

# Characterizing variability in nonlinear, recurrent neuronal networks

Guillaume Hennequin<sup>@1</sup> and Máté Lengyel<sup>1</sup>

<sup>1</sup>Computational and Biological Learning Lab, Department of Engineering, University of  
Cambridge, Cambridge, UK

<sup>@</sup> g.hennequin@eng.cam.ac.uk

October 12, 2016

## Abstract

In this note, we develop semi-analytical techniques to obtain the full correlational structure of a stochastic network of nonlinear neurons described by rate variables. Under the assumption that pairs of membrane potentials are jointly Gaussian – which they tend to be in large networks – we obtain deterministic equations for the temporal evolution of the mean firing rates and the noise covariance matrix that can be solved straightforwardly given the network connectivity. We also obtain spike count statistics such as Fano factors and pairwise correlations, assuming doubly-stochastic action potential firing. Importantly, our theory does not require fluctuations to be small, and works for several biologically motivated, convex single-neuron nonlinearities.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Theoretical results</b>	<b>2</b>
2.1	Model setup . . . . .	3
2.2	Temporal evolution of the moments of the voltage . . . . .	3
2.3	Calculation of nonlinear moments . . . . .	5
2.4	Extension to temporally correlated input noise . . . . .	7
2.5	Summary of equations . . . . .	7
2.6	Firing rate correlations . . . . .	8
2.7	Spike count statistics . . . . .	11
<b>3</b>	<b>Numerical validation</b>	<b>13</b>
3.1	Weakly connected random network . . . . .	13
3.2	Strongly connected, inhibition-stabilized random network . . . . .	15

# 1 Introduction

In this technical note, we develop a novel theoretical framework for characterising across-trial (and temporal) variability in stochastic, nonlinear recurrent neuronal networks with rate-based dynamics. We consider networks in which momentary firing rates  $\mathbf{r}$  are given by a nonlinear function of the underlying (coarse-grained) membrane potentials  $\mathbf{u}$ . In particular, we treat the case of the threshold-power-law nonlinearity  $r = k[u]_+^n$  (where  $n$  is an arbitrary positive integer), which has been shown to approximate the input-output function of real cortical neurons (with  $n$  ranging from 1 to 5; [Priebe et al., 2004](#); [Miller and Troyer, 2002](#)). This model is of particular interest as it captures many of the nonlinearities observed in the trial-averaged responses of primary visual cortex (V1) neurons to visual stimuli ([Ahmadian et al., 2013](#); [Rubin et al., 2015](#)), as well as the stimulus-induced suppression of (co-)variability observed in many cortical areas ([Hennequin et al., submitted](#)).

We derive assumed density filtering equations that describe the temporal evolution of the first two moments of the joint probability distribution of membrane potentials and firing rates in the network. These equations are based on the assumption that pairs of membrane potentials are jointly Gaussian at all times, which tends to hold in large networks. This approach allows us to solve the temporal evolution of the mean firing rates and the across-trial noise covariance matrix, given the model parameters (network connectivity, feedforward input to the network, and statistics of the input noise). We also obtain the full firing rate cross-correlogram for any pair of neurons, which allows us to compute both Fano factors and pairwise spike count correlations in arbitrary time windows, assuming doubly stochastic (inhomogeneous Poisson) spike emission. Importantly, our theory does not require fluctuations to be small, and is therefore applicable to physiologically relevant regimes where spike count variability is well above Poisson variability (implying large fluctuations of  $\mathbf{u}$ , which is therefore often found below the threshold of the nonlinearity).

This note is structured as follows. We first provide all the derivations, together with a summary of equations and some practical details on implementation. We then demonstrate the accuracy of our approach on two different networks of excitatory and inhibitory neurons: a random, weakly connected network, and a random, strongly connected but inhibition-stabilized network. In the latter case, the stochastic dynamics are dominated by balanced (nonnormal) amplification, leading to the emergence of strong correlations between neurons ([Murphy and Miller, 2009](#); [Hennequin et al., 2014](#)) which could in principle make the membrane potential distribution non-Gaussian – the only condition that could break the accuracy of our theory. Even in this regime, we obtain good estimates of the first and second-order moments of the joint distributions of both  $\mathbf{u}$  and  $\mathbf{r}$ .

## 2 Theoretical results

**Notations** We use  $\cdot^T$  to denote the vector/matrix transpose, bold letters to denote column vectors ( $\mathbf{v} \equiv (v_1, \dots, v_N)^T$ ) and bold capitals for matrices (e.g.  $\mathbf{W} = \{W_{ij}\}$ ). The notation  $\text{diag}(\mathbf{v})$  is used for the diagonal matrix with entries  $v_1, v_2, \dots, v_N$  along the diagonal.

## 2.1 Model setup

We consider a network of  $N$  interconnected neurons, whose continuous-time, stochastic and nonlinear dynamics follow

$$du_i = \frac{dt}{\tau_i} \left( -u_i(t) + h_i(t) + \sum_j W_{ij} r_j(t) \right) + d\chi_i \quad (1)$$

where the momentary firing rate  $r_j(t)$  is given by a positive, nonlinear function  $f$  of  $u_j(t)$ :

$$r_j(t) = f[u_j(t)] \quad (2)$$

The first few derivations below hold for arbitrary nonlinearity  $f$ ; we will commit to the threshold-powerlaw nonlinearity  $f(u) \propto [u]_+^n$  only later. In Equation 1,  $\tau_i$  is the intrinsic (“membrane”) time constant of unit  $i$ ,  $h_i(t)$  denotes a deterministic and perhaps time-varying feedforward input,  $W_{i\ell}$  is the synaptic weight from unit  $\ell$  onto unit  $i$ , and  $d\chi$  is a multivariate Wiener process with covariance  $\langle \chi_i(t) \chi_j(t+s) \rangle = \Sigma_{ij}^x \delta(s)$ . Note that the elements of  $\Sigma^x$  have units of  $(\text{mV})^2/s$ . We will later extend Equation 1 to the case of temporally correlated input noise.

Most of our results can be derived from the finite-difference version of Equation 1, given for very small  $\epsilon$  by

$$u_i(t) - u_i(t - \epsilon) = \frac{\epsilon}{\tau_i} \left[ -u_i(t - \epsilon) + h_i(t) + \sum_{\ell} W_{i\ell} r_{\ell}(t - \epsilon) \right] + \sqrt{\epsilon} \chi_i(t) \quad (3)$$

Here  $\chi(t)$  is drawn from a multivariate normal distribution with covariance matrix  $\Sigma^x$ , independently of the state of the system in the previous time step  $(t - \epsilon)$ .

We now show how to derive deterministic equations of motion for the first and second-order moments of  $\mathbf{u}$ , then explain how these can be made self-consistent using the Gaussian assumption mentioned in the introduction. Solving these equations is straightforward, and returns the mean membrane potential of each neuron as well as the full matrix of pairwise covariances. We also show how the moments of  $\mathbf{r}$  can be derived from those of  $\mathbf{u}$ , and how Fano factors and spike count correlations can then be obtained from those.

## 2.2 Temporal evolution of the moments of the voltage

We are interested in the moments of the membrane potential variables, defined as

$$\mu_i(t) \equiv \langle u_i(t) \rangle \quad (4a)$$

$$\Sigma_{ij}(t, s) \equiv \langle \tilde{u}_i(t) \tilde{u}_j(t + s) \rangle \quad (4b)$$

where  $\langle \cdot(t) \rangle$  denotes ensemble averaging (or “trial-averaging”, i.e. an average over all possible realisations of the noise processes  $\chi(t')$  for  $0 \leq t' \leq t$ ), and with the notation  $\tilde{z}(t) \equiv z(t) - \langle z(t) \rangle$ . Note that  $\langle \cdot(t) \rangle$  will coincide with temporal averages in the stationary case when the dynamics has a single fixed point (ergodicity). Before we proceed, let us introduce similar notations for other moments:

$$\nu_i(t) \equiv \langle r_i(t) \rangle \quad (5a)$$

$$\Gamma_{ij}(t, s) \equiv \langle \tilde{u}_i(t) \tilde{r}_j(t + s) \rangle \quad (5b)$$

$$\Lambda_{ij}(t, s) \equiv \langle \tilde{r}_i(t) \tilde{r}_j(t + s) \rangle \quad (5c)$$

Taking the ensemble average of Equation 3 and the limit  $\epsilon \rightarrow 0$  yields a differential equation for the membrane potential mean:

$$\tau_i \frac{d\mu_i}{dt} = -\mu_i(t) + h_i(t) + \sum_j W_{ij} \nu_j(t) \quad (6)$$

