Emergent rate-based dynamics in duplicate-free populations of spiking neurons

Valentin Schmutz,[1,2,*] Johanni Brea,[1] and Wulfram Gerstner[1]

[1]*School of Life Sciences and School of Computer and Communication Sciences,*
École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
[2]*UCL Queen Square Institute of Neurology,*
University College London, WC1E 6BT London, United Kingdom

1

Abstract

Can Spiking Neural Networks (SNNs) approximate the dynamics of Recurrent Neural Networks (RNNs)? Arguments in classical mean-field theory based on laws of large numbers provide a positive answer when each neuron in the network has many "duplicates", i.e. other neurons with almost perfectly correlated inputs. Using a disordered network model that guarantees the absence of duplicates, we show that duplicate-free SNNs can converge to RNNs, thanks to the concentration of measure phenomenon. This result reveals a general mechanism underlying the emergence of rate-based dynamics in large SNNs.

Neurons in the brain interact via spikes – short and stereotyped membrane potential deflections – commonly modelled as Dirac pulses [1, 2]. SNNs with recurrent connectivity are simplified models of real networks with the essential biological feature of spike-based neuronal communication. On the contrary, traditional RNNs are continuous dynamical systems where abstract rate neurons directly transmit their firing rate to other neurons, a type of communication which is not biological. Despite their inferior realism, RNNs continue to play a central role in theoretical neuroscience because they can be trained by modern machine learning methods [3], they can be analysed using tools from statistical physics [4–7], and because biological networks are believed to implement computations by approximation of continuous dynamical systems [8]. Closing the gap between the more biological SNNs and the more tractable RNNs requires identifying the conditions under which the continuous dynamics of RNNs can be approximated by SNNs [9]. This problem is important for neuroscience as it is closely linked to a fundamental question: given the discontinuous and stochastic nature of spike-based communication [10], how can biological networks produce noise-robust population dynamics? The question is nontrivial because the microscopic dynamics of SNNs can be unstable with respect to single-spike perturbations [11, 12].

To clearly state the problem, let us consider an SNN composed of N Poisson neurons (linear-nonlinear-Poisson neurons [13] or nonlinear Hawkes processes [14]). For each neuron index i, the spike times $\{t_i^k\}_k$ of neuron i, which define the neuron's spike train $S_i(t) = \sum_k \delta(t - t_i^k)$, are generated by an inhomogeneous Poisson process with instantaneous firing rate $\phi(h_i(t))$, where $h_i(t)$ represents the neuron's potential and ϕ is a positive-valued nonlinear transfer function. The potential $h_i(t)$ is a leaky integrator of the recurrent inputs coming from

neurons $j \neq i$ and the external input $I_i^{\text{ext}}(t)$:

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} h_i(t) = -h_i(t) + \sum_{j=1}^{N} J_{ij} S_j(t) + I_i^{\text{ext}}(t), \tag{1}$$

where τ is the integration (or membrane) time constant and J_{ij} is the synaptic weight from neuron j to neuron i (by convention, $J_{ii} = 0$). While the spike-based model described here is biologically simplistic, it is mathematically convenient as it has a straightforward rate-based counterpart. If we replace the spike trains $\{S_j(t)\}_j$ in (1) by the corresponding instantaneous firing rates $\{\phi(h_j(t))\}_j$ (i.e. neurons communicate their firing rate directly), we get the rate-based dynamics

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} x_i(t) = -x_i(t) + \sum_{j=1}^{N} J_{ij} \phi(x_j(t)) + I_i^{\text{ext}}(t), \tag{2}$$

which defines an RNN with N rate units. To avoid confusion, we write $h_i(t)$ for the potentials of the SNN (1) and $x_i(t)$ for the potentials of the RNN (2). While the mapping from the SNN to the RNN looks simple at first glance, the spike-based stochastic process (1) and the rate-based dynamical system (2) describe very different kinds of systems and the SNN potentials $h_i(t)$ are not guaranteed, in general, to be equal or even close to the RNN potentials $x_i(t)$, even if both networks receive the same external input. From a probabilistic perspective, the RNN (2) also describes the *expectation* of the stochastic SNN dynamics (1) over Poisson spike noise, when the two systems share the same initial conditions and the same external input. This implies that the problem of approximating rate-based dynamics can be reduced to the problem of controlling the deviation (due to spike noise) of the SNN potentials $h_i(t)$ from their expected trajectories $x_i(t)$; in short, approximating rate-based dynamics is equivalent to producing network dynamics that is robust to spike noise.

If the neurons are uncoupled (i.e. $J_{ij} = 0$ for all i, j), the SNN potentials $h_i(t)$ are trivially equal to the RNN potentials $x_i(t)$. Therefore, comparing the SNN and the RNN is meaningful only if the coupling does not vanish. For nontrivial coupling, there are two known types of scaling limits where the SNN potentials $h_i(t)$ converge to the RNN potentials $x_i(t)$:

(i) *Spatial averaging over neuronal duplicates*: Consider networks of increasing size N where neurons are localized in some fixed space such that two neurons assigned to the same point are duplicates, i.e. they always share the same recurrent and external input. If the synaptic weights are scaled by $1/N$, we can take the mean-field limit

$N \to \infty$ [15]. The fixed space can be either discrete and finite [16] or continuous and finite-dimensional [17], e.g. a ring [18]. These classical mean-field limits rely on a strong form of redundancy: the existence of large ensembles of neuronal duplicates receiving (almost) the same recurrent and external input.

(ii) *Temporal averaging over single-neuron spikes*: In (1), we can replace the transfer function ϕ and the weights $\{J_{ij}\}_{i \neq j}$ by $b\,\phi$ and $\{J_{ij}/b\}_{i \neq j}$, respectively (for $b > 0$), and take the limit $b \to \infty$ [19]. This limit entails arbitrarily high firing rates in the SNN, which is biologically unrealistic since two spikes have to be separated by at least 1 to 2 milliseconds (the absolute refractory period) [1, 20]. Alternatively, but to a similar effect, we can take the limit $\tau \to \infty$ in both (1) and (2) while re-scaling the weights $\{J_{ij}\}_{i \neq j}$ and the external inputs $I_i^{\text{ext}}(t)$ by $1/\tau$. This last limit entails arbitrarily slow network dynamics, which is incompatible, for example, with human visual processing speed (less than 150 milliseconds) [21].

In this letter, we address the following question: can large SNNs, as defined in (1), converge to *equally large* RNNs in the absence of neuronal duplicates and without temporal averaging?

Temporal averaging as in (ii) is made impossible if we impose, as we do, that

$$\max \phi \leq 1/\tau. \tag{3}$$

Under this constraint, leaky integration by the potential (1) is too fast to average out the Poisson noise of individual input spike trains and neither of the two scalings mentioned under (ii) can be applied.

To quantify the amount of duplication in an RNN, we look at the distribution of correlations [22] between pairs of distinct neurons $i \neq j$

$$C_{ij} := \lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{(x_i(t) - \overline{x_i})(x_j(t) - \overline{x_j})}{\sigma_{x_i} \sigma_{x_j}} dt, \tag{4}$$

where $\overline{x_i}$ and $\sigma_{x_i}^2$ are, respectively, the time average and fluctuation of $x_i(t)$ ($\overline{x_i} := \lim_{T \to \infty} \frac{1}{T} \int_0^T x_i(t) dt$ and $\sigma_{x_i}^2 := \lim_{T \to \infty} \frac{1}{T} \int_0^T (x_i(t) - \overline{x_i})^2 dt$.

Based on the correlation distribution, we say that duplicates accumulate in large networks if, as $N \to \infty$, the limit pairwise correlation density has a strictly positive mass around 1. To illustrate how duplicates accumulate in classical mean-field models, let us consider a toy example with a ring structure. Let the N units of an RNN be uniformly and independently
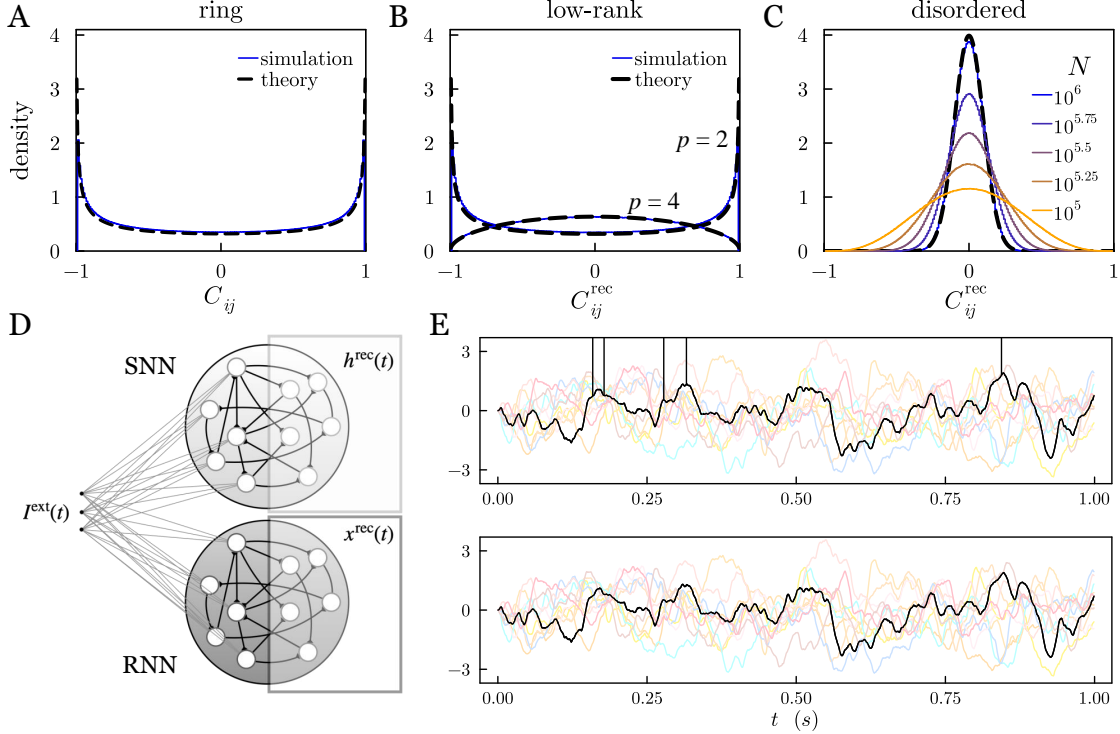
FIG. 1. (**A**) Limit correlation density for the ring model. The histogram of a simulation with $N = 10^6$ units (solid blue line) is compared to the theoretical limit density (5) (dashed black line). (**B**) Limit correlation density (of units receiving recurrent input only) for low-rank networks $p = 2$ and $p = 4$. Simulations of RNNs with $N = 10^6$ units (solid blue line) are compared to the corresponding theoretical limit "Gegenbauer" density (11) (dashed black line). (**C**) Concentration around 0 of the correlation distribution in disordered networks with $p = \alpha N$ and fixed load $\alpha = 10^{-4}$. The correlation distribution (of units receiving recurrent input only) of simulations of networks of increasing size N converge to a centred normal density with variance $1/p = 1/(\alpha N)$ (the dashed black line indicates the normal density for $N = 10^6$). (**D**) Setup for comparing SNNs and RNNs. (**E**) Trajectories of single-neuron potentials in the SNN ($h_i^{\mathrm{rec}}(t)$, upper panel) and in the RNN ($x_i^{\mathrm{rec}}(t)$, lower panel) during a one-second simulation of the setup shown in (D). The networks have $N = 10^6$ neurons and load $\alpha = 10^{-4}$, as in (C). The same randomly chosen 11 neurons are recorded in the SNN and in the RNN and each colour corresponds to a different neuron (the colours in the upper and lower panels correspond). For one neuron (the black trace), the spike times of the neuron in the SNN are indicated by vertical bars. (**B-E**) Neuronal parameters: $\tau = 10$ ms and $\phi(x) = \frac{1}{2\tau}\left(\tanh(x - b) + 1\right)$ with $b = 2$; input standard deviation is $\sigma = 0.5$ (see (9)).

