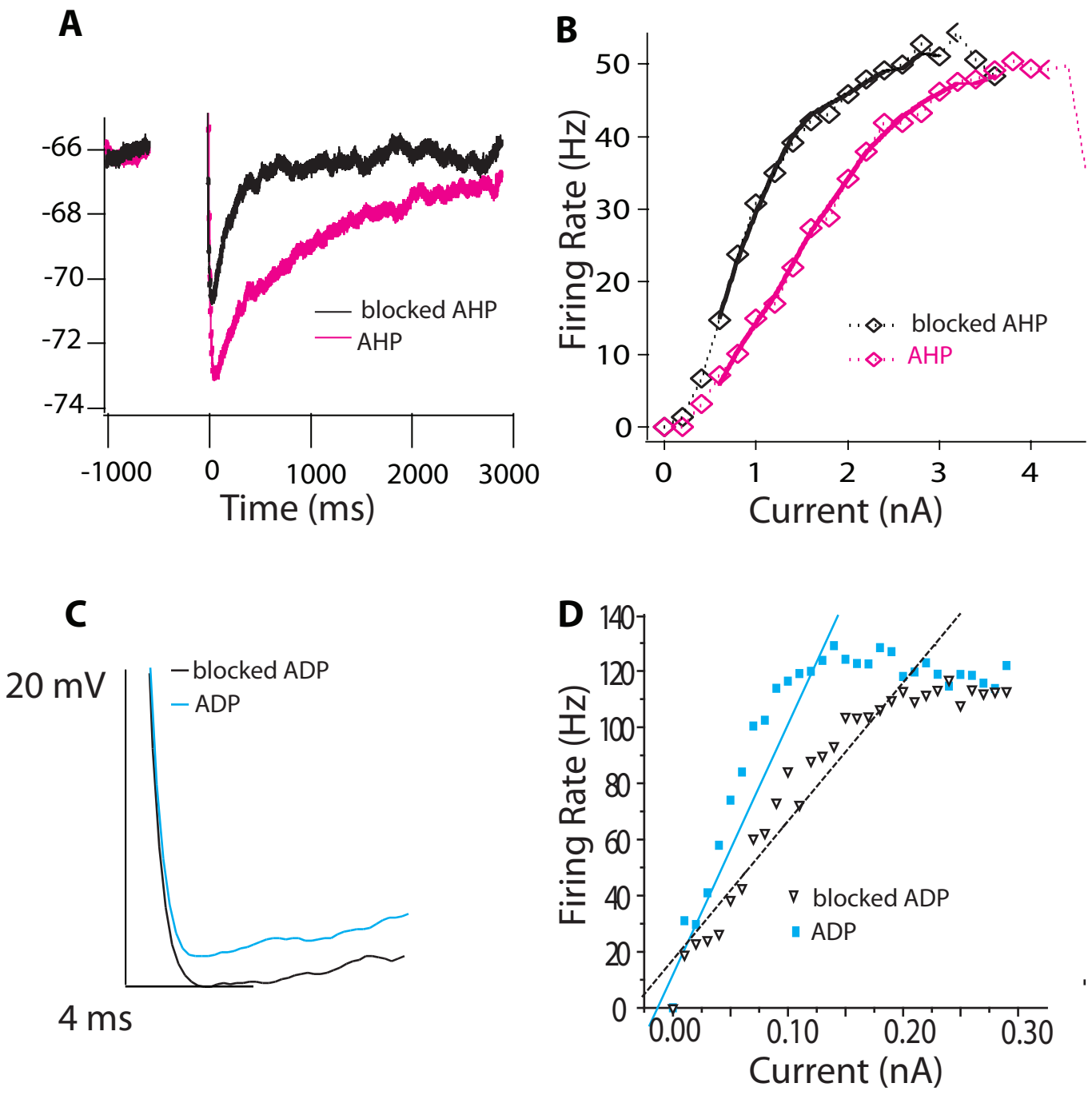

Figure 1: **Effects of ADP's and AHP's.** (A) Chemical inhibition of AHP reveals hyperpolarization present in control. (B) Gain (slope of FI curves) changes induced by amplifying an AHP with dynamic clamp. (C) Chemical inhibition of ADP reveals depolarizing effect present in control. (D) Gain changes induced by suppressing ADP. Panels A and B modified from figures donated by M. Higgs and W. Spain; C and D modified from figures donated by H. Mehaffey.

2.0 THE SINGLE SELF-COUPLED NEURON

2.1 DERIVATION OF THE MODEL

While this study is primarily concerned with a pair of self-coupled neurons, we first introduce the underlying concepts for a simple single neuron system. We represent our idealized neurons using extensions to the the classic leaky integrate-and-fire (LIF) model:

$$\tau_m V' = \mu - V + I(t) \tag{2.1}$$

Here, V represents the membrane potential of the neuron, while τ_m denotes the membrane time constant. Throughout our calculations and simulations, we set τ_m=10 msec. The terms μ and $I(t)$ respectively denote constant and time-varying currents. The model is equipped with the standard reset rule:

$$V(t) \geq V_{threshold} \implies V(t) = V_{reset} \tag{2.2}$$

which signifies the firing of an action potential and instantaneously resets the soma potential to its hyper-polarized state. Additionally, we assume these neurons have an absolute refractory period of duration τ_R, which briefly fixes the neuron's voltage at its hyper-polarized state after any spike time t_i so that:

$$V'(t) = 0 \quad t \in [t_i, t_i + \tau_R] \tag{2.3}$$

For values of μ below 1, our neuron lies in an excitable regime - the neuron's voltage will never reach the spiking threshold when $I(t) = 0$. We focus solely on scenarios where μ lies

in the excitable regime and our time-dependent input $I(t)$ is a stochastic process. In this case, our voltage equation is:

$$\tau_m V' = \mu - V(t) + \sigma\xi(t) \tag{2.4}$$

where $\xi(t)$ represents our time-dependent noise source with variance σ^2. When the stochastic input is viewed as a noise term, it is reasonable that its intensity should be correlated with the strength of the baseline current (i.e. we assume the presynaptic pool has Poisson statistics). We therefore parameterize μ and σ by the normalized parameter θ, writing:

$$\theta \in [0, 1] \tag{2.5}$$

$$\mu = \mu_0 + k_\mu \theta \tag{2.6}$$

$$\sigma = \sigma_0 + k_\sigma \theta \tag{2.7}$$

Throughout this study, we choose these constant values to be: μ_0= .511, σ_0=0.3, k_μ=0.35, k_σ=0.31. These values have been chosen to satisfy several constraints: to produce a range of output firing rates over approximately 5-50 Hz, to guarantee that that the neuron remains in the excitable regime for all values of θ, and to ensure that σ values remain sufficiently large to avoid numerical difficulties during evaluation.

In order to incorporate the effects of spike-driven feedback on voltage, one important extension to the model is required. The implementation of such a feedback mechanism should adhere to several intuitive principles. First, in the absence of spiking events (due to prolonged hyperpolarization), there should be zero current due to spike-driven feedback. Once a spike is emitted, the neuron should receive an influx of current after a short transmission time τ_D. Following this initial influx, the spike should continue to contribute current to the neuron, but this effect decays rapidly with time with a rate of $\alpha = \frac{1}{\tau_s}$. We follow [13], which proposes using an alpha function to represent current due to afterpotentials.

If we consider a sequence of k spikes occurring at times $\{t_1, t_2, ..., t_k\}$, we can model this time-dependent feedback current as:

$$x(t) = g\alpha^2 \sum_{j=1}^{k} (t - t_j - \tau_D) e^{-\alpha(t-t_j-\tau_D)} H(t - t_j - \tau_D) \tag{2.8}$$

where H denotes the Heaviside step function and g specifies the strength of the feedback. Throughout this study, we consider both positive g (corresponding to an ADP) and negative g (corresponding to an AHP). We note we do not consider these perturbations symmetrically: our lower bound for g is -2, while our upper bound for g is 1. The reason for this asymmetry is that $g > 0$ constitutes positive feedback into the system which introduces numerical instability in later calculations. We have therefore confined positive g to a range where it does not pose problems. Figure 2 illustrates the effect of this feedback on the voltage of an LIF neuron in the immediate wake of one spike. Negative values of g exert an inhibitory influence by causing hyperpolarization of the neuron and thus slowing subsequent rises in voltage as compared to those of a neuron without feedback (i.e. $g = 0$). Conversely, positive values of g have an excitatory influence, accelerating the depolarization of the neuron relative to the $g = 0$ case. For the parameter values we have chosen, the afterpotential's effect on the voltage following a single spike is barely visible beyond a period of approximately 25 milliseconds. We remark that when $g > 0$, care must be taken to ensure the neuron remains in the excitable regime: the combined baseline and mean feedback currents should not exceed the threshold at which the neuron will fire in the absence of noise.

Differentiating x with respect to t, we can formulate our resulting model as a system of stochastic first-order ODE's:

$$\begin{aligned} V' &= \tfrac{1}{\tau_m} (\mu - V + gx + \sigma\xi(t)) \\ x' &= y \\ y' &= -\alpha^2 x - 2\alpha y + \alpha^2 \textstyle\sum_j \delta(t - t_j - \tau_D) \end{aligned} \tag{2.9}$$

where we have factored g out of $x(t)$ and inserted it explicitly into the voltage equation.

For the simulation component of this study, we solve these equations numerically using an Euler-Maruyama integration scheme [19]. Figure 3 depicts the schematic of our single neuron system, using sample input and output from simulation.

2.2 SPIKE TRAIN STATISTICS

In addition to simulating our model to gauge the effects of afterpotentials on spike trains, we can also compute theoretical estimates of the relevant statistics. This theory offers a check for our simulations, provides a more efficient method of obtaining results, and provides a basis for our later analytical treatment of correlation modulation. Throughout this section, we denote a sequence of spike times $\{t_j\}$ for a given neuron i as a series of delta functions: $y_i(t) = \sum_j \delta(t - t_j)$. We now discuss several intermediate results which will be necessary to calculate the statistics we are ultimately interested in.

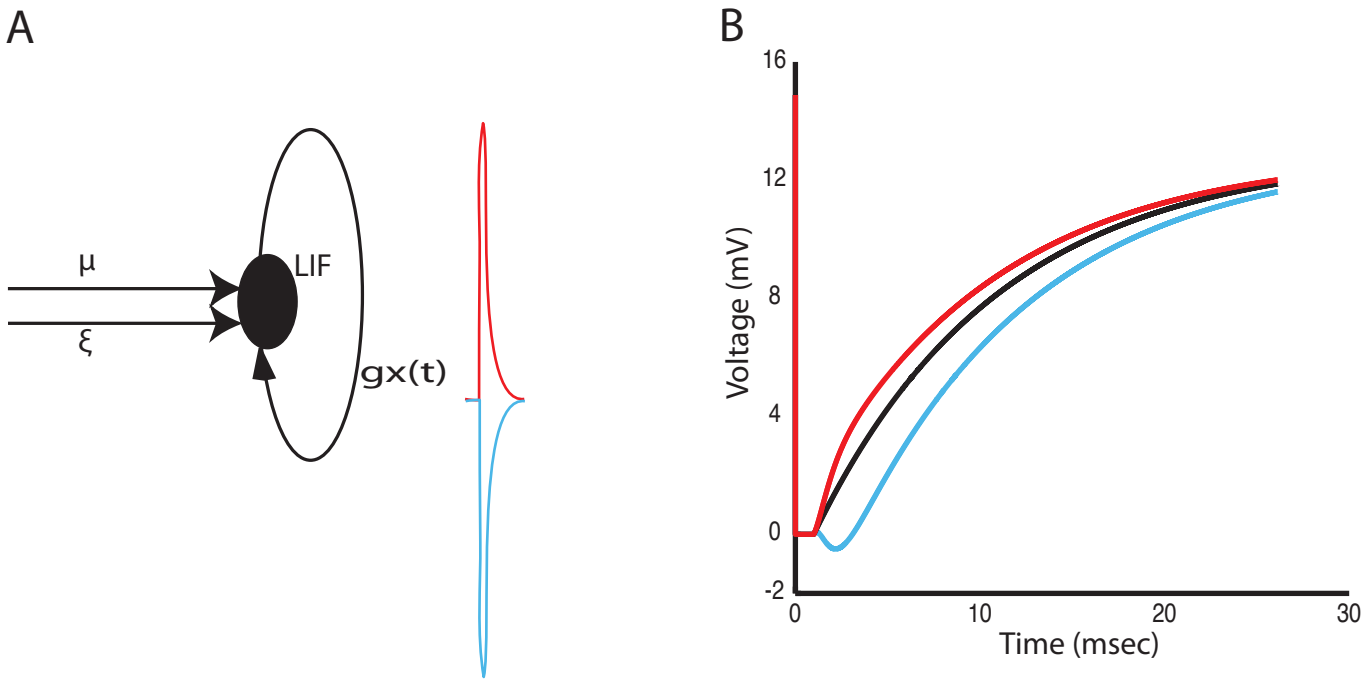

Figure 2: **Aftercurrents and afterpotentials for a self-coupled neuron.** (A) Our LIF neuron is driven by constant current μ, stochastic noise ξ, and a delayed feedback current $gx(t)$ after each spike. Aftercurrents are shown for g=1 (red) and g=-1 (blue). (B) Voltage deflections induced by afterpotentials in our stochastic ODE model 2.9 with $\theta = 1$ (i.e. μ=0.861, σ=0.61) with varying feedback strengths: g=-2 (blue), g=0 (black), and g=1 (red). Our model's normalized voltage outputs have been scaled by a factor of 15 for comparison to realistic neural voltages.

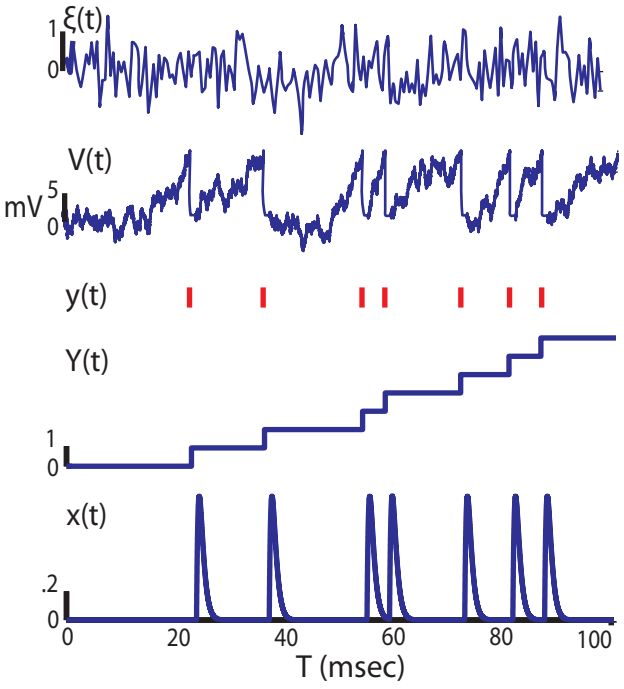

Figure 3: **Schematic of the single neuron scenario**. Gaussian white noise $\xi(t)$ drives the neuron, eliciting changes in voltage. The spike train y(t) records points at which action potentials are fired. Y(t) records the cumulative spike counts over time, while x(t) shows the level of feedback current to the neuron due to its prior spiking. Here g=1 and θ=1 (i.e. μ=0.861, σ=0.61). Our model's normalized voltage outputs have been scaled by a factor of 15 for comparison to realistic neural voltages.

