

## Moment neural network and an efficient numerical method for modeling irregular spiking activity

Yang Qi \*

*Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China;*  
*Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University),*  
*Ministry of Education, Shanghai 200433, China;*  
*and MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200433, China*

 (Received 3 June 2024; accepted 1 July 2024; published 20 August 2024)

Continuous rate-based neural networks have been widely applied to modeling the dynamics of cortical circuits. However, cortical neurons in the brain exhibit irregular spiking activity with complex correlation structures that cannot be captured by mean firing rate alone. To close this gap, we consider a framework for modeling irregular spiking activity, called the moment neural network, which naturally generalizes rate models to second-order moments and can accurately capture the firing statistics of spiking neural networks. We propose an efficient numerical method that allows for rapid evaluation of moment mappings for neuronal activations without solving the underlying Fokker-Planck equation. This allows simulation of coupled interactions of mean firing rate and firing variability of large-scale neural circuits while retaining the advantage of analytical tractability of continuous rate models. We demonstrate how the moment neural network can explain a range of phenomena including diverse Fano factor in networks with quenched disorder and the emergence of irregular oscillatory dynamics in excitation-inhibition networks with delay.

DOI: [10.1103/PhysRevE.110.024310](https://doi.org/10.1103/PhysRevE.110.024310)

### I. INTRODUCTION

The cortex in the brain forms a complex network of neurons that communicate via spike trains. Being both high-dimensional and nonlinear, these spiking neural networks are hard to analyze. This has motivated the development of continuous firing rate models that can be thought of as describing the coarse-grained activity of some underlying spiking neural network. These phenomenological models, being more analytically tractable than their spiking counterpart, are widely used for the theoretical modeling of cortical networks and brain functioning [1–3]. However, the spiking activity of cortical neurons in the brain is highly irregular and cannot be captured by the mean firing rate alone. Such fluctuating neural activity often exhibits diverse spike count Fano factors close to one [4,5] and noise correlations with complex spatiotemporal structures [6,7]. Recordings of cortical neurons *in vitro* have further revealed that neural pairwise correlations are nonlinearly coupled to the mean firing rate [8]. Correlated variability has been suggested to have a significant impact on the coding properties of neural populations, which depending on the structure of the correlation can be both beneficial or adverse [9,10]. Understanding how neurons in the brain

process noisy spikes with correlated fluctuations is a key step to unveiling the inner working of the brain [6,11].

To analyze the nonlinear noise coupling of spiking neurons, a number of mathematical techniques have been developed. One of them is the master equation approach considering a network of binary neurons given a transition probability. This results in a closed, self-consistent system of equations involving the second-order moments of the system [12,13]. For more biologically realistic spiking neuron models with continuous membrane dynamics, an analytical technique known as the diffusion approximation has been developed [14], in which the synaptic current generated by presynaptic spikes is replaced by a Gaussian white noise with the same mean and variance. By solving the first passage time problem associated with the firing threshold, the mean and variance of the postsynaptic spike train can then be derived [15–19]. Another technique is linear response theory used to obtain the temporal and pairwise covariance of spiking neurons [8,20–22]. These analytical techniques have been applied to studying asynchronous and correlated states in balanced networks [23–25], firing statistics in spiking networks with heterogeneous connectivity [26], and the emergence of spatiotemporal patterns in neural circuits [27,28].

While mean-field approaches based on the diffusion approximation are typically used to model the statistical properties of neural populations, Feng and Lu have proposed a closed, self-consistent system of equations describing the nonlinear coupling of mean firing rate and firing covariability in a neural network with generic synaptic weights [16,21]. This type of model, known as the moment neural network, can be thought of as a natural generalization of Wilson-Cowan firing

\*Contact author: yang\_qi@fudan.edu.cn

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

rate neural network model to second-order moments. Unlike the standard Wilson-Cowan formalism, which typically considers phenomenological neural activation (such as sigmoidal activation), neuronal activation in the moment neural network, referred to as the moment activation (MA), is derived from the underlying spiking neural network with noisy inputs on a mathematically rigorous ground. Meanwhile, the MNN also retains the analytical tractability of continuous rate models, enabling analysis otherwise infeasible with spiking neural networks.

Despite the availability of closed-form analytical expressions for the MNN, numerical evaluation of the moment mappings faces a multitude of challenges. Foremost, the MA contains a group of ill-conditioned Dawson-like functions that make it numerically intractable [21]. Specifically, these Dawson-like functions involve products of exploding and vanishing terms, causing their direct numerical evaluation to be prone to errors. Moreover, these ill-conditioned functions occur in nested integrals, which are slow to evaluate even for input range where they are well behaved. Although methods such as threshold-integration schemes can be used to evaluate the MA by numerically solving the associated Fokker-Planck equation [18,29], these methods are computationally cumbersome and unsuitable for large population sizes. These challenges have limited the practical usage of the MNN for analyzing and simulating the dynamics of correlated neural variability in spiking neural circuits.

In this study, we develop an efficient numerical scheme for evaluating the MA ensuring both reliability and speed through a combination of techniques including asymptotic expansion and Chebyshev polynomial approximation. The proposed method leads to an accurate and reliable evaluation of the MA orders of magnitude faster than brute-force methods. Powered by this efficient numerical method, we demonstrate how the moment neural network can be used to model a range of phenomena including diverse Fano factor in networks with quenched disorder and the emergence of irregular oscillatory dynamics in excitation-inhibition networks with delay. The MNN can thus serve as a powerful tool for investigating the firing statistics and correlated variability in large-scale spiking neural circuits and their impact on neural coding.

The remaining parts of the paper are outlined as follows. We first present the theoretical framework of the moment neural network and an efficient numerical method for evaluating the moment activation for arbitrary input statistics, in particular, by deriving asymptotic expansions for the Dawson-like functions. Detailed benchmark results about the accuracy and efficiency of the moment activation are provided in Appendixes C and D. Next, we use the MA to investigate the extent to which the linear response theory holds for approximating the correlation map. Lastly, we apply the MA for modeling the coupled interaction of mean firing rate and firing variability in large-scale neural circuits and reveal the emergence of diverse Fano factors due to quenched disorder and delay-induced irregular oscillatory dynamics, and further show how the MA can be used to investigate the coding property of neural populations driven by correlated noisy inputs.

## II. SPIKING NEURAL NETWORK MODEL

Consider the leaky integrate-and-fire (LIF) neuron [30] whose membrane potential dynamics is described by

$$\frac{dV_i}{dt} = -LV_i(t) + I_i(t), \quad (1)$$

where the subthreshold membrane potential  $V_i(t)$  of a neuron  $i$  is driven by the total synaptic current  $I_i(t)$  and  $L = 0.05 \text{ ms}^{-1}$  is the conductance. When the membrane potential  $V_i(t)$  exceeds a threshold  $V_{\text{th}} = 20 \text{ mV}$  a spike is emitted, represented by a Dirac delta function. Afterwards, the membrane potential  $V_i(t)$  is reset to the resting potential  $V_{\text{res}} = 0 \text{ mV}$ , followed by a refractory period  $T_{\text{ref}} = 5 \text{ ms}$ . The synaptic current takes the form

$$I_i(t) = \sum_{ij} w_{ij} S_j(t) + I_i^{\text{ext}}(t), \quad (2)$$

where  $S_j(t) = \sum_k \delta(t - t_j^k)$  represents the spike train generated by presynaptic neurons.

The mean firing rate  $\mu_i$  and firing covariability  $C_{ij}$  of the SNN are defined as [16,21]

$$\mu_i = \lim_{\Delta t \rightarrow \infty} \frac{\mathbb{E}[N_i(\Delta t)]}{\Delta t} \quad (3)$$

and

$$C_{ij} = \lim_{\Delta t \rightarrow \infty} \frac{\text{Cov}[N_i(\Delta t), N_j(\Delta t)]}{\Delta t}, \quad (4)$$

where  $N_i(\Delta t)$  is the spike count of neuron  $i$  over a time window  $\Delta t$ . The type of correlation described by Eq. (4) should be considered as noise correlation, which measures the correlation between the temporal fluctuations of pairs of neurons, and should be distinguished from other types of correlations reported in the literature [9].

## III. THE MOMENT NEURAL NETWORK

To capture coupled interactions between mean firing rate and correlated neural variability, we consider a mathematical formalism called the moment neural network (MNN) proposed by Feng and Lu [16,21], which naturally generalizes the rate-based Wilson-Cowan model to the second order. Specifically, the moment neural network summarizes neural spike trains into second-order moments of the spike count,  $\mu_i$  and  $C_{ij}$ , which are referred to as the mean firing rate and the firing covariability, respectively. The moments of the synaptic current in Eq. (2) are calculated as

$$\bar{\mu}_i = \sum_k w_{ik} \mu_k + \bar{\mu}_i^{\text{ext}}, \quad (5)$$

$$\bar{C}_{ij} = \sum_{kl} w_{ik} C_{kl} w_{jl} + \bar{C}_{ij}^{\text{ext}}, \quad (6)$$

where  $w_{ik}$  are synaptic weights, and  $\bar{\mu}_i^{\text{ext}}$  and  $\bar{C}_{ij}^{\text{ext}}$  are the mean and covariance of an external input current. The dynamics of the statistical moments of neural activity in this recurrent circuit can then be described by the following closed system

of equations:

$$\tau \frac{d\mu_i}{dt} = -\mu_i + \phi_\mu(\bar{\mu}_i, \bar{\sigma}_i), \quad (7)$$

$$\tau \frac{d\sigma_i}{dt} = -\sigma_i + \phi_\sigma(\bar{\mu}_i, \bar{\sigma}_i), \quad (8)$$

$$\tau \frac{d\rho_{ij}}{dt} = -\rho_{ij} + \chi(\bar{\mu}_i, \bar{\sigma}_i)\chi(\bar{\mu}_j, \bar{\sigma}_j)\bar{\rho}_{ij}, \quad i \neq j, \quad (9)$$

where  $\tau$  is a time constant and the correlation coefficient  $\rho_{ij}$  is related to the covariance by  $C_{ij} = \sigma_i\sigma_j\rho_{ij}$ . This model known as the moment neural network (MNN) naturally generalizes rate-based Wilson-Cowan models to second-order statistical moments and can be considered as the minimalistic model capable of describing the dynamics of fluctuating neural activity with nontrivial (e.g., noiseless or independent Poisson) probability distributions [16,21]. The functions  $\phi_\mu$  and  $\phi_\sigma$  together map the mean and variance of the input current to that of the output spikes, whose specific functional form depends on the type of underlying spiking neuron model. The function  $\chi$ , which we refer to as the linear perturbation coefficient, is derived using a linear perturbation analysis around  $\bar{\rho}_{ij} = 0$  [8,21]. This approximation is justified because pairwise correlations between neurons in the brain are typically weak. The mappings  $\phi_\mu$ ,  $\phi_\sigma$ ,  $\chi$  together form what we refer to as the moment activation (MA), which is described in detail below.

#### IV. THE MOMENT ACTIVATION

In this work, we consider the MA for the leaky integrate-and-fire (LIF) spiking neuron model [Eq. (1)] [30], though the general principle applies to other spiking neuron models. The first two components of the MA describe the statistical input-output relation of a single neuron [16], in which case we drop the neuronal index for clarity,

$$\mu = \phi_\mu(\bar{\mu}, \bar{\sigma}) = \frac{1}{T_{\text{ref}} + \mathbb{E}[T]}, \quad (10)$$

$$\sigma^2 = \phi_\sigma(\bar{\mu}, \bar{\sigma}) = \mu^3 \text{Var}[T]. \quad (11)$$

The mean and variance of the interspike interval  $T$  are given by

$$\mathbb{E}[T] = \frac{2}{L} \int_{I_{\text{lb}}}^{I_{\text{ub}}} g(u) du = \frac{2}{L} [G(I_{\text{ub}}) - G(I_{\text{lb}})], \quad (12)$$

$$\text{Var}[T] = \frac{8}{L^2} \int_{I_{\text{lb}}}^{I_{\text{ub}}} h(u) du = \frac{8}{L^2} [H(I_{\text{ub}}) - H(I_{\text{lb}})], \quad (13)$$

with upper and lower integration bounds  $I_{\text{ub}}(\bar{\mu}, \bar{\sigma}) = \frac{V_{\text{th}}L - \bar{\mu}}{\sqrt{L\bar{\sigma}}}$  and  $I_{\text{lb}}(\bar{\mu}, \bar{\sigma}) = \frac{V_{\text{res}}L - \bar{\mu}}{\sqrt{L\bar{\sigma}}}$ , respectively. The four Dawson-like functions that appear in Eqs. (12) and (13) are

$$g(x) = e^{x^2} \int_{-\infty}^x e^{-u^2} du, \quad (14)$$

$$h(x) = e^{x^2} \int_{-\infty}^x e^{-u^2} [g(u)]^2 du, \quad (15)$$

$$G(x) = \int_0^x g(u) du, \quad (16)$$

$$H(x) = \int_{-\infty}^x h(u) du. \quad (17)$$

The last component of the MA is linear response coefficients

$$\chi(\bar{\mu}, \bar{\sigma}) = \frac{\bar{\sigma}}{\sigma} \frac{\partial \mu}{\partial \bar{\mu}}, \quad (18)$$

from which the correlation map between a pair of neurons can be calculated [8,21]:

$$\rho_{ij} = \chi(\bar{\mu}_i, \bar{\sigma}_i)\chi(\bar{\mu}_j, \bar{\sigma}_j)\bar{\rho}_{ij}, \quad i \neq j. \quad (19)$$

Here  $\rho_{ij}$  and  $\bar{\rho}_{ij}$  correspond to spike count correlation and input current correlation, respectively. The three components of the MA, namely, the mean firing rate  $\mu$  [Eq. (10)], the firing variability  $\sigma$  [Eq. (11)], and the linear response coefficient  $\chi$  [Eq. (18)], are shown in Figs. 1(a)–1(c). The family of Dawson-like functions are illustrated in Fig. 1(d).

The evaluation of the MA based on these integral representations becomes problematic in both reliability and speed. First, the Dawson-like functions [Eqs. (14)–(17)] are ill-conditioned so that direct evaluation of these integrals may fail catastrophically. To illustrate this point, consider  $g(x)$  in Eq. (14). When  $x$  becomes increasingly negative, the exponential function outside the integral explodes, whereas the exponential function inside the integral vanishes, resulting in a numeric instability of type “ $\infty \cdot 0$ ” even for moderately negative values of  $x$ . This scenario is frequently encountered in practice as negative values of  $x$ , corresponding to when  $\bar{\mu} > V_{\text{res}}L$ , happen to be in the biological range. The same kind of issue is further amplified in  $h(x)$  as the integrand itself depends on  $g(x)$ . Second, even for the input range over which the functions are well behaved, direct evaluation of the MA is slow as it involves double or triple integrals. In the following, we present an efficient numerical scheme that overcomes these difficulties.