5

positioned on a ring, θ_i denoting the angular position of unit i. If the correlation between unit i and j is given by $C_{ij} := \cos(\theta_i - \theta_j)$, then, as $N \to \infty$, the limit distribution of the pairwise correlation, which we call the limit correlation density, is

$$\rho(z) = \mathbb{1}_{[-1,1]}(z) \frac{1}{\pi} \frac{1}{\sqrt{1-z^2}} \tag{5}$$

(Fig 1A; proof in SM Sec. I [23]). This limit density has a strictly positive mass around 1, reflecting the fact that each unit has a number of duplicates that grows linearly with N.

Conversely, we say that large networks of a given model are *duplicate-free* if, for any threshold $\varepsilon > 0$, the probability that there exists a pair of distinct neurons $i \neq j$ such that their absolute correlation is greater or equal to ε tends to 0 as $N \to \infty$, i.e. for all $\varepsilon > 0$,

$$\mathbb{P}\left(\exists\,(i,j) \text{ with } i \neq j \text{ s.t. } |C_{ij}| \geq \varepsilon\right) \xrightarrow[N \to \infty]{} 0. \tag{6}$$

In the following, we propose a disordered network model where large networks are duplicate-free.

Input-driven disordered network model.— We construct the connectivity matrix $\mathbf{J} = \{J_{ij}\}_{i,j}$ as a sum of random rank-one matrices (minus self-interaction terms), a construction similar to that of Hopfield networks [24–26]. For any number of units N and any number of patterns p, let $\boldsymbol{\xi}$ be a random $N \times p$-matrix with *i.i.d.*, zero-mean, unit-variance, normally distributed entries $\{\xi_{i\mu}\}_{i,\mu}$. We choose a connectivity matrix given by

$$J_{ij} := \frac{1}{cN} \sum_{\mu=1}^{p} \xi_{i\mu} \left(\phi(\xi_{j\mu}) - a\right) \quad \text{for all } i \neq j, \tag{7}$$

and $J_{ii} := 0$ for all i, where $a := \int_{-\infty}^{\infty} \mathcal{D}z\, \phi(z)$ and $c := \int_{-\infty}^{\infty} \mathcal{D}z\, (\phi(z)-a)^2$ are fixed constants ($\mathcal{D}z$ denotes the standard Gaussian measure).

A well-known feature of this type of connectivity is that, exchanging the order of summation, the dynamics of the SNN (1) can be re-written in terms of p overlap variables $\{m_\mu(t)\}_\mu$ [27–29]: for all $i = 1, \ldots, N$ and for all $\mu = 1, \ldots, p$,

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} h_i(t) = -h_i(t) + \sum_{\nu=1}^{p} \xi_{i\nu} m_\nu(t) - \gamma_i S_i(t) + I_i^{\text{ext}}(t),$$

$$m_\mu(t) = \frac{1}{cN} \sum_{j=1}^{N} \left(\phi(\xi_{j\mu}) - a\right) S_j(t). \tag{8}$$

where the $\gamma_i := \frac{1}{cN} \sum_{\nu=1}^{p} \xi_{i\nu} \left(\phi(\xi_{i\nu}) - a\right)$ are virtual self-interaction weights. An analogous reformulation holds for the RNN (2). In the language of latent variable models in neuroscience,

the overlap variables $m_\mu(t)$ would be called "factors" [30]. The reformulation in terms of overlaps/factors clearly shows that if $p \ll N$, the γ_i are small and therefore the recurrent drive $\{\sum_{j=1}^{N} J_{ij} S_j(t)\}_i$ is approximately restricted to the p-dimensional subspace spanned by the p columns of the random matrix $\boldsymbol{\xi}$. To force the recurrent drive to visit all p dimensions homogeneously over time in a single stationary and ergodic process, we inject the following p-dimensional external input to half of the neurons:

$$I_i^{\text{ext}}(t) = \frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^{p} \xi_{i\mu} \eta_\mu(t) \quad \text{if } i \leq N/2, \tag{9}$$

where the $\eta_1(t), \ldots, \eta_p(t)$ are the formal derivatives of independent Wiener processes (or standard Brownian motions) $B_1(t), \ldots, B_p(t)$, i.e. $\eta_\mu(t) = \mathrm{d}B_\mu(t)/\mathrm{d}t$, and $\sigma > 0$ is the standard deviation of the input. For clarity, we use the superscript 'in' to emphasize that, for all $i \leq N/2$, the neurons $h_i^{\text{in}}(t)$ and the units $x_i^{\text{in}}(t)$ receive external as well as recurrent inputs, and we use the superscript 'rec' to emphasize that, for all $i > N/2$, the neurons $h_i^{\text{rec}}(t)$ and the units $x_i^{\text{rec}}(t)$ receive recurrent input only. Note that the $x_i^{\text{rec}}(t)$ can be seen as linear readouts of the recurrent drive.

Assuming that for any N and p the input (9) leads to a stationary ergodic process, we find, for any N and p, that the correlation between two units satisfies the bound

$$C_{ij} \leq \frac{\sum_{\mu=1}^{p} \xi_{i\mu}\xi_{j\mu}}{\sqrt{\sum_{\mu=1}^{p} \xi_{i\mu}^2}\sqrt{\sum_{\mu=1}^{p} \xi_{j\mu}^2}}, \tag{10}$$

because of the radial symmetry of the joint stationary distribution of the variables $\{m_\mu\}_{\mu=1}^{p}$. The bound (10) is tight, if the virtual self-interaction weights $\gamma_i = 0$, that is, if $J_{ii} = \frac{1}{cN}\sum_{\mu=1}^{p} \xi_{i\mu}\left(\phi(\xi_{i\mu}) - a\right)$. The bound (10) also holds for the pairwise correlations in the SNN (1) since the addition of spike noise can only reduce correlations. Henceforth, we write C_{ij}^{rec} for the pairwise correlation (4) between the 'rec' units x_i and x_j.

If the number of patterns p is kept constant as $N \to \infty$, the limit RNN is a low-rank mean-field model (without disorder) [6, 31]. In such low-rank models, the bound (10) is asymptotically tight for the C_{ij}^{rec}, since the virtual self-interaction weights $\gamma_i \to 0$ as $N \to \infty$. However, duplicates accumulate as $N \to \infty$ because the distribution of correlations $\{C_{ij}^{\text{rec}}\}_{i<j}$ converges to a limit density with a strictly positive mass around 1, indicating that the number of duplicates per unit grows linearly with N. More specifically, for any fixed $p > 1$, the limit

7

density is given by the so-called "Gegenbauer distribution" with parameter p,

$$\rho(z) = \mathbb{1}_{[-1,1]}(z)\frac{\Gamma(p/2)}{\sqrt{\pi}\,\Gamma((p-1)/2)}(1-z^2)^{\frac{p-3}{2}}, \tag{11}$$

which is the orthogonal projection of the uniform distribution on the unit sphere $\mathbb{S}^{p-1}$ onto its diameter (Fig 1B; proof in SM Sec. I [23]). Intuitively, the explanation for the accumulation of duplicates is the same as in the ring model presented above, except that, here, the fixed space is not a ring but $\mathbb{R}^p$: unit i has coordinate $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{ip})$ and units with similar coordinates receive similar recurrent and external inputs. Therefore, if p is kept fixed as $N \to \infty$, we fall again in the case of spatial averaging over neuronal duplicates (i) leading to neural field equations (see SM Sec. II [32]).

To prevent duplicates from accumulating as $N \to \infty$, we make the number of patterns p grow linearly with N, taking $p = \alpha N$ for some fixed load $\alpha > 0$, as in the Hopfield model [25]. With this choice of scaling, weights $\{J_{ij}\}_{i,j}$ scale as $\mathcal{O}(1/\sqrt{N})$ (as in random RNNs [5, 33–35]), whence the name "disordered network" for this model. First (I), we will show that for any fixed $\alpha > 0$, large networks are duplicate-free (as defined above). Second (II), we will show that large SNNs converge to large RNNs, with convergence rate $\sqrt{\alpha}$, as $\alpha \to 0$ (this means that for arbitrarily small $\alpha > 0$, large SNNs behave almost exactly like large RNNs). To compare the SNN (1) with the RNN (2), we will inject the same time-dependent external input (9) in both networks (Fig 1D) and compare the trajectories $h_i^{\text{rec}}(t)$ of the SNN with the trajectories $x_i^{\text{rec}}(t)$ of the RNN (Fig 1E).