2.2.1 FIRING RATE ESTIMATION

The simplest statistic we can compute for a spike train is its asymptotic firing rate ν: the number of spikes produced per unit of time. Using our definition of the spike train $y(t)$, we can count the number of spikes observed prior to time t:

$$Y(t) = \int_0^t y(t')dt' \tag{2.10}$$

The asymptotic firing rate of the system is then given by:

$$\nu = \lim_{T\to\infty} \frac{Y(T)}{T} \tag{2.11}$$

For an LIF neuron driven with constant current μ and Gaussian white noise of variance σ^2, the following closed form expression for firing rate is given in [40]:

$$\nu_0(\mu, \sigma) = \left[\tau_R + \sqrt{\pi} \int_{\frac{(\mu - V_T)}{\sigma}}^{\frac{(\mu - V_R)}{\sigma}} e^{z^2} dz \right]^{-1} \tag{2.12}$$

In this expresion τ_R represents the absolute refractory period, while V_T and V_R represent the voltage threshold and reset values, respectively. In our system, however, we must also account for the effect of feedback on firing rates. We employ the strategy of [24, 12, 14], using the mean feedback-induced current $\langle x(t) \rangle$ to approximate the "effective current" as a static correction term to the baseline current:

$$\mu_{eff} = \mu + \frac{g}{\tau_m} \langle x(t) \rangle = \mu + \frac{g}{\tau_m} \nu \tag{2.13}$$

Our corrected firing rate ν is thus given by the transcendental equation

$$\nu = \nu_0(\mu_{eff}, \sigma) = \nu_0(\mu + \frac{g}{\tau_m} \nu, \sigma) \tag{2.14}$$

In actual calculations, we solve for ν numerically with a root-finding method, using $\nu_0(\mu, \sigma)$ as our initial estimate. Figure 4 compares firing rates estimated using this process to those computed from spike trains generated by simulations of our model.

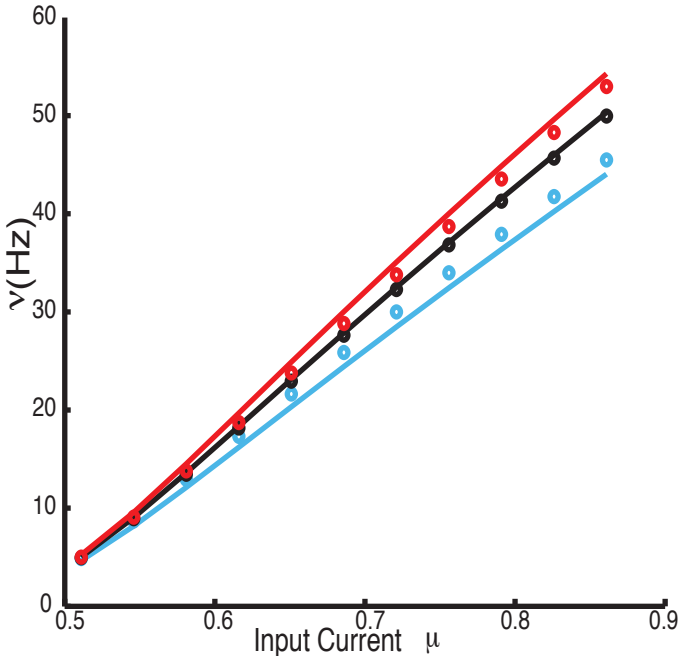

Figure 4: **FI curves: theory (lines) vs simulation (circles).** Here g=-2 (blue), g=0 (black), g=1 (red). The x-axis shows $\mu(\theta)$ as θ ranges over [0,1] .

2.2.2 LINEAR RESPONSE THEORY

Viewing changes in a neuron's membrane potential as a stochastic process allows us to apply useful theoretical frameworks for discussing correlations. As briefly outlined in appendix A.0.1, the response statistics of our system are given by a Fokker-Planck equation (FPE) associated with equation 2.4. Linear response theory, which has been developed as a way to study perturbations to FPE's [33], provides a useful way to model the perturbations due to feedback. To derive formulas for dynamic gain (and ultimately correlation susceptibility for multi-neuron scenarios), we apply linear response theory as demonstrated in [6, 24]. To do so, we define a baseline "unperturbed" neuron, and then consider weak feedback terms as a small perturbation to this baseline. Our unperturbed system encompasses the leak term, the baseline current μ, and the noise source $\xi(t)$. Furthermore, we also incorporate the mean of the feedback current, $\langle x(t) \rangle$, into the unperturbed neuron. Having done this, only

the mean-corrected time-varying component of the feedback current $x(t) - \langle x(t) \rangle$ remains unaccounted for, and it is this quantity which will constitute the perturbation to the system.

Since later calculations are more convenient in the Fourier domain, we compute the transformed spike train: $\tilde{y}(\omega) = \frac{1}{\sqrt{T}} \int_0^T e^{i\omega t}(y(t) - \nu)dt$, where we have subtracted off the mean firing rate of the neuron, ν. To continue building our ansatz, we now compute the Fourier transform of $x(t)$, the feedback current as defined in equation 2.8:

$$F(\omega) = g \frac{e^{i\omega\tau_D}}{(1 - \frac{i\omega}{\alpha})^2} \tag{2.15}$$

Introducing a linear transfer function $A(\omega)$ which incorporates the spike-induced feedback into the spike train, we can write our final linear ansatz for the perturbed self-coupled neuron's spike train as:

$$\tilde{y}(\omega) = \tilde{y}_0(\omega) + A(\omega)F(\omega)\tilde{y}(\omega) \tag{2.16}$$

$A(\omega)$ describes the susceptibility of $\tilde{y}(\omega)$ of being perturbed from equilibrium by inputs with intensity $F(\omega)$. The derivation of $A(\omega)$ requires extensive calculations which are detailed in [23, 6]. A simplified sketch of this derivation is provided in appendix A to demonstrate the key ideas behind it. The resulting definition of $A(\omega)$ is given by:

$$A(\omega, \mu, D) = \frac{i\nu\omega\sqrt{(D)}}{i\omega - 1} \frac{\mathcal{D}_{i\omega-1}\left(\frac{\mu - V_T}{\sqrt{D}}\right) - e^{\Delta}\mathcal{D}_{i\omega-1}\left(\frac{\mu - V_R}{\sqrt{D}}\right)}{\mathcal{D}_{i\omega}\left(\frac{\mu - V_T}{\sqrt{D}}\right) - e^{\Delta i\omega\tau_R}\mathcal{D}_{i\omega}\left(\frac{\mu - V_R}{\sqrt{D}}\right)} \tag{2.17}$$

$$\Delta = \frac{V_R^2 - V_T^2 + 2\mu(V_T - V_R)}{4D} \tag{2.18}$$

in which the $\mathcal{D}_x$ terms denote the parabolic cylinder functions and the argument $D = \frac{\sigma^2}{2}$.

2.2.3 DYNAMIC GAIN

Neuronal gain is conventionally defined to be the rate of change in firing rate with respect to input current, i.e. the quantity $d\nu/d\mu$. In this study, we are interested not only in how gain is affected by spike-driven feedback, but also how gain is affected over a *window* of time following spiking events. Below we demonstrate both computational and theoretical approaches to computing dynamic gain.

2.2.3.1 COMPUTATIONAL APPROACH The conventional definition of gain can be easily rewritten in terms of spike counts:

$$\frac{d\nu}{d\mu} = \lim_{\epsilon \to 0} \lim_{T \to \infty} \frac{1}{T} \frac{\langle Y_T^{\mu+\epsilon} \rangle - \langle Y_T^{\mu} \rangle}{\epsilon} \tag{2.19}$$

Where Y_T^{μ} denotes the spike count over a train produced by a neuron with input μ, and $\langle . \rangle$ denotes averaging over many simulations of duration T. We can estimate this derivative numerically by simulating our ODE model with very large values of T and tiny values of ϵ and computing a finite difference quotient. Dynamic gain generalizes this concept: rather than observing changes in asymptotic firing rates, we observe firing rate changes over a narrower window of time immediately following a perturbation in μ. To write this formally, we make a slight change to our definition of spike count Y, providing control over the bounds of the window of integration:

$$Y(t_0, t_1) = \int_{t_0}^{t_1} y(t) dt \tag{2.20}$$

We can incorporate a perturbation in μ into each simulation as a jump discontinuity at some time t^*:

$$\hat{\mu}(t) = \begin{cases} \mu & t < t^* \\ \mu + \epsilon & t \geq t^* \end{cases} \tag{2.21}$$

Dynamic gain over an integration window T can then be computed as:

$$G_T^{\mu} = \lim_{\epsilon \to 0} \lim_{t^* \to \infty} \frac{1}{\epsilon} \left(\frac{\langle Y^{\hat{\mu}}(t^*, t^* + T) \rangle}{T} - \frac{\langle Y^{\hat{\mu}}(0, t^*) \rangle}{t^*} \right) \tag{2.22}$$

where $\langle . \rangle$ again denotes averages over many simulations of our model with different noise realizations. In actual simulations t^* should, of course, be chosen large enough to guarantee that the neuron has converged to its asymptotic firing rate before the jolt is delivered. Figure 5 illustrates the use of this approximation scheme on simulation results.

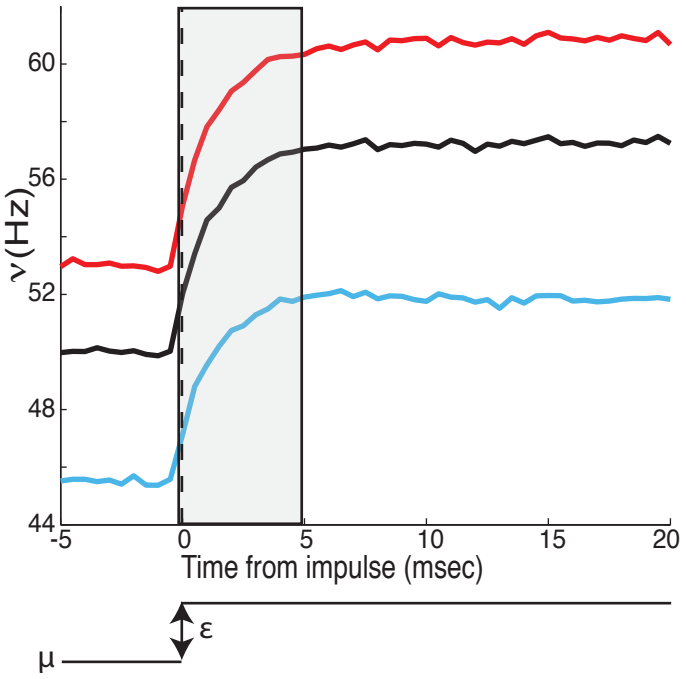

Figure 5: **Numerical approximation of dynamic gain for θ=1** (i.e. μ=0.861, σ=0.61). The shaded region illustrates the window of integration when T=5 ms. The constant current μ is delivered up to $t^* = 0$, at which point the current is increased by ϵ=0.1. Firing rates are plotted for g=-2 (blue), g=0 (black), and g=1 (red), and were computed over bins of 0.5 ms width, averaged over 10 million noise realizations. Note that firing rates are seen to rise slightly before t=0: this is an artifact of plot interpolation for continuous values of t between the discrete bins at t=-0.5 ms and t=0 ms.

2.2.3.2 THEORETICAL APPROACH In the continuous domain of our theory, it is natural to implement windows of time as a kernel that places heavy weights on spikes inside the target window, and rapidly vanishes outside it. We employ such a windowing scheme, described in [10], by defining a window of width T in the time-domain as:

$$k_T(t) = \frac{T - |t|}{T} \tag{2.23}$$

Computing the Fourier transform of k_T, we obtain

$$K_T(\omega) = \frac{4\sin^2(\frac{\omega T}{2})}{\omega^2 T} \tag{2.24}$$

Noting that

$$\lim_{\omega \to 0} A(\omega) = d\nu/d\mu \tag{2.25}$$

we propose that $A(\omega)$ is a natural dynamic analog of gain and write:

$$G_T = \int_0^\infty |A(\omega)| K_T(\omega) d\omega \tag{2.26}$$

Indeed, as T approaches infinity $K_T(\omega)$ converges to $\delta(\omega)$, so by 2.25 we see that our dynamic gain converges to conventional gain, as desired:

$$\lim_{T \to \infty} G_T = \int_0^\infty |A(\omega)| \delta(\omega) d\omega = \frac{d\nu}{d\mu} \tag{2.27}$$

One further change is necessary to properly account for the effects of feedback on gain. Throughout our theory, we have addressed the feedback-induced current by the approximation $\mu_{eff} = \mu + \frac{g}{\tau_m}\nu(\mu_{eff})$. To discuss the feedback-influenced gain $d\nu/d\mu$ requires a simple application of the chain rule, as pointed out in [37, 14]:

$$\frac{d\mu_{eff}}{d\mu} = \frac{1}{1 - \frac{g}{\tau_m}\frac{d\nu}{d\mu_{eff}}} \tag{2.28}$$

$$\frac{d\nu}{d\mu} = \frac{d\nu}{d\mu_{eff}}\frac{d\mu_{eff}}{d\mu} = \frac{\frac{d\nu}{d\mu_{eff}}}{1 - \frac{g}{\tau_m}\frac{d\nu}{d\mu_{eff}}} \tag{2.29}$$

where $d\nu/d\mu_{eff}$ is obtained by differentiating the formula for ν given in equation 2.12. Analogously, replacing $d\nu/d\mu_{eff}$ with $A(\omega)$ yields our expression for feedback-corrected dynamic gain:

$$G_T(g) = \int_0^\infty \frac{|A(\omega)|}{1 - \frac{g}{\tau_m}|A(\omega)|} K_T(\omega) d\omega \tag{2.30}$$

We now use both our theoretical and computational methods to predict the gain modulation due to positive and negative feedback. It is convenient to express our gain changes relative to the baseline gain which is obtained in the $g = 0$ case, as shown in figure 6. Here we see general agreement between the predictions from theory and simulation. The sign of the gain changes matches that of g, and is roughly 10% for both positive and negative g. Importantly, both theory and simulation show that gain modulation is relatively uniform across all time windows.