#### V. EFFICIENT NUMERICAL METHOD FOR THE MOMENT ACTIVATION

To achieve a reliable and fast numerical evaluation of the MA for arbitrary input values, our overall strategy is to look for direct numerical approximations to the Dawson-like functions  $g(x)$ ,  $h(x)$ ,  $G(x)$ , and  $H(x)$ . This allows us to efficiently evaluate the interspike interval in Eqs. (12) and (13) by computing  $G(x)$  and  $H(x)$  only at the integration bounds, thereby significantly reducing the computational complexity compared to explicit evaluation of the nested integrals. These approximations also enable efficient evaluation of the linear response coefficient  $\chi$  [Eq. (18)] and the derivatives of the MA. The next step of our strategy is to divide the entire input domain of the MA into five regimes, namely, the mean-dominant regime, the extended balanced regime, the weak fluctuation regime, the subthreshold regime, and the fluctuation-dominant regime. Figure 1(e) shows a schematic diagram of different regimes over the input domain. Different regimes are indicated by color patches, except for the weak fluctuation limit, which is indicated by the thick solid line corresponding to  $\bar{\sigma} \rightarrow 0$ . The dashed lines signify that the division into these regimes cover the entire upper half-plane. These regimes intercept at  $\bar{\mu} = V_\alpha L$  and  $\bar{\sigma} = 0$  where  $V_\alpha$  corresponds to either the firing threshold  $V_{\text{th}}$  or the reset potential  $V_{\text{res}}$ . They correspond to two sets of divisions depending on whether the Dawson-like functions are evaluated at the

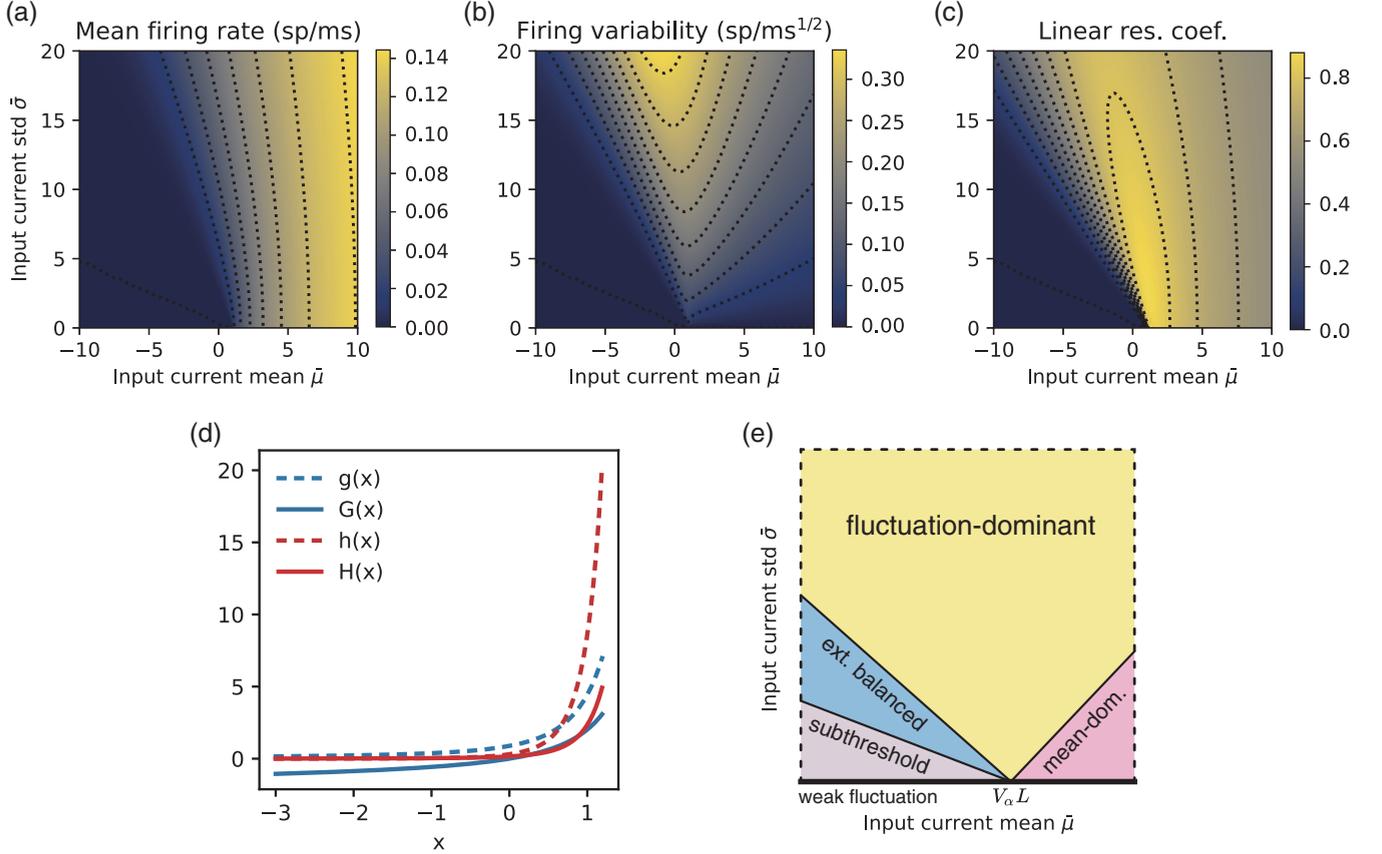


FIG. 1. The moment activation (MA). The MA maps the statistical moments of the input synaptic current to those of the output spikes. The three components of MA are (a) the mean firing rate  $\mu$ , (b) the firing variability  $\sigma$ , and (c) the linear response coefficient  $\chi$  used for the correlation map. Dotted lines represent contours. (d) The family of Dawson-like functions appearing in the MA exhibiting faster-than-exponential growth. (e) Schematic diagram showing a decomposition of the entire input domain of the MA into five regimes. The regimes intercept at  $\bar{\mu} = V_\alpha L$  and  $\bar{\sigma} = 0$  where  $L$  is the leak conductance and  $V_\alpha$  corresponds to either the firing threshold  $V_{th}$  or the reset potential  $V_{res}$ . The slopes of the boundaries depend on which one of the Dawson-like functions is being evaluated.

lower or upper integration bounds [Eqs. (12) and (13)]. The exact slopes of the boundaries separating the regimes also depend on which one of the Dawson-like functions is being evaluated. The above classification of input regimes serves two purposes: first, it provides a conceptual framework for interpreting neural response properties under different types of noisy inputs and, second, it provides practical guidance for designing efficient numerical strategies that suit the best to each input regime.

The mean-dominant, extended balanced, and subthreshold regimes correspond to when the magnitude of the input current mean is much larger than the input current standard deviation, that is, when  $\frac{|\bar{\mu} - V_\alpha L|}{\sqrt{L\bar{\sigma}}} \gg 1$ . In these regimes, the Dawson-like functions [Eqs. (14)–(17)] vanish or explode, as shown in Fig. 1(d), rendering direct numerical integration intractable. To overcome this, we construct asymptotic expansions for each of the Dawson-like functions  $g(x)$ ,  $h(x)$ ,  $G(x)$ , and  $H(x)$  with a suitable truncation. The weak fluctuation regime corresponds to when  $\bar{\sigma}$  is close to zero regardless of the value of  $\bar{\mu}$ , in which case we derive explicit analytical expressions for the MA. The fluctuation-dominant regime corresponds to the input range outside the aforementioned three regimes. For this regime, Chebyshev polynomial approximations with look-up tables for the coefficients are used.

In the following, we present details of these approximations for the MA under each of these regimes and their physical significance is also discussed. The derivatives of the MA are presented in Appendix A.

### A. Mean-dominant regime

We first consider the mean-dominant regime when neural firing is largely driven by positive input current mean, that is, when  $\frac{\bar{\mu} - V_\alpha L}{\sqrt{L\bar{\sigma}}} \gg 1$ , resulting in regular activity with high firing rates. Here we present asymptotic expansions of the Dawson-like functions [Eqs. (14)–(17)] as  $x \rightarrow -\infty$ , allowing us to efficiently and reliably evaluate the MA. To our knowledge, only the asymptotic expansion of  $g(x)$  has been previously reported in the literature.

The asymptotic expansion for  $g(x)$  as  $x \rightarrow -\infty$  is [31]

$$g(x) \sim \sum_{n=0}^{\infty} (-1)^{n+1} \frac{(2n-1)!!}{2^{n+1} x^{2n+1}}. \quad (20)$$

In fact, the function  $g(x)$  is related to the scaled complementary error function as  $g(x) = \frac{\sqrt{\pi}}{2} \operatorname{erfcx}(-x)$ , which has been implemented previously using a different approach based on continued fraction expansions [32].

The asymptotic expansion for  $G(x)$  is found by rewriting the integral form in Eq. (16) to an equivalent differential equation whose solution can be expressed as a generalized hypergeometric function. This in turn allows us to find its asymptotic expansion as  $x \rightarrow -\infty$ ,

$$G(x) \sim -\frac{1}{4}\gamma_e - \frac{1}{2}\log(-2x) + \sum_{n=1}^{\infty} \left(-\frac{1}{2}\right)^{n+2} \frac{(2n-1)!!}{n} \frac{1}{x^{2n}}, \quad (21)$$

where  $\gamma_e$  is Euler's constant. It is worth noting that  $G(x)$  is well behaved for  $x < 0$  as the leading term in the asymptotic expansion is logarithmic.

To find the asymptotic expansion for  $h(x)$ , we substitute the asymptotic expansion [Eq. (20)] for  $g(x)$  into Eq. (15) and formally expand the series. Then, by applying integration by parts to each term, we obtain the asymptotic expansion for  $h(x)$  as  $x \rightarrow -\infty$ ,

$$h(x) \sim \sum_{n=0}^{\infty} \frac{a_n}{x^{2n+3}}, \quad (22)$$

with coefficients found to be

$$a_n = \sum_{k=0}^n \sum_{l=0}^k \left(-\frac{1}{2}\right)^{n+3} \frac{(2l-1)!!(2k-2l-1)!!(2n+1)!!}{(2k+1)!!}. \quad (23)$$

Next, by integrating the asymptotic expansion of  $h(x)$  term by term, we obtain the asymptotic expansion for  $H(x)$  as

$$H(x) \sim \sum_{n=0}^{\infty} \frac{-a_n}{2n+2} \frac{1}{x^{2n+2}}, \quad (24)$$

where  $a_n$  is the same coefficients in Eq. (23).

Note that for numerical implementation, an appropriate level of truncation is applied to the asymptotic expansion to achieve a balance between accuracy and applicable input range. The mean firing rate  $\mu$  and the firing variability  $\sigma^2$  of the MA can then be evaluated by combining the approximations for  $G(x)$  and  $H(x)$  with Eqs. (10)–(13). For the correlation mapping, we evaluate the linear response coefficient  $\chi$  [Eq. (18)] using the derivative of the mean firing rate [see Eq. (A1) in Appendix A].

### B. Extended balanced regime

Next, we consider the extended balanced regime when the input current mean and std compete with each other, i.e., when  $\frac{\mu - V_{th}L}{\sqrt{L}\bar{\sigma}} \ll -1$ , to produce spiking activity at a low rate (only a few spikes per second). Here we present exact transformation formulas for the Dawson functions [Eqs. (14)–(17)] from the input domain of  $x < 0$  to that of  $x > 0$ , and also derive the leading terms as  $x \rightarrow +\infty$ .

The following identity is used to evaluate  $g(x)$  for  $x > 0$

$$g(x) = \sqrt{\pi}e^{x^2} - g(-x). \quad (25)$$

It is evident that the leading term is  $g(x) \sim \sqrt{\pi}e^{x^2}$  as  $x \rightarrow +\infty$ .

For  $G(x)$  we derive the following identity found by integrating both sides of Eq. (25):

$$G(x) = \frac{\pi}{2}\text{erfi}(x) + G(-x), \quad (26)$$

where  $\text{erfi}(x)$  is the imaginary error function, a well-known special function with existing numerical implementations. The leading term of Eq. (26) is found to be  $G(x) \sim \frac{\sqrt{\pi}}{2} \frac{e^{x^2}}{x}$ , which has been previously reported in [33].

For  $h(x)$  we derive the following identity:

$$h(x) = \sqrt{\pi}e^{x^2} \left[ \frac{1}{2} \log 2 + G(x) + G(-x) \right] - h(-x), \quad (27)$$

with the leading term found to be  $h(x) \sim \frac{\pi}{2} \frac{e^{2x^2}}{x}$ .

For  $H(x)$  the following identity holds for  $x > 0$ :

$$H(x) = \frac{\pi \log 2}{4} \text{erfi}(x) + \sqrt{\pi} \int_0^x e^{u^2} [G(u) + G(-u)] du + H(-x). \quad (28)$$

No practically useful simplification is found for Eq. (28). Therefore, we approximate  $H(x)$  with its leading term  $H(x) \sim \frac{\pi^2}{8} [\text{erfi}(x)]^2 \sim \frac{\pi}{8} \frac{e^{2x^2}}{x^2}$ , as  $x \rightarrow +\infty$ .

### C. Weak fluctuation regime

The weak fluctuation regime corresponds to when the input current variability  $\bar{\sigma}$  is close to zero, regardless of the value of the input current mean  $\bar{\mu}$ . In this scenario, the integration bounds in Eqs. (12) and (13) contain singularities as the input current variability  $\bar{\sigma} \rightarrow 0$ , making it unsuitable for numerical implementation. However, these singularities are removable as the moment activation is well behaved when  $\bar{\sigma} \rightarrow 0$ . By considering the leading terms of  $g(x)$ ,  $G(x)$ ,  $h(x)$ , and  $H(x)$  as  $x \rightarrow +\infty$ , we find that the corresponding limits exist and have simple analytical expressions as presented below.

The limit for the mean firing rate  $\mu$  is

$$\lim_{\bar{\sigma} \rightarrow 0} \mu(\bar{\mu}, \bar{\sigma}) = \begin{cases} 0, & \text{for } \bar{\mu} \leq V_{th}L, \\ \frac{1}{T_{ref} - \frac{1}{L} \log\left(1 - \frac{V_{th}L}{\bar{\mu}}\right)}, & \text{for } \bar{\mu} > V_{th}L. \end{cases} \quad (29)$$

This limit is consistent with the solution of the leaky integrate-and-fire neuron model receiving a constant input current [30], as can be obtained by integrating Eq. (1) directly.