(I) Large networks are duplicate-free.— Assuming that the bound (10) is tight for C_{ij}^{rec}, for any distinct pair of units $i \neq j$, we have, by the central limit theorem, that the scaled correlation $\sqrt{p}\, C_{ij}^{\text{rec}}$ converges in law to a standard normal random variable as $p \to \infty$. Then, if $p \to \infty$ as $N \to \infty$, we can prove that the distribution of correlations $\{C_{ij}^{\text{rec}}\}_{i<j}$ converges to a centered normal distribution with variance $1/p$ as $N \to \infty$ (Fig. 1C; proof in SM Sec. I [23]). More importantly, using the bound (10) and the fact that we have for all $N, p > 1$ a Gegenbauer distribution with parameter p (11), we can derive a bound for the probability of a duplicate:

$$\mathbb{P}\left(\exists\,(i,j) \text{ with } i \neq j \text{ s.t. } |C_{ij}| \geq \varepsilon\right) \leq \frac{N(N-1)}{2}\frac{(1-\varepsilon^2)^{\frac{p-1}{2}}}{\sqrt{\pi}}, \tag{12}$$

for all $0 < \varepsilon < 1$ (see SM Sec. III [36]). Since we have $p = \alpha N$, the bound tends to 0 as $N \to \infty$, which confirms that large networks are duplicate-free. In fact, the bound guarantees

8

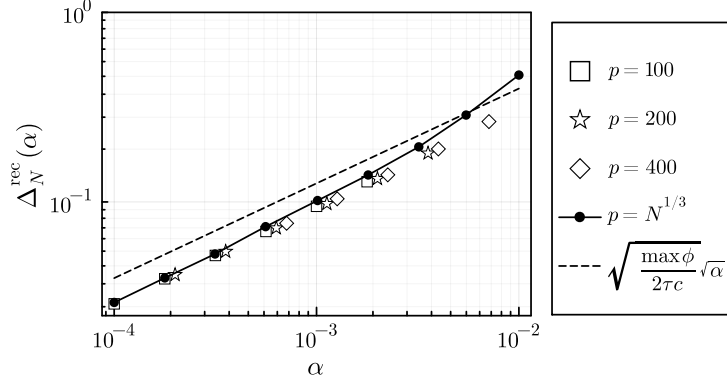FIG. 2. Large N simulations of the distance $\Delta_N^{\mathrm{rec}}(\alpha)$ as a function of α for $p = 100$ (rectangles), $p = 200$ (stars), and $p = 400$ (diamonds). Simulations of $\Delta_N^{\mathrm{rec}}(\alpha)$ for the duplicate-free sequence of networks $p = N^{1/3}$ (full circles). Theoretical $\mathcal{O}(\sqrt{\alpha})$ bound predicated by the feedforward model simplification (dashed line). Same parameters as in Fig. 1B-E.

that the sublinear scaling $p = \alpha N^\beta$ with $0 < \beta < 1$ suffices to get duplicate-free networks as $N \to \infty$.

(II) Large SNNs converge to large RNNs at rate $\sqrt{\alpha}$, as $\alpha \to 0$.— For any fixed load α, we define the average distance between the SNN and the RNN as

$$\Delta_N^{\mathrm{rec}}(\alpha) := \frac{2}{N} \sum_{i=N/2+1}^{N} \lim_{T \to \infty} \frac{1}{T} \int_0^T \left| h_i^{\mathrm{rec}}(t) - x_i^{\mathrm{rec}}(t) \right| \mathrm{d}t$$

and the large N limit distance as $\Delta^{\mathrm{rec}}(\alpha) := \lim_{N \to \infty} \Delta_N^{\mathrm{rec}}(\alpha)$. Numerical estimates of the limit distance $\Delta^{\mathrm{rec}}(\alpha)$ as a function of α indicates that

$$\Delta^{\mathrm{rec}}(\alpha) = \mathcal{O}(\sqrt{\alpha}), \quad \text{as } \alpha \to 0 \tag{13}$$

(Fig. 2). This scaling implies that large SNNs can converge to equally large RNNs despite the fact that (i) there is no duplicate averaging (Fig. 1C and (12)) and (ii) no temporal averaging (3). For example, if we take the sublinear scaling $p = N^{1/3}$, which implies $\alpha = N^{-2/3}$, the distance $\Delta_N^{\mathrm{rec}}(\alpha) \to 0$ vanishes (Fig. 2, full circles) and networks are duplicate-free (12), as $N \to \infty$.

In the absence of averaging of types (i) and (ii), the concentration of measure phenomenon [37, 38] can explain the convergence of SNNs to RNNs. In our case, the concentration of measure is controlled by the ℓ^2 norm $\|\mathbf{J}_i\|_2 = \sqrt{\sum_{j=1}^{N} J_{ij}^2}$ of each neuron i's incoming weights.
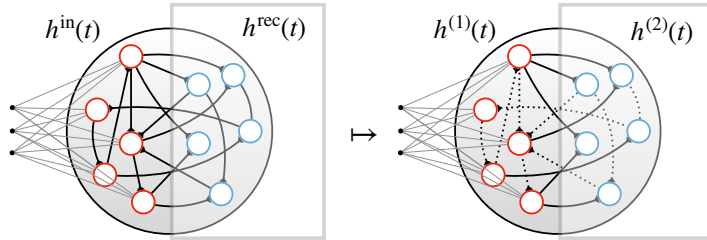
FIG. 3. The feedforward model as a simplification of the input-driven model. Dotted lines indicate removed connections

When $p = \alpha N$, for a typical neuron i, we find the convergence in probability

$$\|\mathbf{J}_i\|_2 \xrightarrow[N\to\infty]{\mathbb{P}} \sqrt{\frac{\alpha}{c}}, \tag{14}$$

(proof in SM Sec. IV [39]) which means that the ℓ^2 norms of each neuron's incoming weights concentrate around $\sqrt{\alpha/c}$ (where c is defined after (7)). The numerical scaling (13) of the limit distance between large SNNs and large RNNs, as $\alpha \to 0$, corresponds to the theory-based scaling (14) of the limit ℓ^2 norm of a typical neuron's incoming weights.

Although we do not have an exact theory linking the limit ℓ^2 norm (14) with the limit distance $\Delta^{\mathrm{rec}}(\alpha)$ (13), a simplified feedforward model, which is analytically tractable, offers a good intuition for the concentration of measure phenomenon at play. This simplified model is obtained by keeping network connections from 'in' to 'rec' neurons and removing all the other connections (Fig. 3). After this pruning procedure, we are left with two-layer feedforward networks where the 'in' neurons make up the first layer, $(h_i^{\mathrm{in}}(t), x_i^{\mathrm{in}}(t)) \mapsto (h_i^{(1)}(t), x_i^{(1)}(t))$, and the 'rec' neurons make up the second layer, $(h_i^{\mathrm{rec}}(t), x_i^{\mathrm{rec}}(t)) \mapsto (h_i^{(2)}(t), x_i^{(2)}(t))$. In general, for any $N > 1$ and for *any* connectivity matrix $\mathbf{J}$, we find that, for any neuron i in the second layer, the single-neuron distance satisfies the bound

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \left| h_i^{(2)}(t) - x_i^{(2)}(t) \right| \mathrm{d}t \le \sqrt{\frac{\max\phi}{2\tau}} \|\mathbf{J}_i\|_2,$$

(proof in SM Sec. V [40]) which is indeed controlled by the ℓ^2 norm $\|\mathbf{J}_i\|_2$. Then, if $p = \alpha N$, using the convergence of the ℓ^2 norm (14), we get the expected scaling for the limit distance, as $N \to \infty$, between a spiking neuron $h_i^{(2)}$ and a rate unit $x_i^{(2)}$:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \left| h_i^{(2)}(t) - x_i^{(2)}(t) \right| \mathrm{d}t \le \sqrt{\frac{\max\phi}{2\tau c}} \sqrt{\alpha}$$

(cf. Fig. 2). Therefore, the feedforward model provides an intuition for how the vanishing ℓ^2 norms of the incoming weights (as $\alpha \to 0$) cause concentration of measure in duplicate-free, large SNNs. An exact theory of the original recurrent networks would require a full-fledged dynamical mean-field theory [41] for our input-driven disordered network model, which remains an open problem.

Concentration of measure [38] has been shown to be instrumental for the theory of spin glasses [42] and the theory of infinite-width artificial neural networks [43]. By contrast, this probabilistic notion has barely permeated the theory of biological neural networks. The standard perspective has been that, to produce spike-noise-robust population dynamics, large networks have to perform averages over the spike activity of many neuronal duplicates [2, 9, 29, 44], i.e. mechanism (i). Recently, DePasquale *et al.* [30] proposed after extensive numerical simulations the perspective that "weighted averages" over heterogeneous neurons could lead to population-level latent factors with noise-robust dynamics, which can then produce the illusion of rate-based dynamics. We provide a theoretical foundation for this idea by showing that the concentration of measure phenomenon explains how the dynamics of large, duplicate-free SNNs can converge to the dynamics of equally large RNNs. Note that, while we have focused our analysis on networks of linear-nonlinear-Poisson neurons, analogous results still hold, for example, when diffusive membrane noise is added to the potential dynamics (see SM Sec. VI [45]). In the absence of neuronal duplicates, rate-based population dynamics in networks of spiking neurons can only be understood as an emergent behaviour: neurons interact via spikes, but they behave *as if* they were interacting via their unique, time-varying firing rates. This notion of emergent rate-based dynamics sheds new light on the long-standing spike vs. rate debate in computational neuroscience [2, 9, 44].

---

* Corresponding author: v.schmutz@ucl.ac.uk

[1] F. Rieke, D. Warland, R. d. R. Van Steveninck, and W. Bialek, *Spikes: exploring the neural*

code (MIT press, 1999).

[2] W. Gerstner and W. M. Kistler, *Spiking neuron models: Single neurons, populations, plasticity* (Cambridge University Press, 2002).

[3] O. Barak, Recurrent neural networks as versatile tools of neuroscience research, Curr. Opin. Neurobiol. **46**, 1 (2017).

[4] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Storing infinite numbers of patterns in a spin-glass model of neural networks, Phys. Rev. Lett. **55**, 1530 (1985).

[5] H. Sompolinsky, A. Crisanti, and H.-J. Sommers, Chaos in random neural networks, Phys. Rev. Lett. **61**, 259 (1988).

[6] F. Mastrogiuseppe and S. Ostojic, Linking connectivity, dynamics, and computations in low-rank recurrent neural networks, Neuron **99**, 609 (2018).

[7] U. Pereira-Obilinovic, J. Aljadeff, and N. Brunel, Forgetting leads to chaos in attractor networks, Physical Review X **13**, 011009 (2023).

[8] S. Vyas, M. D. Golub, D. Sussillo, and K. V. Shenoy, Computation through neural population dynamics, Ann. Rev. Neurosci. **43**, 249 (2020).

[9] R. Brette, Philosophy of the spike: rate-based vs. spike-based theories of the brain, Front. Syst. Neurosci. **9**, 151 (2015).

[10] A. A. Faisal, L. P. Selen, and D. M. Wolpert, Noise in the nervous system, Nat. Rev. Neurosci. **9**, 292 (2008).

[11] M. London, A. Roth, L. Beeren, M. Häusser, and P. E. Latham, Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex, Nature **466**, 123 (2010).

[12] M. Monteforte and F. Wolf, Dynamic flux tubes form reservoirs of stability in neuronal circuits, Phys. Rev. X **2**, 041007 (2012).

[13] E. Chichilnisky, A simple white noise analysis of neuronal light responses, Network **12**, 199 (2001).

[14] P. Brémaud and L. Massoulié, Stability of nonlinear Hawkes processes, Ann. Probab. **24**, 1563 (1996).

[15] W. Gerstner, Time structure of the activity in neural network models, Phys. Rev. E **51**, 738 (1995).

[16] S. Ditlevsen and E. Löcherbach, Multi-class oscillating systems of interacting neurons, Stochastic Process. Appl. **127**, 1840 (2017).

[17] J. Chevallier, A. Duarte, E. Löcherbach, and G. Ost, Mean field limits for nonlinear spatially extended hawkes processes with exponential memory kernels, Stochastic Process. Appl. **129**, 1 (2019).