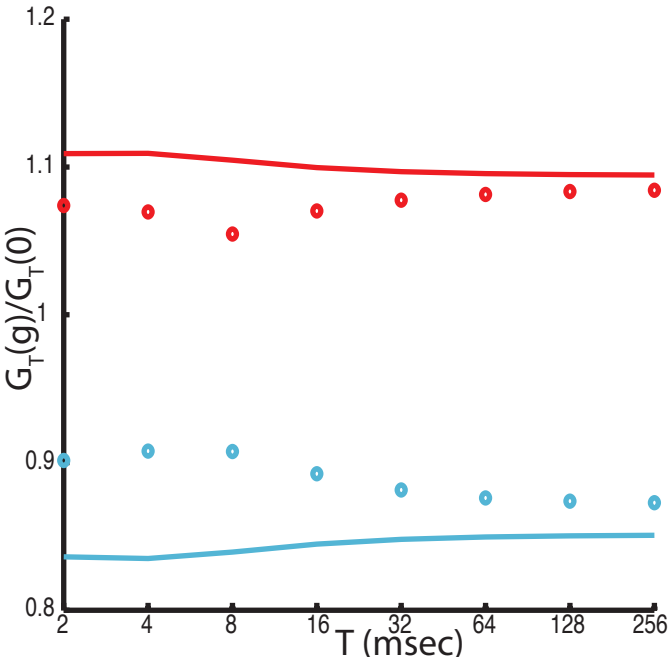

Figure 6: **Temporal changes in dynamic gain in theory and simulation for θ=1** (i.e. μ=0.861, σ=0.61). Results are expressed as ratios relative to the g=0 case, for g=-2 (blue) and g=1 (red) for both theory (lines) and simulation (circles).

3.0 NETWORKS OF PAIRED SELF-COUPLED NEURONS

3.1 DERIVATION OF THE MODEL

Having discussed isolated self-coupled neurons with spike-influenced feedback, we begin to consider networks of these neurons. Generally, such networks are constructed such that each neuron receives spike-influenced feedback from the other neurons in the network. However, such a setup conflates the effects from feedback with those of multi-neuron networks. To distinguish these effects, we study an even simpler "network": a pair of self-coupled neurons. Each neuron in the pair receives feedback from its own spikes, but does not communicate with its counterpart. If the noise inputs to these neurons are completely independent, there is nothing to study; the network would be nothing more than a collection of independent self-coupled neurons. We therefore alter slightly our previous formulation, using two sources of stochastic input as done in [24, 38, 16, 5, 35, 30]. As in our single-neuron formulation, each neuron will receive its own unique source of stochastic input, $\xi_i(t)$. We now add another source of stochastic input which is common to all neurons in the network, $\xi_c(t)$. Our parameter σ will represent the intensity of the *combined* sources, and we use the normalized parameter c to control the distribution of this intensity between the two sources. Since we would like to maintain the condition that $\sigma^2 = \sigma_i^2 + \sigma_c^2$, we define our distribution rule as:

$$\sigma_i = \sigma\sqrt{1-c} \tag{3.1}$$

$$\sigma_c = \sigma\sqrt{c} \tag{3.2}$$

At one extreme, setting c=0 eliminates the common noise, while setting c=1 eliminates all individual noise sources. In our studies, we are interested in weakly correlated inputs

where c is small, so individual noise is considerably stronger than common noise. In all of our work with paired neurons, we take $c = 0.2$.

It is simple to adapt our earlier model 2.9 to handle this new scheme. Our feedback function $x(t)$ requires no change from the single-neuron case, other than an index to specify which spike train it is summating. To distinguish the spike times from the distinct trains, we denote the time of the j-th spike of neuron i as t_{ij}. Assigning each neuron its own state variables for feedback current levels and voltage, and adding the new noise source to the voltage equations, we arrive at the model:

$$\tau_m V_i' = \mu - V_i + g x_i + \sqrt{c}\sigma\xi_c(t) + \sqrt{1-c}\sigma\xi_i(t) \tag{3.3}$$

$$x_i' = y_i \tag{3.4}$$

$$y_i' = -\alpha^2 x_i - 2\alpha y_i + \alpha^2 \textstyle\sum_j \delta(t - t_{ij} - \tau_D) \tag{3.5}$$

As before, we solve these equations numerically using an Euler-Maruyama integration scheme. Figure 7 shows a schematic for the network of paired, self-coupled neurons using sample input and output from this system.

3.2 SPIKE TRAIN STATISTICS

We now update our spike train statistic calculations for the paired, self-coupled neuron scenario.

3.2.1 FIRING RATE ESTIMATION

Our firing rate estimation formulas from the single neuron case continue to hold here. Note that when evaluating firing rate $\nu_0(\mu, \sigma)$, σ represents the total variance which is distributed between both the shared and individual noise sources. Our correction for μ_{eff} also remains unchanged.

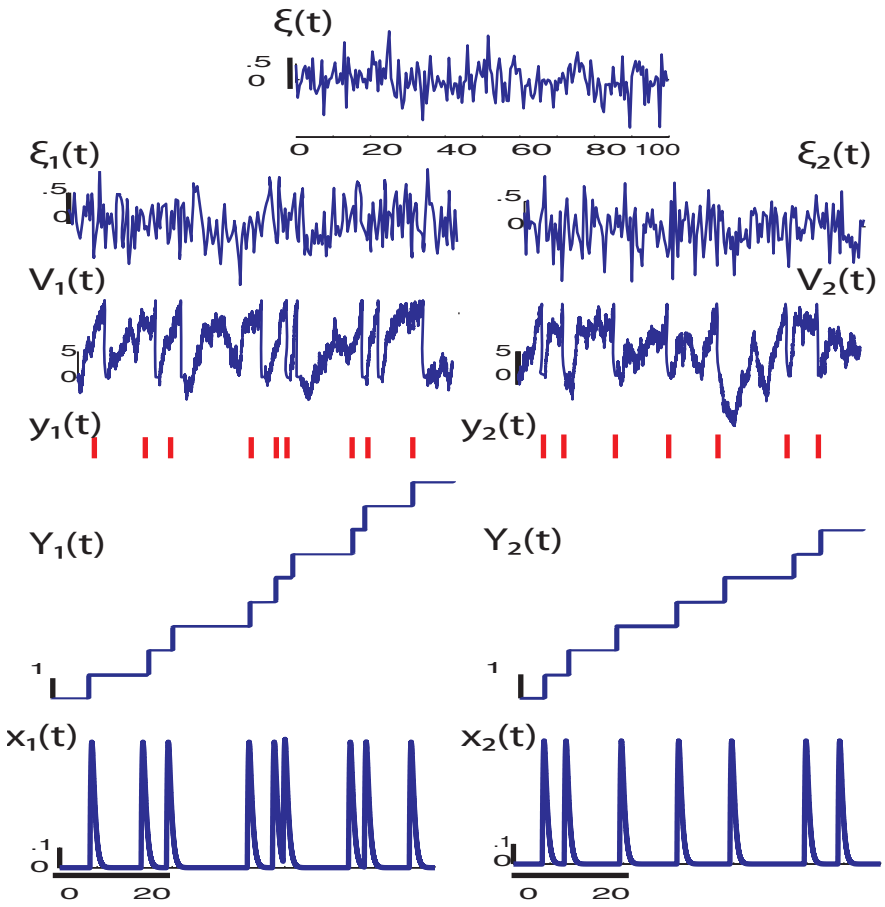

Figure 7: **Schematic of paired self-coupled neurons**

3.2.2 SPIKE TRAIN CORRELATIONS AND LINEAR RESPONSE THEORY

When working with pairs of neurons, we can now consider additional statistics that were not relevant in the single neuron scenario. Specifically, we would like to study the degree of correlation between the spike trains produced by each neuron of the pair. To arrive at formulas for correlation, we continue to build on the linear response theory we used for the single neuron. We make straightforward modifications to 2.16, adding terms for the new noise sources to obtain:

$$\tilde{y}_i(\omega) = \tilde{y}_{0,i}(\omega) + A(\omega)\left(\sqrt{c}\tilde{\xi}_c(\omega) + F(\omega)\tilde{y}_i(\omega)\right) \tag{3.6}$$

We point out that only the common noise ξ_c appears explicitly in the expression above, even though each neuron is also receiving an individual noise input ξ_i. The reason for this is that, as in our previous ansatz for the single neuron case, the formula for y_0 (the "unperturbed" neuron's spike train) accounts for the individual noise input as well as the baseline current, the leak, and the mean of the feedback current. In this paired neuron scenario with both common and individual noise, our perturbation now accounts for both the common noise input and the time-varying component of the feedback current.

With this approximation, we can now compute power spectra and cross-spectra of our transformed spike trains:

$$S_i = \lim_{T\to\infty} \langle \tilde{y_i}^* \tilde{y_i} \rangle \equiv \text{power spectrum of spike train of neuron i} \tag{3.7}$$

$$S_{ij} = \lim_{T\to\infty} \langle \tilde{y_i}^* \tilde{y_j} \rangle \equiv \text{cross spectrum of spike trains from neurons i and j} \tag{3.8}$$

$$X_{ik} = \langle \tilde{y_i} \tilde{\xi_k} \rangle \equiv \text{cross spectrum of spike train from neuron i and noise source k} \tag{3.9}$$

In the equations above, the $\langle . \rangle$ refers to averaging over phases. For the simple case of two separate, self-coupled neurons subject to both shared and individual stimuli, we can use the above definitions to obtain expressions for power and cross spectra. Expanding out the inner products above, we obtain a system of linear equations to be solved simultaneously for the variables S_i, S_{ij}, and X. Many of the resulting terms vanish: since the unperturbed power is uncorrelated with the noise terms, we have $\langle \tilde{y_0}^*, \tilde{\xi_i} \rangle = \langle \tilde{y_0}^*, \tilde{\xi_c} \rangle = 0$. Furthermore, the unperturbed spike trains themselves are independent of each other, so that $\langle \tilde{y_{0,i}}^*, \tilde{y_{0,j}} \rangle = 0$ for $i \neq j$. Eliminating these terms and rewriting the expanded products, we find:

$$S_i(\omega) = \frac{S_{0,i}(\omega) + c|A(\omega)|^2 S_{ST}(\omega)}{|1 - A(\omega)F(\omega)|^2} \text{ i=1,2} \tag{3.10}$$

$$S_{12}(\omega) = \frac{c|A(\omega)|^2 S_{ST}(\omega)}{|1 - A(\omega)F(\omega)|^2} \tag{3.11}$$

where $S_{ST} = \langle \tilde{\xi_c}^* \tilde{\xi_c} \rangle$ denotes the power spectrum of the noise source. Considering that our problem involves two unfiltered Gaussian sources leads us to alter these formulas. Since the power spectrum of Gaussian white noise is well-known to be $S_{ST}(\omega, \sigma) = \sigma^2$, we could substitute this result above. However, we also observe that unfiltered Gaussian white noise

violates the assumptions of our ansatz - it does not have a finite cutoff frequency. To deal with this, we adopt a strategy suggested in [24], which we briefly summarize. First, we note that since both sources are Gaussian white noise, they can be treated as a single noise source with combined intensity $Q = c\frac{\sigma^2}{2} + (1-c)\frac{\sigma^2}{2} = \sigma^2$. Next, we point out that the numerator of 3.10 is a linear approximation of the unperturbed power spectrum with combined noise, $S_{0,i}(\omega, Q)$. Since this is the point at which we evaluate the linear response, we argue it suffices to replace the numerator of 3.10 with $S_{0,i}(\omega, Q)$. Following this argument, we should now also replace references to $A(\omega)$ with $A(\omega, Q)$, and compute firing rates with 2.12 using $\sigma = Q$. Strictly speaking, this is no longer a true linear response approximation, as the point around which linearization is done will depend on the strength of the signal. This approach yields the modified formulas:

$$S_i(\omega) = \frac{S_{0,i}(\omega,Q)}{|1-A(\omega,Q)F(\omega,Q)|^2} \quad i = 1, 2 \tag{3.12}$$

$$S_{12}(\omega) = \frac{c|A(\omega,Q)|^2\sigma^2}{|1-A(\omega,Q)F(\omega,Q)|^2} \tag{3.13}$$

Figure 8 compares the accuracy of our theoretical approximations of the spectra against spectra calculated numerically for our simulation output using the Chronux software package [29, 1].

3.3 CORRELATION SUSCEPTIBILITY

When studying multiple-neuron networks, we can consider the correlation between the spike trains produced by the neurons. Correlation between spike trains may exist over a wide range of time scales, and we are interested in how spike-driven feedback affects this structure. For any chosen time window T, we can use our definition of the spike count Y_T^i for neuron i to compute the variance for each neuron's spike train, as well as the covariance between them:

$$Cov_T = \langle \left(Y_T^1 - \langle Y_T^1 \rangle\right)\left(Y_T^2 - \langle Y_T^2 \rangle\right) \rangle \tag{3.14}$$

$$Var_T^i = \langle \left(Y_T^i - \langle Y_T^i \rangle\right)^2 \rangle \tag{3.15}$$

where $\langle . \rangle$ denotes an average over trains obtained from many realizations of the noisy inputs.