For the variance mapping, we note that the limit of the Fano factor as  $\bar{\sigma} \rightarrow 0$  is simply the Heaviside step function

$$\lim_{\bar{\sigma} \rightarrow 0} \frac{\sigma^2}{\mu} = \begin{cases} 1, & \text{for } \bar{\mu} \leq V_{th}L, \\ 0, & \text{for } \bar{\mu} > V_{th}L. \end{cases} \quad (30)$$

Combining this result and Eq. (29), we conclude that

$$\lim_{\bar{\sigma} \rightarrow 0} \sigma(\bar{\mu}, \bar{\sigma}) = 0.$$

For the linear response coefficient  $\chi$  in Eq. (18), the limit as  $\bar{\sigma} \rightarrow 0$  is

$$\lim_{\bar{\sigma} \rightarrow 0} \chi(\bar{\mu}, \bar{\sigma}) = \begin{cases} 0, & \text{for } \bar{\mu} \leq V_{th}L, \\ \sqrt{\frac{2}{L}} \frac{1}{\sqrt{T_{ref} - \frac{1}{L} \log\left(1 - \frac{V_{th}L}{\bar{\mu}}\right)} \sqrt{\frac{2\bar{\mu}}{V_{th}L} - 1}}, & \text{for } \bar{\mu} > V_{th}L. \end{cases} \quad (31)$$

These limits can then be used to approximate the moment activation when  $\bar{\sigma}$  is very close to zero, in which case

Eqs. (12) and (13) become numerically intractable. Similar limits can be derived for the gradient of the moment activation (see Appendix A).

#### D. Subthreshold regime

The subthreshold regime corresponds to when both the input current mean  $\bar{\mu}$  as well as the variability  $\bar{\sigma}$  are weak so that the neuron receiving the input ceases firing. Concretely, this corresponds to when  $I_{ub}(\bar{\mu}, \bar{\sigma}) = \frac{V_{th}L - \bar{\mu}}{\sqrt{L\bar{\sigma}}} > \theta$  for some sufficiently large positive number  $\theta$ . In this scenario, the integrals in Eqs. (12) and (13) explode and all components of the moment activation including  $\mu$ ,  $\sigma$ , and  $\chi$  vanish. The quantity  $\theta$  can thus be viewed as a form of generalized firing threshold, below which the output is simply set to  $\mu = \sigma = \chi = 0$ .

#### E. Fluctuation-dominant regime

The fluctuation-dominant regime is when neural firing is largely driven by fluctuations in the input current. This corresponds to the input range outside the aforementioned regimes. In this regime, direct numerical integration for the Dawson-like functions is possible but slow. To overcome this, we follow the strategy previously used for implementing the scaled complementary error function by using Chebyshev polynomial approximations with lookup tables for the coefficients [32].

For  $x \leq 0$ , we first apply the transformation  $x' = \frac{4}{4-x}$ , which maps the input  $x \in (-\infty, 0]$  to the unit interval  $x' \in (0, 1]$ . We then divide the unit interval into  $N$  subintervals of equal length and fit (in the least-square sense) the function over each subinterval with a Chebyshev polynomial of an appropriate degree. The coefficients of the polynomial expansion are then saved to a look-up table which can then be used for fast evaluation of each Dawson-like function. Special identities [Eqs. (25)–(27)] can then be used to evaluate the functions for  $x > 0$ . This general strategy is applied to all of  $g(x)$ ,  $G(x)$ ,  $h(x)$ ,  $H(x)$  but with a couple of exceptions. First, for  $G(x)$  the subdivision is applied directly over the interval  $x \in [-c, 0]$  for some constant  $c$  without the transformation because  $G(x)$  does not vanish as  $x \rightarrow -\infty$  and grows logarithmically to negative infinity. Second, since there is no special identity relating  $H(x)$  with  $H(-x)$ , we apply the Chebyshev polynomial approximation to  $x > 0$  as well, which is done by fitting  $\tilde{H}(x) = H(x)e^{-2x^2}$  to a Chebyshev polynomial for each subintervals over  $x \in (0, c]$ .

As mentioned earlier, the MA [Eqs. (10)–(18)] is derived directly from the LIF neuron model [Eq. (1)] through a series of mathematically rigorous approximations. There are three potential sources of error due to these approximations, namely, the diffusion approximation for the synaptic current, the assumption for stationary process, and the linear response approximation for correlation mapping. We benchmark the accuracy of the MA for approximating the LIF neuron model [Eq. (1)] for a single neuron in Appendix C and investigate the conditions under which the linear response theory is valid for calculating the correlation map in Sec. VI. Benchmark results for the computational efficiency of the MA are presented in Appendix D.

## VI. NONLINEAR RESPONSE PROPERTIES IN THE EXTENDED BALANCED REGIME

As mentioned earlier, one of the approximations used for deriving the MA is the linear response theory for obtaining the pairwise correlation map of LIF neurons [21]. Conceptually, the linear response theory provides a first-order approximation to the correlation map near  $\bar{\rho}_{ij} = 0$ , and is thus the most accurate for weakly correlated neural activity. However, at a quantitative level, there is currently a lack of understanding about the conditions under which this approximation is valid. The numerical method developed in this work enables us to systematically investigate this problem, particularly for those ill-conditioned input regimes where the MA could not be reliably evaluated using previous methods. As we show below through numerical simulations, the linear response approximation is accurate for the vast majority of input regimes but breaks down around the extended balanced regime. Since we are concerned only with pairwise correlations, it is sufficient to consider two neurons without loss of generality. We treat the aggregated postsynaptic currents as correlated Gaussian random variables without explicitly modeling the input spike trains. This treatment allows us to separate the effect due to the linear response approximation from that due to the diffusion approximation. Concretely, the input currents received by the pair of neurons are

$$I_i(t) = \bar{\mu}_i + \bar{\sigma}_i \xi_i(t),$$

where  $\bar{\mu}_i$  and  $\bar{\sigma}_i$  are the mean and standard deviation of the input current for neuron  $i \in \{1, 2\}$ , and  $\xi_i(t)$  are Gaussian white noise with a correlation coefficient equal to  $\bar{\rho}$ .

Our goal is to perform a parameter sweep through the full space spanned by the input current mean/std of pairs of neurons, plus the correlation between them, and to compare the theoretical correlation map with empirical sample estimates from simulated spiking neurons. Although sweeping through this five-dimensional space is computationally prohibitive for spiking neurons, we show that by avoiding redundancy one only needs to sweep through a 3D space. See Appendix B for details.

To establish some basic intuition about the correlation maps, we first show the correlation maps along specific slices through the mean/std parameter space. First, we fix  $\bar{\mu}_2 = 0$  and  $\bar{\sigma}_2 = 2.5$  while varying  $\bar{\mu}_1$  and  $\bar{\sigma}_1$  [Fig. 2(a)], and second, we vary both inputs at the same time with  $\bar{\mu}_1 = \bar{\mu}_2$  and  $\bar{\sigma}_1 = \bar{\sigma}_2$  [Fig. 2(b)]. For most input values, the linear response theory provides accurate predictions (solid lines) to LIF neurons (dashed lines) even for correlation coefficients away from zero. For some input values, the predictions based on linear response theory (solid lines) deviate away from the LIF neurons (dashed lines) for  $|\bar{\rho}| > 0$ . This deviation becomes more apparent for inputs closer to  $\bar{\rho} = 1$ . Based on these observations, it is evident that the MA based on linear response theory provides reasonably accurate predictions of the correlation mapping for most regions of the input space, but the quality of approximation degrades for other regions.

To effectively visualize the high-dimensional results, we calculate the L2 distance between the theoretical and empirical correlation maps for each pair of input mean/std to obtain a 4D heat map. We visualize this 4D heat map as a 2D array

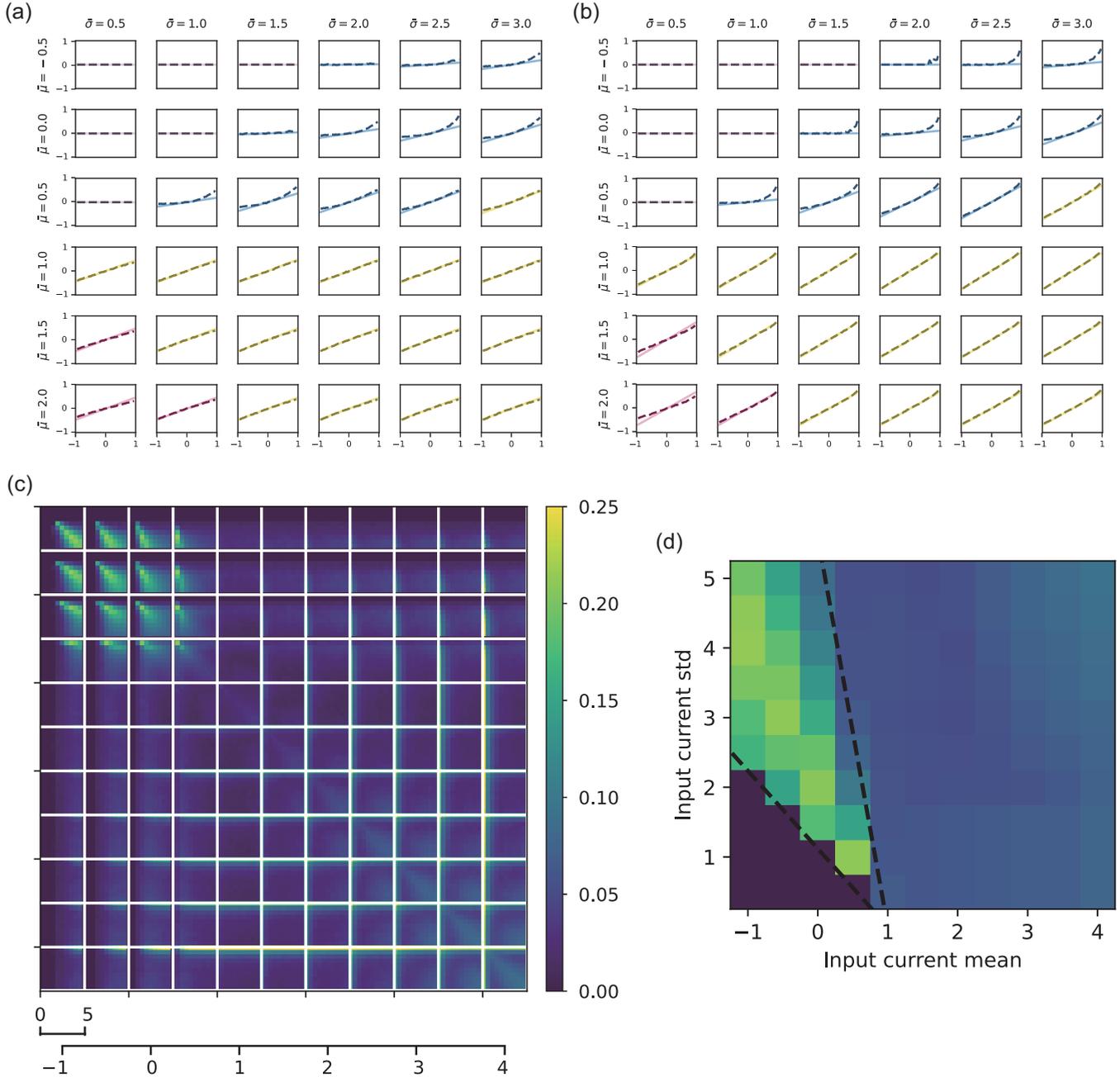


FIG. 2. Correlation map of the moment activation (MA). (a), (b) Correlation map between a pair of LIF neurons receiving correlated input currents with varying statistics. In each panel, horizontal and vertical axes correspond to input and output correlation, respectively. In (a) we fix  $\bar{\mu}_2 = 0$  mV/ms and  $\bar{\sigma}_2 = 2.5$  mV/ms<sup>1/2</sup> (extended balanced regime) while varying  $\bar{\mu}_1$  and  $\bar{\sigma}_1$ . In (d) we vary both inputs with  $\bar{\mu}_1 = \bar{\mu}_2$  and  $\bar{\sigma}_1 = \bar{\sigma}_2$ . In both cases the linear response theory (solid line) is largely consistent with simulations of LIF neurons (dashed line), but shows deviation for  $|\bar{\rho}| > 0$  particularly in the extended balanced regime. The color code indicates different input regimes. (c) A 4D heatmap showing the L2 distance between the empirical and theoretical correlation maps. The inner dimensions (axes in each tile) correspond to the input std  $\bar{\sigma}$  of two neurons, whereas the outer dimensions (axes across tiles) correspond to the input mean  $\bar{\mu}$  of two neurons. (d) The maximum of L2 distance given the input mean and std of one neuron, over all input mean and std of the other neuron in the range shown in (c). Dashed line indicates the boundaries between the subthreshold and extended balanced regime, and between the fluctuation-dominant and extended balanced regime. Color scale is as in (c).

of 2D heat maps, as shown in Fig. 2(c). The inner dimensions (the axes of each panel) correspond to the input std of two neurons, whereas the outer dimensions (the axes across panels) correspond to the input mean of two neurons. Let us first focus on the outer dimensions. It can be seen that for the

vast majority of the input space, i.e., input mean  $\bar{\mu} \geq V_{th}L = 1.0$  mV/ms for any one of the input neurons, the MA based on linear response theory accurately predicts the correlation mapping. Inaccuracies in the correlation map due to the linear approximation are concentrated in the range where the input

mean  $\bar{\mu} < 1$  mV/ms. Next, let us turn our attention to individual panels within this range. Within each panel, inaccuracies due to linear approximation only begin to appear beyond a minimum threshold in the input std  $\bar{\sigma}$ , and this threshold decreases with the input mean  $\bar{\mu}$ . In terms of the taxonomy presented in Fig. 1(e), the region below this threshold corresponds to the subthreshold regime, in which neurons do not fire and the correlation is zero. As it turns out, the input regime in which the linear response theory breaks down largely overlaps with the extended balanced regime, in which input fluctuations and negative input mean compete to produce spiking activity with low firing rates. This can be seen by marginalizing the 4D heat map to the 2D plane spanned by the input mean/std of one of the neurons, as shown in Fig. 2(d). Moreover, we find that the amount of deviation of the theoretical correlation map from the empirical result is primarily determined by the effective input drive  $\gamma = \frac{\sqrt{L}\bar{\sigma}}{V_{th}L - \bar{\mu}}$ . As shown in Fig. 2(d), the L2 distance along the line defined by the effective drive  $\gamma$  is at a maximum around  $\gamma = 0.5$  and then quickly decreases as  $\gamma$  increases. We find that for  $\gamma > 1.25$  the correlation map derived from the linear response theory agrees with the empirical ground truth. The large deviation occurring in the range of  $0.25 < \gamma < 1.25$  suggests that nonlinear effects dominate in these regimes. It is worth noting that even in this regime, the linear response theory still provides a reasonably accurate approximation to the correlation map for weakly correlated inputs, as can be seen from Figs. 2(a) and 2(b), but care must be taken when the input correlation becomes strong.