[18] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, Theory of orientation tuning in visual cortex., Proceedings of the National Academy of Sciences **92**, 3844 (1995).

[19] T. G. Kurtz, Limit theorems for sequences of jump markov processes approximating ordinary differential processes, J. Appl. Probab. **8**, 344 (1971).

[20] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of neural science*, Vol. 4 (McGraw-hill New York, 2000).

[21] S. Thorpe, D. Fize, and C. Marlot, Speed of processing in the human visual system, Nature **381**, 520 (1996).

[22] A. Renart, J. De La Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris, The asynchronous state in cortical circuits, Science **327**, 587 (2010).

[23] See Supplemental Material at [URL will be inserted by publisher] Sec. I.

[24] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).

[25] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Spin-glass models of neural networks, Phys. Rev. A **32**, 1007 (1985).

[26] U. Pereira and N. Brunel, Attractor dynamics in networks with learning rules inferred from in vivo data, Neuron **99**, 227 (2018).

[27] D. J. Amit, *Modeling brain function: The world of attractor neural networks* (Cambridge university press, 1989).

[28] J. Hertz, A. Krogh, R. G. Palmer, and H. Horner, *Introduction to the theory of neural computation* (American Institute of Physics, 1991).

[29] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal dynamics: From single neurons to networks and models of cognition* (Cambridge University Press, 2014).

[30] B. DePasquale, D. Sussillo, L. Abbott, and M. M. Churchland, The centrality of population-level factors to network computation is demonstrated by a versatile approach for training spiking networks, Neuron **111**, 631 (2023).

[31] M. Beiran, A. Dubreuil, A. Valente, F. Mastrogiuseppe, and S. Ostojic, Shaping dynamics with multiple populations in low-rank recurrent networks, Neural Comput. **33**, 1572 (2021).

[32] See Supplemental Material at [URL will be inserted by publisher] Sec. II for additional explanations.

[33] J. Kadmon and H. Sompolinsky, Transition to chaos in random neuronal networks, Phys. Rev. X **5**, 041030 (2015).

[34] A. van Meegen and S. J. van Albada, Microscopic theory of intrinsic timescales in spiking neural networks, Phys. Rev. Research **3**, 043077 (2021).

[35] D. G. Clark, L. Abbott, and A. Litwin-Kumar, Dimension of activity in random neural networks, Phys. Rev. Lett. **131**, 118401 (2023).

[36] See Supplemental Material at [URL will be inserted by publisher] Sec. III for the proof.

[37] M. Talagrand, A new look at independence, Ann. Probab. , 1 (1996).

[38] M. Ledoux, *The concentration of measure phenomenon*, 89 (American Mathematical Soc., 2001).

[39] See Supplemental Material at [URL will be inserted by publisher] Sec. IV.

[40] See Supplemental Material at [URL will be inserted by publisher] Sec. V.

[41] M. Helias and D. Dahmen, *Statistical Field Theory for Neural Networks* (Springer, 2020).

[42] M. Talagrand, *Mean field models for spin glasses: Volume I: Basic examples*, Vol. 54 (Springer Science & Business Media, 2010).

[43] S. Mei and A. Montanari, The generalization error of random features regression: Precise asymptotics and the double descent curve, Comm. Pure Appl. Math. **75**, 667 (2022).

[44] 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).

[45] See Supplemental Material at [URL will be inserted by publisher] Sec. VI.

Supplementary Material on

Emergent rate-based dynamics in duplicate-free populations of spiking neurons

Valentin Schmutz,[1,2,*] Johanni Brea,[1] and Wulfram Gerstner[1]

[1]*School of Life Sciences and School of Computer and Communication Sciences,*

École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

[2]*UCL Queen Square Institute of Neurology,*

University College London, WC1E 6BT London, United Kingdom

CONTENTS

* Corresponding author: v.schmutz@ucl.ac.uk

I. LIMIT CORRELATION DENSITY

We prove two theorems characterizing the limit distribution of pairwise correlations between distinct neurons $\{C_{ij}\}_{i \neq j}$, as $N \to \infty$, assuming that the correlation C_{ij} between a pair of distinct neurons $i \neq j$ is given by

$$C_{ij} = \frac{1}{p} \sum_{\mu=1}^{p} \tilde{\xi}_{i\mu} \tilde{\xi}_{j\mu}, \tag{S1}$$

where $\tilde{\xi}_{i\mu} := \xi_{i\mu} / \sqrt{\frac{1}{p} \sum_{\mu=1}^{p} \xi_{i\mu}^2}$ and the $\{\xi_{i\mu}\}_{1 \leq i \leq N, 1 \leq \mu \leq p}$ are $i.i.d.$ are zero-mean, unit-variance normal random variables. More explicitly, we prove the convergence of the random empirical measure

$$\rho_N(\mathrm{d}\mathbf{z}) := \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} \delta_{\frac{1}{p} \sum_{\mu=1}^{p} \tilde{\xi}_{i\mu} \tilde{\xi}_{j\mu}}(\mathrm{d}z), \tag{S2}$$

to a deterministic probability measure as $N \to \infty$. In (S2), $\delta_{\frac{1}{p} \sum_{\mu=1}^{p} \tilde{\xi}_{i\mu} \tilde{\xi}_{j\mu}}(\mathrm{d}z)$ denotes the Dirac measure centered on the point $\frac{1}{p} \sum_{\mu=1}^{p} \tilde{\xi}_{i\mu} \tilde{\xi}_{j\mu}$.

Let us recall a general mathematical framework for studying the convergence of random empirical measures.

Definition 1. *Let $\mathcal{P}(\mathbb{R})$ denote the space of probability measures on $\mathbb{R}$ endowed with the topology of weak convergence. We say that a sequence of random probability measures $\{\rho_N\}_{N=1}^{\infty}$ in $\mathcal{P}(\mathbb{R})$ converges in probability to a deterministic probability measure $\rho \in \mathcal{P}(\mathbb{R})$ if, for all continuous and bounded functional $F : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$,*

$$\mathbb{E}[F(\rho_N)] \to \mathbb{E}[F(\rho)], \quad as \ N \to \infty.$$

In the field the interacting particle systems, It is well known that to show the convergence defined as in Definition 1, it suffices to show that for all continuous and bounded functions $f : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}\left[\left| \int f(z) \rho_N(\mathrm{d}z) - \int f(z) \rho(\mathrm{d}z) \right| \right] \to 0, \quad as \ N \to \infty;$$

see for example [1, Proposition 4.2]. (The functions f are usually called "test functions".)

We distinguish two cases: first, the case where $p \in \mathbb{N}_+$ is fixed as $N \to \infty$ (Theorem 1); second , the case where $p = \alpha N$, for $\alpha > 0$, as $N \to \infty$ (Theorem 2). To prove both theorems, we will use the following lemma which does not depend on the scaling of p as $N \to \infty$.

Lemma 1. *For any $N > 1$ and $p > 0$, let $\{\zeta_{i,\mu}\}_{1\leq i\leq N, 1\leq \mu\leq p}$ be a collection of i.i.d. random variables. For all bounded functions $f : \mathbb{R} \to \mathbb{R}$,*

$$\mathrm{Var}\left(\frac{2}{N(N-1)} \sum_{1\leq i<j\leq N} f\left(\frac{1}{p}\sum_{\mu=1}^{p} \zeta_{i\mu}\zeta_{j\mu}\right)\right) \leq \frac{2(2N-3)}{N(N-1)}\|f\|_\infty^2 \qquad (S3)$$

In particular, the variance bound (S3) does not depend on p and it vanishes as $N \to \infty$.

Proof. We use the shorthand $f(i,j) := f\left(\frac{1}{p}\sum_{\mu=1}^{p}\zeta_{i\mu}\zeta_{j\mu}\right)$.

$$\mathrm{Var}\left(\frac{2}{N(N-1)}\sum_{1\leq i<j\leq N} f(i,j)\right)$$

$$= \frac{4}{N^2(N-1)^2}\left\{\mathbb{E}\left[\left(\sum_{1\leq i<j\leq N} f(i,j)\right)^2\right] - \left(\mathbb{E}\left[\sum_{1\leq i<j\leq N} f(i,j)\right]\right)^2\right\}$$

$$= \frac{4}{N^2(N-1)^2}\sum_{1\leq i<j\leq N}\sum_{1\leq k<l\leq N}\mathbb{E}[f(i,j)f(k,l)] - \mathbb{E}[f(i,j)]\mathbb{E}[f(k,l)]. \qquad (S4)$$

If $\{i,j\}\cap\{k,l\} = \varnothing$, $f(i,j)$ and $f(k,l)$ are independent and $\mathbb{E}[f(i,j)f(k,l)] - \mathbb{E}[f(i,j)]\mathbb{E}[f(k,l)] = 0$. If $\{i,j\}\cap\{k,l\} \neq \varnothing$, $\mathbb{E}[f(i,j)f(k,l)] - \mathbb{E}[f(i,j)]\mathbb{E}[f(k,l)]$ can be upper-bounded by $\|f\|_\infty^2$. Since, in the double sum (S4), $\{i,j\}\cap\{k,l\} = \varnothing$ for $N(N-1)/2 \times (N-2)(N-3)/2$ of the summands, we know that $\{i,j\}\cap\{k,l\} \neq \varnothing$ for

$$\frac{N(N-1)}{2}\frac{N(N-1)}{2} - \frac{N(N-1)}{2}\frac{(N-2)(N-3)}{2} = \frac{N(N-1)}{2}(2N-3)$$

of the summands. Hence, we get that

$$\frac{4}{N^2(N-1)^2}\sum_{1\leq i<j\leq N}\sum_{1\leq k<l\leq N}\mathbb{E}[f(i,j)f(k,l)] - \mathbb{E}[f(i,j)]\mathbb{E}[f(k,l)]$$

$$\leq \frac{4}{N^2(N-1)^2}\frac{N(N-1)}{2}(2N-3)\|f\|_\infty^2 = \frac{2(2N-3)}{N(N-1)}\|f\|_\infty^2.$$

$\square$

Theorem 1. *For any fixed $p > 1$, the sequence of empirical measures*

$$\rho_N(\mathrm{d}z) := \frac{2}{N(N-1)}\sum_{1\leq i<j\leq N}\delta_{\frac{1}{p}\sum_{\mu=1}^{p}\tilde{\xi}_{i\mu}\tilde{\xi}_{j\mu}}(\mathrm{d}z) \qquad \forall N > 1,$$

converges in probability (according to Definition 1), as $N \to \infty$, to the deterministic probability measure $\rho \in \mathcal{P}(\mathbb{R})$ given by

$$\rho(\mathrm{d}z) := \mathbb{1}_{[-1,1]}(z)\frac{\Gamma(p/2)}{\sqrt{\pi}\,\Gamma((p-1)/2)}(1-z^2)^{\frac{p-3}{2}}\mathrm{d}z.$$