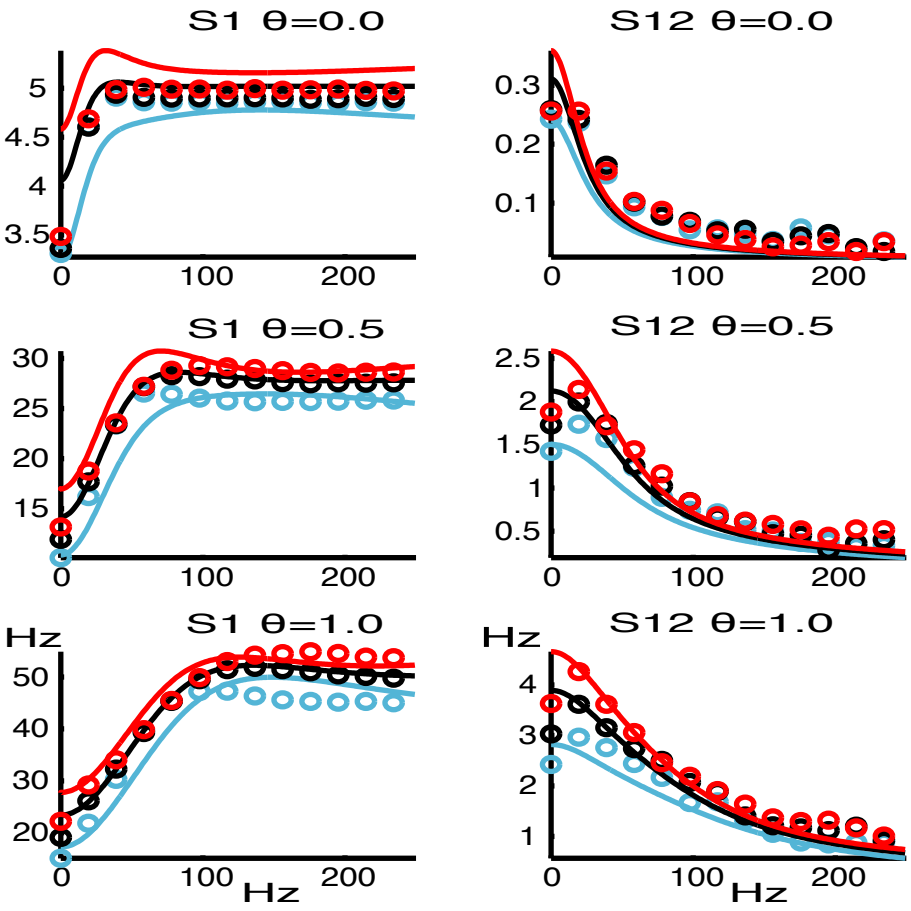

Figure 8: **Spectral estimates.** Theoretical (solid) and numerical estimates (circles) of power-spectra (S1) and cross-spectra (S12) of spike trains from paired self-coupled neurons at different values of θ, for g=-2 (blue), g=0 (black), g=1 (red).

Equipped with these definitions, we can then compute the correlation between spike trains over window T:

$$\rho_T = \frac{Cov_T}{\sqrt{Var_T^1 Var_T^2}} \tag{3.16}$$

Finally, we compute the correlation susceptibility S_T [11], which describes the sensitivity of output correlations to changes in our input correlation c:

$$S_T = \frac{\partial \rho_T}{\partial c} \approx \frac{\rho_T}{c} \text{ for c } << 1 \tag{3.17}$$

Taking advantage of the fact that our noise sources are weakly-correlated ($c = 0.2$), this approximation acts as a linear approximation to the true susceptibility.

These same notions can be extended to the continuous realm of our linear response theory. For an LIF neuron with noisy inputs and a window length of T, we again employ $K_T(\omega)$, the Fourier-transform of the triangular time-window given in equation 2.24. The relevant definitions for computing covariance, variance, correlation and susceptibility in the Fourier domain are provided in [35] as:

$$Cov_T(\omega) = T\int_{-\infty}^{\infty} S_{12}(\omega)K_T(\omega)d\omega \tag{3.18}$$

$$Var_T^i(\omega) = T\int_{-\infty}^{\infty} S_i(\omega)K_T(\omega)d\omega \tag{3.19}$$

$$\rho_T(\omega) = \frac{Cov_T(\omega)}{\sqrt{Var_T^1(\omega)Var_T^2(\omega)}} \tag{3.20}$$

$$S_T(\omega) = \frac{\rho_T(\omega)}{c} \tag{3.21}$$

In the course of actual calculation, it is convenient to make use of the fact that S_1 and S_2 will always be identical in the self-coupled neuron case we have presented. Several figures below present comparisons between theory and simulation results. Figure 9 shows the primary result we are interested in for a fixed θ: changes in these statistics relative to baseline over a range of time windows. Figure 10 shows raw changes in the statistics over a broad range of θ values, while figure 11 shows the same data relative to the non-feedback case.

In these figures, we see that there is strong qualitative agreement between theory and simulation, but there is visible quantitative error (particularly in the variance and covariance plots) that merits explanation. Our theory consistently overemphasizes the effect of feedback for both variance and covariance, though these errors tend to cancel out when we compute susceptibilities. This overestimation arises from the magnitude of the feedback term g which is used. Inspection of our alpha function $x(t)$ also reveals that g's effect is further amplified by the time scale under consideration: the faster the feedback is, the more feedback current is received per unit of time. As a result of these factors, the feedback currents are not truly tiny perturbations, and test the limits of the linear response approximation. There is a tradeoff

in this choice of g: smaller g values are more amenable to linear response but evoke changes which are much harder to see, and will not produce the degree of voltage deflection that is reported from ADP/AHP experiments. In the interest of producing afterpotentials that are consistent with the literature and obtaining more distinguishable modulation, we have chosen larger values of g at the cost of some prediction accuracy. This problem should be less pronounced in networks of coupled neurons, where each neuron receives aftercurents which are more diffuse over time, and realistic voltage deflections can be obtained by summing over multiple feedback sources with smaller g values.

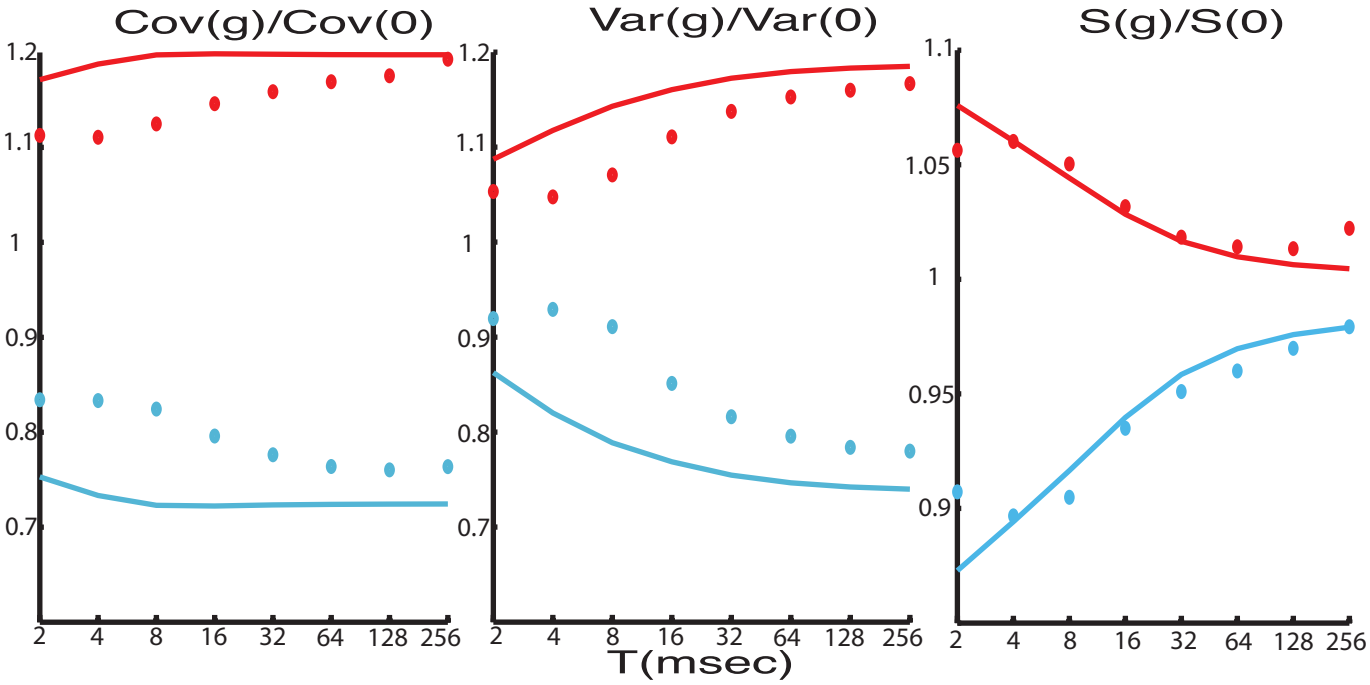

Figure 9: **Temporal changes in covariance, variance, and susceptibility ratios due to feedback at θ=1** (i.e. μ=0.861, σ=0.61). Theory(solid) and simulation statistics (circles) at g=-2 (blue) and g=1 (red) are shown here relative to results for the baseline g=0.

3.4 ANALYSIS OF RESULTS

Figure 9 clearly shows correlation modulation due to self-coupling in both theory and simulation. We see that positive feedback increases susceptibility relative to baseline, while

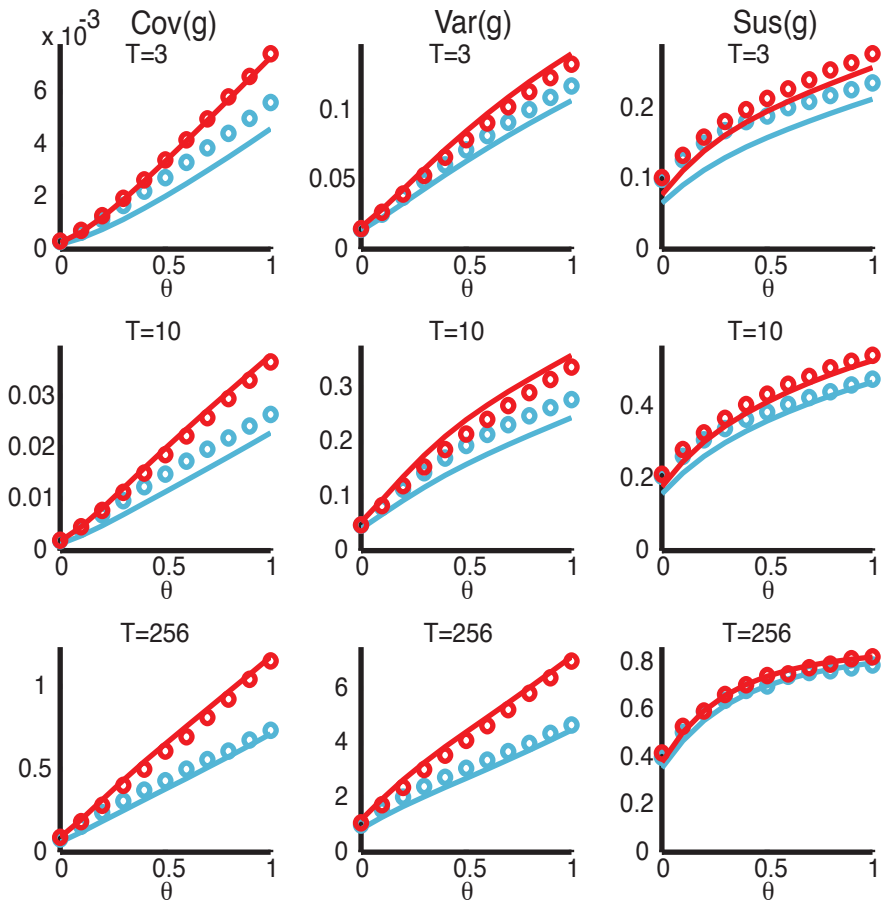

Figure 10: **Susceptibility changes due to feedback over our full range of θ, for short (T=3 ms), medium (T=10 ms), and very large time windows (T=256 ms).** Theory (solid) and simulation (circles) are shown at g=-2 (blue) and g=1 (red).

negative feedback decreases susceptibility. Furthermore, in both cases this effect is most pronounced for small values of the time window T, and these effects taper off as T is increased. Decomposing susceptibility into its constituent variance and covariance terms, we see that for very small T the spike count covariance is slightly more affected by g than the variance term is. In order to gain a better understanding of the cause of this effect, we analyze how variance, covariance, and susceptibilities depend on g and T. In particular, we pose the question of whether these effects can be attributed *solely* to changes in firing rate arising from feedback, and answer this question with the perturbative argument presented

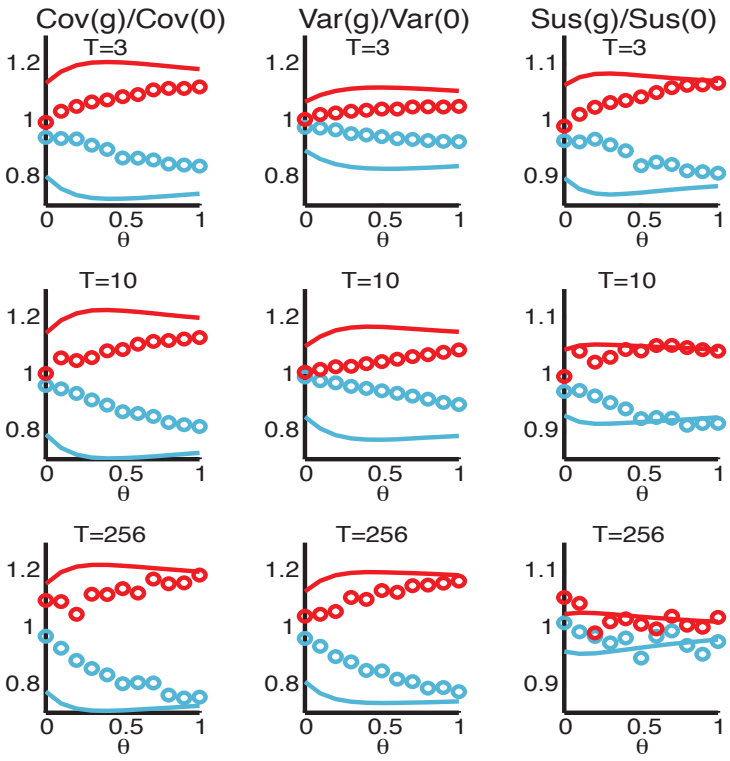

Figure 11: **Susceptibility *ratio* changes due to feedback over our full range of θ, for short (T=3 ms), medium (T=10 ms), and very large time windows (T=256 ms).** Theory (solid) and simulation (circles) at g=-2 (blue) and g=1 (red) are shown relative to baseline g=0.

below. First, recall the definition of suceptibility (where we omit the dependence on θ for clarity):

$$S_T(g) = \frac{Cov(g)}{Var(g)} = \frac{\int_{-\infty}^{\infty} S_{12}(\omega, g) K_T(\omega) d\omega}{\int_{-\infty}^{\infty} S_1(\omega, g) K_T(\omega) d\omega} \tag{3.22}$$

The ratio of susceptibilities plotted in figure 9 can then be defined as $R_T(g)$:

$$R_T(g) = \frac{S_T(g)}{S_T(0)} = \frac{Cov(g)Var(0)}{Var(g)Cov(0)} \tag{3.23}$$