We now subtract its average from Equation 3:

$$\tilde{u}_i(t) - \tilde{u}_i(t - \epsilon) = \frac{\epsilon}{\tau_i} \left[ -\tilde{u}_i(t - \epsilon) + \sum_j W_{ij} \tilde{r}_j(t - \epsilon) \right] + \sqrt{\epsilon} \chi_i(t) \quad (7)$$

The equations of motion for the variances and covariances can be obtained in a number of ways. Here we adopt a simple approach based on the finite differences of Equation 7, observing that

$$\tilde{u}_i(t) \tilde{u}_j(t) - \tilde{u}_i(t - \epsilon) \tilde{u}_j(t - \epsilon) = \tilde{u}_i(t) [\tilde{u}_j(t) - \tilde{u}_j(t - \epsilon)] + \tilde{u}_j(t - \epsilon) [\tilde{u}_i(t) - \tilde{u}_i(t - \epsilon)] \quad (8)$$

Substituting Equation 7 into Equation 8, we obtain

$$\begin{aligned} \tilde{u}_i(t) \tilde{u}_j(t) - \tilde{u}_i(t - \epsilon) \tilde{u}_j(t - \epsilon) &= -\frac{\epsilon}{\tau_i} \tilde{u}_i(t - \epsilon) \tilde{u}_j(t - \epsilon) - \frac{\epsilon}{\tau_j} \tilde{u}_i(t) \tilde{u}_j(t - \epsilon) \\ &\quad + \epsilon \sum_\ell \tilde{r}_\ell(t - \epsilon) \left( \frac{W_{i\ell}}{\tau_i} \tilde{u}_j(t - \epsilon) + \frac{W_{j\ell}}{\tau_j} \tilde{u}_i(t) \right) \\ &\quad + \sqrt{\epsilon} (\chi_i(t) \tilde{u}_j(t) + \chi_j(t) \tilde{u}_i(t - \epsilon)) \end{aligned} \quad (9)$$

We now take ensemble expectations on both sides. The l.h.s. of Equation 9 averages to  $\Sigma_{ij}(t) - \Sigma_{ij}(t - \epsilon)$ . The term  $\chi_j(t) \tilde{u}_i(t - \epsilon)$  averages to zero for any  $(i, j)$  pair, because the input noise terms at time  $t$  are independent of the state of the system at time  $t - \epsilon$ . However, we expect the equal-time product  $\chi_i(t) \tilde{u}_j(t)$  to average to something small ( $\mathcal{O}(\sqrt{\epsilon})$ ) but non-zero. Let us compute it, by recalling that

$$\tilde{u}_j(t) = \tilde{u}_j(t - \epsilon) + \frac{\epsilon}{\tau_j} [\text{something at time } (t - \epsilon)] + \sqrt{\epsilon} \chi_j(t) \quad (10)$$

Multiplying both sides by  $\chi_i(t)$  and taking expectations, again all products of  $\chi_i(t)$  with quantities at time  $t - \epsilon$  vanish, and we are left with

$$\langle \tilde{u}_j(t) \chi_i(t) \rangle = \Sigma_{ij}^x \sqrt{\epsilon} \quad (11)$$

Thus, when averaged, Equation 9 becomes

$$\frac{\Sigma_{ij}(t) - \Sigma_{ij}(t - \epsilon)}{\epsilon} = -\frac{\Sigma_{ij}(t - \epsilon, 0)}{\tau_i} - \frac{\Sigma_{ij}(t, -\epsilon)}{\tau_j} + \sum_\ell \left( \frac{W_{i\ell}}{\tau_i} \Gamma_{j\ell}(t - \epsilon, 0) + \frac{W_{j\ell}}{\tau_j} \Gamma_{i\ell}(t, -\epsilon) \right) + \Sigma_{ij}^x \quad (12)$$

Now taking the limit of  $\epsilon \rightarrow 0$ , and by continuity of both  $\mathbf{u}$  and  $\mathbf{r}$ , we obtain the desired equation of motion of the zero-lag covariances:

$$\frac{d\Sigma_{ij}(t, 0)}{dt} = \Sigma_{ij}^x + \frac{1}{\tau_i} \left( -\Sigma_{ij}(t, 0) + \sum_\ell W_{i\ell} \Gamma_{j\ell}(t, 0) \right) + \frac{1}{\tau_j} \left( -\Sigma_{ij}(t, 0) + \sum_\ell W_{j\ell} \Gamma_{i\ell}(t, 0) \right) \quad (13)$$

In the special case of *constant* input  $h_i(t)$ , the covariance matrix of the steady-state, stationary distribution of potentials satisfies Equation 13 with the l.h.s. set to zero. This system of equations can be seen as a nonlinear extension of the classical Lyapunov equation for the multivariate (linear) Ornstein-Uhlenbeck process (Gardiner, 1985), in which  $W_{i\ell} \Sigma_{j\ell}$  would replace  $W_{i\ell} \Gamma_{j\ell}$

inside the sum. Such linear equations have been derived previously in the context of balanced networks of threshold binary units (Renart et al., 2010; Barrett, 2012; Dahmen et al., 2016).

We can also obtain the lagged cross-covariances by integrating the following differential equation – obtained using similar methods as above – over  $s > 0$ :

$$\frac{d\Sigma_{ij}(t, s)}{ds} = \frac{1}{\tau_j} \left( -\Sigma_{ij}(t, s) + \sum_{\ell} W_{j\ell} \Gamma_{i\ell}(t, s) \right) \quad (14)$$

Covariances for negative time lags are then obtained from those at positive lags, since:

$$\Sigma(t, -s) = \Sigma(t - s, s)^T. \quad (15)$$

In the stationary case, we have simply  $\Sigma(\infty, -s) = \Sigma(\infty, s)^T$ . Equation 14 can be integrated with initial condition  $\Sigma(t, 0)$  which is itself obtained by integrating Equations 6 and 13.

### 2.3 Calculation of nonlinear moments

To be able to integrate Equations 6, 13 and 14, we need to express the nonlinear moments  $\nu$  and  $\Gamma$  as a function of  $\mu$  and  $\Sigma$ . This is a moment closure problem, which in general cannot be solved exactly. Here, we will approximate  $\nu$  and  $\Gamma$  by making a Gaussian process assumption for  $\mathbf{u}$ : we assume that for any pair of neurons  $(i, j)$  and any pair of time points  $(t, t + s)$ , the potentials  $u_i(t)$  and  $u_j(t + s)$  are jointly Gaussian, i.e.

$$\forall(i, j), \forall(t, s) \quad \begin{pmatrix} u_i(t) \\ u_j(t + s) \end{pmatrix} \sim \mathcal{N} \left[ \begin{pmatrix} \mu_i(t) \\ \mu_j(t + s) \end{pmatrix}, \begin{pmatrix} \Sigma_{ii}(t, 0) & \Sigma_{ij}(t, s) \\ \Sigma_{ij}(t, s) & \Sigma_{jj}(t + s, 0) \end{pmatrix} \right] \quad (16)$$

In other words, we systematically and consistently ignore all moments of order 3 or higher. This is the strongest assumption we make here, but its validity can always be checked empirically by running stochastic simulations. For certain firing rate nonlinearities  $f$  (in particular, threshold-power law functions), the Gaussian process assumption will enable a direct and exact calculation of  $\nu$  and  $\Gamma$  given  $\mu$  and  $\Sigma$ , with no need to linearise the dynamics, as detailed below.

From now on, we will drop the time dependence from the notations to keep the derivations uncluttered, with the understanding that in using the results that follow to compute second-order moments such as  $\Gamma_{ij}(t, s)$  or  $\Lambda_{ij}(t, s)$ , the quantities  $\mu_i$  and variances  $\Sigma_{ii}$  that regard neuron  $i$  will have to be evaluated at time  $t$ , those that regard neuron  $j$  evaluated at time  $t + s$ , and any covariance  $\Sigma_{ij}$  will have to be understood as  $\Sigma_{ij}(t, s)$  as defined in Equation 4b.