Proof. As mentioned above, to show the convergence of ρ_N to ρ in probability, it suffices to show that for all continuous and bounded functions $f : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}\left[\left|\int f(z)\rho_N(\mathrm{d}z) - \int f(z)\rho(\mathrm{d}z)\right|\right] \to 0, \quad \text{as } N \to \infty.$$

By Jensen's inequality [2, Theorem 5.1 p. 132] and Lemma 1,

$$\mathbb{E}\left[\left|\int f(z)\rho_N(\mathrm{d}z) - \mathbb{E}\left[\int f(z)\rho_N(\mathrm{d}z)\right]\right|\right] \leq \left[\mathrm{Var}\left(\int f(z)\rho_N(\mathrm{d}z)\right)\right]^{1/2} \xrightarrow[N\to\infty]{} 0.$$

Hence, it only remains to show that, for all $N > 1$,

$$\mathbb{E}\left[\int f(z)\rho_N(\mathrm{d}z)\right] = \int f(z)\rho(\mathrm{d}z) = \int f(z)\frac{\Gamma(p/2)}{\sqrt{\pi}\,\Gamma((p-1)/2)}(1-z^2)^{\frac{p-3}{2}}\mathrm{d}z. \tag{S5}$$

Indeed, we have

$$\mathbb{E}\left[\int f(z)\rho_N(\mathrm{d}z)\right] = \mathbb{E}\left[\frac{2}{N(N-1)}\sum_{1\leq i<j\leq N} f\left(\frac{1}{p}\sum_{\mu=1}^p \tilde{\xi}_{i\mu}\tilde{\xi}_{j\mu}\right)\right] = \mathbb{E}\left[f\left(\frac{1}{p}\sum_{\mu=1}^p \tilde{\xi}_{1\mu}\tilde{\xi}_{2\mu}\right)\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[f\left(\frac{1}{p}\sum_{\mu=1}^p \tilde{\xi}_{1\mu}\tilde{\xi}_{2\mu}\right) \,\bigg|\, \tilde{\xi}_{1,1},\ldots,\tilde{\xi}_{1,p}\right]\right]. \tag{S6}$$

By definition, the random vector $\frac{1}{\sqrt{p}}(\tilde{\xi}_{1,1},\ldots,\tilde{\xi}_{1,p})^T$ lies on the $(p-1)$-sphere of unit radius, $\mathbb{S}^{p-1}$, and, by symmetry, the conditional expectation in (S6) is constant for all $\frac{1}{\sqrt{p}}(\tilde{\xi}_{1,1},\ldots,\tilde{\xi}_{1,p})^T \in \mathbb{S}^{p-1}$. Since the vector $\frac{1}{\sqrt{p}}(\tilde{\xi}_{2,1},\ldots,\tilde{\xi}_{2,p})^T$ is uniformly distributed on $\mathbb{S}^{p-1}$, we have, by the Funk-Hecke formula (see [3, Theorem 1.2.9] for a reference),

$$\mathbb{E}\left[\mathbb{E}\left[f\left(\frac{1}{p}\sum_{\mu=1}^p \tilde{\xi}_{1\mu}\tilde{\xi}_{2\mu}\right) \,\bigg|\, \tilde{\xi}_{1,1},\ldots,\tilde{\xi}_{1,p}\right]\right] = \int_{-1}^1 f(z)\frac{\Gamma(p/2)}{\sqrt{\pi}\,\Gamma((p-1)/2)}(1-z^2)^{\frac{p-3}{2}}\mathrm{d}z,$$

which concludes the proof of (S5). $\qquad\square$

Corollary 1. *If $\theta_1,\ldots,\theta_N$ are i.i.d. and uniformly distributed on the interval $[0, 2\pi[$, the sequence of empirical measures*

$$\rho_N(\mathrm{d}z) := \frac{2}{N(N-1)}\sum_{1\leq i<j\leq N} \delta_{\cos(\theta_i-\theta_j)}(\mathrm{d}z)$$

converges in probability (according to Definition 1), as $N \to \infty$, to the deterministic probability measure $\rho \in \mathcal{P}(\mathbb{R})$ given by

$$\rho(\mathrm{d}z) := \mathbb{1}_{[-1,1]}(z) \frac{1}{\pi} \frac{1}{\sqrt{1-z^2}} \mathrm{d}z.$$

Proof. Notice that the collection of random variables $\{\cos(\theta_i - \theta_j)\}_{i<j}$ has the same joint law as $\{\frac{1}{2}(\tilde{\xi}_{i,1}\tilde{\xi}_{j,1} + \tilde{\xi}_{i,2}\tilde{\xi}_{j,2})\}_{i<j}$. Thus, the result follows from the Theorem 1 with $p = 2$. $\quad\square$

Theorem 2. *If $p = \alpha N$ for some fixed $\alpha > 0$, the sequence of empirical measures*

$$\widehat{\rho_N}(\mathrm{d}z) := \frac{2}{N(N-1)} \sum_{1 \le i < j \le N} \delta_{\frac{1}{\sqrt{p}} \sum_{\mu=1}^{p} \tilde{\xi}_{i\mu}\tilde{\xi}_{j\mu}}(\mathrm{d}z) \qquad \forall\, N > 1,$$

converge in probability (according to Definition 1), as $N \to \infty$, to the deterministic probability measure $\rho \in \mathcal{P}(\mathbb{R})$ given by

$$\rho(\mathrm{d}z) := \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \mathrm{d}z.$$

Proof. As in the proof of Theorem 1, to show that the convergence in probability of ρ_N to ρ, it suffices to show that for all continuous and bounded function $f : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}\left[\left| \int f(z)\widehat{\rho_N}(\mathrm{d}z) - \int f(z)\rho(\mathrm{d}z) \right| \right] \to 0, \quad \text{as } N \to \infty.$$

By triangular inequality,

$$\mathbb{E}\left[\left| \int f(z)\widehat{\rho_N}(\mathrm{d}z) - \int f(z)\rho(\mathrm{d}z) \right| \right]$$

$$\le \mathbb{E}\left[\left| \int f(z)\widehat{\rho_N}(\mathrm{d}z) - \mathbb{E}\left[\int f(z)\widehat{\rho_N}(\mathrm{d}z) \right] \right| \right] + \left| \mathbb{E}\left[\int f(z)\widehat{\rho_N}(\mathrm{d}z) \right] - \int f(z)\rho(\mathrm{d}z) \right|. \quad \text{(S7)}$$

By Jensen's inequality and Lemma 1, we know the first term on the right-hand side of (S7) vanishes as $N \to \infty$ and it only remains to show that

$$\mathbb{E}\left[\int f(z)\widehat{\rho_N}(\mathrm{d}z) \right] \to \int f(z)\rho(\mathrm{d}z), \quad \text{as } N \to \infty. \quad \text{(S8)}$$

Since

$$\mathbb{E}\left[\int f(z)\widehat{\rho_N}(\mathrm{d}z) \right] = \mathbb{E}\left[f\left(\frac{1}{\sqrt{p}} \sum_{\mu=1}^{p} \tilde{\xi}_{1\mu}\tilde{\xi}_{2\mu} \right) \right],$$

showing (S8) is equivalent to showing

$$\mathbb{E}\left[f\left(\frac{1}{\sqrt{p}}\sum_{\mu=1}^{p}\tilde{\xi}_{1\mu}\tilde{\xi}_{2\mu}\right)\right] \to \int f(z)\rho(\mathrm{d}z), \quad \text{as } N \to \infty,$$

which in turn is equivalent to showing that the random variables $\widetilde{X}_N := \frac{1}{\sqrt{p}}\sum_{\mu=1}^{p}\tilde{\xi}_{1\mu}\tilde{\xi}_{2\mu}$ converge in law to a standard normal variable $X \sim \mathcal{N}(0,1)$ as $N \to \infty$. We recall that $p = \alpha N$. As $N \to \infty$, we have, on the one hand, that $\sqrt{\frac{1}{p}\sum_{\mu=1}^{p}\xi_{i\mu}^2}$ converges in probability to 1 by the weak law of large numbers and the continuous mapping theorem [2, Theorem 10.3 p. 245], and on the other hand, $X_N := \frac{1}{\sqrt{p}}\sum_{\mu=1}^{p}\xi_{1\mu}\xi_{2\mu}$ converges in law to X by the central limit theorem. Hence, $\widetilde{X}_N$ converges in law to X by Slutsky's theorem [2, Theorem 11.4 p. 249]. This proves (S8) and concludes the proof. $\qquad \square$

II. FROM LOW-RANK SNNS TO NEURAL FIELD EQUATIONS

For any number of neurons N and any number of patterns p, let $\boldsymbol{\xi}$ be a random $N \times p$-matrix with $i.i.d.$, zero-mean, unit-variance, normally distributed entries $\{\xi_{i\mu}\}_{i,\mu}$. We choose a connectivity matrix given by

$$J_{ij} := \frac{1}{cN} \sum_{\mu=1}^{p} \xi_{i\mu} \left(\phi(\xi_{j\mu}) - a \right) \quad \text{for all } i \neq j,$$

and $J_{ii} := 0$ for all i. As in the main text, $a := \int_{-\infty}^{\infty} \mathcal{D}z \, \phi(z)$ and $c := \int_{-\infty}^{\infty} \mathcal{D}z \, (\phi(z) - a)^2$, where $\mathcal{D}z$ is the standard Gaussian measure $\mathcal{D}z = dz$. This type of connectivity matrix was introduced by Brunel and Pereira [4] in the context of Hopfield networks.

When the number of patterns p is kept constant as the number of neurons N tends to infinity, we say that the connectivity matrix $\mathbf{J} = \{J_{ij}\}_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}}$ is low-rank since its rank remains finite as $N \to \infty$. Although large, random, low-rank networks of neurons [5, 6] are usually not presented as spatially structured networks, they do have an implicit spatial structure. By identifying this spatial structure, we can derive exact neural field equations describing the dynamics of large networks, as we show here for the connectivity matrix $\mathbf{J}$.