In the limit of small T and g, we can obtain a very simple approximation to the integral for $Var(g)$. First, we replace $K_T(\omega)$ with its first-order Taylor expansion about $T = 0$:

$$K_T(\omega) = T + O(T^2) \tag{3.24}$$

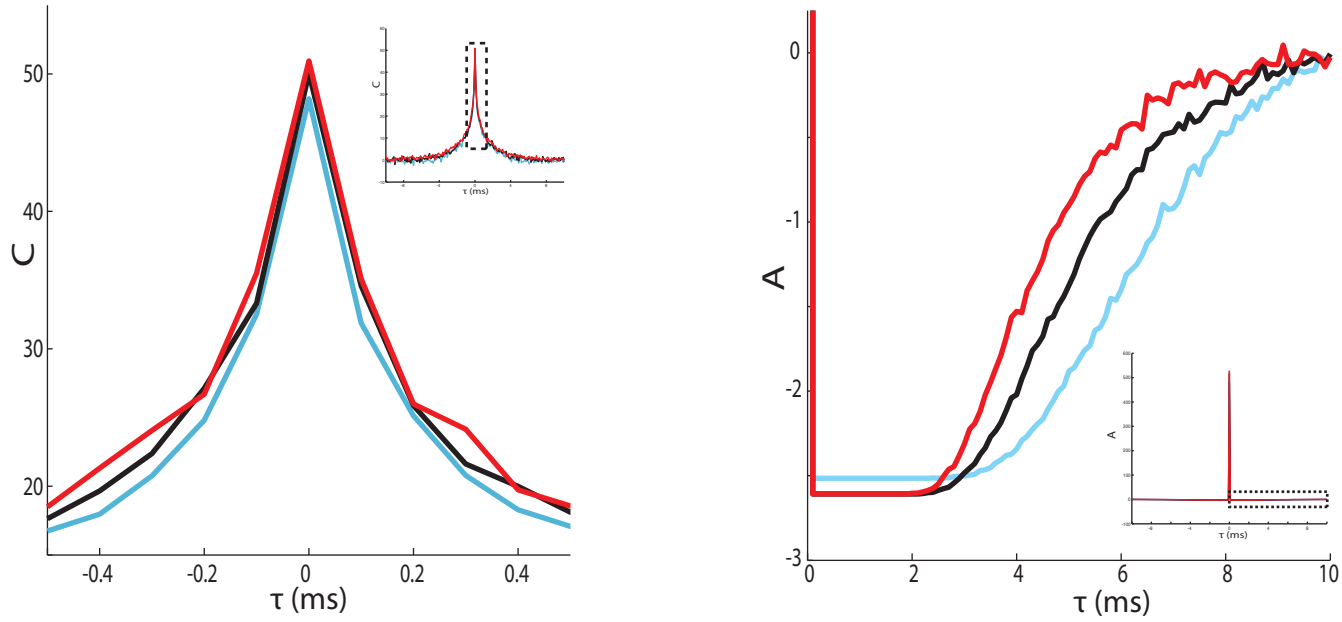

Figure 12: **Cross-correlation and autocorrelation for θ=1 (i.e. μ=0.861, σ=0.61), where firing rate matching has been performed with μ.** These statistics are computed from simulations, and are shown for g=-2 (blue), g=0 (black), and g=1 (red).

Furthermore, we know the power spectrum S_1 rapidly converges to the firing rate ν as ω grows large [23], so our integral over infinite frequencies will be completely dominated by this ν dependence. For T very small, we can approximate the integral up to a finite frequency $\omega^* = \frac{2\pi}{T}$:

$$\lim_{T\to 0} Var(g) = \lim_{T\to 0} \int_0^\infty S_1(\omega, g) K_T(\omega) d\omega \approx T\nu \int_0^{\omega^*} d\omega = 2\pi\nu \tag{3.25}$$

We conclude that for T near zero, g acts on Var only through its effect on the firing rate ν. With this knowledge, we can eliminate the need to consider Var terms if we control for g's effect on firing rate. To do so, we can redefine $\mu(\theta)$ and $\sigma(\theta)$ to also become functions of g, defined to compensate for the effects of g so that a constant firing rate is maintained at all values of g. It is possible to consider various definitions of $\mu(g)$ and $\sigma(g)$ that enforce this matching, some of which are more amenable to analysis than others. Before introducing

specific definitions for $\mu(g)$ and $\sigma(g)$, we proceed with our simplification of R_T, which is independent of the matching scheme chosen. When firing rates have been matched, result 3.25 tells us $Var(0) \approx Var(g)$ for small T, and since $K_T(\omega) \approx T$ for small T, the ratio can be simplified to

$$R_T(g) = \frac{Cov(g)}{Cov(0)} = \frac{\int_{-\infty}^{\infty} S_{12}(\omega, g) d\omega}{\int_{-\infty}^{\infty} S_{12}(\omega, 0) d\omega} \tag{3.26}$$

Before proceeding, we clarify some of the notation that will be used. We remark that the transfer function A under firing-rate matching has explicit dependencies $A(\omega, \mu_{eff}(\mu(\theta, g), g), \sigma(\theta, g))$. As above, we continue to suppress the θ dependence for simplicity, and we now suppress the g dependence as well, writing $A(\omega, \mu_{eff}, \sigma)$. Next, it is helpful to factor the g term out of $F(g, \omega)$ so that we can work with an $\hat{F}(\omega)$ that is independent of g, defined as:

$$F(g, \omega) = g\hat{F}(\omega) \tag{3.27}$$

Lastly, we introduce functions for the real and imaginary components of the complex functions A and $\hat{F}$:

$$A(\omega, \mu_{eff}, \sigma) = A_R(\omega, \mu_{eff}, \sigma) + iA_I(\omega, \mu_{eff}, \sigma) \tag{3.28}$$

$$\hat{F}(\omega) = \hat{F}_R(\omega) + i\hat{F}_I(\omega) \tag{3.29}$$

Since the formula for S_{12} derived in 3.13 is not readily conducive to analysis, we now rewrite its numerator N_{S12} and denominator D_{S12} as Taylor expansions about $g = 0$. Making use of the fact that $\mu_{eff}(\mu, 0) = \mu$, we obtain:

$$
\begin{aligned}
S_{12}(\omega, g) &= c\sigma^2 N_{S12}(\omega, g) D_{S12}(\omega, g) && (3.30)\\
N_{S12}(\omega, g) &= |A\left(\omega, \mu_{eff}(g), \sigma(g)\right)|^2 \\
&= A_R\left(\omega, \mu_{eff}(g), \sigma(g)\right)^2 + A_I\left(\omega, \mu_{eff}(g), \sigma(g)\right)^2 \\
&= |A(\omega, \mu, \sigma)|^2 + \\
&\quad 2gA_R\left(\omega, \mu, \sigma\right) \frac{d}{dg} A_R\left(\omega, \mu, \sigma\right) + \\
&\quad 2gA_I\left(\omega, \mu.\sigma\right) \frac{d}{dg} A_I\left(\omega, \mu, \sigma\right) + O\left(g^2\right) && (3.31)\\
D_{S12}(\omega, g) &= \frac{1}{|1 - gA\left(\omega, \mu_{eff}(g), \sigma(g)\right) \hat{F}(\omega)|^2} \\
&= \frac{1}{1 - 2gRe\left(A\left(\omega, \mu_{eff}(g), \sigma(g)\right) \hat{F}(\omega)\right) + O\left(g^2\right)} \\
&= 1 + 2gRe\left(A\left(\omega, \mu_{eff}(0), \sigma(0)\right) \hat{F}(\omega)\right) + O\left(g^2\right) \\
&= 1 + 2gRe\left(A(\omega, \mu, \sigma)\hat{F}(\omega)\right) + O\left(g^2\right) && (3.32)
\end{aligned}
$$

Expanding out the product of 3.30 and using $S_{12}(\omega, 0) = c\sigma^2 |A(\omega, \mu, \sigma)|^2$, to first order in g we have:

$$
\begin{aligned}
S_{12}(\omega, g) &= S_{12}(\omega, 0) + \\
&\quad c\sigma^2 \Big[2g|A(\omega, \mu, \sigma)|^2 Re\left(A(\omega, \mu, \sigma)\hat{F}(\omega)\right) + \\
&\quad 2gA_R(\omega, \mu, \sigma) \frac{d}{dg} A_R(\omega, \mu, \sigma) + \\
&\quad 2gA_I(\omega, \mu, \sigma) \frac{d}{dg} A_I(\omega, \mu, \sigma) \Big] && (3.33)
\end{aligned}
$$

For correlation susceptibility modulation beyond that expected from firing rate changes, we must show $R_T > 1$ for $g > 0$ and $R_T < 1$ for $g < 0$. By 3.26, this is equivalent to showing

$$
\int_{-\infty}^{\infty} S_{12}(\omega, g) - S_{12}(\omega, 0) d\omega > 0 \text{ for } g > 0 \quad (3.34)
$$

$$
\int_{-\infty}^{\infty} S_{12}(\omega, g) - S_{12}(\omega, 0) d\omega < 0 \text{ for } g < 0 \quad (3.35)
$$

Since the subtractive terms above eliminate the $S_{12}(\omega, 0)$ term in 3.33, the direction of the change in susceptibility is determined by the sign of

$$2c\sigma^2 g \int_{-\infty}^{\infty} A_R(\omega,\mu,\sigma)\frac{d}{dg}A_R(\omega,\mu,\sigma)+A_I(\omega,\mu,\sigma)\frac{d}{dg}A_I(\omega,\mu,\sigma)+|A(\omega,\mu,\sigma)|^2 Re\left(A(\omega,\mu,\sigma)\hat{F}(\omega)\right)d\omega \tag{3.36}$$

Futhermore, since the signs of the target conditions 3.34 correspond to the signs of g in each case, it suffices to show that the following condition holds for all g:

$$\int_{-\infty}^{\infty} A_R(\omega,\mu,\sigma)\frac{d}{dg}A_R(\omega,\mu,\sigma)+A_I(\omega,\mu,\sigma)\frac{d}{dg}A_I(\omega,\mu,\sigma)+|A(\omega,\mu,\sigma)|^2 Re\left(A(\omega,\mu,\sigma)\hat{F}(\omega)\right)d\omega > 0 \tag{3.37}$$

Since the $\frac{dA}{dg}$ terms above will depend on $\mu(\theta, g)$ and $\sigma(\theta, g)$, we must now commit to a definition of our firing rate matching scheme to proceed. Below we match firing rates with μ alone, which turns out to be more convenient than matching with both μ and σ (a strategy which is explored in appendix B). We choose our scheme such that for $g = 0$, our standard definition holds for $\mu(\theta)$ and $\sigma(\theta)$. For $g \neq 0$, σ remains untouched, while we calculate an offsetting $\mu(g)$ such that the effective current $\mu_{eff}(g)$ matches the $g = 0$ current. This is easily arranged, as shown below. Our definition of firing rate matching means the following condition must be met:

$$\nu\left(\mu(0),\sigma\right) - \nu\left(\mu_{eff}(g),\sigma\right) = 0 \tag{3.38}$$

Equation 3.38 can only be satisfied when $\mu_{eff}(g) = \mu(0)$. We can make use of this fact and the definition of the feedback-adjusted current 2.13 to determine μ for nonzero g:

$$\begin{aligned}
\mu(0) = \mu_{eff}(g) &= \mu(g) + \frac{g}{\tau_m}\nu\left(\mu_{eff}(g),\sigma\right) && (3.39)\\
&= \mu(g) + \frac{g}{\tau_m}\nu\left(\mu(0),\sigma\right) && (3.40)\\
\implies \mu(g) &= \mu(0) - \frac{g}{\tau_m}\nu\left(\mu(0),\sigma\right) && (3.41)
\end{aligned}$$

Differentiating both A_I and A_R by the chain rule and using $\frac{d\sigma}{dg} = 0$, we obtain for each A_j

$$\frac{dA_j\left(\omega, \mu_{eff}, \sigma\right)}{dg} = \frac{\partial A_j}{\partial \mu_{eff}} \frac{d\mu_{eff}}{\partial g} \tag{3.42}$$

$$\frac{d\mu_{eff}}{dg} = \frac{d\mu}{dg} + \frac{1}{\tau_m}\nu(\mu(0), \sigma) + \frac{g}{\tau_m}\frac{d\nu}{dg} \tag{3.43}$$

$$\frac{d\mu}{dg} = -\frac{\nu(\mu(0), \sigma)}{\tau_m} \tag{3.44}$$

Substituting 3.44 into 3.43 and noting that firing rate matching by definition entails $\frac{d\nu}{dg} = 0$, we have:

$$\frac{\partial \mu_{eff}}{\partial g} = 0 \tag{3.45}$$

so by 3.42 we see

$$\frac{dA_R}{dg} = \frac{dA_I}{dg} = 0 \tag{3.46}$$

With this result, the required condition 3.37 reduces to

$$\int_{-\infty}^{\infty} |A(\omega, \mu, \sigma)|^2 Re\left(A(\omega, \mu, \sigma)\hat{F}(\omega)\right) d\omega > 0 \tag{3.47}$$

Our analysis thus shows that correlation modulation can be achieved without changing firing rates when inequality 3.47 is satisfied. The direction of the modulation follows the sign of g, and the magnitude of the effect is determined by the Fourier transforms of both the transfer function A and the feedback kernel F. Unlike firing rate, which is driven by the static mean statistics (μ and σ), A and F are functions of frequency. This suggests that correlation modulation is sensitive to the time-varying aspects of the noise, stimulus, and the time scale of the synapse itself. Indeed, this expression implies that correlation modulation is very dependent on the feedback's time scale. Inspecting the definition of F (2.15), we see that F increases as the feedback decay rate $\alpha = \frac{1}{\tau_s}$ increases. Conversely, as α shrinks to zero, F vanishes. From this, we draw the lesson that feedback modulation of correlation will be most significant for synapses with fast rise and decay rates. Both linear response calculations and simulation corroborate these findings. Figure 13 shows susceptibility ratio changes as a function of T under firing rate matching. Here we have solved for the $\mu(g)$ values that

elicit 50 Hz firing rates for each value of g. While the shape of the result remains the same between matched and non-matched scenarios, we see that rate-matching has reduced the magnitude of the effect somewhat compared to the unmatched case. This demonstrates that although firing rates changes are not strictly necessary to cause correlation modulation, they can amplify it. As expected, figure 14 shows correlation modulation being greatly attenuated for large τ_s. Here we see modulation at $\alpha = 0.1$ (i.e. $\tau_s = 10$), as compared to our previous plots at $\alpha = 2$ (i.e. $\tau_s = 0.5$). Plotting our predictions over a range of feedback time-scales (figure 15) puts this into perspective, illustrating the spectrum of correlation changes that can be achieved. Figure 16 helps visualize the behavior of the condition we have derived for modulation. We see that this quantity is positive as required, and decreases as τ_s increases. This figure also shows that as we increase the correlation time-window T, moving further from the underlying assumptions of our derivation, the decay of this quantity is becoming linearized. We caution that these results were derived for neurons in excitable regimes; it is not necessarily true that they generalize to oscillatory regimes.