Using the Gaussian assumption, mean firing rates become Gaussian integrals:

$$\nu_i = \int \mathcal{D}z \cdot f(\mu_i + z\sqrt{\Sigma_{ii}}) \quad (17)$$

where  $\mathcal{D}z$  denotes the standard Gaussian measure, and  $f(\cdot)$  is the firing rate nonlinearity (cf. Equation 2). Similarly,

$$\Gamma_{ij} = \iint du_i du_j \cdot \mathcal{N} \left[ \begin{pmatrix} u_i \\ u_j \end{pmatrix}; \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix}, \begin{pmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ij} & \Sigma_{jj} \end{pmatrix} \right] \cdot (u_i - \mu_i) \cdot f(u_j) \quad (18)$$

To calculate  $\Gamma_{ij}$ , we make use of the fact that the elliptical Gaussian distribution with correlation  $c_{ij} = \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$  in Equation 18 can be turned into a spherical Gaussian via a change of variable:

$$\Gamma_{ij} = \iint \mathcal{D}z \mathcal{D}z' \cdot \sqrt{\Sigma_{ii}} \left( z'c_{ij} + z\sqrt{1 - c_{ij}^2} \right) \cdot f(\mu_j + z'\sqrt{\Sigma_{jj}}) \quad (19)$$

Now, the integral over  $z$  can be performed inside the other one, and clearly vanishes. We are left with

$$\Gamma_{ij} = c_{ij} \sqrt{\Sigma_{ii}} \int \mathcal{D}u \cdot u \cdot f(\mu_j + u \sqrt{\Sigma_{jj}}) \quad (20)$$

Finally, integrating by part (assuming the relevant integrals exist), we obtain a simpler form:

$$\Gamma_{ij} = \Sigma_{ij} \gamma_j \quad (21)$$

with

$$\gamma_j \equiv \int \mathcal{D}u \cdot f'(\mu_j + u \sqrt{\Sigma_{jj}}) \quad (22)$$

The one-dimensional integrals in Equations 17 and 22 turn out to have closed-form solutions for a number of nonlinearities  $u \mapsto f(u)$ , including the exponential  $f(u) \propto \exp(u)$  (Buesing et al., 2012) and the class of threshold power-law nonlinearities of the form  $f(u) = k[u]_+^n$  for any integer exponent  $n \geq 1$ , which closely match the behavior of real cortical neurons under realistic noise conditions (Priebe et al., 2004; Miller and Troyer, 2002).

In the threshold-powerlaw case  $r = k[u]_+^n$ , integration by parts yields the following recursive formulas:

$$\nu_i^{(n)} = \begin{cases} k\mu_i \psi\left(\frac{\mu_i}{\sqrt{\Sigma_{ii}}}\right) + k\sqrt{\Sigma_{ii}} \phi\left(\frac{\mu_i}{\sqrt{\Sigma_{ii}}}\right) & \text{if } n = 1 \\ \mu_i \nu_i^{(1)} + k\Sigma_{ii} \psi\left(\frac{\mu_i}{\sqrt{\Sigma_{ii}}}\right) & \text{if } n = 2 \\ \mu_i \nu_i^{(n-1)} + (n-1)\Sigma_{ii} \nu_i^{(n-2)} & \text{otherwise} \end{cases} \quad (23)$$

$$\gamma_i^{(n)} = \begin{cases} k\psi\left(\frac{\mu_i}{\sqrt{\Sigma_{ii}}}\right) & \text{if } n = 1 \\ n \nu_i^{(n-1)} & \text{otherwise} \end{cases} \quad (24)$$

where  $\phi$  and  $\psi$  denote the standard Gaussian probability density function and its cumulative density function respectively.

Plugging the moment-conversion results of Equations 23 and 24 into the equations of motion (Equations 6, 13 and 14) yields a system of self-consistent, *deterministic* differential equations for the temporal evolution of the potential distribution (or its moments). These equations can be integrated straightforwardly given the initial moments at time  $t = 0$ .

If we are interested in the stationary distribution of  $\mathbf{u}(t)$  or  $\mathbf{r}(t)$ , i.e. in the case where  $h_i(t) = \text{constant}$ , we can start from any valid initial condition  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  with  $\boldsymbol{\Sigma} \succ 0$ , and let the integration of Equations 6 and 13 converge to a fixed point. In the ergodic case, this will indeed converge to a unique stationary distribution, independent of the initial conditions used to integrate the equations of motion. If there are several fixed points, this procedure yields the moments of the stationary distribution of  $\mathbf{u}$  conditioned on the initial conditions, i.e. will discover only one of the fixed points and return the moments of the fluctuations around it. Our approach cannot capture multistability explicitly, that is, it ignores the possibility that the network could change its set point with non-zero probability.

Finally, let us emphasise that we have never required that fluctuations be small. As long as the Gaussian assumption holds for the membrane potentials (Equation 16), we expect to obtain accurate solutions which is confirmed below in our numerical examples. This is because we have been able to express the moments of  $\mathbf{r}$  as a function of those of  $\mathbf{u}$  in closed form, *without approximation*.

## 2.4 Extension to temporally correlated input noise

So far, we have considered spatially correlated, but temporally white, external noise sources. We now extend our equations to the case of noise with spatiotemporal correlations, assuming space and time are separable, and temporal correlations in the input fall off exponentially with time constant  $\tau_\eta$ :

$$\tau_i \frac{du_i}{dt} = -u_i(t) + h_i(t) + \sum_{\ell} W_{i\ell} r_{\ell}(t) + \eta_i(t) \quad (25)$$

with  $\langle \eta_i(t) \rangle = 0$  and  $\langle \eta_i(t) \eta_j(t+s) \rangle \equiv \Sigma_{ij}^{\eta} e^{-|s|/\tau_\eta}$ . The equation for the mean voltages (Equation 6) does not change. For the voltage covariances, however, Equation 13 becomes

$$\begin{aligned} \frac{d\Sigma_{ij}(t, 0)}{dt} &= \frac{1}{\tau_i} \left( -\Sigma_{ij}(t, 0) + \sum_{\ell} W_{i\ell} \Gamma_{j\ell}(t, 0) + \Sigma_{ij}^*(t) \right) \\ &+ \frac{1}{\tau_j} \left( -\Sigma_{ij}(t, 0) + \sum_{\ell} W_{j\ell} \Gamma_{i\ell}(t, 0) + \Sigma_{ji}^*(t) \right) \end{aligned} \quad (26)$$

with the definition  $\Sigma_{ij}^*(t) \equiv \langle \eta_i(t) \tilde{u}_j(t) \rangle$ . These moments can also be obtained by simultaneously integrating the following:

$$\frac{d\Sigma_{ij}^*(t)}{dt} = -\frac{\Sigma_{ij}^*(t)}{\tau_\eta} + \frac{1}{\tau_j} \left( -\Sigma_{ij}^*(t) + \Sigma_{ij}^{\eta} + \sum_{\ell} W_{j\ell} \Gamma_{i\ell}^*(t) \right) \quad (27)$$

with an analogous definition  $\Gamma_{ij}^*(t) \equiv \langle \eta_i(t) \tilde{r}_j(t) \rangle = \Sigma_{ij}^*(t) \gamma_j(t)$ , which can be expressed self-consistently as a function of  $\mu_j(t)$ ,  $\Sigma_{jj}(t)$ , and  $\Sigma_{ij}^*(t)$  according to Equations 21 and 24. Altogether, Equations 6, 26 and 27 form a set of closed and coupled differential equations for  $\{\mu_i(t)\}_i$  and  $\{\Sigma_{ij}(t, 0)\}_{i \geq j}$  which can be integrated straightforwardly.

Temporal correlations are given by (for  $s > 0$ ):

$$\frac{d\Sigma_{ij}(t, s)}{ds} = \frac{1}{\tau_j} \left[ -\Sigma_{ij}(t, s) + \sum_{\ell} W_{j\ell} \Gamma_{i\ell}(t, s) + \Sigma_{ji}^*(t) \exp\left(-\frac{s}{\tau_\eta}\right) \right] \quad (28)$$

## 2.5 Summary of equations

The equations of motion for the moments of  $\mathbf{u}$  can be summarised in matrix form as follows.