It has been theorized that cortical neurons in the brain operate in a balanced regime, in which the average excitation and inhibition roughly cancel out, and that neural activity is primarily driven by fluctuations in the input. Our analysis shows that part of this balanced regime overlaps with the input regime where the linear response theory breaks down, indicating the necessity of developing higher-order approximations of the correlation map. The strong dependence of nonlinear effects in the correlation map on effective drive  $\gamma$  also suggests that  $\gamma$  could be a more appropriate quantifier for dynamical regimes of fluctuation-driven spiking activity than simple E-I balance, and that correlated variability could play an important role in the computational properties of balanced state in neural circuits [25].

## VII. MOMENT INTERACTIONS IN LARGE-SCALE NEURAL CIRCUIT

Having established the efficient numerical scheme for the MA, we now demonstrate how the MNN can be used for modeling the coupled interactions of mean firing rate and firing variability in large-scale neural circuits. For this purpose, the usage of the MNN as presented in this work has a number of advantages. First, the MNN enables a closed and self-consistent description of fluctuating neural dynamics up to the second order. This is a significant improvement over mean field analysis which commonly imposes additional constraints such as Poisson firing statistics (i.e., Fano factor equals one) [15,27]. As a result, the MNN is capable of expressing a wide range of Fano factors as is consistent with experimental observations of cortical neurons. Second,

mean-field analysis often performs ensemble averaging over independent realizations of randomly connected neural networks, erasing potential contributions from synaptic in-degree heterogeneity, i.e., quenched disorder. The efficiency of the numerical scheme developed in this work enables full simulation of large-scale neural circuits at the level of individual neurons, allowing for the investigation of dynamical effects associated with quenched disorder.

### A. Diverse Fano factor and irregular oscillatory activity

We consider a large, sparsely connected network of excitatory and inhibitory neurons with settings similar to those of [15]. We find from numerical simulations that the spiking activity in this model is uncorrelated, likely due to the sparsity of the network; therefore we restrict this section to the uncorrelated MNN, that is,  $\rho_{ij} = 0$  for  $i \neq j$ . The omission of correlation also greatly reduces the computational cost, allowing us to simulate the interaction of second-order moments of a very large network at single-neuron resolution. To incorporate synaptic delay, we replace Eqs. (5) and (6) for the postsynaptic current with

$$\bar{\mu}_i^\alpha(t) = \sum_{\beta,j} w_{\alpha\beta} a_{ij}^{\alpha\beta} \mu_j^\beta(t - D) + w_{\alpha,\text{ext}} \mu_{\text{ext}}, \quad (32)$$

$$(\bar{\sigma}_i^\alpha)^2(t) = \sum_{\beta,j} w_{\alpha\beta}^2 a_{ij}^{\alpha\beta} (\sigma_j^\beta)^2(t - D) + w_{\alpha,\text{ext}}^2 \mu_{\text{ext}}, \quad (33)$$

where  $\alpha, \beta \in \{E, I\}$  are indices for excitatory and inhibitory populations and  $D$  is the synaptic delay. We consider a recurrent network consisting of  $N_E = 10\,000$  excitatory neurons and  $N_I = 2500$  inhibitory neurons and the connectivity matrix  $a_{ij}^{\alpha\beta} = 1$  with probability  $p = 0.1$  and zero otherwise. The excitatory synaptic weights are fixed at  $w_{EE} = w_{IE} = 0.1$  mV and the inhibitory synaptic weights are  $w_{EI} = w_{II} = -g w_{EE}$ , where  $g$  is the inhibition-to-excitation synaptic weight ratio (IE ratio). Self-connections are excluded. For this section, the neuronal parameters for the MA are changed to  $V_{\text{res}} = 10$  mV and  $T_{\text{ref}} = 2$  ms to be consistent with [15]. We model external inputs as Poisson spike trains with a mean firing rate  $\mu_{\text{ext}}$  that is homogeneous across all neurons and set the weights to  $w_{\alpha,\text{ext}} = w_{EE} = 0.1$  mV. We fix the time constant in Eqs. (7)–(9) to be  $\tau = 1$  a.u. and the delay to be  $D = 0.5$  a.u., but we will show how they can be calibrated with respect to physical time unit in Appendix E. All other parameters for the MA remain unchanged.

To investigate the dynamics of this recurrent moment neural network with delay, we systematically vary the external input rate  $\mu_{\text{ext}}$  and the IE ratio  $g$  and construct phase diagrams using the population-averaged mean firing rate and Fano factor, which are calculated by evolving the model for a sufficiently long period of time to obtain the stable fixed points and then average them across all neurons, and over a large time window in the case of oscillations. As shown in Fig. 3(a), three distinct phases are visible: a mean-dominant phase with high firing rate and vanishing Fano factor when inhibition is weak, a fluctuation-dominant phase with lower firing rate and large Fano factor when inhibition is stronger, and an irregular oscillatory phase with low mean firing rate and low but nonvanishing Fano factor. To better understand

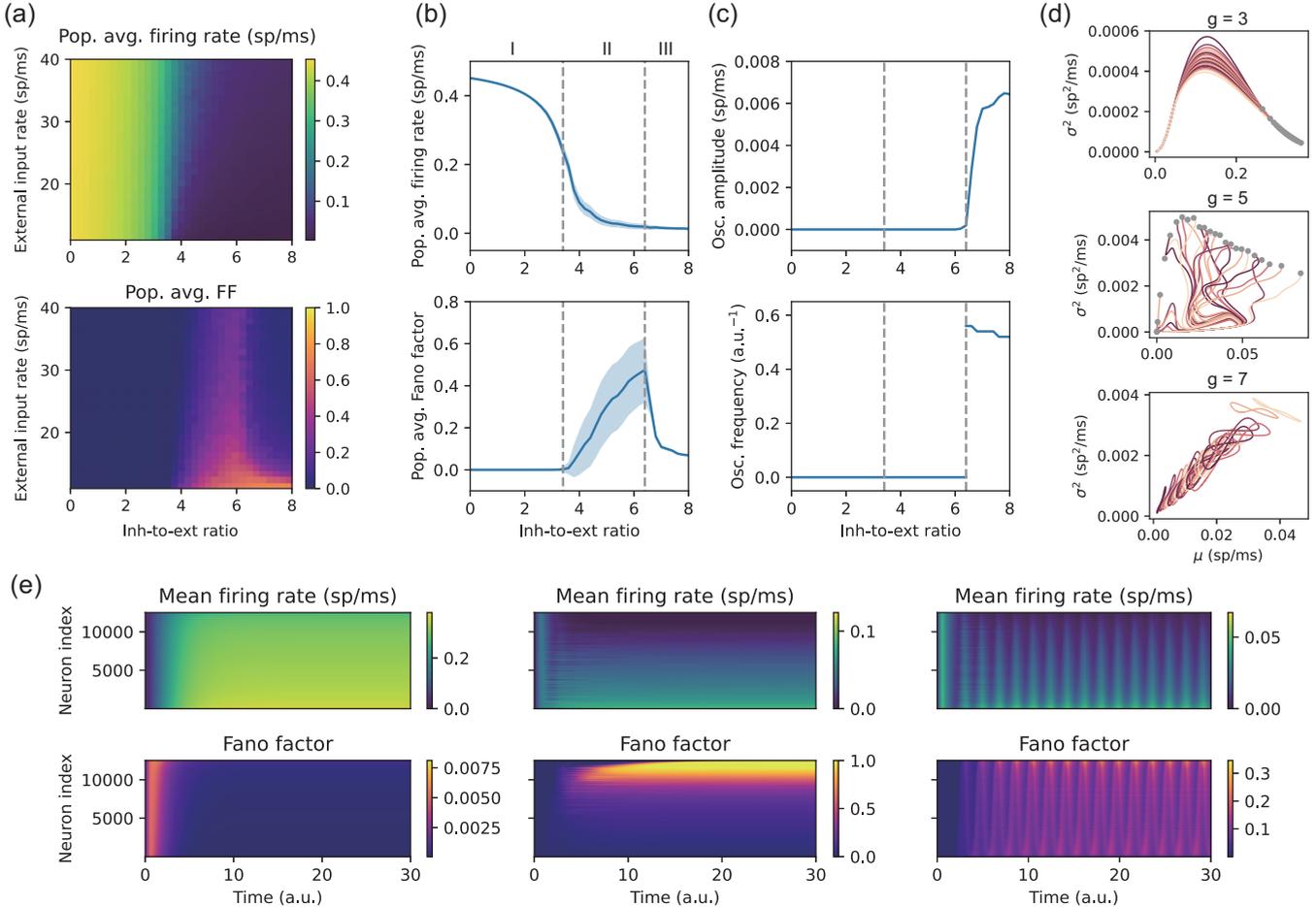


FIG. 3. Dynamics of recurrent moment neural network with delay. (a). Phase diagrams with varying excitation-to-inhibition weight ratio  $g$  and external input rate  $\mu_{\text{ext}}$ . Three distinct phases correspond to mean-dominant, fluctuation-dominant, and irregular oscillatory activity. Here we fix delay  $D = 0.5$  a.u. (b). A slice of the phase diagram along  $\mu_{\text{ext}} = 20$  sp/ms. The solid line and shades indicate population averages  $\pm 0.5$  std. The dashed lines mark the critical points at  $g = 3.4$  and  $g = 6.4$  between three phases: (I) mean-dominant, (II) fluctuation-dominant, (III) irregular oscillatory. (c). Oscillation amplitude and frequency of population averaged firing rate along the same slice. (d) Typical examples of temporal trajectories of neural activity projected onto the mean-var plane. Top, middle, and bottom panels correspond to mean-dominant ( $g = 3$ ), fluctuation-dominant ( $g = 5$ ), and irregular oscillatory ( $g = 7$ ) activity; each curve corresponds to a representative neuron in the network; dots represent stable fixed points. For  $g = 7$ , the stable limit cycles are shown. Color is for visual contrast only. (e) Spatiotemporal activity patterns at  $g = 3$  (homogeneous activity),  $g = 5$  (heterogeneous activity), and  $g = 7$  (oscillations). Neurons are ranked based on their mean firing rates.

these phases, we show in Fig. 3(b) a slice of the phase diagram along  $\mu_{\text{ext}} = 20$  sp/ms. For the mean-dominant phase, the population-averaged mean firing rate decreases with increasing IE ratio, whereas the Fano factor remains at zero, indicating highly regular spiking activity. A critical phase transition occurs at  $g = 3.4$  beyond which the system enters fluctuation-dominant phase. Near the phase transition, the mean firing rate appears to decrease smoothly with  $g$ , but a sharp boundary separating the two phases is revealed by the Fano factor which starts to increase beyond the critical point. This sudden change in the behavior of the system is not limited to the population-averaged Fano factor but also a sudden surge in its variability across neurons as indicated by the shades in Fig. 3(b).

As the IE ratio increases further, the population-averaged mean firing rate continues to decrease while the population-averaged Fano factor and the strength of its het-

erogeneity increase until reaching a peak at  $g = 6.4$ . This is the critical point at which the system transitions into irregular oscillatory phase, as characterized by low firing rate and nonvanishing Fano factor that oscillate over time. This transition to oscillatory activity is more clearly summarized in Fig. 3(c) showing the oscillation amplitude of the population-averaged instantaneous firing rate  $\mu(t)$ , calculated as  $A = \frac{1}{2}[\max_t \mu(t) - \min_t \mu(t)]$ , and the oscillation frequency, which does not vary significantly with the IE ratio. In spiking neural networks, this kind of activity corresponds to global oscillations in the population-averaged firing rate but irregular spiking activity at the level of individual neurons [34–36]. Such irregular spiking activity and collective oscillations are ubiquitous features of cortical neurons in the brain and may appear paradoxical at first glance. The moment neural network presented here provides an elegant mathematical formalism for describing the coexistence of

irregular firing (diverse instantaneous Fano factor) and collective oscillation (instantaneous mean firing rate and firing variability together vary over time), and for explaining how collective oscillatory dynamics can emerge in a network driven by noisy inputs even when individual neurons fire irregularly.

To provide further intuition to the coupled interactions of the mean firing rate and firing variability, we show typical examples of neural activity state for each of the three phases. As shown in Fig. 3(d), temporal trajectories of neural activity are projected onto the mean-var plane of the neural activity state. We find that for the mean-dominant phase ( $g = 3$ ), the transient trajectories for different neurons start from the initial value at the origin and evolve toward closely packed stable fixed points in an ordered, laminar fashion. In contrast, for the fluctuation-dominant phase ( $g = 5$ ), the transient trajectories for different neurons appear to be irregular and turbulent, crossing each other consistently in this 2D projection. The resulting stable fixed points are scattered over a large region of the mean-var plane, forming a smoothly shaped manifold, which reflects the broad distribution of the Fano factor. For the irregular oscillatory phase ( $g = 7$ ), we show the trajectories of the neural activity state after they have converged, revealing limit cycles forming the shape of a figure “8.” Additional plots of the spatiotemporal activity patterns corresponding to these examples are shown in Fig. 3(e) where neurons are ranked based on their mean firing rates. It can be seen that the neural activity in the mean-dominant phase is largely homogeneous across all neurons, but in the fluctuation-dominant phase it is heterogeneous, as indicated by the diverse Fano factor ranging from 0 to 1. For the irregular oscillatory phase, the instantaneous mean firing rate and Fano factor of different neurons oscillate synchronously over time. Curiously, there appears to be some sort of phase splitting in the Fano factor of different neurons: oscillations in neurons with the highest and the lowest mean firing rates exhibit opposite phases, whereas for all other neurons in between the oscillations are superposition of these two phases with a continuous shift.