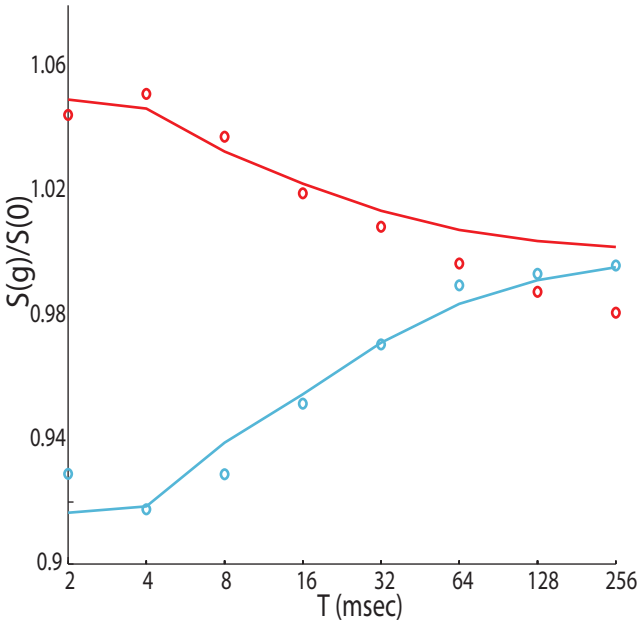

Figure 13: **Changes in susceptibility ratios with firing rate matching for θ=1**(i.e. μ=0.861, σ=0.61). The baseline θ=1 produces 50Hz firing for g=0. The g=-2 (blue) and g=1 (red) show susceptibilities relative to susceptibility at g=0.

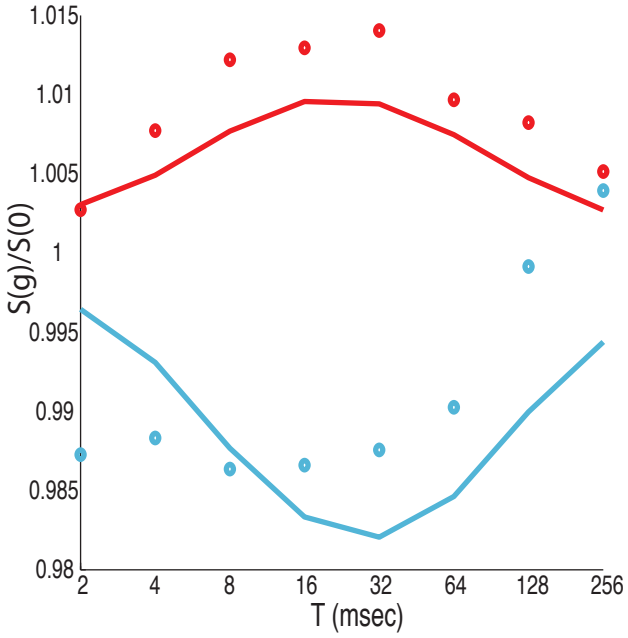

Figure 14: **Changes in susceptibility ratios with μ-based firing rate matching for θ=1 (i.e. μ=0.861, σ=0.61) at α=0.1** ($\tau_s = 10$).

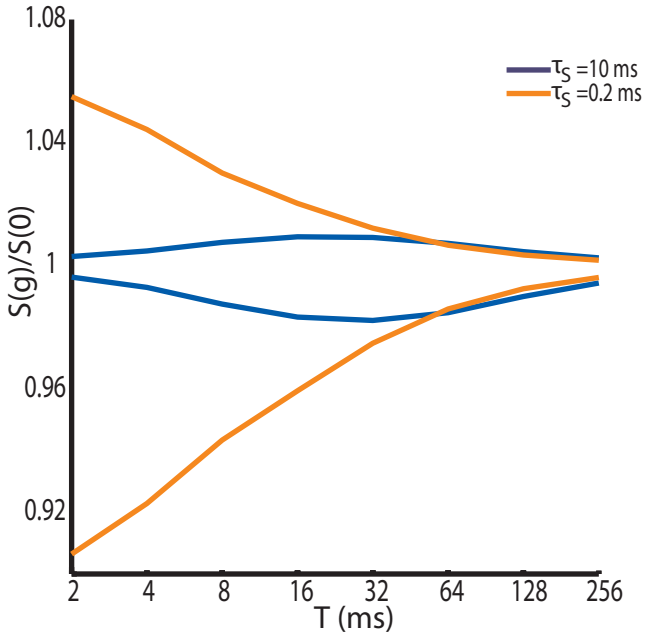

Figure 15: **Range of correlation modulation as a function of τ_s.** Here we use μ-based firing rate matching for θ=1 (i.e. μ=0.861, σ=0.61).

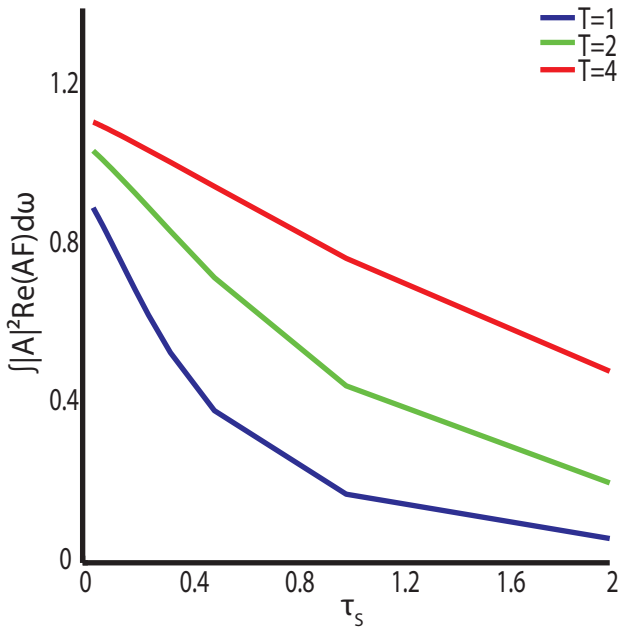

Figure 16: **Extinction of theoretical condition for correlation modulation as feedback time scale slows.**

4.0 CONCLUSIONS

We have shown that spike driven feedback influences both neural gain and correlation in self-coupled neurons. The profiles of these effects are markedly different, however. Figure 6 shows a maximum change in dynamic gain of roughly $\pm$ 10%, fairly uniformly over all time windows. In contrast, figure 9 illustrates marked temporal dependence in the modulation of correlation while producing maximum effects of comparable magnitude. Other studies [11] have shown that correlation increases with firing rate, and provided an analytical treatment of this effect the limit of large T. It is therefore natural to ask if the correlation changes we observe are solely due to changes in firing rate induced by feedback. By fixing firing rates in the presence of feedback, we have shown that firing rate changes are not the sole cause of this effect.

These findings have strong consequences for efforts to manipulate gain and correlation by altering firing rate. We see that these properties cannot be modulated completely independently - experimental attempts to modulate one will affect the other. Furthermore, the time scales associated with these effects imply constraints on how simultaneous modulatory targets can be achieved. Gain is modulated by feedback fairly uniformly over all time scales, while correlation shaping depends strongly on the time scale of the feedback. Modulation of gain with minimal effect on correlation should be possible for neurons with relatively slow intrinsic feedback. This goal, however, would prove difficult to for neurons with fast intrinsic feedback, requiring manipulation of some other more orthogonal mechanism. Conversely, any attempt to modulate correlation entails changes in gain, regardless of the time scale considered.

Our results can be connected to previously reported experimental results, and offer some predictions for future work. In the presence of positive feedback we predict gain will increase

regardless of the time-scale τ_s of the feedback. This prediction is consonant with the gain increases due to ADP's reported in [28] and shown in panels C and D of figure 1. Likewise, our theory predicts that AHP's will decrease gain, which concurs with the observations for AHP's in [18], shown in panels A and B of figure 1. Furthermore, our theory would predict that in both cases, the observed gain modulation would persist over any chosen time window T. Unlike gain, we predict that correlation changes are dependent on the time-scale of the feedback. Previous modeling work done with this ADP data [28, 13] report $\alpha = 0.24$, corresponding to $\tau_s = 4.16$. This τ_s is relatively large compared to the values that we see maximize synchrony changes in our theory; we would predict a very mild increase in synchrony for this experiment. The AHP study discussed above uses a dynamic clamp setup to emulate an AHP, where the decay of the feedback is 1 second. For such slow a feedback time scale, our theory would predict virtually no correlation modulation should be observed.

Our theory predicts that feedback alters correlations primarily over short time scales, which should be relevant to bursting dynamics. Prior studies in electric fish [31] have explored the role of ADP's in as a mechanism for bursting, and proposed the viability of bursting as a coding scheme for sensory stimuli. These experiments were also modeled using an LIF setup very similar to ours for single neurons [13]. As expected, our work reproduces the finding that bursting is increased by ADP's in paired neurons. Additionally, our results suggest that due to their limited time-scale of effectiveness in altering correlation, ADP's are in fact well-suited to controlling bursting by changing short-time scale correlation without altering longer time window correlations. This further supports the notion that self-feedback could serve as an important mechanism for implementing coding schemes based on bursting. Recent in vivo experiments in electric fish [7] have studied bursting changes in populations of neurons under both narrow and wide spatial field stimuli. They report that narrow field stimulation of neurons produces increases in bursting and overall correlations, consistent with our predictions. They also report that wide field stimulation increases the occurrence of near-synchronous spikes as we would predict, but they find that overall correlation is actually decreased. Our theory as it stands will not replicate this finding without invoking changes in g or c. Alternative theories using conductance-based models have been proposed to explain these in vivo results [22]. We propose that the reported changes are due to the complex

interplay of mechanisms which are present in vivo, and predict that such effects would not be observed in vitro. Our results may also have implications for constraints on the design of neural codes - other researchers have presented theoretical arguments for the significance of correlation with respect to population coding. In particular, it has been pointed out that noise in the brain has its own correlation structure and this may limit how correlated firing can implement information coding and decoding across populations of neurons [3]. We acknowledge that our study is confined to self-coupled neurons, which do not represent the most common form of spiking feedback in the brain. However, such architectures do arise in nature, for example in autaptic neurons: neurons which form synapses with themselves. Such synapses have been found to exist in many parts of the brain, including the neocortex and striatum [4].

Lastly, we note the potential utility of this result for understanding gain and correlation modulation in larger, cross-coupled networks. When studying such networks, it is difficult to distinguish whether changes in outputs are evoked purely by the presence of spike-driven feedback (regardless of its source), or whether they arise only from interactions between distinct neurons. Self-coupled neurons provide a reference point which separates the intrinsic effects of spiking feedback from network effects, and serve as a logical bridge between studies of uncoupled neurons and studies of networked populations [26]. Careful comparison of changes seen at each step in the progression from uncoupled neurons to self-coupled neurons to cross-coupled neurons (as depicted in figure 17) may clarify our understanding of the driving factors behind our results. Future work may explore this topic in more detail.

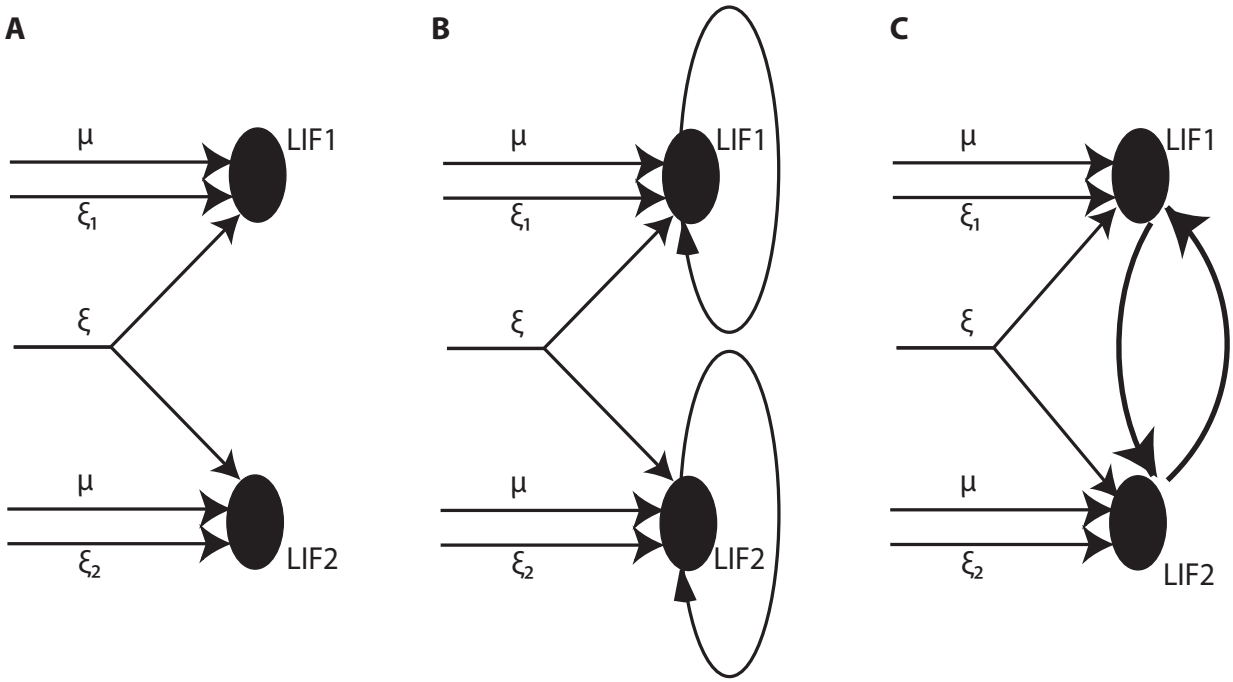

Figure 17: **Progression from (A) uncoupled neurons, to (B) self-coupled neurons, to (C) cross-coupled neurons.**

APPENDIX A

DERIVATION OF SPECTRA AND TRANSFER FUNCTION

The linear response theory used throughout this thesis relies heavily upon approximations to the transfer function $A(\omega)$, the stationary firing rate $\nu(\mu, \sigma)$, and the unperturbed power spectrum S_0. In the following sections, we sketch the main points of the derivations of these expressions as they are presented in [23].