For simplicity, let us first consider an SNN of N neurons without external input. The N neurons are assigned to N points in $\mathbb{R}^p$ corresponding to the rows of $\boldsymbol{\xi}$, i.e., neuron i is assigned to the point $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{ip})$. Writing $u(t, \boldsymbol{\xi}_i) = h_i(t)$ for all $i = 1, \ldots, N$, we can interpret $u(t, \boldsymbol{\xi}_i)$ as being the potential of a neuron at location $\boldsymbol{\xi}_i \in \mathbb{R}^p$ at time t. Without loss of generality, let us consider the dynamics of neuron $i = 1$, which can be rewritten

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} u(t, \boldsymbol{\xi}_1) = -u(t, \boldsymbol{\xi}_1) + \sum_{j=2}^{N} \underbrace{\frac{1}{cN} \sum_{\mu=1}^{p} \xi_{1\mu} \left(\phi(\xi_{j\mu}) - a \right)}_{=J_{1j}} s_j(t),$$

$$= -u(t, \boldsymbol{\xi}_1) + \frac{1}{c} \sum_{\mu=1}^{p} \xi_{1\mu} \frac{1}{N} \sum_{j=2}^{N} \left(\phi(\xi_{j\mu}) - a \right) s_j(t),$$

where $s_j(t)$ has instantaneous firing rate $\phi(u(t, \boldsymbol{\xi}_j))$. By spatial averaging [7, 8], we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{j=2}^{N} \left(\phi(\xi_{j\mu}) - a \right) s_j(t) = \lim_{N \to \infty} \frac{1}{N} \sum_{j=2}^{N} \left(\phi(\xi_{j\mu}) - a \right) \phi(u(t, \boldsymbol{\xi}_j))$$

$$= \int_{\mathbb{R}^p} \underbrace{\lim_{N \to \infty} \frac{1}{N} \sum_{j=2}^{N} \delta_{\boldsymbol{\xi}_j}(d\mathbf{z})}_{= \mathcal{D}\mathbf{z}} \left(\phi(z_\mu) - a \right) \phi(u(t, \mathbf{z}))$$

$$= \int_{\mathbb{R}^p} \mathcal{D}\mathbf{z} \left(\phi(z_\mu) - a \right) \phi(u(t, \mathbf{z})), \tag{S9}$$

where $\delta_{\boldsymbol{\xi}_j}(d\mathbf{z})$ denotes the Dirac measure centered on the point $\boldsymbol{\xi}_j$ and $\mathcal{D}\mathbf{z}$ denotes the p-dimensional standard Gaussian measure

$$\mathcal{D}\mathbf{z} = dz_1 \ldots dz_p \frac{1}{(2\pi)^{p/2}} e^{-\|\mathbf{z}\|_2^2}.$$

Intuitively, the equalities (S9) simply reflect the fact that as $N \to \infty$, the empirical measure $\frac{1}{N} \sum_{i=1}^{N} \delta_{\boldsymbol{\xi}_i}$ on $\mathbb{R}^p$, which summarizes the locations $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_N$ of neurons in $\mathbb{R}^p$, converges to the standard Gaussian measure $\mathcal{D}\mathbf{z}$ on $\mathbb{R}^p$ (since the p-dimensional random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_N$ are *i.i.d.* with distribution $\mathcal{D}\mathbf{z}$), turning the summation over neurons into an integration over a Gaussian measure. Therefore, since the spatial distribution of neurons in $\mathbb{R}^p$ converges to a continuous Gaussian distribution as $N \to \infty$, we find that the continuous field $u(t, \mathbf{y})$, with $\mathbf{y} \in \mathbb{R}^p$, solves the integro-differential equation

$$\tau \frac{\partial}{\partial t} u(t, \mathbf{y}) = -u(t, \mathbf{y}) + \frac{1}{c} \sum_{\mu=1}^{p} y_\mu \int_{\mathbb{R}^p} \mathcal{D}\mathbf{z} \left(\phi(z_\mu) - a \right) \phi(u(t, \mathbf{z})), \qquad \forall \mathbf{y} \in \mathbb{R}^p.$$

Defining the connectivity kernel $w : \mathbb{R}^p \to \mathbb{R}^p$

$$w(\mathbf{y}, \mathbf{z}) := \frac{1}{c} \sum_{\mu=1}^{p} y_\mu \left(\phi(z_\mu) - a \right), \tag{S10}$$

we get a neural field equation (see [9] and Eq. 6.129 in [10, p. 244]):

$$\tau \frac{\partial}{\partial t} u(t, \mathbf{y}) = -u(t, \mathbf{y}) + \int_{\mathbb{R}^p} \mathcal{D}\mathbf{z} \, w(\mathbf{y}, \mathbf{z}) \phi(u(t, \mathbf{z})), \qquad \forall \mathbf{y} \in \mathbb{R}^p. \tag{S11}$$

While the arguments presented here are informal, the neural field equation is the exact mean-field limit of the low-rank SNN: using the embedding of the low-rank SNN in $\mathbb{R}^p$, the convergence of the SNN to the neural field equation (S11) is guaranteed by rigorous results for spatially structured SNNs [11].

10

Of course, by the same arguments, the corresponding low-rank RNN also converges to the neural field equation (S11).

In the case where half of the neurons receive an external input $\frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^{p} \xi_{i\mu} \eta_\mu(t)$, as defined in the main text, we can describe the population dynamics with two neural field equations. Following the same steps as above, we get that for each neuron i receiving external input, $\lim_{N \to \infty} h_i^{\text{in}}(t) = \lim_{N \to \infty} x_i^{\text{in}}(t) = u^{\text{in}}(t, \boldsymbol{\xi}_i)$, and for all neuron i' receiving no external input but only recurrent input, $\lim_{N \to \infty} h_{i'}^{\text{rec}}(t) = \lim_{N \to \infty} x_{i'}^{\text{rec}}(t) = u^{\text{rec}}(t, \boldsymbol{\xi}_{i'})$, where the fields u^{in} and u^{rec} are the solutions to the system

$$
\tau \frac{\partial}{\partial t} u^{\text{in}}(t, \mathbf{y}) = -u^{\text{in}}(t, \mathbf{y}) + \frac{1}{2} \int_{\mathbb{R}^p} \mathcal{D}\mathbf{z}\, w(\mathbf{y}, \mathbf{z}) \phi(u^{\text{in}}(t, \mathbf{z}))
$$
$$
+ \frac{1}{2} \int_{\mathbb{R}^p} \mathcal{D}\mathbf{z}\, w(\mathbf{y}, \mathbf{z}) \phi(u^{\text{rec}}(t, \mathbf{z})) + \frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^{p} y_\mu \eta_\mu(t), \qquad \forall \mathbf{y} \in \mathbb{R}^p,
$$
$$
\tau \frac{\partial}{\partial t} u^{\text{rec}}(t, \mathbf{y}) = -u^{\text{in}}(t, \mathbf{y}) + \frac{1}{2} \int_{\mathbb{R}^p} \mathcal{D}\mathbf{z}\, w(\mathbf{y}, \mathbf{z}) \phi(u^{\text{in}}(t, \mathbf{z}))
$$
$$
+ \frac{1}{2} \int_{\mathbb{R}^p} \mathcal{D}\mathbf{z}\, w(\mathbf{y}, \mathbf{z}) \phi(u^{\text{rec}}(t, \mathbf{z})), \qquad \forall \mathbf{y} \in \mathbb{R}^p,
$$

where the $\{\eta_\mu(t)\}_{\mu=1}^{p}$ are the formal derivatives of independent Wiener processes (or standard Brownian motions) $B_1(t), \ldots, B_p(t)$, i.e. $\eta_\mu(t) = \mathrm{d}B_\mu(t)/\mathrm{d}t$.

III. BOUND ON THE PROBABILITY OF A DUPLICATE

Let us assume that the correlations $\{C_{ij}\}_{1\leq i,j\leq N}$ are given by (S1).

Lemma 2. *For any $N, p > 1$ and for any $0 < \varepsilon < 1$, we have the bound*

$$\mathbb{P}(\exists\, i \neq j \text{ such that } |C_{ij}| \geq \varepsilon) \leq \frac{N(N-1)}{2}\frac{\left(1-\varepsilon^2\right)^{\frac{p-1}{2}}}{\sqrt{\pi}}.$$

Proof. By union bound, we have

$$\mathbb{P}(\exists\, i \neq j \text{ such that } |C_{ij}| \geq \varepsilon) \leq \frac{N(N-1)}{2}\mathbb{P}(|C_{12}| \geq \varepsilon).$$

Then, we use the fact that C_{12} follows a Gegenbauer distribution with parameter p, whence

$$\mathbb{P}(|C_{12}| \geq \varepsilon) = 2\int_\varepsilon^1 \frac{\Gamma(p/2)}{\sqrt{\pi}\,\Gamma((p-1)/2)}(1-z^2)^{\frac{p-3}{2}}\mathrm{d}z.$$

First, we notice that

$$\frac{\Gamma(p/2)}{\Gamma((p-1)/2)} \leq \frac{\Gamma((p+1)/2)}{\Gamma((p-1)/2)} = \frac{p-1}{2};$$

second, using the change of variable $y^2 = 1 - z^2$,

$$\int_\varepsilon^1 (1-z^2)^{\frac{p-3}{2}}\mathrm{d}z = \int_0^{\sqrt{1-\varepsilon^2}} \left[y^2\right]^{\frac{p-3}{2}}\frac{y}{\sqrt{1-y^2}}\mathrm{d}y \leq \int_0^{\sqrt{1-\varepsilon^2}} y^{p-2}\mathrm{d}y = \frac{\left(1-\varepsilon^2\right)^{\frac{p-1}{2}}}{p-1}.$$

Hence, we have shown that

$$\mathbb{P}(|C_{12}| \geq \varepsilon) \leq \frac{\left(1-\varepsilon^2\right)^{\frac{p-1}{2}}}{\sqrt{\pi}},$$

which concludes the proof. $\qquad\square$

IV.   CONCENTRATION OF THE ℓ^2 NORMS OF THE INCOMING WEIGHTS

We recall that the ℓ^2 norm of the incoming weights to neuron i is $\|\mathbf{J}_i\|_2 = \sqrt{\sum_{j=1}^N J_{ij}^2}$ and that the weights are given by $J_{ij} := \frac{1}{cN} \sum_{\mu=1}^p \xi_{i\mu} \left(\phi(\xi_{j\mu}) - a\right)$, for all $i \neq j$, and $J_{ii} := 0$, where $a = \int_{-\infty}^\infty \mathcal{D}z\, \phi(z)$ and $c = \int_{-\infty}^\infty \mathcal{D}z\, (\phi(z) - a)^2$ are constants ($\mathcal{D}z$ denotes the standard Gaussian measure).

Theorem 3. *When $p = \alpha N$ for some fixed $\alpha > 0$, the ℓ^2 norm of the incoming weights of a typical neuron i converges in probability to $\sqrt{\alpha/c}$ as $N \to \infty$, i.e.*

$$\|\mathbf{J}_i\|_2 \xrightarrow[N\to\infty]{\mathbb{P}} \sqrt{\frac{\alpha}{c}}.$$

Proof. Without loss of generality, we can take $i = 1$. The proof can be decomposed in two steps. First (Step 1), we will show that

$$\lim_{N\to\infty} \mathbb{E}\left[\|\mathbf{J}_1\|_2^2\right] = \frac{\alpha}{c}, \tag{S12}$$

then (Step 2), we will show that

$$\mathrm{Var}\left(\|\mathbf{J}_1\|_2^2\right) = \frac{2\alpha(1+\alpha)}{c^2 N} + \mathcal{O}\left(\frac{1}{N^2}\right). \tag{S13}$$

If (S12) and (S13) are verified, $\|\mathbf{J}_1\|_2^2$ converges in probability to α/c and, by the continuous mapping theorem, $\|\mathbf{J}_1\|_2$ converges in probability to $\sqrt{\alpha/c}$.