For the type of homogeneous random network (Erdős-Rényi network) considered here, the synaptic in-degrees are Poisson distributed, so it is commonly assumed that as the population size grows the distribution of the in-degree should become increasingly concentrated around its mean. This motivates the idea that for this type of homogeneous random network, the dynamics of the system should be self-averaging in the limit of large system size such that the resulting neural activity is also homogeneous. This is the assumption underlying many mean field analyses aimed at describing the population-averaged neural activity, as if all neurons inside a population have identical firing properties. So the emergence of strong heterogeneity in the fluctuation-dominant phase comes at quite a surprise. This raises the question of what the mechanism is for the emergence of diverse neural variability (i.e., Fano factor) in the fluctuation-dominant phase. To address this problem, we fix the average synaptic in-degree while decreasing its variance and find that the strength of heterogeneity in the Fano factor decreases. When the in-degrees are equal to a constant, the resulting neural activity becomes completely homogeneous. This suggests that the heterogeneous activity is due to quenched disorder and a potential explanation for the emergence of diverse Fano factor is sym-

metry breaking associated with quenched disorder, causing the self-averaging assumption to no longer apply. The detailed mechanism warrants further investigation but is beyond the scope of the present paper.

To enable meaningful interpretations of these results, we must resolve the issue of the undetermined time constant  $\tau$  in the MNN. The aim is to estimate  $\tau$  so that the oscillation frequency in the MNN matches that in the SNN. The main difficulty is that the oscillation frequency is influenced by both the time constant and the delay so that they must be determined simultaneously. Here we propose a calibration procedure based on a scaling argument about the frequency-delay curve, allowing us to uniquely determine the calibration factor that works for all delay values, given that other parameters are fixed (see Appendix E for details). For the model considered here, we find that the calibration factor is equal to  $\beta = 1.168$  ms/a.u. which gives a calibrated time constant of  $\tau' = 1.168$  ms, significantly shorter than the membrane time constant  $\tau_m = 20$  ms of an isolated spiking neuron. As shown in Fig. 7(c), the frequency-delay curves for the SNN and the calibrated MNN are in agreement for all delay values, and the MNN predicts a critical delay value around  $D^* = 0.5$  ms above which oscillations emerge, as consistent with the SNN model.

## B. Correlated variability in neural circuits

In this section we consider a recurrent network consisting of  $N_e = 0.8N$  excitatory neurons and  $N_i = 0.2N$  inhibitory neurons. Synaptic weights  $w_{ij}^{\alpha\beta}$  are drawn randomly from a normal distribution with mean  $J^{\alpha\beta}/N_\alpha$  and standard deviation  $K/N_\alpha$ . For this model, self-connection is excluded. The synaptic weight parameters are set to  $J^{EE} = 120$ ,  $J^{EI} = -190$ ,  $J^{IE} = 108$ ,  $J^{II} = -178$ , and  $K = 40$ ; the connection probability is set to  $p = 0.2$ .

The external input is characterized by its mean  $\bar{\mu}_{\text{ext}}(s)$  and covariance  $\bar{C}_{\text{ext}}(s)$ , which in general can depend on a latent feature  $s$  of the stimulus. For illustrative purposes, here we set the external input mean to be  $\bar{\mu}_{\text{ext}}(s) = \kappa s$  where  $s$  is a scalar parameter and  $\kappa = 1$  mVms<sup>-1</sup>. We then compare the neural response and coding properties under two different types of input covariance. One of these input covariances is a spatially unstructured, uniform covariance defined as  $(\bar{C}_{\text{ext}})_{ij} = \bar{\sigma}_{\text{ext}}^2[(1-c)\delta_{ij} + c]$ , where  $\bar{\sigma}_{\text{ext}}^2$  is the noise strength and  $0 \leq c \leq 1$  controls the relative strength of independent and shared noise sources. The other type of input covariance is spatially structured cosine-shaped covariance defined as  $(\bar{C}_{\text{ext}})_{ij} = \bar{\sigma}_{\text{ext}}^2[(1-c)\delta_{ij} + c \cos(\theta_i - \theta_j)]$ . Here we have assigned the inputs received by excitatory neurons with spatial coordinates evenly spaced on the unit circle  $\theta_i \in [0, 2\pi)$ . For inputs to inhibitory neurons, uniform covariance is used. In this paper, we fix the stimulus value to be  $s = 1.825$  and the external noise strength to be  $\bar{\sigma}_{\text{ext}}^2 = 32.6$  mV<sup>2</sup> ms<sup>-1</sup>. The steady state of the model is obtained by evolving Eqs. (7)–(9) for a duration of  $T = 20$  a.u. with a time step  $\delta t = 0.1$  a.u.

The neural response properties of the MNN model that receives these two types of input are shown in Fig. 4(a). Both the case with uniform input covariance (upper panel) and the case with cosine-shaped input covariance (lower panel) show a mean firing rate within the biologically realistic range (around

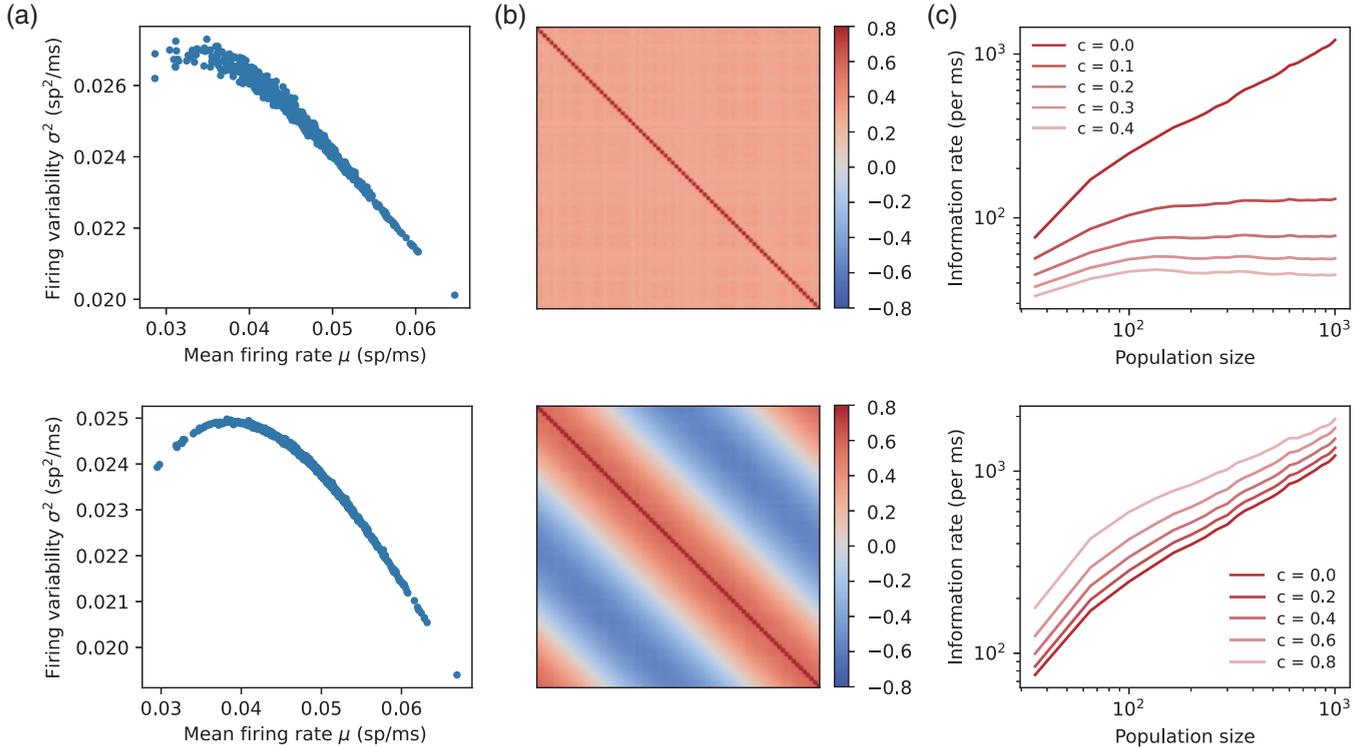


FIG. 4. Neural response and coding properties of large-scale neural circuits modeled with the MA. (a) Scatter plot showing the mean firing rate  $\mu$  and firing variability  $\sigma^2$  of neurons in a recurrent network receiving spatially uniform input covariance (upper panel) and cosine-shaped covariance (lower panel); in both cases the strength of input correlation is set to be  $c = 0.4$ . Both cases exhibit biologically plausible firing rate and diverse fluctuations with Fano factors ranging from 0.3 to 1. (b) The correlation coefficient of excitatory neurons in the recurrent network. (c) The scaling behavior of the linear Fisher information rate as the population size grows. For uniform input covariance, information quickly saturates as population size grows except for uncorrelated inputs ( $c = 0$ ); for cosine-shaped covariance, the information does not saturate with population size regardless of the strength of input correlation.

50 sp/s) and fluctuations with diverse Fano factor ranging from 0.3 to 1.0, similar to that found in cortical neurons in the brain [4]. The correlation coefficients  $\rho_{ij}$  of the neural response in the MNN are shown in Fig. 4(b).

The variable of interest here is the amount of information encoded by the neural activity in the recurrent network about the latent feature  $s$  in the stimulus. This can be quantified by the linear Fisher information rate  $I = \left(\frac{d\mu}{ds}\right)^T C^{-1} \frac{d\mu}{ds}$ , where  $\mu$  and  $C$  are the mean and covariance of neural activity. The linear Fisher information rate describes the amount of information per unit time about the latent feature  $s$ , which can potentially be extracted by a linear decoder. Previously, it has been suggested that information could saturate as population size increases due to a type of correlation called differential correlation [37,38]. Here we use the MNN model to investigate whether information saturates with population size  $N$  under different input covariance structures.

We find that these two types of correlation structures lead to distinct scaling behaviors in the linear Fisher information [Fig. 4(c)]. For uniform correlation, the linear Fisher information quickly saturates as the population size grows, except when the input noise is independent (i.e., the strength of shared noise is zero). The linear Fisher information also decreases as the strength of shared noise increases. This result is consistent with previous findings based on direct numerical simulations of the spiking neural network model [37] and

analysis based on recordings of large neural ensembles in the rodent brain [39]. In contrast, for spatially structured correlation, the linear Fisher information does not saturate with population size regardless of the relative strengths of independent and correlated noises in the stimulus. Moreover, the information increases with the relative strength of the correlated input noise (controlled by the parameter  $c$ ). These results show that neural correlation can potentially be exploited to enhance neural coding rather than to limit it. Full details of the role of correlation in neural coding are beyond the scope of this paper and are discussed elsewhere [40].

## VIII. DISCUSSION

In this study, we have developed an efficient numerical method for the moment activation (MA) through a combination of strategies that provide both reliability and speed. The proposed numerical scheme overcomes the numerical instability caused by a group of ill-conditioned Dawson-like integrals in the MA through asymptotic approximation, allowing for a reliable evaluation of the MA for arbitrary input range. Moreover, the proposed method circumvents multiple nested integrals in the MA and reduces the computation to finite series expansion, thus vastly reducing the cost of evaluating the MA. The proposed method is thus more effective than previous methods for evaluating neural firing

statistics which require numerically solving the associated Fokker-Planck equation [18,29]. We have also demonstrated the effectiveness of the MA for modeling large-scale neural circuits and for investigating the role of correlation in neural coding. The numerical method for evaluating the Dawson-like functions may also find application in studying other physical systems where these integrals naturally arise.

The method developed in this study provides a comprehensive numerical tool for evaluating the moment mapping for spiking neuron models under the diffusion formalism and has a number of key advantages. First, the MA deals with all three components of the moment mapping of a spiking neuron as well as their derivatives, whereas previous methods often focus only on the mean firing rate mapping while omitting variance and correlation mappings [17,29]. The latter two are crucial for enabling a closed, self-consistent system of equations that incorporates the first- and second-order moments of neural activity [16,21]. Second, the MA is highly reliable for the entirety of its input domain (the closed upper half-plane), due to its efficient usage of asymptotic approximations to the family of ill-conditioned Dawson-like integrals. Although the idea of using asymptotic approximation for analyzing neural spike statistics has been explored previously [33,41], it has considered only the large firing threshold limit (equivalent to the subthreshold regime in this paper). Here we have completed the missing pieces by including new results for the mean-dominant regime and exact transformation formulas linking it to the subthreshold regime. Third, the computational time for evaluating the MA is not only low but also highly consistent across different input regimes.

The MA powered by the proposed numerical scheme has potential applications in a number of areas of computational neuroscience. Derived from spiking neuron models on a mathematically rigorous ground, the MA faithfully captures correlated fluctuations of neural spikes and provides an ideal tool for modeling correlated neural variability. Specifically, the computational efficiency and scalability of the proposed numerical method can enable simulations of correlation propagation through large-scale cortical circuits and provide new insights about cortical computation previously unobtainable with direct simulation of spiking neurons or simplified firing rate models. The efficient implementation of the derivatives of the MA also provides a tool for a semianalytical approach to investigating the dynamical properties of correlated neural fluctuations in neural circuits.

The observations of strongly irregular firing activity of cortical neurons have also led to the idea that neural computation is fundamentally probabilistic. A number of theories for probabilistic neural computation have been proposed, such as probabilistic population code and neural sampling [42–46]. However, theoretical investigation of probabilistic neural computation involving correlated neural activity faces significant challenges in that high-dimensional joint probability distribution of neural activity in a large network is analytically and computationally intractable. As a result, it is common to assume independent Poisson spikes in the asynchronous regime to facilitate theoretical analysis. The proposed numerical method for evaluating the MA provides an alternative approach to model probabilistic neural computation through a

moment-space representation, without requiring explicit calculation of the joint probability density or random sampling.

The approach developed in this study can potentially be extended in two directions. First, the MA considered here is based on a particular type of spiking neuron model, that is, the current-based leaky integrate-and-fire neuron model. In the future, the proposed method may be extended to other types of neuron models to incorporate biological features such as synaptic conductance [17–19] and multiple types of synapses with slow/fast timescales [22]. Moreover, the present numerical method for the MA considers pairwise covariance without temporal lag and future works may extend this to incorporating cross-covariance to fully capture the rich spatiotemporal covariance structure of cortical networks [22].

The code for the proposed numerical method is available at [47].

### ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (Grant No. 62306078) and STI2030-Major Projects (No. 2021ZD0200204).