A.0.1 POPULATION DENSITY AND FOKKER-PLANCK TREATMENT

The central idea underlying these derivations is the approach of modeling neural firing as a first-passage time problem. The behavior of our model neurons is dictated by evolution of voltages over time: current enters the neuron, raising voltage until a spike triggers a voltage reset and the cycle continues. Since our system is in the excitable regime, we require $\sigma > 0$ to ensure spiking events, and therefore we study the aggregate behavior of a large population of neurons which are subject to noise. Let $P(v, t)$ denote the probability that a neuron in this ensemble has voltage v at time t. The mean firing rate of the ensemble then corresponds to the flux of population voltages across firing threshold over a given time window. After a spike, the neuron's voltage will be reset to v_R, but first the neuron will enter a refractory state, which persists for duration τ. We let $P_T(t)$ denote the probability that a neuron is in the refractory state at time t, and note that each neuron either has voltage in $[-\infty, v_T]$, or it is currently in the refractory state, i.e:

$$\int_{-\infty}^{v_T} P(v, t)dv + P_T(t) = 1 \tag{A.1}$$

We assume that the voltage of each neuron evolves according to a fixed input current μ, with a leak v, and a noisy input ξ with intensity $D = \frac{\sigma^2}{2}$:

$$v'(t) = \mu - v + \sqrt{2D}\xi(t) \tag{A.2}$$

Since neurons stay in the refractory state for duration τ, the probability P_T that a neuron is refractory corresponds to the probability that the neuron's voltage was pushed above threshold over the last τ time units:

$$P_T(t) = -D \int_{t-\tau}^{t} \frac{\partial P(v, t')}{\partial v} dt' \Big|_{v=v_T} \tag{A.3}$$

Next, we note that a neuron that hits the threshold instantly transitions to the refractory state, so we have:

$$P(v_T, t) = 0 \tag{A.4}$$

Furthermore, once a neuron reaches v_T, it will spend τ time units in the refractory state, and will then immediately be set to v_R. The influx of neurons (J_+) with $v = v_R$ will match the outflux of neurons (J_-) with $v = v_T$ exactly τ time units ago:

$$J_+(t) = J_-(t - \tau) = J(v_T, t - \tau) \tag{A.5}$$

We can summarize this in the Fokker-Planck equation:

$$\partial_t P(v, t) = \partial_v(v - \mu + D\partial_v)P(v, t) + J_+(t)\delta(v - v_R) = -\partial_v J(v, t) \tag{A.6}$$

where we stipulate that $\lim_{v \to -\infty} P(v, t) = 0$ to avoid non-physical solutions. Making the change of variables $x = v - \mu$, we obtain the final form:

$$\partial_t P(x, t) = \partial_x(x + D\partial_x)P(x, t) = -\partial_x J(x, t) \tag{A.7}$$

A.0.2 STATIONARY FIRING RATE

This thesis makes frequent use of the stationary firing rate of neurons, $\nu(\mu, \sigma)$, which can be derived using the Fokker-Planck formulation described above. Stationary implies $\partial_t P(x,t) = 0$ and therefore a constant flux $J(v,t) = J_0$. Neglecting the trivial solution of zero current, we find the solution for finite currents to be

$$P_0(x) = c_1 e^{-x^2/(2D)} \int_x^{c_2} e^{y^2/(2D)} dy \tag{A.8}$$

Using the change of variable from v to x in the previous section, we denote $x_- = v_T - \mu$ as the threshold at which neurons are put into the refractory state. Likewise, we denote $x_+ = v_R - \mu$ as the reset point at which neurons resume evolution after their refractory periods. Since we require $P(x_-, t) = 0$, we see $c_2 = x_-$. Furthermore, differentiating gives $\frac{dP_0}{dx}(x_-) = c1$. At the threshold x_- this should match the outgoing current, $\frac{J_0}{D}$. However, since current flowing out at x_- reenters at x_+ after the refractory period, our expression holds only over $[x_+, x_-]$. For $x < x_+$ we can take $P_0(x) = c_3 e^{-x^2/(2D)}$. We know that the jumps in $\frac{\partial P_0}{\partial x}$ at x_- and x_+ must be equal, which yields

$$c_3 = \frac{J_0}{D} \int_{x_+}^{x_-} e^{y^2/(2D)} dy \tag{A.9}$$

Our solution has now taken the form:

$$P_0(x) = \frac{J_0}{D} e^{-x^2/(2D)} \int_x^{x_-} e^{y^2/(2D)} H(y - x_+) dy \tag{A.10}$$

$$= \frac{J_0}{D} e^{-x^2/(2D)} \begin{cases} \int_{x_+}^{x_-} e^{y^2/(2D)} dy & , x < x_+ \\ \int_x^{x_-} e^{y^2/(2D)} dy & , x_+ < x < x_- \end{cases} \tag{A.11}$$

We can compute the probability of being in the refractory state by evaluating A.3 and recalling $J(v_T, t) = J_0$, which yields $P_T = J_0 \tau$. Substituting back into A.1, subsequent renormalization and simplification yields the final formula for the stationary firing rate:

$$\nu = \left(\tau + \sqrt{\pi} \int_{\frac{-x_-}{\sqrt{2D}}}^{\frac{-x_+}{\sqrt{2D}}} e^{z^2} \operatorname{erfc}(z) dz \right)^{-1} \tag{A.12}$$

A.0.3 POWER SPECTRA

The Wiener-Khinchin theorem tells us that the power spectrum is the Fourier transform of the autocorrelation of a spike train $y(t)$. Furthermore, the spiking of an LIF neuron constitutes a renewal process with an interspike interval (ISI) density $\rho(t)$, so we can write:

$$S(\omega) = \int_{-\infty}^{\infty} \langle y(t)y(t+\tau)\rangle e^{i\omega\tau} d\tau \tag{A.13}$$

$$= \nu \frac{1-|\hat{\rho}(\omega)|^2}{|1-\hat{\rho}(\omega)|^2} + 2\pi\nu\delta(\omega) \tag{A.14}$$

Above we continue to denote the firing rate with ν, and $\hat{\rho}$ is the Fourier transform of the ISI density $\rho(t)$, which is seen to be:

$$\hat{\rho}(\omega) = \int_{-\infty}^{\infty} \rho(t) e^{-i\omega t} dt \tag{A.15}$$

$$= \int_{0}^{\infty} \rho(t) e^{-st} dt \tag{A.16}$$

$$= L[-i\omega, \rho(t)] \tag{A.17}$$

where we have substituted $s = i\omega$ when switching to the Laplace transform.

The quantity $L[-i\omega, \rho(t)] = L[-i\omega, \delta(t-\tau)]L[-i\omega, \rho_{FP}]$ requires the Laplace transform of the absolute refractory period τ and the Laplace transform of the probability density of our first passage time problem. The first transform is trivially found:

$$L[-i\omega, \delta(t-\tau)] = e^{i\omega t} \tag{A.18}$$

The transform of the first passage density is more involved; it is given by [17] as:

$$L[-i\omega, \rho_{FP}] = e^{\Delta} \frac{\mathcal{D}_{i\omega}\left(\frac{(\mu-v_R)\sqrt{2}}{\sigma}\right)}{\mathcal{D}_{i\omega}\left(\frac{(\mu-v_T)\sqrt{2}}{\sigma}\right)} \tag{A.19}$$

$$\Delta = \frac{v_R^2 - v_T^2 + 2\mu(v_T - v_R)}{4D} \tag{A.20}$$

where the $\mathcal{D}_{i\omega}$ terms denote the parabolic cylinder functions.

The result A.19 can then be substituted into A.14 to obtain the power spectrum of the unperturbed neuron:

$$S(\omega) = \nu \frac{\left|\mathcal{D}_{i\omega}\left(\frac{(\mu-v_T)\sqrt{2}}{\sigma}\right)\right|^2 - e^{2\Delta}\left|\mathcal{D}_{i\omega}\left(\frac{(\mu-v_R)\sqrt{2}}{\sigma}\right)\right|^2}{\left|\mathcal{D}_{i\omega}\left(\frac{(\mu-v_T)\sqrt{2}}{\sigma}\right) - e^{\Delta}e^{i\omega\tau}\mathcal{D}_{i\omega}\left(\frac{(\mu-v_R)\sqrt{2}}{\sigma}\right)\right|^2} + 2\pi\nu^2\delta(\omega) \tag{A.21}$$

A.0.4 TRANSFER FUNCTION

Noting that weak feedback to a neuron could be approximated by some Fourier series, we consider the simplest such approximation: a single cosine term. Making this approximation, the neuron's potential is given by:

$$V' = \frac{1}{\tau_m}\left(\mu + \epsilon\cos\omega t + \sigma\xi(t)\right) \tag{A.22}$$

We can rewrite this as a Langevin equation by making the change of variables:

$$x(t) = v - \mu + \epsilon(Ae^{-i\omega t}) \tag{A.23}$$

$$A = \frac{e^{-i\phi}}{2i\omega - 2} \tag{A.24}$$

and then differentiating to obtain:

$$x' = -x + \sigma\xi(t) \tag{A.25}$$

This result can now be expressed by the Fokker-Planck equation with boundary conditions as given below:

$$\partial_t P(x,t) = \partial_x(x + \frac{\sigma^2}{2}\partial_x)P(x,t) \tag{A.26}$$

$$P(x_t(t),t) = 0 \tag{A.27}$$

$$[P(x,t)]_{x_{r(t)}} = 0 \tag{A.28}$$

$$\left[\frac{\partial P(x,t)}{\partial x}\right]_{x_r(t)} = \frac{\partial P(x,t-\tau)}{\partial x}\Big|_{x_t(t-\tau)} \tag{A.29}$$

We can approximate this by

$$P(x,t) = P_0(x) + \epsilon(Ae^{-i\omega t}p(x))e^{(x_+^2 - x^2)/2\sigma^2} \tag{A.30}$$

Substituting this into the Fokker-Planck equation, we get an ODE:

$$\frac{\sigma^2}{2}p''(x) - \left(\frac{x^2}{2\sigma^2} - i\omega - \frac{1}{2}\right)p(x) = 0 \tag{A.31}$$

whose solution can be written in terms of the Kummer's function U as

$$\psi(x) = U(-i\omega - \frac{1}{2}, \frac{-x\sqrt{2}}{\sigma}) \tag{A.32}$$

Expanding the boundary conditions to linear terms yields:

$$p_- = \frac{\nu}{De^{\Delta}} \tag{A.33}$$

$$[p_+] = \frac{\nu}{D} \tag{A.34}$$

$$[p']_+ - e^{\Delta} e^{i\omega\tau} p'_- = \frac{\nu}{2D^2}\left(e^{i\omega\tau} x_- - x_+\right) \tag{A.35}$$

where $\Delta = \frac{\left(x_+^2 - x_-^2\right)}{4D}$

We obtain:

$$|p(x)| = \begin{cases} \left(\frac{\nu}{De^{\Delta}\psi_-} - \frac{\nu}{D\psi_+} + kY_+\right)\psi(x) \\ \frac{\nu}{De^{\Delta}}\frac{\psi(x)}{\psi_-} + kY(x)\psi_+ \end{cases} \tag{A.36}$$

with $Y(x) = \psi(x)\psi_2 - \psi - \psi_2(x)$.

Our final expression for the transfer function becomes:

$$A(\omega, \mu, D) = \frac{i\nu\omega\sqrt{(D)}}{i\omega - 1} \frac{\mathcal{D}_{i\omega-1}\left(\frac{\mu - V_T}{\sqrt{D}}\right) - e^{\Delta}\mathcal{D}_{i\omega-1}\left(\frac{\mu - V_R}{\sqrt{D}}\right)}{\mathcal{D}_{i\omega}\left(\frac{\mu - V_T}{\sqrt{D}}\right) - e^{\Delta i\omega\tau_R}\mathcal{D}_{i\omega}\left(\frac{\mu - V_R}{\sqrt{D}}\right)} \tag{A.37}$$

$$\Delta = \frac{V_R^2 - V_T^2 + 2\mu(V_T - V_R)}{4D} \tag{A.38}$$

where the $\mathcal{D}_{i\omega}$ terms denote the parabolic cylinder functions.

APPENDIX B

ANALYSIS OF FIRING RATE MATCHING WITH θ

In section 3.4, we first derived a general condition for understanding the correlation modulation caused by feedback in a firing-rate matched scenario. We then showed further simplifications that were possible when firing rate matching was achieved using only μ. Since we have linked μ and σ via the parameter θ throughout this work, it is natural to investigate the consequences of rate matching which manipulates both μ and σ through θ. This appendix shows the derivation for this scenario, which leads to a more cumbersome result.

This derivation follows the μ-matched scenario up through result 3.37. We then define a new firing rate matching scheme:

$$\mu(\theta, g) = \mu_0 + k_\mu \theta_g(\theta, g) \tag{B.1}$$

$$\sigma(\theta, g) = \sigma_0 + k_\sigma \theta_g(\theta, g) \tag{B.2}$$

For any given choice of θ and g, we can treat θ_g as a parameter using the original definition of μ and θ:

$$\mu(\theta_g) = \mu_0 + k_\mu \theta_g \tag{B.3}$$

$$\sigma(\theta_g) = \sigma_0 + k_\sigma \theta_g \tag{B.4}$$

Rate matching is achieved by solving for the θ_g that satisfies the root finding problem

$$\nu(\mu_{eff}(\mu(\theta_g)), \sigma(\theta_g)) - \nu(\mu(\theta, 0), \sigma(\theta, 0)) = 0 \tag{B.5}$$

$$\tag{B.6}$$

In this approach, θ_g acts as a "corrected" θ which balances out the effects of feedback on ν. Clearly, $\theta_g > \theta$ for negative g, while $\theta_g < \theta$ for positive g. Figure 18 illustrates continuous firing-rate matched curves in parameter space. Note that although all (θ, g) coordinates along a curve produce identical firing rates, their gains are not identical. Furthermore, changes in gain saturate as firing rate rises; at coarse resolution it is difficult to visually distinguish between the gain curves above roughly ν=30 Hz.