### Temporally white input noise

$$\text{DYNAMICS} \quad d\mathbf{u} = \mathbf{T}^{-1} [-\mathbf{u}(t) + \mathbf{h}(t) + \mathbf{W}\mathbf{r}(t)] dt + d\boldsymbol{\chi} \quad (29)$$

$$\text{MEAN} \quad \frac{d\boldsymbol{\mu}(t)}{dt} = \mathbf{T}^{-1} [-\boldsymbol{\mu}(t) + \mathbf{h}(t) + \mathbf{W}\boldsymbol{\nu}(t)] \quad (30)$$

$$\text{COVARIANCE} \quad \frac{d\boldsymbol{\Sigma}(t, 0)}{dt} = \boldsymbol{\Sigma}^\chi + \boldsymbol{\mathcal{J}}(t) \boldsymbol{\Sigma}(t, 0) + \boldsymbol{\Sigma}(t, 0) \boldsymbol{\mathcal{J}}(t)^T \quad (31)$$

$$\text{LAGGED COV.} \quad \frac{d\boldsymbol{\Sigma}(t, s)}{ds} = \boldsymbol{\Sigma}(t, s) \boldsymbol{\mathcal{J}}(t+s)^T \quad (32)$$

where we have defined  $\mathbf{T} = \text{diag}(\tau_1, \dots, \tau_N)$ , and  $\boldsymbol{\mathcal{J}}(t) \equiv \mathbf{T}^{-1} [\mathbf{W} \text{diag}(\gamma_1(t), \dots, \gamma_N(t)) - \mathbf{I}]$ . In these equations,  $\boldsymbol{\nu}(t)$  and  $\boldsymbol{\gamma}(t)$  are given in closed form as functions of  $\boldsymbol{\mu}(t)$  and  $\boldsymbol{\Sigma}(t, 0)$  according to Equations 23 and 24.

### Temporally correlated input noise

$$\text{DYNAMICS} \quad \frac{d\mathbf{u}}{dt} = \mathbf{T}^{-1} [-\mathbf{u}(t) + \mathbf{h}(t) + \mathbf{W}\mathbf{r}(t) + \boldsymbol{\eta}(t)] \quad (33)$$

$$\text{MEAN} \quad \frac{d\boldsymbol{\mu}(t)}{dt} = \mathbf{T}^{-1} [-\boldsymbol{\mu}(t) + \mathbf{h}(t) + \mathbf{W}\boldsymbol{\nu}(t)] \quad (34)$$

$$\text{COVARIANCE} \quad \frac{d\boldsymbol{\Sigma}(t, 0)}{dt} = [\mathbf{T}^{-1}\boldsymbol{\Sigma}^*(t)] + [\mathbf{T}^{-1}\boldsymbol{\Sigma}^*(t)]^T + \boldsymbol{\mathcal{J}}(t)\boldsymbol{\Sigma}(t, 0) + \boldsymbol{\Sigma}(t, 0)\boldsymbol{\mathcal{J}}(t)^T \quad (35)$$

$$\frac{d\boldsymbol{\Sigma}^*(t)}{dt} = -\frac{1}{\tau_\eta}\boldsymbol{\Sigma}^*(t) + \boldsymbol{\Sigma}^\eta\mathbf{T}^{-1} + \boldsymbol{\Sigma}^*(t)\boldsymbol{\mathcal{J}}(t)^T \quad (36)$$

$$\text{LAGGED COV.} \quad \frac{d\boldsymbol{\Sigma}(t, s)}{ds} = e^{-s/\tau_\eta} [\mathbf{T}^{-1}\boldsymbol{\Sigma}^*(t)]^T + \boldsymbol{\Sigma}(t, s)\boldsymbol{\mathcal{J}}(t+s)^T \quad (37)$$

with the same definitions of  $\mathbf{T}$  and  $\boldsymbol{\mathcal{J}}(t)$  as above, and with  $\boldsymbol{\nu}(t)$  and  $\boldsymbol{\gamma}(t)$  again given by Equations 23 and 24 as functions of  $\boldsymbol{\mu}(t)$  and  $\boldsymbol{\Sigma}(t, 0)$ .

**Note on implementation** We favor using the equations of motion in their matrix form, as we can then use highly efficient libraries for vectorised operations, especially matrix products (we use the OpenBLAS library). We integrate Equations 30 to 32 and Equations 34 to 37 using the classical Euler method with a small time step  $\delta_t = 0.1$  ms. When interested in the stationary moments, a good initial condition from which to start the integration is given by the case  $\mathbf{W} = 0$ , i.e.  $\boldsymbol{\mu}(0) \equiv \mathbf{h}$ ,  $\boldsymbol{\Sigma}^*(0) \equiv \boldsymbol{\Sigma}^\eta \text{diag}\left(\frac{1}{1+\tau/\tau_\eta}\right)$ , and  $\boldsymbol{\Sigma}(0, 0)$  obtained by solving a simple Lyapunov equation.

Care should be taken in integrating the covariance flow of Equations 31 and 35 to preserve the positive definiteness of  $\boldsymbol{\Sigma}$  at all times. We do this for Equation 31 using the following integrator (Bonnabel and Sepulchre, 2012):

$$\boldsymbol{\Sigma}(t + \delta_t, 0) = [\mathbf{I} + \delta_t\boldsymbol{\mathcal{J}}(t)]\boldsymbol{\Sigma}(t, 0)[\mathbf{I} + \delta_t\boldsymbol{\mathcal{J}}(t)]^T + \delta_t\boldsymbol{\Sigma}^\chi \quad (38)$$

and analogously for Equation 35. The complexity is  $\mathcal{O}(TN^3)$  where  $T$  is the number of time bins, and  $N$  is the number of neurons. In comparison, stochastic simulations of Equation 1 cost  $\mathcal{O}(KTN^2)$ , where  $K$  is a certain number of independent trials that must be simulated to get an estimate of activity variability. This complexity is only quadratic in  $N$ , but in practice  $K$  will have to be large for the moments to be accurately estimated. Moreover, the generation of random numbers in Monte-Carlo simulations is expensive. In all the numerical examples given below, theoretical values were obtained at least 10 times faster than the corresponding Monte-Carlo estimates, given a decent accuracy criterion (required number of trials). Where appropriate, one could also apply low-rank reduction techniques to reduce the complexity of the equations of motion to  $\mathcal{O}(TN^2)$  (e.g. in the spirit of Bonnabel and Sepulchre, 2012); this is left for future work.

## 2.6 Firing rate correlations

So far we have obtained results for pairwise membrane potential covariances  $\Sigma_{ij}(t, s)$ , but, as indicated above, these can also be translated into covariances between the corresponding rate variables,  $\Lambda_{ij}(t, s) \equiv \langle \tilde{r}_i(t) \tilde{r}_j(t+s) \rangle$ , which we will need later to compute spike count statistics (e.g. Fano factors or correlations). To shorten the notation, we again drop the time dependence, with the same understanding that in using the equations that follow to compute  $\Lambda_{ij}(t, s)$ , the



quantities  $\mu_i$  and variances  $\Sigma_{ii}$  that regard neuron  $i$  will have to be evaluated at time  $t$ , those that regard neuron  $j$  evaluated at time  $t + s$ , and the covariance  $\Sigma_{ij}$  will have to be understood as  $\Sigma_{ij}(t, s)$ .

Under the same Gaussian process assumption as above, we have

$$\Lambda_{ij} = \iint du du' \cdot \mathcal{N} \left[ \begin{pmatrix} u \\ u' \end{pmatrix}; \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix}, \begin{pmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ij} & \Sigma_{jj} \end{pmatrix} \right] \cdot (f(u) - \nu_i) \cdot (f(u') - \nu_j) \quad (39)$$

Calculating this double integral exactly seems infeasible. However, as detailed below, we were able to derive an analytical approximation that is highly accurate over a broad range of physiologically relevant values (Figure 1), for the class of threshold power-law nonlinearities  $f(u) = k[u]_+^n$ .