### APPENDIX A: DERIVATIVES OF THE MOMENT ACTIVATION

In the following, we supply formulas for the derivatives of the MA. First, for the mean firing rate  $\mu$ , by differentiating Eq. (10) with respect to  $\bar{\mu}$  and  $\bar{\sigma}$  we obtain the corresponding partial derivatives

$$\frac{\partial \mu}{\partial \bar{\mu}} = \frac{2}{L\sqrt{L}} \frac{\mu^2}{\bar{\sigma}} [g(I_{ub}) - g(I_{lb})] \quad (A1)$$

and

$$\frac{\partial \mu}{\partial \bar{\sigma}} = \frac{2}{L} \frac{\mu^2}{\bar{\sigma}} [g(I_{ub})I_{ub} - g(I_{lb})I_{lb}], \quad (A2)$$

respectively. Second, for the firing variability  $\sigma$ , by differentiating Eq. (11) with respect to  $\bar{\mu}$  and  $\bar{\sigma}$  we obtain

$$\frac{\partial \sigma}{\partial \bar{\mu}} = \frac{3}{L\sqrt{L}} \frac{\sigma}{\bar{\sigma}} \mu [g(I_{ub}) - g(I_{lb})] - \frac{1}{2\sqrt{L}} \frac{\sigma}{\bar{\sigma}} \frac{h(I_{ub}) - h(I_{lb})}{H(I_{ub}) - H(I_{lb})} \quad (A3)$$

and

$$\frac{\partial \sigma}{\partial \bar{\sigma}} = \frac{3}{L} \frac{\sigma}{\bar{\sigma}} \mu [g(I_{ub})I_{ub} - g(I_{lb})I_{lb}] - \frac{1}{2} \frac{\sigma}{\bar{\sigma}} \frac{h(I_{ub})I_{ub} - h(I_{lb})I_{lb}}{H(I_{ub}) - H(I_{lb})}, \quad (A4)$$

respectively. Third, for the linear response coefficient, the derivatives are

$$\begin{aligned} \frac{\partial \chi}{\partial \bar{\mu}} &= \frac{1}{2} \frac{\chi}{\mu} \frac{\partial \mu}{\partial \bar{\mu}} - \frac{\sqrt{2}}{L} \sqrt{\frac{\mu}{\Delta H}} [I_{ub}g(I_{ub}) - I_{lb}g(I_{lb})] \frac{1}{\bar{\sigma}} \\ &\quad + \frac{1}{2\sqrt{L}} \chi \frac{\Delta h}{\Delta H} \frac{1}{\bar{\sigma}} \end{aligned} \quad (A5)$$

and

$$\begin{aligned} \frac{\partial \chi}{\partial \bar{\sigma}} &= \frac{1}{2} \frac{\chi}{\mu} \frac{\partial \mu}{\partial \bar{\sigma}} - \frac{\chi}{\bar{\sigma}} \frac{2(I_{ub})^2 g(I_{ub}) - 2(I_{lb})^2 g(I_{lb}) + I_{ub} - I_{lb}}{\Delta g} \\ &\quad + \frac{1}{2} \frac{\chi}{\bar{\sigma}} \frac{I_{ub}h(I_{ub}) - I_{lb}h(I_{lb})}{\Delta H}, \end{aligned} \quad (A6)$$

where the short-hand notation  $\Delta$  denotes the difference between a function evaluated at  $I_{\text{ub}}$  and  $I_{\text{lb}}$ .

We also find analytical expressions of these derivatives in the weak fluctuation regime as  $\bar{\sigma} \rightarrow 0$ . First, for the mean firing rate  $\mu$ , by differentiating Eq. (29) we obtain

$$\lim_{\bar{\sigma} \rightarrow 0} \frac{\partial \mu}{\partial \bar{\mu}} = \begin{cases} 0, & \text{for } \bar{\mu} \leq V_{\text{th}}L, \\ \frac{V\mu^2}{\bar{\mu}(\bar{\mu}-V_{\text{th}}L)}, & \text{for } \bar{\mu} > V_{\text{th}}L. \end{cases} \quad (\text{A7})$$

For the derivative of  $\mu$  with respect to  $\bar{\sigma}$ , the limit is found to be zero everywhere except for an isolated singularity at  $\bar{\mu} = V_{\text{th}}L$ . For practical purposes we simply set it to zero. Second, for the firing variability  $\sigma$ , the gradient with respect to  $\bar{\mu}$  is zero at  $\bar{\sigma} = 0$  except that it is not well defined at  $\bar{\mu} = V_{\text{th}}L$ . For numerical purposes we set it to zero,

$$\left. \frac{\partial \sigma}{\partial \bar{\mu}} \right|_{\bar{\sigma}=0} = 0, \quad (\text{A8})$$

for all  $\bar{\mu}$ . The analytical limit for the derivative of  $\sigma$  with respect to  $\bar{\sigma}$  is

$$\lim_{\bar{\sigma} \rightarrow 0} \frac{\partial \sigma}{\partial \bar{\sigma}} = \frac{1}{\sqrt{2L}} \mu^{\frac{3}{2}} \sqrt{\frac{1}{(V_{\text{th}}L - \bar{\mu})^2} - \frac{1}{\bar{\mu}^2}}. \quad (\text{A9})$$

Third, for the linear response coefficient  $\chi$ , the limits of its derivatives are found to be

$$\lim_{\bar{\sigma} \rightarrow 0} \frac{\partial \chi}{\partial \bar{\mu}} = \frac{1}{\sqrt{2L}} \frac{1}{\sqrt{\mu \left( \frac{2}{V_{\text{th}}L} \bar{\mu} - 1 \right)}} \frac{\partial \mu}{\partial \bar{\mu}} - \sqrt{\frac{2}{L}} \frac{1}{V_{\text{th}}L} \mu^{\frac{1}{2}} \left( \frac{2}{V_{\text{th}}L} \bar{\mu} - 1 \right)^{-\frac{3}{2}}, \quad (\text{A10})$$

for  $\bar{\mu} > V_{\text{th}}L$  and zero otherwise, and

$$\lim_{\bar{\sigma} \rightarrow 0} \frac{\partial \chi}{\partial \bar{\sigma}} = 0. \quad (\text{A11})$$

Note that in all cases, the derivatives vanish for sufficiently large  $I_{\text{ub}}$ .

### APPENDIX B: CALCULATING EMPIRICAL CORRELATION MAP FROM SPIKING NEURON SIMULATIONS

The calculation of empirical correlation map from spiking neuron simulations becomes computationally prohibitive considering that its input space is 5D. Here we show that by avoiding redundancy one only needs to sweep through a 3D space. This is done by simulating a population of neurons whose input statistics cover the 2D space spanned by  $(\bar{\mu}, \bar{\sigma})$  while setting the correlation coefficient between these inputs all to the same value of  $\bar{\rho}$ . Output correlation under different combinations of input mean/std pairs can then be obtained for free. An additional step is required to make this method work for  $\bar{\rho} < 0$ , since the correlation matrix would not be positive definite. This is done by duplicating the neural population with the same input mean/std, such that the input correlations within each population remain positive whereas the input correlations between these two populations are negative. This duplication trick is also used to obtain the correlation of two neurons that receive input with the same mean/std for all

$-1 < \bar{\rho} < 1$ . We simulate spiking neurons for 10 s (with the first 100 ms discarded) over  $10^4$  independent trials in order to obtain accurate estimates of sample correlation. GPU acceleration is used for efficient parallelization and for generating correlated Gaussian inputs. Note that for very high input mean and low input std, the output std becomes too small, resulting in highly inaccurate sample estimates of correlation. These cases have been excluded from the above analysis.

### APPENDIX C: BENCHMARKING ACCURACY OF THE MA

One of the assumptions for deriving the MA is the diffusion approximation which replaces the synaptic current  $I_i(t)$  in [Eq. (1)] representing input spikes with a Gaussian white noise with the same mean and variance. The mean and variance of the output spike train can then be derived by solving the first passage time problem associated with the firing threshold [16]. Theoretically, this input-output relationship predicted by the diffusion approximation should converge to the exact result of the LIF neuron model when the neuron receives a sufficient number of spikes and when the contributions from individual spikes are small.

We validate the mean-variance mapping of the MA with a single LIF neuron receiving a controlled synaptic current of the form

$$I(t) = w_e S_e(t) - w_i S_i(t),$$

where  $S_{e,i}(t) = \sum_k \delta(t - t_{e,i}^k)$  represents the excitatory and inhibitory input spike trains and  $w_{e,i}$  are the corresponding synaptic weights. The mean and variance of the input current are  $\bar{\mu} = w_e \mu_e - w_i \mu_i$  and  $\bar{\sigma}^2 = w_e^2 \sigma_e^2 + w_i^2 \sigma_i^2$ , respectively. The goal here is to determine the statistics of the input spike trains for any given input current statistics. It is tempting to assume Poisson spike trains and solve the linear system for the Poisson rate (assuming  $w_e$  and  $w_i$  are known), but this can easily result in unrealistic firing rates. Moreover, spiking activity of cortical neurons exhibits diverse variability beyond Poisson statistics, with Fano factors both below and above one. Therefore, we generate input spike trains by drawing independently interspike intervals from a gamma distribution with mean  $\mathbb{E}[T_{e,i}]$  and variance  $\text{Var}[T_{e,i}]$ . To simplify, we assume  $w_e = w_i = w$  and  $\sigma_e = \sigma_i$ , and fix the strength of inhibitory current to a constant  $w_i \mu_i = c$ . The interspike interval statistics are then calculated as  $\mathbb{E}[T_e] = w/(\bar{\mu} + c)$ ,  $\mathbb{E}[T_i] = w/c$ , and  $\text{Var}[T_{e,i}] = \mathbb{E}[T_{e,i}]^3 \bar{\sigma}^2 / (2w^2)$ . In the following, we fix  $c = 1.2$  and vary the synaptic weight  $w$  and the input current statistics  $\bar{\mu}, \bar{\sigma}^2$ . We then simulate the LIF spiking neuron model [Eq. (1)] under these settings and calculate the trial-averaged mean firing rate  $\mu$  and firing variability  $\sigma$  using a finite but large time window  $\Delta t$  [see Eqs. (3) and (4)]. According to the diffusion approximation, the output spike statistics of the MA should approach to that of the LIF neuron model when  $w_{e,i}$  is sufficiently small and when  $S_{e,i}(t)$  contains a sufficiently large number of spikes for a given period of time.

We first show the mean firing rate  $\mu$  and the firing variability  $\sigma$  against the input current mean  $\bar{\mu}$  for different values of input current variability  $\bar{\sigma}$ , by setting the synaptic weight at  $w = 0.1$ . As shown in Fig. 5(a), predictions of the MA (solid curves) as implemented using our numerical scheme agree largely with the simulation results of the LIF neuron

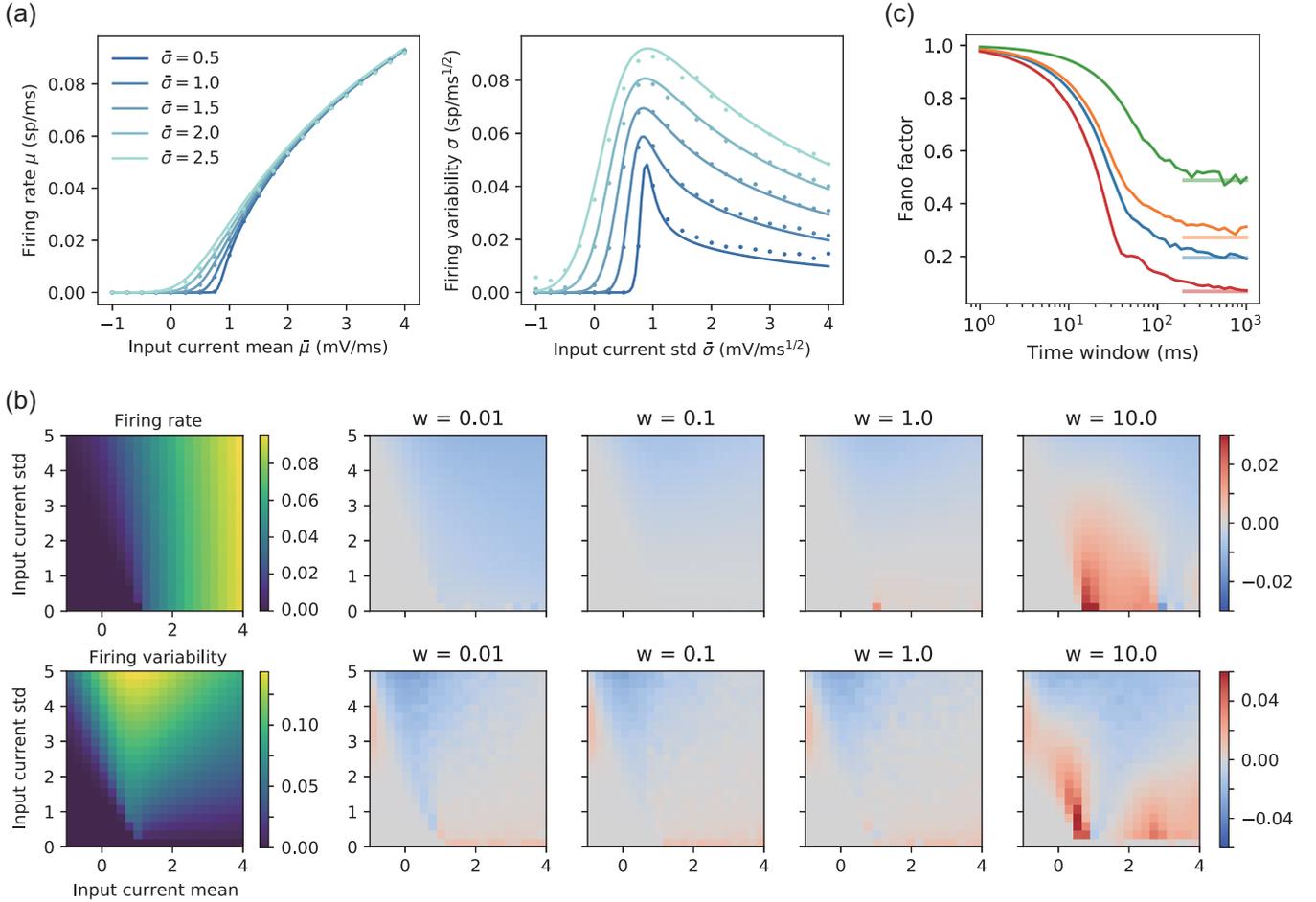


FIG. 5. Moment activation (MA) for approximating the firing statistics of spiking neuron model. (a) The mean firing rate  $\mu$  and firing variability  $\sigma$  as a function of the input current mean  $\bar{\mu}$  for different input current variability  $\bar{\sigma}$ . The MA (solid line) agrees with simulation results of the spiking neuron model (dots) driven by synthetic spike trains ( $w = 0.1$ ). (b) MA (left panels) and the difference between simulation results of the spiking neuron model and theoretical predictions for different synaptic weight strength. (c) Empirical Fano factor (solid curves) decreases with the spike count time window and agrees with the analytical predictions of MA (horizontal bars) for time windows larger than 200 ms. Each curve corresponds to a representative neuron in the network.

model (dots) for both  $\mu$  and  $\sigma$ . Next, we perform a parameter sweep over different input statistics and synaptic weight values and compare the difference between the simulation results of the spiking neuron model and the theoretical prediction by the MA. As shown in Fig. 5(b), the MA accurately captures the statistics of the spiking neuron model for most of the input space, but errors may occur under certain conditions, which are summarized as follows: First, the diffusion approximation fails when the synaptic weight is too large (e.g.,  $w = 10$ , which is half as much as the firing threshold). Second, the MA apparently overestimates the firing rate and variability when the input current std is large. We find that this is due to unrealistically high variability in the interspike interval of the input spike train as required to produce the desired input current variability. Third, the MA appears to underestimate the firing variability when the input current std is close to zero, which turns out to be inaccuracies for estimating spike count variance from spiking neuron simulations over finite time window. Lastly, some of these errors appear to be amplified when  $w$  is very small, due to the unrealistically high input firing

rate and variability required to produce these input current statistics. Note that these errors do not occur if we replace the spiking input with Gaussian noise. In conclusion, the MA accurately captures the statistical response properties of the spiking neuron model for the vast majority of the biologically realistic input space. This is further validated in a recurrent neural circuit under realistic settings, as presented in the main text.