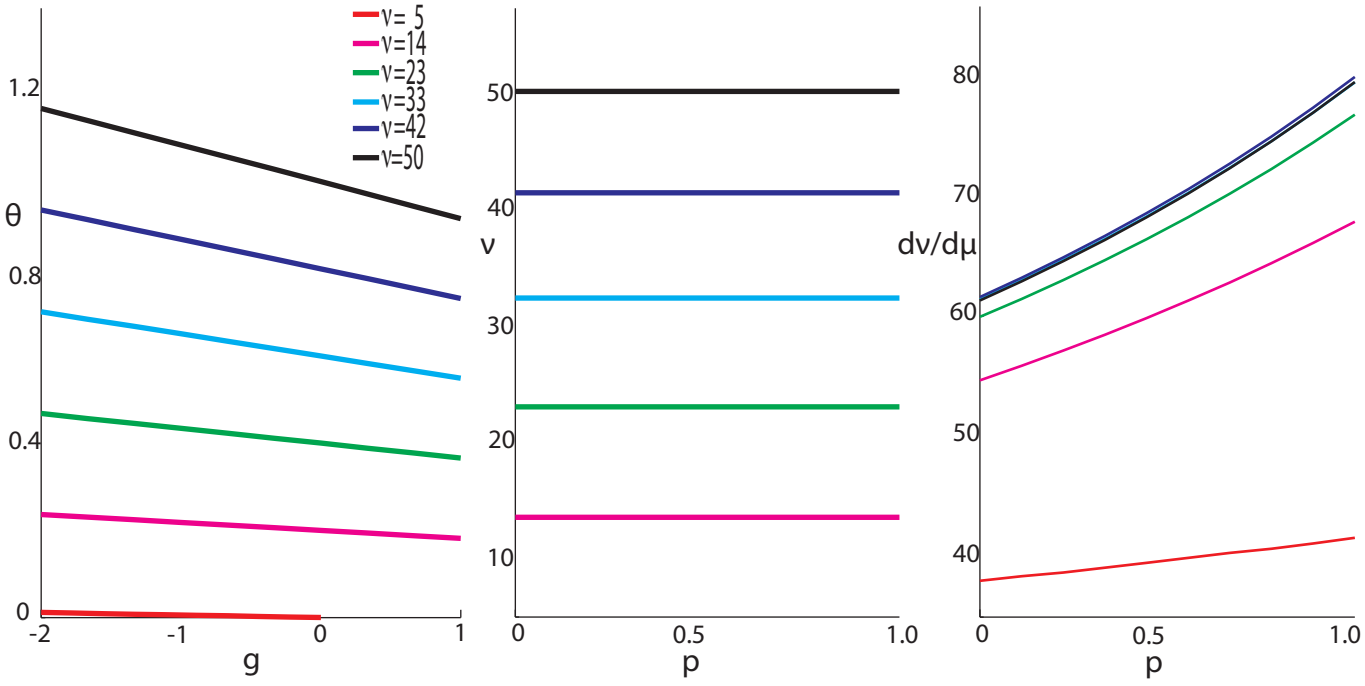

Figure 18: **Firing rate matching for θ=1.** We introduce a parameter "p" for traversing firing rate isolines in (θ, g) space. Panel (a) shows the curves in parameter space which produce identical firing rates. Panel (b) shows the firing rates corresponding to the isolines in panel (a). Panel (c) shows the gain at each point along the firing rate isolines.

As before, we verify that correlation modulation in this scenario is similar to that observed without rate matching. Figure 19 shows susceptibility ratio changes as a function of T in

a firing rate matched scenario. Here we have solved for the θ_g values that elicit 50 Hz firing rates for each value of g. Just as with our μ-based rate-matching scheme, figure 19 reproduces our original findings: correlation changes are maximized at low T, and taper off as T increases. Once again we also observe that the magnitude of the effect is diminished by using rate matching. This confirms that the correlation modulation is not simply a result of changes in firing rate: firing rate changes merely amplify the effect.

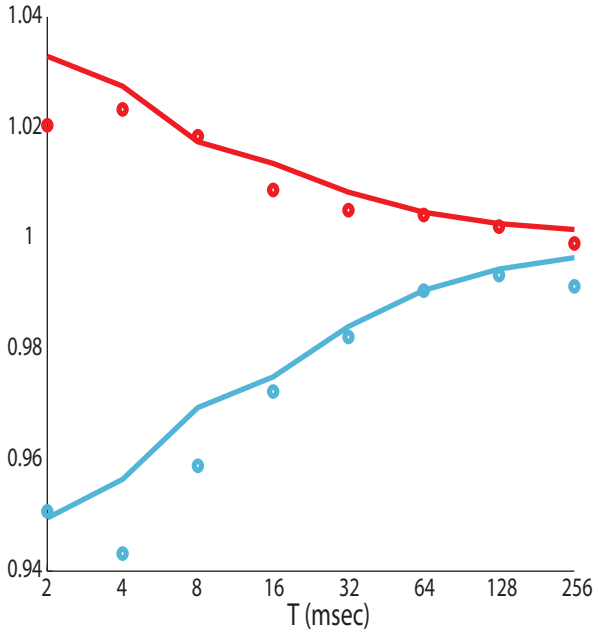

Figure 19: **Changes in susceptibility ratios with firing rate matching for θ=1.** The baseline θ=1 produces 50Hz firing for g=0. The g=-2 (blue) and g=1 (red) show susceptibilities relative to susceptibility at g=0. Firing rate matching is achieved with θ_g=1.168 (g=-2) and θ_g=.915 (g=1).

We now consider the form of the $\frac{dA}{dg}$ terms in condition 3.37. Firing rate matching introduces greater chain rule dependencies as both μ and σ now change with g. Applying the chain rule, and using the definition of μ_{eff} we have:

$$\frac{dA}{dg}(\omega, \mu_{eff}, \sigma) = \frac{\partial A}{\partial \mu_{eff}} \frac{d\mu_{eff}}{dg} + \frac{\partial A}{\partial \sigma} \frac{d\sigma}{dg} \tag{B.7}$$

From our definition of $\mu(\theta)$ and $\sigma(\theta)$ in 2.5, it is obvious that

$$\frac{d\sigma}{d\theta_g} = k_\sigma \tag{B.8}$$

$$\frac{d\mu}{d\theta_g} = k_\mu \tag{B.9}$$

Computing the relevant derivatives gives:

$$\frac{d\mu_{eff}}{dg}(\mu, g) = \frac{d\mu}{dg} + \nu(\mu_{eff}) + g\frac{d\nu}{\partial g} \tag{B.10}$$

$$= \left(\frac{\partial\mu}{\partial\theta_g}\frac{d\theta_g}{dg} + \frac{\partial\mu}{\partial g}\right) + \nu(\mu_e f f) \tag{B.11}$$

$$= k_\mu\frac{d\theta_g}{dg} + \nu(\mu_e f f) \tag{B.12}$$

$$\frac{d\sigma}{dg}(\mu, g) = \frac{\partial\sigma}{\partial\theta_g}\frac{d\theta_g}{dg} + \frac{\partial\sigma}{\partial g} \tag{B.13}$$

$$= k_\sigma\frac{d\theta_g}{dg} \tag{B.14}$$

Assembling the results for A_R and A_I, we have:

$$\frac{dA_j}{dg} = \frac{\partial A_j}{\partial g}\left(k_\mu\frac{d\theta_g}{dg} + \nu(\mu_e f f)\right) + \frac{\partial A_j}{\partial\sigma}\left(k_\sigma\frac{d\theta_g}{dg}\right) \tag{B.15}$$

Since $\frac{d\theta_g}{dg}$ is determined by the firing-rate matching requirement, we can compute it:

$$\frac{d\nu}{dg} = 0 = \frac{\partial\nu}{\partial\mu_{eff}}\left(\frac{\partial\mu_{eff}}{\partial\theta_g}\frac{d\theta_g}{dg} + \frac{\partial\mu_{eff}}{\partial g}\right) + \frac{\partial\nu}{\partial\sigma}\left(\frac{\partial\sigma}{\partial\theta_g}\frac{d\theta_g}{dg} + \frac{\partial\sigma}{\partial g}\right) \tag{B.16}$$

$$\implies \frac{d\theta_g}{dg} = \frac{-\nu(\mu_{eff})\frac{\partial\nu}{\partial\mu_{eff}}}{k_\mu\frac{\partial\nu}{\partial\mu_{eff}} + k_\sigma\frac{d\nu}{d\sigma}} \tag{B.17}$$

Unlike the simpler μ-based matching scheme we studied in the text, the $\frac{dA}{dg}$ terms do not vanish in this scheme, yielding a modulation condition which is much less illuminating. Recognizing that this result does not offer much insight, we do not pursue this analysis further, but we note again that figure 19 confirms the existence of our central result even with this more complicated rate-matching scheme.

BIBLIOGRAPHY

[1] Chronux software package. http://chronux.org/.

[2] B. B. Averbeck. Noise correlations and information encoding and decoding. In K. Josic, J. Rubin, M. Matias, and R. Romo, editors, *Coherent Behavior in Neuronal Networks*, chapter 11, pages 207–228. Springer New York, New York, NY, 2009.

[3] B. B. Averbeck, P. E. Latham, and A. Pouget. Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–366, May 2006.

[4] A. Bacci, J. R. Huguenard, and D. A. Prince. Functional autaptic neurotransmission in fast-spiking interneurons: A novel form of feedback inhibition in the neocortex. *J. Neurosci.*, 23(3):859–866, February 2003.

[5] M. D. Binder and R. K. Powers. Relationship between simulated common synaptic input and discharge synchrony in cat spinal motoneurons. *J Neurophysiol*, 86(5):2266–2275, November 2001.

[6] N. Brunel and V. Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural computation*, 11(7):1621–1671, October 1999.

[7] M. J. Chacron and J. Bastian. Population coding by electrosensory neurons. *J Neurophysiol*, 99(4):1825–1835, April 2008.

[8] M. R. Cohen and J. H. R. Maunsell. Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, 12(12):1594–1600, November 2009.

[9] B. W. Connors and M. J. Gutnick. Intrinsic firing patterns of diverse neocortical neurons. *Trends in Neurosciences*, 13(3):99 – 104, 1990.

[10] D. R. Cox and P. A. W. Lewis. *Statistical Analysis of Series of Events*. Monographs on Applied Probability & Statistics. Methuen young books, 1966.

[11] J. de La Rocha, B. Doiron, E. Shea-Brown, K. Josic, and A. Reyes. Correlation between neural spike trains increases with firing rate. *Nature*, 448(7155):802–806, August 2007.

[12] B. Doiron, B. Lindner, A. Longtin, L. Maler, and J. Bastian. Oscillatory activity in electrosensory neurons increases with the spatial correlation of the stochastic input stimulus. *Physical Review Letters*, 93(4):048101+, Jul 2004.

[13] B. Doiron, A.-M. M. Oswald, and L. Maler. Interval Coding. II. Dendrite-Dependent Mechanisms. *J Neurophysiol*, 97(4):2744–2757, 2007.

[14] B. Ermentrout. Linearization of f-i curves by adaptation. *Neural Computation*, 10(7):1721, Oct 1998.

[15] D. Ferster and K. D. Miller. Neural mechanisms of orientation selectivity in the visual cortex. *Annu Rev Neurosci*, 23:441–471, 2000.

[16] R. F. Galan, N. Fourcaud-Trocme, G. B. Ermentrout, and N. N. Urban. Correlation-induced synchronization of oscillations in olfactory bulb neurons. *J. Neurosci.*, 26(14):3646–3655, April 2006.

[17] B. Gluss. A model for neuron firing with exponential decay of potential resulting in diffusion equations for probability density. *Bulletin of Mathematical Biology*, 29(2):233–243, June 1967.

[18] M. H. Higgs, S. J. Slee, and W. J. Spain. Diversity of Gain Modulation by Noise in Neocortical Neurons: Regulation by the Slow Afterhyperpolarization Conductance. *J. Neurosci.*, 26(34):8787–8799, 2006.

[19] P. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations (Applications of Mathematics)*. Springer, October 1997.

[20] A. Kohn and M. A. Smith. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25(14):3661–3673, April 2005.

[21] A. Kohn, A. Zandvakili, and M. A. Smith. Correlations and brain states: from electrophysiology to functional imaging. *Current Opinion in Neurobiology*, 19(4):434–438, August 2009.

[22] A. L. Kumar, M. J. Chacron, and B. Doiron. Background synaptic activity modulates spike train correlation. Frontiers in Neuroscience, February 2010. Conference abstract: Computational and Systems Neuroscience doi: 10.3389/conf.fnins.2010.03.00114.

[23] B. Lindner. *Coherence and Stochastic Resonance In Nonlinear Dynamical Systems*. PhD thesis, Humboldt University, Berlin, 2001.

[24] B. Lindner, B. Doiron, and A. Longtin. Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. *Physical Review E*, 72(6):061919+, Dec 2005.

[25] C. Ly and B. Doiron. Divisive gain modulation with dynamic stimuli in integrate-and-fire neurons. *PLoS Comput Biol*, 5(4):e1000365+, April 2009.

[26] C. Ly and G. B. Ermentrout. Synchronization dynamics of two coupled neural oscillators receiving shared and unshared noisy stimuli. *Journal of Computational Neuroscience*, 26(3):425–443, November 2008.

[27] C. J. McAdams and J. H. Maunsell. Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron*, 23(4):765–773, August 1999.

[28] W. H. Mehaffey, B. Doiron, L. Maler, and R. W. Turner. Deterministic Multiplicative Gain Control with Active Dendrites. *J. Neurosci.*, 25(43):9968–9977, 2005.

[29] P. Mitra and H. Bokil. *Observed Brain Dynamics*. Oxford University Press, USA, 1 edition, December 2007.

[30] R. Moreno-Bote and N. Parga. Auto- and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses. *Physical Review Letters*, 96(2):028101+, Jan 2006.

[31] A. M. Oswald, B. Doiron, and L. Maler. Interval coding. i. burst interspike intervals as indicators of stimulus intensity. *Journal of Neurophysiology*, 97(4):2731–2743, April 2007.

[32] J. H. Reynolds and D. J. Heeger. The normalization model of attention. *Neuron*, 61(2):168–185, January 2009.

[33] H. Risken. *The Fokker-Planck Equation: Methods of Solutions and Applications*. Springer Series in Synergetics. Springer, second edition, September 1996.

[34] E. Salinas and P. Thier. Gain modulation: a major computational principle of the central nervous system. *Neuron*, 27(1):15–21, July 2000.

[35] E. Shea-Brown, K. Josić, Jaime, and B. Doiron. Correlation and synchrony transfer in integrate-and-fire neurons: Basic properties and consequences for coding. *Physical Review Letters*, 100(10):108102+, 2008.

[36] P. N. Steinmetz, A. Roy, P. J. Fitzgerald, S. S. Hsiao, K. O. Johnson, and E. Niebur. Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature*, 404(6774):187–190, March 2000.

[37] C. Sutherland, B. Doiron, and A. Longtin. Feedback-induced gain control in stochastic spiking networks. *Biological cybernetics*, 100(6):475–489, June 2009.

[38] G. Svirskis and J. Hounsgaard. Influence of membrane properties on spike synchronization in neurons: theory and experiments. *Network: Computation in Neural Systems (Bristol, England)*, 14(4):747–763, November 2003.

[39] Y. Trotter and S. Celebrini. Gaze direction controls response gain in primary visual-cortex neurons. *Nature*, 398(6724):239–242, March 1999.

[40] H. C. Tuckwell. *Introduction to Theoretical Neurobiology: Volume 1, Linear Cable Theory and Dendritic Structure (Cambridge Studies in Mathematical Biology)*. Cambridge University Press, April 1988.

[41] D. E. Winkowski and E. I. Knudsen. Top-down gain control of the auditory space map by gaze control circuitry in the barn owl. *Nature*, 439(7074):336–339, 2006.