Numerical explorations of the behaviour of  $\Lambda_{ij}$  as a function of the moments of  $u_i$  and  $u_j$  suggest the following *ansatz*:

$$\Lambda_{ij} = \alpha_{ij}^{(3)} c_{ij}^3 + \alpha_{ij}^{(2)} c_{ij}^2 + \alpha_{ij}^{(1)} c_{ij} \quad (40)$$

where  $c_{ij} = \Sigma_{ij} / \sqrt{\Sigma_{ii}\Sigma_{jj}}$  is the correlation coefficient between  $u_i$  and  $u_j$  (at time  $t$  and lag  $s$ ), and the three coefficients  $\alpha_{ij}^{(\cdot)}$  do not depend on  $c_{ij}$  (though they depend on the marginals over  $u_i$  and  $u_j$ , as detailed below) and can be computed exactly. We focus on  $\Lambda_{ij}$  as a function of  $c_{ij}$  – we abuse the notation of Equation 5c and write this dependence as  $\Lambda_{ij}(c_{ij})$ . Clearly,  $\Lambda_{ij}(0) = 0$ . Next, we note that

$$\alpha_{ij}^{(2)} = [\Lambda_{ij}(+1) + \Lambda_{ij}(-1)] / 2 \quad (41a)$$

$$\alpha_{ij}^{(1)} + \alpha_{ij}^{(3)} = [\Lambda_{ij}(+1) - \Lambda_{ij}(-1)] / 2 \quad (41b)$$

$$\alpha_{ij}^{(1)} = \left. \frac{d\Lambda_{ij}(c_{ij})}{dc_{ij}} \right|_{c_{ij}=0} \quad (41c)$$

where, specialising to the threshold-power law nonlinearity  $f(u) = k[u]_+^n$ ,

$$\Lambda_{ij}(\pm 1) = -\nu_i \nu_j + k^2 \int \mathcal{D}u [\mu_i \pm u \sqrt{\Sigma_{ii}}]_+^n [\mu_j \pm u \sqrt{\Sigma_{jj}}]_+^n \quad (42)$$

After some algebra, Equation 41c yields

$$\alpha_{ij}^{(1)} = \left( \gamma_i \sqrt{\Sigma_{ii}} \right) \left( \gamma_j \sqrt{\Sigma_{jj}} \right) \quad (43)$$

where  $\Gamma_{ii}$  and  $\Gamma_{jj}$  were derived previously in Equations 21 and 24. To compute  $\Lambda_{ij}(+1)$ , let us define more generally

$$A_{ij}^{(n,m)} \equiv k^2 \int \mathcal{D}u [\mu_i + u \sqrt{\Sigma_{ii}}]_+^n [\mu_j + u \sqrt{\Sigma_{jj}}]_+^m \quad (44)$$

keeping in mind that we are ultimately interested in  $A_{ij}^{(n,n)}$ , since  $\Lambda_{ij}(+1) = -\nu_i \nu_j + A_{ij}^{(n,n)}$ . In the following, we assume that  $\mu_i / \sqrt{\Sigma_{ii}} \geq \mu_j / \sqrt{\Sigma_{jj}}$ . If the opposite holds, then the indices  $i$  and  $j$  must be swapped at this stage (this is just a matter of notation). Using techniques similar to the integral calculations carried out previously (mostly, integration by parts), we derived the following recursive formula valid for  $0 \leq n \leq m$ :

$$A_{ij}^{(n,m)} = \begin{cases} k\nu_j^{(m)} & \text{if } n = 0 \\ k\mu_i \nu_j^{(m)} + k\sqrt{\frac{\Sigma_{ii}}{\Sigma_{jj}}} \Gamma_{jj}^{(m)} & \text{if } n = 1 \\ \mu_i A_{ij}^{(n-1,m)} + (n-1)\Sigma_{ii} A_{ij}^{(n-2,m)} + m\sqrt{\Sigma_{ii}\Sigma_{jj}} A_{ij}^{(n-1,m-1)} & \text{otherwise} \end{cases} \quad (45)$$

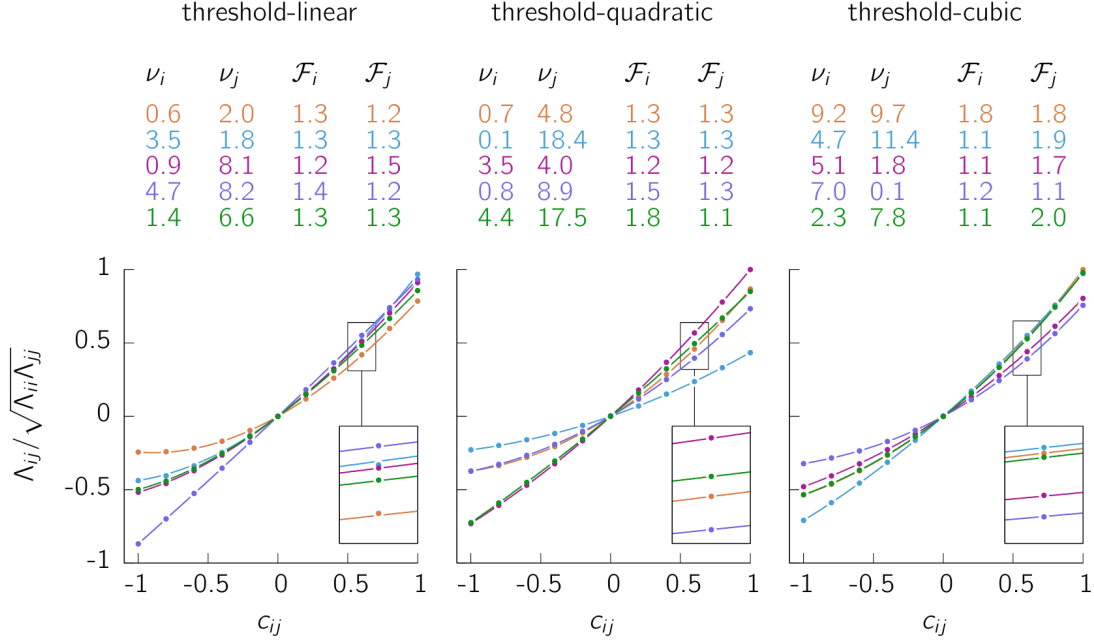


Figure 1: Numerical validation of our ansatz for rate covariances. Shown in 5 different colors are the results of 5 different simulations in which  $\sqrt{\Sigma_{ii}}$  (resp.  $\sqrt{\Sigma_{jj}}$ ) was drawn uniformly between 1 and 4, and  $\mu_i$  (resp.  $\mu_j$ ) was chosen so as to achieve a mean firing rate  $\nu_i$  (resp.  $\nu_j$ ) drawn from a Gamma distribution with a mean 5 Hz and a shape parameter of 1. Dots show  $\Lambda_{ij}$  as a function of  $c_{ij}$ , as estimated via Monte-Carlo integration of Equation 39 with a million samples. Solid lines show Equation 40, which is by construction always exact at  $c_{ij} = \pm 1$ , and always has the right slope at  $c_{ij} = 0$ . The corresponding tables show the mean firing rates and the Fano factors of both neurons (Fano factors were computed using Equation 54 with  $\tau_A = 50$  ms and  $T = 100$  ms). The figures indicate that our approximation is highly accurate over a range of parameters corresponding to physiological values of firing rates and Fano factors. *Parameters:*  $k = 3$  (threshold-linear),  $k = 0.3$  (threshold-quadratic),  $k = 0.02$  (threshold-cubic).

where  $\nu_j^{(m)}$  and  $\Gamma_j^{(m)}$  were calculated previously (cf. Equations 23 and 24).

The calculation of  $\Lambda_{ij}(-1)$  is slightly more involved, but follows a similar logic. This time we assume that  $\mu_i/\sqrt{\Sigma_{ii}} > -\mu_j/\sqrt{\Sigma_{jj}}$  (otherwise,  $\Lambda_{ij}(-1) = 0$ ), and we define

$$B_{ij}^{(n,m)} \equiv k^2 \int \mathcal{D}u \, [\mu_i - u\sqrt{\Sigma_{ii}}]_+^n [\mu_j + u\sqrt{\Sigma_{jj}}]_+^m \quad (46)$$

Similar to Equation 45, we have the following recursive formula for  $1 < n \leq m$ :

$$B_{ij}^{(n,m)} = \mu_i B_{ij}^{(n-1,m)} + (n-1)\Sigma_{ii} B_{ij}^{(n-2,m)} - m\sqrt{\Sigma_{ii}\Sigma_{jj}} B_{ij}^{(n-1,m-1)} \quad (47)$$

However, now the boundary conditions must also be computed recursively:

$$B_{ij}^{(0,m)} = \begin{cases} k^2 [\psi_i + \psi_j - 1] & \text{if } m = 0 \\ \mu_j B_{ij}^{0,0} + k^2 \sqrt{\Sigma_{jj}} (\phi_j - \phi_i) & \text{if } m = 1 \\ \mu_j B_{ij}^{(0,m-1)} - \sqrt{\Sigma_{jj}} \left[ k^2 \phi_i R_{ij}^{m-1} - (m-1) \sqrt{\Sigma_{jj}} B_{ij}^{(0,m-2)} \right] & \text{otherwise} \end{cases} \quad (48)$$

$$B_{ij}^{(1,m)} = \mu_i B_{ij}^{(0,m)} + \sqrt{\Sigma_{ii}} \begin{cases} k^2 (\phi_i - \phi_j) & \text{if } m = 0 \\ k^2 \phi_i R_{ij}^m - m \sqrt{\Sigma_{jj}} B_{ij}^{(0,m-1)} & \text{otherwise} \end{cases} \quad (49)$$

where we have used the shorthands  $\phi_\ell \equiv \phi(\mu_\ell/\sqrt{\Sigma_{\ell\ell}})$ ,  $\psi_\ell \equiv \psi(\mu_\ell/\sqrt{\Sigma_{\ell\ell}})$ , and  $R_{ij} \equiv \mu_j + \mu_i \sqrt{\Sigma_{jj}/\Sigma_{ii}}$ .