Another note of caution is that the Fano factor  $\sigma^2/\mu$  as computed by the MA corresponds to the infinite-time Fano factor, as reflected by the limit in the spike count time window  $\Delta t$  in Eqs. (3) and (4). In practice, the Fano factor of event count in a renewal process depends on the time window  $\Delta t$  such that the Fano factor is always one at  $\Delta t = 0$  and converges to a finite value after a sufficiently large time window [48]. To quantify how large is sufficient, we simulate a homogeneous recurrent network using the LIF neuron model with random synaptic weights and investigate the dependence of the Fano factor on the size of spike count time window  $\Delta t$ . Specifically, the neural network consists of

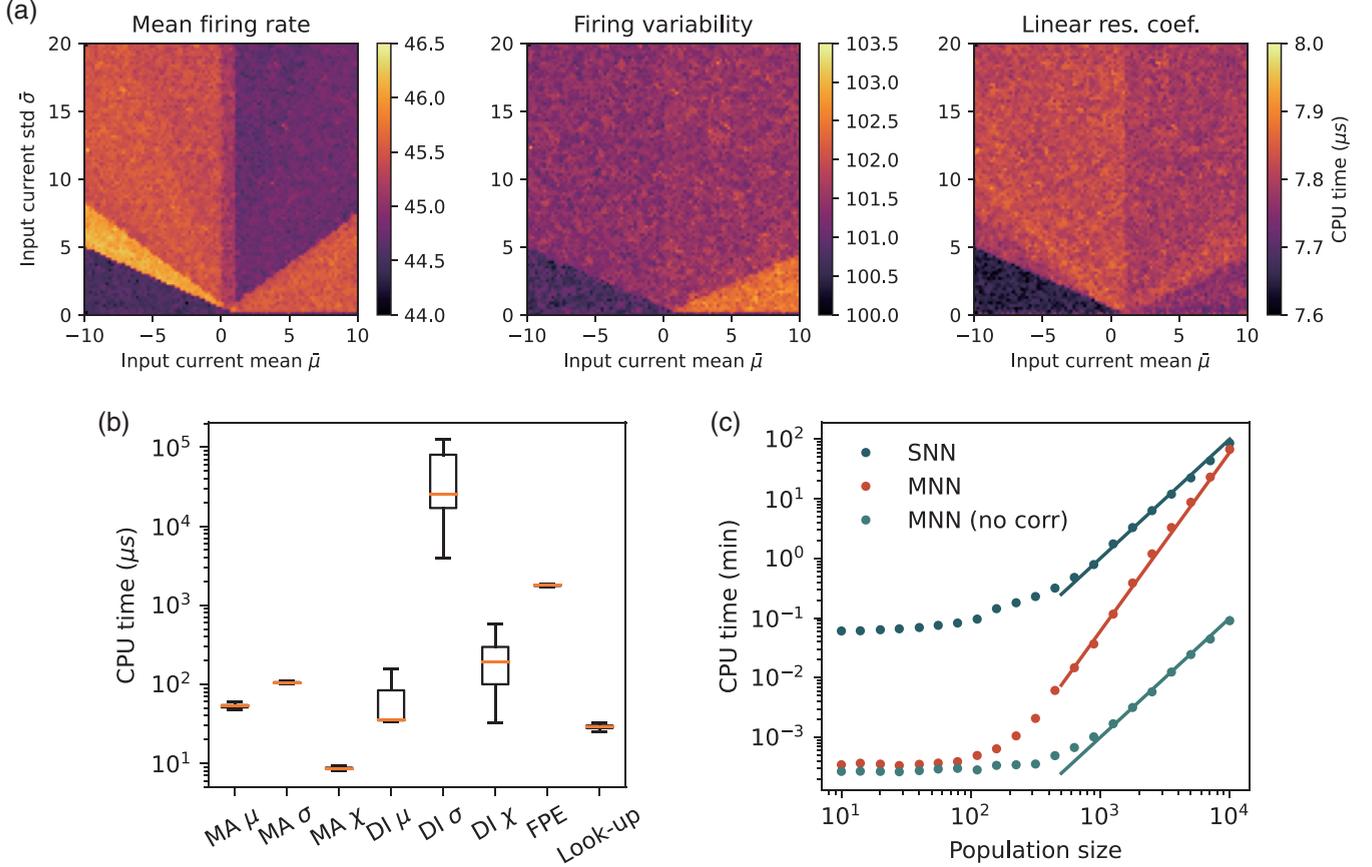


FIG. 6. Benchmarking the computational efficiency of the moment activation (MA). (a) The CPU time for calculating the MA for different input current statistics. Highly consistent performance are obtained across different regimes of input domain. (b) The CPU time for evaluating the moment mappings using different methods, over the input domain shown in (a); box plot indicates quartiles with whiskers indicating 1.5 interquartile range. MA: moment activation using our method; DI: direction integration; FPE: mean firing rate obtained by solving the time-independent Fokker Planck equation using a backward integration scheme; Look-up: look-up table with interpolation. (c) Performance comparison of different approaches for large-scale neural network simulation. SNN: direct simulation using spiking neuron models; MNN: moment neural network; MNN (no corr): same as MNN but with correlation assumed to be zero. Solid lines indicate quadratic growth for SNN and MNN (no corr) and cubic growth for MNN. All experiments were run on a desktop CPU with code implemented in Python.

$N_e = 100$  excitatory neurons and  $N_i = 100$  inhibitory neurons with a synaptic connection probability of  $p = 0.3$ . The synaptic weights are drawn randomly from normal distributions such that  $w_e \sim \mathcal{N}(0.2, 0.1)$  and  $w_i \sim \mathcal{N}(0.4, 0.2)$ . In addition to recurrent connections, each neuron in the network receives external feedforward currents in the form of Gaussian white noise with mean  $\bar{\mu}_{\text{ext}}$  and standard deviation  $\bar{\sigma}_{\text{ext}}$  which are drawn randomly for each neuron as  $\bar{\mu}_{\text{ext}} \sim \mathcal{N}(1, 0.2)$  and  $\bar{\sigma}_{\text{ext}} \sim \mathcal{N}(1, 0.2)$ .

As shown in Fig. 5(c), the Fano factor (solid curve, each for a different neuron) computed from the spiking neuron simulation is equal to one for  $\Delta t = 0$  and gradually decreases as  $\Delta t$  increases. After a sufficiently large time window, the Fano factor eventually converges to the theoretical limit (solid bars) predicted by the MA. We find that reasonably accurate approximations are achieved for time windows larger than  $\Delta t = 200$  ms. This result also implies that the assumption for stationary process can be relaxed to weakly nonstationary processes [49], that is, processes with statistics that slowly change over a time scale much larger than  $\Delta t = 200$  ms.

#### APPENDIX D: BENCHMARKING EFFICIENCY OF THE MA

In the following, we provide detailed benchmark results for the MA and compare its performance with a number of representative methods in the literature [18,50]. We first individually benchmark the three components of the MA, namely, the mappings for mean firing rate  $\mu$ , firing variability  $\sigma$ , and linear response coefficient  $\chi$ . For this purpose, we evaluate these mappings using our numerical method for different inputs and record the CPU time averaged over 100 independent trials. All experiments in this section are run on a desktop CPU with code implemented in Python. The results of the MA benchmark are shown in Fig. 6(a). For the particular hardware used, each component of the MA can be evaluated within tens of microseconds. Importantly, performance is highly consistent across different regimes of the input domain, indicating the reliability and robustness of our method. The apparent color patches reflect the division of the input domain into different regimes [Fig. 1(e)].

We further benchmark the efficiency of the proposed numerical method for the MA in comparison to other methods. One of them is the brute force method that involves direct integrations of Eqs. (12) and (13) defining the MA. Since it involves ill-conditioned Dawson-like integrals, this method fails for some input regimes (such as the mean-dominant and weak-fluctuation regimes). For input regimes that are within the bound of numerical precision, direction integration can be quite slow due to the nested integrals and the need for finer integration increments for dealing with steep problems. Nonetheless, this method provides a useful baseline for comparison. For numerical implementation, we use Numpy's "quad" function.

Another methods is the backward-integration method for solving the time-independent Fokker-Planck equation [18] associated with the membrane potential distribution from which the firing statistics can be derived. One drawback of this approach is that only the mean firing rate  $\mu$  can be evaluated and no formula is provided for the firing variability  $\sigma$ . Moreover, as the high-dimensional joint probability density is numerically intractable, it is only suitable for modeling neural activity without correlations. We implement this method in Python; an integration step of  $\delta V = 0.04$  mV is used.

In [50] the time-dependent Fokker-Planck equation is numerically solved to obtain the time-varying instantaneous firing rate. For the purpose of calculating the stationary mean firing rate, however, this method is rather slow. Therefore, the results are first calculated on a finite grid and then saved to a lookup table, which can later be interpolated to quickly evaluate the mean firing rate given the input statistics [50]. The main drawback of this method is that a lookup table cannot be used to extrapolate values outside the predetermined input range and a new table must be generated every time neuronal parameters such as the leak conductance and firing threshold are changed. Although the use of a look-up table is similar in spirit to what MA uses in the fluctuation-dominant regime, the MA stores the coefficients of Chebyshev polynomial rather than the function values on a finite grid and thus does not rely on interpolations. In fact, our method works for arbitrarily large inputs  $\bar{\mu}$  and  $\bar{\sigma}$  up to the limit of machine precision. For numerical implementation, we use Scipy's linear interpolation over a regular grid.

Figure 6(b) summarizes the CPU time for evaluating the moment mappings using different methods, over the entire grid of the input domain shown in Fig. 6(a). There are a number of observations worth mentioning. First, the direct integration method ("DI  $\mu$ ," "DI  $\sigma$ ," and "DI  $\chi$ ") not only has the highest computational cost, but is also highly variable across different inputs. In contrast, the method developed in this study ("MA  $\mu$ ," "MA  $\sigma$ ," and "MA  $\chi$ ") is both efficient and highly reliable. For our testing, speed gains of 2.7,  $6.6 \times 10^2$ , and 25 are obtained for each component of the MA relative to direct integration. Second, calculating the mean firing rate mapping by solving the time-independent Fokker-Planck equation using the backward integration method ("FPE") is slower on average than direct integration ("DI  $\mu$ "), with a relative speed gain of 0.082, but is more consistent. Moreover,

numerically solving the FPE only yields the mean firing rate but not the firing variability, leaving the MA the only practical method for calculating the complete moment mapping.

In addition to benchmarking the performance of single calls of the MA, we also benchmark its performance for modeling large-scale neural circuits and compare it to direct simulation of the spiking neural network (SNN) model [Eqs. (1) and (2)]. The main advantage of using the MNN over the SNN is as follows. Since the MNN directly deals with the statistical moments of spiking activity, it does not need to track the fine temporal dynamics of the membrane potential or spike timing as the SNN does. As a result, the moments of steady state activity can be obtained by evolving Eqs. (7)–(9) for a short period of time. In contrast, the SNN needs to be simulated for a much longer period of time (or alternatively over many trials) to obtain a sufficient number of spike counts over small time windows. The situation for SNNs is particularly worse for estimating second-order statistics, such as the Fano factor and correlation coefficient, which could be erroneous unless a large number of sample spikes are collected [48]. The main trade-off for this reduction in temporal complexity in the MNN is the increase in spatial complexity for storing and propagating the covariance matrix.

To benchmark the performance, we consider recurrent networks that receive noisy inputs with uniform covariance, using the same model parameters as in the previous section. For the SNN, the model is simulated for a total duration of  $T = 100$  s with a time step of  $\delta t = 0.1$  ms (corresponding to  $10^6$  integration steps); spike count is then calculated over  $\Delta T = 200$  ms time windows, resulting in 500 samples of spike count. For the MNN, the stationary statistics are obtained by evolving Eqs. (7)–(9) for a total duration of  $T = 20$  a.u. with a time step of  $\delta t = 0.1$  a.u. (corresponding to 200 integration steps). The CPU time as a function of population size is shown in Fig. 6(c). We find that for smaller population sizes ( $N < 1000$ ), the MNN equipped with our numerical method for the MA is orders of magnitude faster than direct simulations of the SNN. Although the exact time taken depends on the total simulation period and the size of the time step, significant speed improvement can generally be expected. We find that as the population size  $N$  grows, the computation time starts to show polynomial growth, which is found to be quadratic for the SNN and cubic for the MNN. The quadratic and cubic complexity is due to matrix multiplication that involves synaptic weights in Eqs. (2) and (6), respectively. As a result, for very large population sizes  $N > 10^4$ , the MNN gradually loses its advantage over the SNN. Nonetheless, for very large sparsely connected neural populations, the correlations in neural activity are generally weak (as is the case in the brain [51]), in which case we can set the correlation coefficients in the MNN to be zero, if one is mainly interested in the variance. Under this treatment, the covariance mapping in Eq. (6) is simplified into a mapping for the variance  $\bar{\sigma}_i^2 = \sum_j w_{ij}^2 \sigma_j^2 + \bar{\sigma}_{i,\text{ext}}^2$ , which only has a quadratic complexity [Fig. 6(c)]. In this scenario, the MNN equipped with our efficient numerical method for the MA retains its speed advantage and can be used to model the dynamics of irregular spiking activity in very large neural populations.