Step 1:

$$\mathbb{E}\left[\|\mathbf{J}_1\|_2^2\right] = \mathbb{E}\left[\sum_{j=1}^N J_{1j}^2\right] = \sum_{j=1}^N \mathbb{E}\left[J_{1j}^2\right] = \frac{1}{c^2 N^2} \sum_{j=2}^N \mathbb{E}\left[\left(\sum_{\mu=1}^p \xi_{1\mu}\left(\phi(\xi_{j\mu}) - a\right)\right)^2\right]$$

$$\overset{(*)}{=} \frac{1}{c^2 N^2} \sum_{j=2}^N \mathbb{E}\left[\sum_{\mu=1}^p \xi_{1\mu}^2 \left(\phi(\xi_{j\mu}) - a\right)^2\right] = \frac{1}{c^2 N^2} \sum_{j=2}^N \sum_{\mu=1}^p \mathbb{E}\left[\xi_{1\mu}^2 \left(\phi(\xi_{j\mu}) - a\right)^2\right]$$

$$= \frac{1}{c^2 N^2} (N-1)\alpha N c \xrightarrow[N\to\infty]{} \frac{\alpha}{c}$$

For the equality $(*)$, we used the fact that the cross terms are null since the columns of the random matrix $\boldsymbol{\xi}$, i.e. the patterns, are independent and when $\mu \neq \nu$,

$$\mathbb{E}\left[\xi_{1\mu}\left(\phi(\xi_{j\mu}) - a\right)\xi_{1\nu}\left(\phi(\xi_{j\nu}) - a\right)\right] = \mathbb{E}\left[\xi_{1\mu}\left(\phi(\xi_{j\mu}) - a\right)\right]\mathbb{E}_{\boldsymbol{\xi}}\left[\xi_{1\nu}\left(\phi(\xi_{j\nu}) - a\right)\right] = 0.$$

15

<u>Step 2:</u>

$$\text{Var}\left(\|\mathbf{J}_1\|_2^2\right) = \text{Var}\left(\sum_{j=1}^{N} J_{1j}^2\right) = \mathbb{E}\left[\left(\sum_{j=1}^{N} J_{1j}^2\right)^2\right] - \mathbb{E}\left[\sum_{j=1}^{N} J_{1j}^2\right]^2. \tag{S14}$$

By the computation in Step 1 above, we have

$$\mathbb{E}\left[\sum_{j=1}^{N} J_{1j}^2\right]^2 = \frac{1}{c^2}\frac{(N-1)^2 p^2}{N^4}. \tag{S15}$$

The second moment can be decomposed as

$$\mathbb{E}\left[\left(\sum_{j=1}^{N} J_{1j}^2\right)^2\right] = \mathbb{E}\left[\sum_{j=1}^{N} J_{1j}^4 + \sum_{j=1}^{N}\sum_{k\neq j} J_{1j}^2 J_{1k}^2\right]$$

$$= (N-1)\mathbb{E}\left[J_{12}^4\right] + (N-1)(N-2)\mathbb{E}\left[J_{12}^2 J_{13}^2\right], \tag{S16}$$

and we compute the terms $\mathbb{E}\left[J_{12}^4\right]$ and $\mathbb{E}\left[J_{12}^2 J_{13}^2\right]$ separately.

$$\mathbb{E}\left[J_{12}^4\right] = \frac{1}{c^4 N^4}\mathbb{E}\left[\left(\sum_{\mu=1}^{p} \xi_{1\mu}\left(\phi(\xi_{2\mu}) - a\right)\right)^4\right]$$

$$= \frac{1}{c^4 N^4}\sum_{\mu=1}^{p}\mathbb{E}\left[\xi_{1\mu}^4\left(\phi(\xi_{2\mu}) - a\right)^4\right]$$

$$+ \frac{1}{c^4 N^4}\sum_{1\leq\mu<\nu\leq p}\binom{4}{2}\mathbb{E}\left[\xi_{1\mu}^2\left(\phi(\xi_{2\mu}) - a\right)^2\xi_{1\nu}^2\left(\phi(\xi_{2\nu}) - a\right)^2\right]$$

where for the second equality, we used the fact that, in the multinomial expansion of

$$\mathbb{E}\left[\left(\sum_{\mu=1}^{p} \xi_{1\mu}\left(\phi(\xi_{2\mu}) - a\right)\right)^4\right],$$

all terms with at least one odd power are null. Hence, we get that

$$\mathbb{E}\left[J_{12}^4\right] = \mathcal{O}\left(\frac{1}{N^3}\right) + \frac{1}{c^4 N^4}\frac{p(p-1)}{2}6c^2 = \frac{3}{c^2 N^4}p(p-1) + \mathcal{O}\left(\frac{1}{N^3}\right). \tag{S17}$$

Similarly,

$$
\begin{aligned}
\mathbb{E}\left[J_{12}^2 J_{13}^2\right] &= \frac{1}{c^4 N^4}\mathbb{E}\left[\left(\sum_{\mu=1}^{p}\xi_{1\mu}\left(\phi(\xi_{2\mu})-a\right)\right)^2\left(\sum_{\mu=1}^{p}\xi_{1\mu}\left(\phi(\xi_{3\mu})-a\right)\right)^2\right] \\
&= \frac{1}{c^4 N^4}\sum_{\mu=1}^{p}\mathbb{E}\left[\xi_{1\mu}^4\left(\phi(\xi_{2\mu})-a\right)^2\left(\phi(\xi_{3\mu})-a\right)^2\right] \\
&\quad + \frac{2}{c^4 N^4}\sum_{1\le\mu<\nu\le p}\mathbb{E}\left[\xi_{1\mu}^2\left(\phi(\xi_{2\mu})-a\right)^2\xi_{1\nu}^2\left(\phi(\xi_{3\nu})-a\right)^2\right] \\
&= \frac{3}{c^2 N^4}p + \frac{1}{c^2 N^4}p(p-1). \tag{S18}
\end{aligned}
$$

From Eqs. (S14), (S15), (S16), (S17) and (S18), we get

$$
\begin{aligned}
\mathrm{Var}\left(\|\mathbf{J}_1\|_2^2\right) &= \mathbb{E}\left[\left(\sum_{j=1}^{N}J_{ij}^2\right)^2\right] - \mathbb{E}\left[\sum_{j=1}^{N}J_{ij}^2\right]^2 \\
&= (N-1)\mathbb{E}\left[J_{12}^4\right] + (N-1)(N-2)\mathbb{E}\left[J_{12}^2 J_{13}^2\right] - \frac{1}{c^2 N^4}(N-1)^2 p^2 \\
&= \frac{3}{c^2}\frac{(N-1)p(p-1)}{N^4} \\
&\quad + \frac{3}{c^2}\frac{(N-1)(N-2)p}{N^4} + \frac{1}{c^2}\frac{(N-1)(N-2)p(p-1)}{N^4} \\
&\quad - \frac{1}{c^2}\frac{(N-1)^2 p^2}{N^4} + \mathcal{O}\left(\frac{1}{N^2}\right).
\end{aligned}
$$

Using the fact that

$$
\begin{aligned}
(N-1)^2 p^2 &= (N-1)(N-2)p^2 + (N-1)p^2 \\
&= (N-1)(N-2)p(p-1) + (N-1)(N-2)p + (N-1)p(p-1) + (N-1)p,
\end{aligned}
$$

and after some rearrangement of terms, we obtain, for large N,

$$
\mathrm{Var}\left(\|\mathbf{J}_1\|_2^2\right) = \frac{2\alpha(1+\alpha)}{c^2 N} + \mathcal{O}\left(\frac{1}{N^2}\right).
$$

$\square$

V. SIMPLIFIED FEEDFORWARD MODEL

To understand mathematically why large SNNs converge to large RNNs as the load $\alpha \to 0$, we consider a simplified feedforward model. This simplified model is obtained by keeping network connections from 'in' to 'rec' neurons and removing all the other connections. After this pruning procedure, we are left with two-layer feedforward networks where the 'in' neurons make up the first layer, $(h_i^{\text{in}}(t), x_i^{\text{in}}(t)) \mapsto (h_i^{(1)}(t), x_i^{(1)}(t))$, and the 'rec' neurons make up the second layer, $(h_i^{\text{rec}}(t), x_i^{\text{rec}}(t)) \mapsto (h_i^{(2)}(t), x_i^{(2)}(t))$ (Fig. S1).
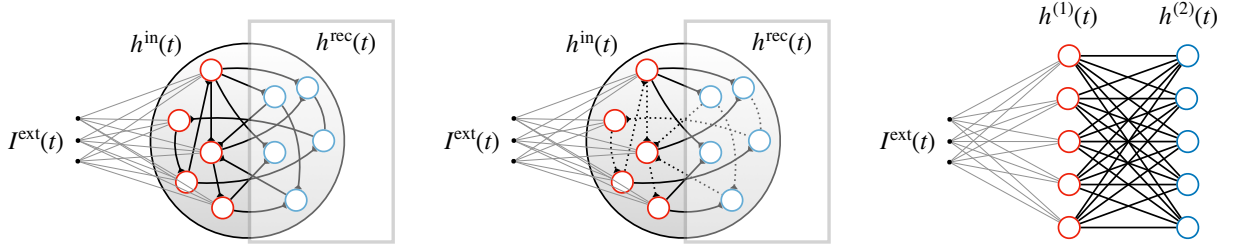

FIG. S1. Feedforward model as an simplification of the input-driven model.

In the feedforward SNN, the dynamics of the neurons of the first layer follows

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} h_j^{(1)}(t) = -h_j^{(1)}(t) + \frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^{p} \xi_{j\mu} \eta_\mu(t), \tag{S19a}$$

and the dynamics of the neurons in the second layer follows

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} h_i^{(2)}(t) = -h_i^{(2)}(t) + \sum_{j=1}^{N/2} J_{ij} S_j^{(1)}(t). \tag{S19b}$$

Analogously, in the feedforward rate network, the dynamics of the units reads

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} x_j^{(1)}(t) = -x_j^{(1)}(t) + \frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^{p} \xi_{j\mu} \eta_\mu(t), \tag{S20a}$$

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} x_i^{(2)}(t) = -x_i^{(2)}(t) + \sum_{j=1}^{N/2} J_{ij} \phi(x_j^{(1)}(t)). \tag{S20b}$$

In the following, we assume that the systems (S19) and (S20) share the same initial conditions at time 0, i.e. $h_j^{(1)}(0) = x_j^{(1)}(0)$ and $h_i^{(2)}(0) = x_i^{(2)}(0)$.