Thus,  $\Lambda_{ij}(c_{ij})$  is approximated in closed-form by a third-order polynomial in  $c_{ij}$ , given  $\mu_i, \mu_j, \Sigma_{ii}, \Sigma_{jj}$  and  $\Sigma_{ij}$ . Our polynomial approximation is very accurate over a broad, physiologically relevant range of parameters (and in fact, even beyond that), for power law exponents in the physiological range (Figure 1).

In combination with the results of Section 2.2, we can then obtain the moments of  $\mathbf{r}$ , which we use in the following section to compute Fano factors and spike count correlations.

## 2.7 Spike count statistics

Under the assumption that each neuron  $i$  emits spikes according to an inhomogeneous Poisson process with rate function  $r_i(t)$  (“doubly stochastic”, or “Cox” process)<sup>1</sup>, we can compute the joint statistics of the spike counts in some time window, which is what electrophysiologists often report. Let  $C_i^T(t)$  denote the number of spikes that are emitted by neuron  $i$  in a window of duration  $T$  starting at time  $t$ , and let  $\kappa_i^T(t) = \int_0^T dt' r_i(t+t')$  be the expected number of spikes in that window, for a given trial and *given* the underlying rate trace  $r_i(t)$ . The Fano factor of the distribution of  $C_i^T(t)$  is given by

$$\mathcal{F}_i^T(t) = \frac{\text{var}[C_i^T(t)]}{\text{mean}[C_i^T(t)]} = 1 + \frac{\text{var}[\kappa_i^T(t)]}{\langle \kappa_i^T(t) \rangle} \quad (50)$$

with

$$\langle \kappa_i^T(t) \rangle = \int_0^T dt' \nu_i(t+t') \quad (51)$$

and

$$\text{var}[\kappa_i^T(t)] = \int_0^T ds \int_0^T ds' \Lambda_{ii}(t+s, s'-s) \quad (52)$$

All expected values such as  $\nu_i$  and  $\Lambda_{ii}$  have been calculated in previous sections. We approximate the integrals above by simple Riemann sums with a discretisation step of 1 ms.

In the special case of constant input  $\mathbf{h}$  leading to a stationary rate distribution, Equation 50 simplifies to

$$\mathcal{F}_i^T = 1 + \frac{1}{T\nu_i} \int_0^T ds \int_0^T ds' \Lambda_{ii}(\cdot, s'-s) \quad (53)$$

(the rate variance  $\Lambda_{ii}(t, \tau)$  no longer depends on  $t$ , hence the notation  $\Lambda_{ii}(\cdot, \tau)$ ). A closed-form approximation can be derived if the rate autocorrelation is well approximated by a Laplacian with decay time constant  $\tau_A$ , i.e.  $\Lambda_{ii}(\cdot, \tau) \approx \Lambda_{ii}(\cdot, 0) \exp(-|\tau|/\tau_A)$ . In this case, Equation 53 evaluates to

$$\mathcal{F}_i^T \approx 1 + \frac{2\tau_A \Lambda_{ii}(\cdot, 0)}{\nu_i} \left[ 1 - \frac{\tau_A}{T} (1 - e^{-T/\tau_A}) \right] \quad (54)$$

This expression is shown in Figure 2 as a function of the counting window  $T$ , for various autocorrelation lengths  $\tau_A$ .

The behaviour of  $\mathcal{F}_i$  as a function of membrane potential sufficient statistics  $\mu_i$  and  $\sqrt{\Sigma_{ii}}$  is depicted in Figure 3, for various exponents of the threshold power-law gain function. In all cases (i.e. independent of the exponent  $n$ ), the Fano factor grows with increasing potential variance. However, the dependence on  $\mu_i$  strongly depends on  $n$ . For  $n = 1$ , the Fano factor decreases with  $\mu_i$ , while for  $n = 3$  it increases. For  $n = 2$ , the Fano factor has only a weak dependence on  $\mu_i$ . This behaviour can be understood qualitatively by linearising the gain function around the

<sup>1</sup>This does not affect the form of the network dynamics which remain rate-based (Equation 1); that is, spikes are generated on top of the firing rate fluctuations given by the rate model.

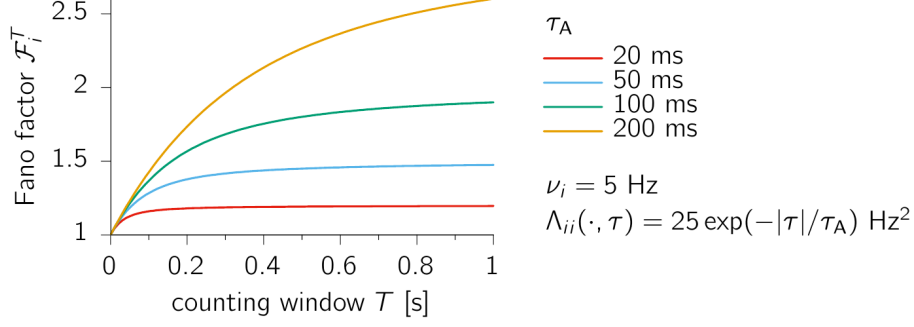


Figure 2: Approximation of the Fano factor when the autocorrelation is a Laplacian with time constant  $\tau_A$  (varied here; cf. color legend) – see also Equation 54.

mean potential, which gives us an idea of how the super-Poisson part of the Fano factor depends on membrane potential statistics:

$$\mathcal{F}_i^T - 1 \propto n \mu_i^{n-2} \Sigma_{ii} \quad (55)$$

Therefore, the iso-Fano-factor lines in Figure 3 are expected to approximately obey  $\sqrt{\Sigma_{ii}} \approx \mu_i^{2-n}$ , which is roughly what we see from Figure 3.

It is instructive to plug in some typical numbers: assuming a mean firing rate of  $\nu_i = 5$  Hz, a counting window  $T = 50$  ms, a Fano factor  $\mathcal{F}_i^T = 1.5$  (characteristic of spontaneous activity in the cortex, Churchland et al., 2010), and an autocorrelation time constant  $\tau_A = 40$  ms (Kenet et al., 2003; Berkes et al., 2011), then Equation 54 tells us that the fluctuations in  $r_i$  must have a standard deviation of  $\sqrt{\Lambda_{ii}} \simeq 8.5$  Hz. This is larger than the assumed mean firing rate (5 Hz), thus implying that the underlying rate variable in physiological conditions is often going to be zero, in turn implying that the membrane potential will often be found below the threshold of the firing rate nonlinearity  $f(u)$ . It is precisely in this regime that it becomes important to perform the nonlinear conversion of the moments of  $\mathbf{u}$  into those of  $\mathbf{r}$ , taking into account the specific form of the nonlinearity; in that same regime, linearisation of the dynamics of Equation 1 can become very inaccurate.