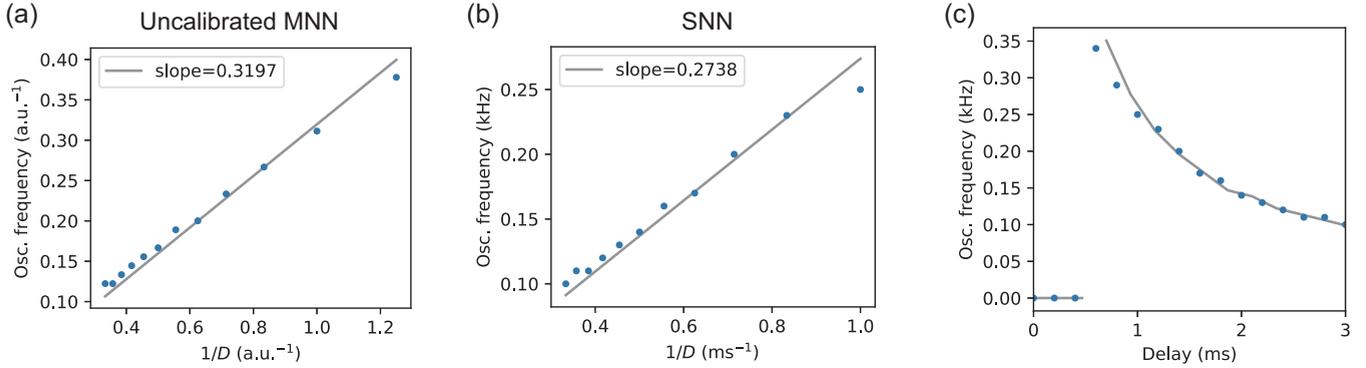


FIG. 7. Time constant calibration for the moment neural network. Empirical frequency-delay curves for the uncalibrated moment neural network with time constant  $\tau = 1$  a.u. and the spiking neural network are shown in (a) and (b), respectively. The calibration factor for the time constant are determined from the fitted slopes. Dots: empirical oscillation frequency; solid line: linear fit. In the example shown, I-E weight ratio is 6.0 and background input rate is 40 sp/ms. (c) Frequency-delay curves in the irregular oscillatory activity (IE ratio  $g = 6$  and input rate  $r = 40$  sp/ms) for the calibrated moment neural network (solid line) and spiking neural network (dots). In both models, the oscillation frequency is found to be inversely proportional to delay, and a critical delay is found around  $D = 0.5$  ms.

#### APPENDIX E: TIME CONSTANT CALIBRATION

In this section we explain how the time constant in the MNN can be calibrated to produce the oscillation frequency observed in spiking neural networks. A naïve approach would be treating the time constant  $\tau$  and the delay  $D$  in the MNN as independent free parameters and then fine-tuning them to generate an oscillation frequency that matches the SNN model. However, this approach cannot uniquely determine a time constant that works for all delay values. Instead, we propose an empirical calibration procedure for the time constant based on a simple scaling argument, which works for all delay values. Our method is based on the observation that the oscillation frequency in MNN is inversely proportional to both the time constant and the delay, such that a coordinate transformation through a linear scaling in time by a factor  $\beta$  would result in a change in oscillation frequency by  $1/\beta^2$ .

We first start with the uncalibrated MNN model whose time constant is set, without loss of generality, to  $\tau = 1$  a.u. We then simulate the model for different  $D$  with other parameters such as the I-E ratio and the background firing rate fixed, and calculate the empirical frequency-delay curve

$$f(D) = A/D,$$

whose coefficient  $A$  can be fitted. Now, consider a coordinate transformation to time  $t' = \beta t$ . Under this scaling, we must also have  $D' = \beta D$  and  $f' = f/\beta^2$ . Applying this coordinate transformation leads to

$$f' = \beta^{-2} f(D'/\beta) = (A/\beta)/D'.$$

We can also simulate SNN with varying delay and establish the empirical frequency-delay curve for the SNN,

$$f'(D') = B/D',$$

whose coefficient  $B$  can also be fitted. Finally, matching the transformed frequency-delay curve of the MNN with the ground truth of the SNN, we conclude that  $\beta = A/B$ . The calibrated time constant for the MNN can then be determined as  $\tau' = \beta\tau$ .

To illustrate, we fix the I-E ratio to 6.0 and the background input rate to 40 sp/ms and simulate the MNN model for  $\tau = 1$  a.u. and variable delay  $D$ , with a duration of  $T = 100$  a.u. and a time increment of  $\delta t = 0.02$  a.u., using the first-order Euler scheme. The oscillation frequencies are then calculated as the peak frequency of the power spectral density of the population-averaged mean firing rate. The same experiments are repeated in the SNN model. The empirical frequency-delay curves for the uncalibrated MNN and the SNN are shown in Fig. 7. In both models, the frequency and inverse delay can well be fitted by a straight line across the origin. The coefficients for these curves are found to be  $A = 0.3197$  and  $B = 0.2738$  and the calibration factor is found to be  $\beta = 1.168$  ms/a.u. The time constant in the MNN is then determined to be  $\tau' = \beta\tau = 1.168$  ms. As can be seen, the effective time constant for the MNN is much shorter than the membrane time constant  $\tau_m = 20$  ms of the spiking neuron model. Similarly, the same calibration factor also applies to the delay and time variables in the MNN according to  $D' = \beta D$  and  $t' = \beta t$ .

- [1] S. Coombes, Next generation neural population models, *Front. Appl. Math. Stat.* **9**, 1128224 (2023).
- [2] C. C. Chow and Y. Karimipناه, Before and beyond the Wilson-Cowan equations, *J. Neurophysiol.* **123**, 1645 (2020).
- [3] A. Byrne, R. D. O'Dea, M. Forrester, J. Ross, and S. Coombes, Next-generation neural mass and field modeling, *J. Neurophysiol.* **123**, 726 (2020).

- [4] W. R. Softky and C. Koch, The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs, *J. Neurosci.* **13**, 334 (1993).
- [5] A. Compte, C. Constantinidis, J. Tegnér, S. Raghavachari, M. V. Chafee, P. S. Goldman-Rakic, and X.-J. Wang, Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task, *J. Neurophysiol.* **90**, 3441 (2003).

- [6] M. R. Cohen and A. Kohn, Measuring and interpreting neuronal correlations, *Nat. Neurosci.* **14**, 811 (2011).
- [7] R. Rosenbaum, M. A. Smith, A. Kohn, J. E. Rubin, and B. Doiron, The spatial structure of correlated neuronal variability, *Nat. Neurosci.* **20**, 107 (2017).
- [8] J. de la Rocha, B. Doiron, E. Shea-Brown, K. Josić, and A. Reyes, Correlation between neural spike trains increases with firing rate, *Nature (London)* **448**, 802 (2007).
- [9] A. Kohn, R. Coen-Cagli, I. Kanitscheider, and A. Pouget, Correlations and neuronal population information, *Annu. Rev. Neurosci.* **39**, 237 (2016).
- [10] S. Panzeri, M. Moroni, H. Safaai, and C. D. Harvey, The structures and functions of correlations in neural population codes, *Nat Rev Neurosci.* **23**, 551 (2022).
- [11] A. E. Urai, B. Doiron, A. M. Leifer, and A. K. Churchland, Large-scale neural recordings call for new insights to link brain and behavior, *Nat. Neurosci.* **25**, 11 (2022).
- [12] M. A. Buice, J. D. Cowan, and C. C. Chow, Systematic fluctuation expansion for neural network activity equations, *Neural Comput.* **22**, 377 (2010).
- [13] M. A. Buice and C. C. Chow, Beyond mean field theory: Statistical field theory for neural networks, *J. Stat. Mech.* (2013) P03003.
- [14] R. M. Capocelli and L. M. Ricciardi, Diffusion approximation and first passage time problem for a model neuron, *Kybernetik* **8**, 214 (1971).
- [15] N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, *J. Comput. Neurosci.* **8**, 183 (2000).
- [16] J. Feng, Y. Deng, and E. Rossoni, Dynamics of moment neuronal networks, *Phys. Rev. E* **73**, 041906 (2006).
- [17] M. J. E. Richardson, Effects of synaptic conductance on the voltage distribution and firing rate of spiking neurons, *Phys. Rev. E* **69**, 051918 (2004).
- [18] M. J. E. Richardson, Firing-rate response of linear and nonlinear integrate-and-fire neurons to modulated current-based and conductance-based synaptic drive, *Phys. Rev. E* **76**, 021919 (2007).
- [19] A. Sanzeni, M. H. Histed, and N. Brunel, Emergence of irregular activity in networks of strongly coupled conductance-based neurons, *Phys. Rev. X* **12**, 011044 (2022).
- [20] I. Ginzburg and H. Sompolinsky, Theory of correlations in stochastic neural networks, *Phys. Rev. E* **50**, 3171 (1994).
- [21] W. Lu, E. Rossoni, and J. Feng, On a Gaussian neuronal field model, *NeuroImage* **52**, 913 (2010).
- [22] R. Moreno-Bote and N. Parga, Auto- and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses, *Phys. Rev. Lett.* **96**, 028101 (2006).
- [23] C. van Vreeswijk and H. Sompolinsky, Chaos in neuronal networks with balanced excitatory and inhibitory activity, *Science* **274**, 1724 (1996).
- [24] M. N. Shadlen and W. T. Newsome, The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding, *J. Neurosci.* **18**, 3870 (1998).
- [25] C. Baker, C. Ebsch, I. Lampl, and R. Rosenbaum, Correlated states in balanced neuronal networks, *Phys. Rev. E* **99**, 052414 (2019).
- [26] M. Veu e and A. Roxin, Firing rate distributions in spiking networks with heterogeneous connectivity, *Phys. Rev. E* **100**, 022208 (2019).
- [27] R. Rosenbaum and B. Doiron, Balanced networks of spiking neurons with spatially dependent recurrent connections, *Phys. Rev. X* **4**, 021039 (2014).
- [28] Y. Gu, Y. Qi, and P. Gong, Rich-club connectivity, diverse population coupling, and dynamical activity patterns emerging from local cortical circuits, *PLoS Comput. Biol.* **15**, e1006902 (2019).
- [29] R. Rosenbaum, A diffusion approximation and numerical methods for adaptive neuron models with stochastic inputs, *Front. Comput. Neurosci.* **10**, 39 (2016).
- [30] A. N. Burkitt, A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input, *Biol. Cybern.* **95**, 1 (2006).
- [31] M. Abramowitz, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables* (Dover Publications, New York, 1974).
- [32] S. G. Johnson, Faddeeva W function implementation, <http://ab-initio.mit.edu/Faddeeva>.
- [33] L. M. Ricciardi and S. Sato, First-passage-time density and moments of the Ornstein-Uhlenbeck process, *J. Appl. Probab.* **25**, 43 (1988).
- [34] N. Brunel and V. Hakim, Fast global oscillations in networks of integrate-and-fire neurons with low firing rates, *Neural Comput.* **11**, 1621 (1999).
- [35] N. Brunel and V. Hakim, Sparsely synchronized neuronal oscillations, *Chaos* **18**, 015113 (2008).
- [36] M. Chalk, B. Gutkin, and S. Den e, Neural oscillations as a signature of efficient coding in the presence of synaptic delays, *eLife* **5**, e13824 (2016).
- [37] R. Moreno-Bote, J. Beck, I. Kanitscheider, X. Pitkow, P. Latham, and A. Pouget, Information-limiting correlations, *Nat. Neurosci.* **17**, 1410 (2014).
- [38] I. Kanitscheider, R. Coen-Cagli, and A. Pouget, Origin of information-limiting noise correlations, *Proc. Natl. Acad. Sci. USA* **112**, E6973 (2015).
- [39] O. Hazon, V. H. Minces, D. P. Tom s, S. Ganguli, M. J. Schnitzer, and P. E. Jercog, Noise correlations in neural ensemble activity limit the accuracy of hippocampal spatial representations, *Nat. Commun.* **13**, 4276 (2022).
- [40] H. Ma, Y. Qi, P. Gong, J. Zhang, W. Lu, and J. Feng, Self-organization of nonlinearly coupled neural fluctuations into synergistic population codes, *Neural Comput.* **35**, 1820 (2023).
- [41] S. Sato, On the moments of the firing interval of the diffusion approximated model neuron, *Math. Biosci.* **39**, 53 (1978).
- [42] E. Y. Walker, R. J. Cotton, W. J. Ma, and A. S. Tolias, A neural basis of probabilistic computation in visual cortex, *Nat. Neurosci.* **23**, 122 (2020).
- [43] R. M. Haefner, P. Berkes, and J. Fiser, Perceptual decision-making as probabilistic inference by neural sampling, *Neuron* **90**, 649 (2016).
- [44] G. Orb n, P. Berkes, J. Fiser, and M. Lengyel, Neural variability and sampling-based probabilistic representations in the visual cortex, *Neuron* **92**, 530 (2016).
- [45] W. J. Ma, J. M. Beck, P. E. Latham, and A. Pouget, Bayesian inference with probabilistic population codes, *Nat. Neurosci.* **9**, 1432 (2006).
- [46] Y. Qi and P. Gong, Fractional neural sampling as a theory of spatiotemporal probabilistic computations in neural circuits, *Nat. Commun.* **13**, 4572 (2022).

- [47] <https://github.com/BrainsoupFactory/moment-neural-network>.
- [48] K. Rajdl, P. Lansky, and L. Kostal, Fano factor: A potentially useful information, *Front. Comput. Neurosci.* **14**, 569049 (2020).
- [49] R. L. Goris, J. A. Movshon, and E. P. Simoncelli, Partitioning neuronal variability, *Nat. Neurosci.* **17**, 858 (2014).
- [50] M. Augustin, J. Ladenbauer, F. Baumann, and K. Obermayer, Low-dimensional spike rate models derived from networks of adaptive integrate-and-fire neurons: Comparison and implementation, *PLoS Comput. Biol.* **13**, e1005545 (2017).
- [51] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, *Nature (London)* **440**, 1007 (2006).