Similar derivations can be done for spike count correlations:

$$\begin{aligned} \rho_{ij}^T(t) &\equiv \frac{\text{cov} [C_i^T(t), C_j^T(t)]}{\sqrt{\text{var} [C_i^T(t)] \text{var} [C_j^T(t)]}} \\ &= \frac{\langle \kappa_i^T(t) \kappa_j^T(t) \rangle - \langle \kappa_i^T(t) \rangle \langle \kappa_j^T(t) \rangle}{\sqrt{(\text{var} [\kappa_i^T(t)] + \langle \kappa_i^T(t) \rangle) (\text{var} [\kappa_j^T(t)] + \langle \kappa_j^T(t) \rangle)}} \\ &= \frac{\iint_{[0:T]^2} ds ds' \Lambda_{ij}(t, s' - s)}{\sqrt{(\text{var} [\kappa_i^T(t)] + \langle \kappa_i^T(t) \rangle) (\text{var} [\kappa_j^T(t)] + \langle \kappa_j^T(t) \rangle)}} \\ &= [\langle \kappa_i^T(t) \rangle \langle \kappa_j^T(t) \rangle \mathcal{F}_i^T(t) \mathcal{F}_j^T(t)]^{-\frac{1}{2}} \iint_{[0:T]^2} ds ds' \Lambda_{ij}(t, s' - s) \end{aligned} \quad (56)$$

In the stationary case, Equation 56 becomes

$$\rho_{ij}^T = \frac{\iint_{[0:T]^2} ds ds' \Lambda_{ij}(\cdot, s' - s)}{T \sqrt{\nu_i \nu_j \mathcal{F}_i^T \mathcal{F}_j^T}} \quad (57)$$

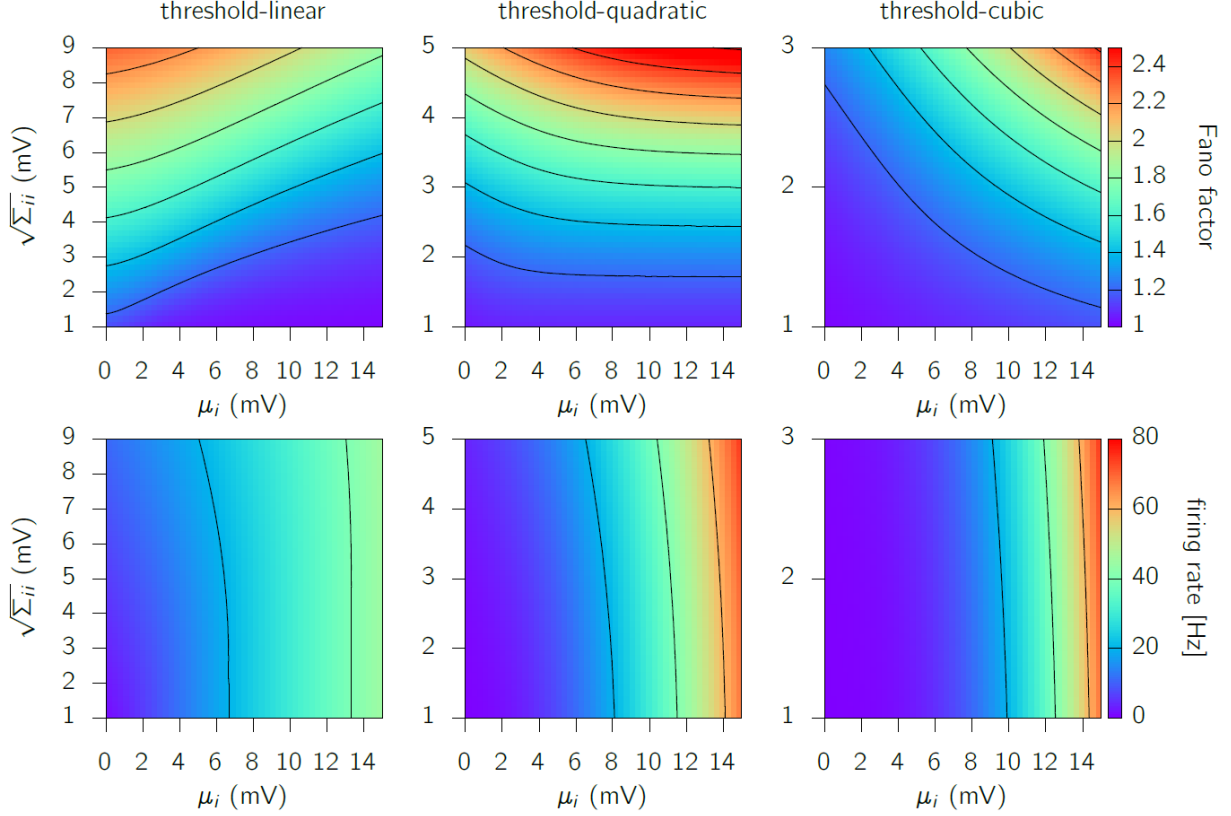


Figure 3: Fano factor  $\mathcal{F}_i$  (top row) and mean firing rate  $\nu_i$  (bottom row) as a function of membrane potential mean  $\mu_i$  and standard deviation  $\sqrt{\Sigma_{ii}}$ , for  $n = 1$  to  $n = 3$  (left to right). To calculate  $\mathcal{F}_i$ , we used Equation 54, assuming an autocorrelation time constant of  $\tau_A = 50$  ms. The multiplicative rescaling  $k$  of the threshold power-law gain function was adjusted so that both Fano factors and firing rates span roughly the same range of values across  $n = 1$  to  $n = 3$ , but the precise value of  $k$  has no impact on the shape of the contour lines.

### 3 Numerical validation

In this section, we demonstrate the validity of the equations derived in Section 2 on two examples: a random E/I network with weak and sparse connections, and a strongly connected, inhibition-stabilized E/I network.

#### 3.1 Weakly connected random network

The network is made of  $N_E = 250$  excitatory and  $N_I = 250$  inhibitory neurons, each with a threshold-quadratic I/O nonlinearity:  $f(u) = 0.3[u]_+^2$ . We set all intrinsic time constants to  $\tau_i = \tau \equiv 20$  ms. The input noise has no spatial, but only temporal, correlations:

$$\langle \eta_i(t) \eta_j(t+s) \rangle = \frac{\sigma_0^2}{1 + \tau/\tau_\eta} \exp\left(-\frac{|s|}{\tau_\eta}\right) \delta_{ij} \quad (58)$$

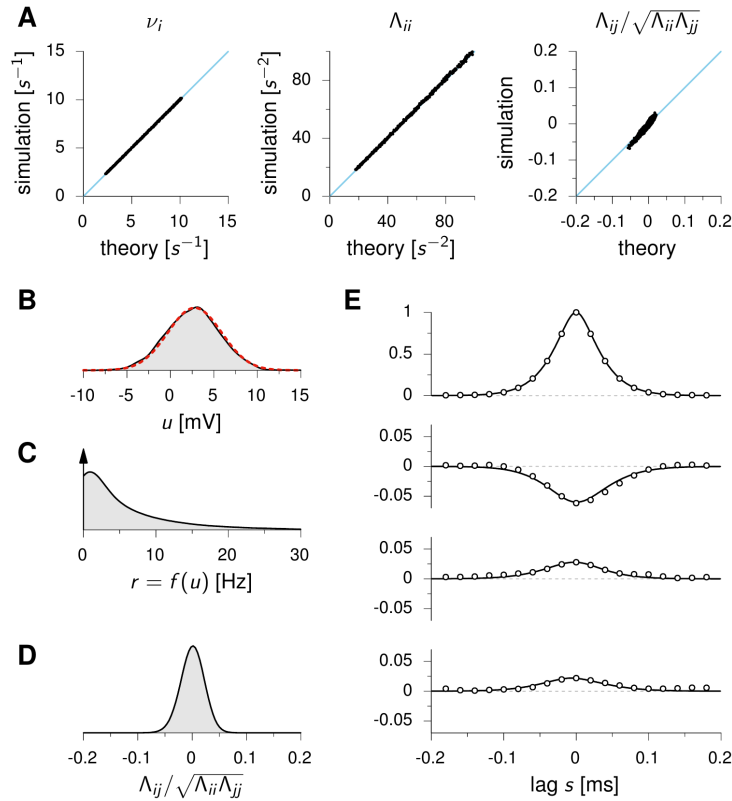
with  $\tau_\eta = 50$  ms and  $\sigma_0 = 3$  mV representing the standard deviation of the membrane potentials if the recurrent connectivity were removed. Synaptic weights are drawn randomly from the following distribution:

$$W_{ij} = \frac{\alpha s_j}{\sqrt{N}} \times \begin{cases} 1 & \text{with probability 0.2} \\ 0 & \text{otherwise} \end{cases} \quad (59)$$

where  $\alpha$  is a global scaling factor (see below),  $s_j$  is a presynaptic signed factor equal to  $+1$  if neuron  $j$  is excitatory ( $1 \leq j \leq N/2$ ), and equal to  $-\gamma$  if neuron  $j$  is inhibitory ( $N/2 < j \leq N$ ). We set  $\gamma = 3$  to place the network in the inhibition-dominated regime where the average pairwise correlation among neurons is weak (Renart et al., 2010; Hennequin et al., 2012). The scaling factor  $\alpha = 2.2$  was chosen such that the network is effectively weakly connected and thus far from instability.

The network is fed with a constant input  $\mathbf{h} = \mathbf{u}^* - \mathbf{W}f(\mathbf{u}^*)$ , such that, in the absence of stochasticity, the network would have a (stable) fixed point at  $\mathbf{u} = \mathbf{u}^*$  – we drew the elements of  $\mathbf{u}^*$  from a uniform distribution between 1 and 4 mV.

We simulated the stochastic dynamics of the network (Equation 25) for 5000 seconds, and computed empirical estimates of the moments in the steady-state, stationary regime (mean firing rates, firing rate variances, firing rate correlations, and full membrane potential cross-correlograms). We found those estimates to agree very well with the semi-analytical solutions obtained by integrating the relevant equations of motion derived in Section 2 under the Gaussian assumption (Figure 4A). Membrane potentials are indeed roughly normally distributed (Figure 4B), with a standard deviation on the same order as the mean, yielding very skewed



**Figure 4: Validation of our theoretical results on a random, weakly connected E/I network.** (A) Mean firing rates  $\{\nu_i\}$ , firing rate variances  $\{\Lambda_{ii}\}$ , and pairwise firing rate correlations, as predicted semi-analytically by the moment dynamics equations of Section 2 (x-axes) and compared to empirical estimates obtained from long stochastic simulations of Equation 25 (y-axes). (B) Distribution of membrane potentials  $u_i(t)$  for a randomly chosen neuron (gray), and its Gaussian approximation  $\mathcal{N}(u; \mu_i, \Sigma_{ii})$  used in our theory. (C) Corresponding firing rate distribution of the same neuron as in (B). (D) Distribution of pairwise firing rate correlations across the network. (E) Example membrane potential normalized auto- (top row) and cross- (bottom rows) correlograms  $(\Sigma_{ij}(\infty, s) / \sqrt{\Sigma_{ii} \Sigma_{jj}})$ , as predicted by the theory (solid) and estimated from simulations (dots).

distributions of firing rates (Figure 4C). Due to the effectively weak connectivity, pairwise correlations among firing rate fluctuations are weak in magnitude (Figure 4D). Membrane potential cross-correlograms have a simple, near-symmetric structure (random, weakly connected networks are close to equilibrium) and are well captured by the theory.

### 3.2 Strongly connected, inhibition-stabilized random network

We now consider a strongly connected E/I network much further away from equilibrium than the weakly-connected network of Figure 4. This network operates in the inhibition-stabilized regime, whereby excitatory feedback is strongly destabilizing on its own, but is dynamically stabilized by feedback inhibition. The details of how we obtained such a network are largely irrelevant here (but see Hennequin et al., 2014). All the parameters of the network will be posted online together with the code to ensure reproducibility.

Our results are summarized in Figure 5, in exactly the same format as in Figure 4. This network strongly amplifies the input noise process along a few specific directions in state space, leading to strong (negative and positive) pairwise correlations among neuronal firing rates Figure 5D. Thus, the central limit theorem – which would in principle justify our core assumption that membrane potentials are jointly Gaussian, as large (and low-pass-filtered) sums of uncorrelated variables – may not apply. Indeed, most membrane potential distributions have substantial (negative) skewness (Figure 5B), leading to some inaccuracies in our semi-analytical calculation of the moments (Figure 5A), which we find are reasonably small nonetheless. We leave the extension of our theory to third-order moments (e.g. Dahmen et al., 2016) for future work. Finally, the strong connectivity gives rise to non-trivial temporal structure in the joint activity of pairs of neurons reflecting non-equilibrium dynamics. Even in this regime, our results capture the membrane potential cross-correlograms well (Figure 5E).

## Acknowledgments

This work was supported by the Swiss National Science Foundation (GH) and the Wellcome Trust (GH, ML).

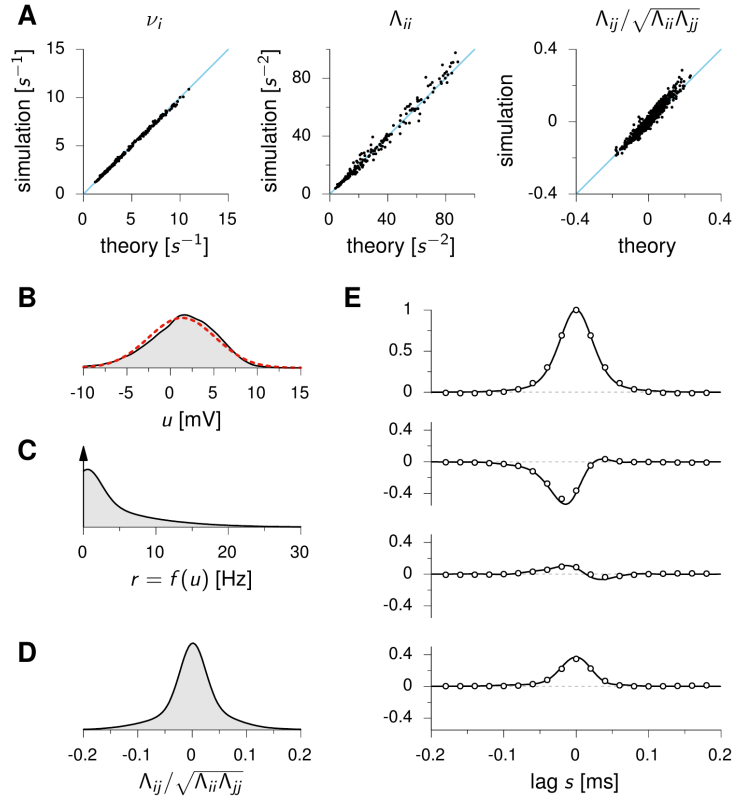


Figure 5: **Validation of our theoretical results on a random, strongly connected, inhibition-stabilized E/I network.** The figure follows the same format as in Figure 4 above; see caption there.



## References

- Ahmadian, Y., Rubin, D. B., and Miller, K. D. (2013). Analysis of the stabilized supralinear network. *Neural Comput.*, 25:1994–2037.
- Barrett, D. (2012). *Computation in balanced networks*. PhD thesis, University College London.
- Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, 331(6013):83–87.
- Bonnabel, S. and Sepulchre, R. (2012). The geometry of low-rank Kalman filters. *arXiv:1203.4049 [cs, math]*.
- Buesing, L., Macke, J., and Sahani, M. (2012). Spectral learning of linear dynamics from generalised-linear observations with application to neural population data. In *Advances in Neural Information Processing Systems 25*, pages 1691–1699.
- Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., Newsome, W. T., Clark, A. M., Hosseini, P., Scott, B. B., Bradley, D. C., Smith, M. A., Kohn, A., Movshon, J. A., Armstrong, K. M., Moore, T., Chang, S. W., Snyder, L. H., Lisberger, S. G., Priebe, N. J., Finn, I. M., Ferster, D., Ryu, S. I., Santhanam, G., Sahani, M., and Shenoy, K. V. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat Neurosci*, 13(3):369–378.
- Dahmen, D., Bos, H., and Helias, M. (2016). Correlated fluctuations in strongly coupled binary networks beyond equilibrium. *Phys. Rev. X*, 6(3):031024.
- Gardiner, C. W. (1985). *Handbook of stochastic methods: for physics, chemistry, and the natural sciences*. Berlin: Springer.
- Hennequin, G., Ahmadian\*, Y., Rubin\*, D. B., Lengyel†, M., and Miller†, K. D. (2016). Stabilized supralinear network dynamics account for stimulus-induced changes of noise variability in the cortex. *Submitted*.
- Hennequin, G., Vogels, T. P., and Gerstner, W. (2012). Non-normal amplification in random balanced neuronal networks. *Phys. Rev. E*, 86:011909.
- Hennequin, G., Vogels, T. P., and Gerstner, W. (2014). Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron*, 82:1394–1406.
- Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., and Arieli, A. (2003). Spontaneously emerging cortical representations of visual attributes. *Nature*, 425(6961):954–956.
- Miller, K. D. and Troyer, T. W. (2002). Neural noise can explain expansive, power-law nonlinearities in neural response functions. *J. Neurophysiol.*, 87(2):653–659.
- Murphy, B. K. and Miller, K. D. (2009). Balanced amplification: A new mechanism of selective amplification of neural activity patterns. *Neuron*, 61:635–648.
- Priebe, N. J., Mechler, F., Carandini, M., and Ferster, D. (2004). The contribution of spike threshold to the dichotomy of cortical simple and complex cells. *Nat. Neurosci.*, 7(10):1113–1122.
- Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. (2010). The asynchronous state in cortical circuits. *Science*, 327:587.
- Rubin, D., VanHooser, S., and Miller, K. (2015). The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, 85(2):402–417.