Theorem 4. *If $\max \phi < \infty$, for any $N > 1$, and for any (feedforward) connectivity matrix $\mathbf{J}$, we have, for all i in the second layer,*

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \left| h_i^{(2)}(t) - x_i^{(2)}(t) \right| \mathrm{d}t \leq \sqrt{\frac{\max \phi}{2\tau}} \|\mathbf{J}_i\|_2. \tag{S21}$$

19

Before proving the theorem, it is useful to rewrite the dynamics of the feedforward SNN (S19) and the feedforward rate network (S20) in a more mathematical form, which will help us to better disentangle the three independent sources of variability present in the model: (i) the quenched disorder $\boldsymbol{\xi}$, (ii) the external input signal $\{\eta_\mu\}_{\mu=1}^p$, and (iii) neuronal spike noise. Let $\{\pi_i\}_{i=1}^{N/2}$ be a collection of $N/2$ independent Poisson random measures on $\mathbb{R}_+ \times \mathbb{R}_+$ with unit intensity [12] (see [13] and [14, Ch. 1.2] for gentle introductions on Poisson random measures in the context of spiking neuron models) and let $\{B_\mu(t)\}_{\mu=1}^p$ be p independent standard Brownian motions. We can rewrite the dynamics of the feedforward SNN (S19) as a system of stochastic differential equations

$$\tau \, dh_j^{(1)}(t) = -h_j^{(1)}(t)dt + \frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^p \xi_{j\mu} dB_\mu(t), \tag{S22a}$$

$$Z_j^{(1)}(t) = \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \le \phi(h_j^{(1)}(s-))} \pi_j(ds, dz), \tag{S22b}$$

$$\tau \, dh_i^{(2)}(t) = -h_i^{(2)}(t)dt + \sum_{j=1}^{N/2} J_{ij} dZ_j^{(1)}(t), \tag{S22c}$$

where $h_j(s-) = \lim_{r \to s-} h_j(r)$ denotes the left-handed limit (which implies that the dynamics are defined in terms of Itô calculus). Formally, the spike train of neuron j, $S_j^{(1)}(t) = \sum_k \delta(t - t_j^k)$ (where the $\{t_j^k\}_k$ are the spike times of neuron j), is simply $S_j^{(1)}(t) = dZ_j^{(1)}(t)/dt$. The mathematical notation (S22), while less common in the theoretical neuroscience literature, has the important advantage of clearly separating intrinsic neuronal noise, which stems from the biophysics of single neurons and which is independent from one neuron to another, from other sources of variability. Here, intrinsic neuronal spike noise is modelled by the independent Poisson random measures π_i [12, 13]. As stated in the main text, conditioned on the potentials $\{h_j^{(1)}(t)\}_j$, the processes $\{S_j^{(1)}(t) = dZ_j^{(1)}(t)/dt\}_j$ are independent inhomogeneous Poisson processes with instantaneous firing rates $\{\phi(h_j^{(1)}(t))\}_j$. To make the notation consistent with the mathematical notation of Itô calculus, we have replaced the formal $\{\eta_\mu(t)dt\}_{\mu=1}^p$ in the external input by $\{dB_\mu(t)\}_{\mu=1}^p$, where $\{B_\mu(t)\}_{\mu=1}^p$ are independent standard Brownian motions. Concurrently, the dynamics of the feedforward rate network (S20) becomes

$$\tau \, dx_j^{(1)}(t) = -x_j^{(1)}(t)dt + \frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^p \xi_{j\mu} dB_\mu(t), \tag{S23a}$$

$$\tau \frac{d}{dt} x_i^{(2)}(t) = -x_i^{(2)}(t) + \sum_{j=1}^{N/2} J_{ij} \phi(x_j^{(1)}(t)). \tag{S23b}$$

20

We emphasize that in the feedforward SNN (S22) and the feedforward rate network (S23), the potentials of the first layers $h_j^{(1)}(t)$ and $x_j^{(1)}(t)$ are equal since they integrate the same external input $\frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^{p} \xi_{j\mu} dB_\mu(t)$.

We can now prove Theorem 4.

Proof. By ergodicity,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \left| h_i^{(2)}(t) - x_i^{(2)}(t) \right| dt = \lim_{T\to\infty} \mathbb{E}_{\{\pi_j\}_{j=1}^{N/2}, \{B_\mu\}_{\mu=1}^{p}} \left[\left| h_i^{(2)}(T) - x_i^{(2)}(T) \right| \right].$$

Above and in the following, expectations are taken over the Poisson random measures $\{\pi_j\}_{j=1}^{N/2}$ and Brownian motions $\{B_\mu\}_{\mu=1}^{p}$ but not over $\boldsymbol{\xi}$ and $\mathbf{J}$ which can be arbitrary. By Jensen's inequality [2, Theorem 5.1 p. 132],

$$\mathbb{E}_{\{\pi_j\}_{j=1}^{N/2}, \{B_\mu\}_{\mu=1}^{p}} \left[\left| h_i^{(2)}(T) - x_i^{(2)}(T) \right| \right] \le \mathbb{E}_{\{\pi_j\}_{j=1}^{N/2}, \{B_\mu\}_{\mu=1}^{p}} \left[\left(h_i^{(2)}(T) - x_i^{(2)}(T) \right)^2 \right]^{1/2}.$$

By a basic property of the conditional expectation,

$$\mathbb{E}_{\{\pi_j\}_{j=1}^{N/2}, \{B_\mu\}_{\mu=1}^{p}} \left[\left(h_i^{(2)}(T) - x_i^{(2)}(T) \right)^2 \right]^{1/2}$$
$$= \mathbb{E}_{\{\pi_j\}_{j=1}^{N/2}, \{B_\mu\}_{\mu=1}^{p}} \left[\mathbb{E}_{\{\pi_j\}_{j=1}^{N/2}} \left[\left(h_i^{(2)}(T) - x_i^{(2)}(T) \right)^2 \Big| \{B_\mu\}_{\mu=1}^{p} \right] \right]^{1/2}.$$

It remains to study the conditional expectation:

$$\mathbb{E}_{\{\pi_j\}_{j=1}^{N/2}} \left[\left(h_i^{(2)}(T) - x_i^{(2)}(T) \right)^2 \Big| \{B_\mu\}_{\mu=1}^{p} \right]$$
$$= \mathbb{E}_{\{\pi_j\}_{i=j}^{N/2}} \left[\left(\sum_{j=1}^{N/2} J_{ij} \int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} dZ_j^{(1)}(s) - \sum_{j=1}^{N/2} J_{ij} \int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \phi(x_j^{(1)}(s)) ds \right)^2 \Big| \{B_\mu\}_{\mu=1}^{p} \right]$$
$$= \sum_{j=1}^{N/2} J_{ij}^2 \, \mathbb{E}_{\pi_j} \left[\left(\int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} dZ_j^{(1)}(s) - \int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \phi(x_j^{(1)}(s)) ds \right)^2 \Big| \{B_\mu\}_{\mu=1}^{p} \right],$$

where in the last equality, we use the fact that the cross terms are null because the Poisson random measures $\{\pi_j^{(1)}\}_{j=1}^{N/2}$ are independent and, for all $j = 1, \ldots, N/2$,

$$\mathbb{E}_{\pi_j} \left[\int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} dZ_j^{(1)}(t) - \int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \phi(x_j^{(1)}(s)) ds \Big| \{B_\mu\}_{\mu=1}^{p} \right] = 0.$$

21

On the other hand, by Itô's isometry for compensated jump processes [15, Lemma 4.2.2 p. 197],

$$\mathbb{E}_{\pi_j}\left[\left(\int_0^T \frac{1}{\tau}e^{-(T-s)/\tau}\mathrm{d}Z_j^{(1)}(s) - \int_0^T \frac{1}{\tau}e^{-(T-s)/\tau}\phi(x_j^{(1)}(s))\mathrm{d}s\right)^2 \Big|\{B_\mu\}_{\mu=1}^p\right]$$

$$= \int_0^T \left(\frac{1}{\tau}e^{-(T-s)/\tau}\right)^2 \phi(x_j^{(1)}(s))\mathrm{d}s.$$

Hence,

$$\mathbb{E}_{\{\pi_j\}_{i=j}^{N/2}}\left[\left(h_i^{(2)}(T) - x_i^{(2)}(T)\right)^2 \Big|\{B_\mu\}_{\mu=1}^p\right] = \sum_{j=1}^{N/2} J_{ij}^2 \int_0^T \left(\frac{1}{\tau}e^{-(T-s)/\tau}\right)^2 \underbrace{\phi(x_j^{(1)}(s))}_{\leq \max\phi}\mathrm{d}s$$

$$\leq \sum_{j=1}^{N/2} J_{ij}^2 \frac{1}{2\tau}(1 - e^{-2T/\tau})\max\phi.$$

In summary, we have

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \left|h_i^{(2)}(t) - x_i^{(2)}(t)\right|\mathrm{d}t = \lim_{T\to\infty}\mathbb{E}_{\{\pi_j\}_{j=1}^{N/2},\{B_\mu\}_{\mu=1}^p}\left[\left|h_i^{(2)}(T) - x_i^{(2)}(T)\right|\right]$$

$$\leq \lim_{T\to\infty}\left(\sum_{j=1}^{N/2} J_{ij}^2 \frac{1}{2\tau}(1 - e^{-2T/\tau})\max\phi\right)^{1/2}$$

$$= \sqrt{\frac{\max\phi}{2\tau}}\|\mathbf{J}_i\|,$$

which concludes the proof. $\qquad\square$

Finally, we can verify that Theorem 3, i.e.

$$\|\mathbf{J}_i\|_2 \xrightarrow[N\to\infty]{\mathbb{P}} \sqrt{\frac{\alpha}{c}},$$

implies that as $N \to \infty$, the typical limit distance between a spiking neuron $h_i^{(2)}(t)$ and the corresponding rate unit $x_i^{(2)}(t)$, $\lim_{T\to\infty}\frac{1}{T}\int_0^T \left|h_i^{(2)}(t) - x_i^{(2)}(t)\right|\mathrm{d}t$, scale as $\mathcal{O}(\sqrt{\alpha})$. Indeed, we have that

$$\forall\varepsilon > 0, \qquad \mathbb{P}\left(\lim_{T\to\infty}\frac{1}{T}\int_0^T \left|h_i^{(2)}(t) - x_i^{(2)}(t)\right|\mathrm{d}t > \sqrt{\frac{\|\phi\|_\infty}{2\tau c}}\sqrt{\alpha} + \varepsilon\right) \xrightarrow[N\to\infty]{} 0,$$

since

$$\mathbb{P}\left(\lim_{T\to\infty}\frac{1}{T}\int_0^T \left|h_i^{(2)}(t) - x_i^{(2)}(t)\right|\mathrm{d}t > \sqrt{\frac{\|\phi\|_\infty}{2\tau c}}\sqrt{\alpha} + \varepsilon\right)$$

$$\leq \mathbb{P}\left(\|\mathbf{J}_i\|_2 > \sqrt{\frac{\alpha}{c}} + \sqrt{\frac{2\tau}{\max\phi}}\varepsilon\right) \xrightarrow[N\to\infty]{} 0.$$

VI. ADDING DIFFUSIVE MEMBRANE NOISE

In this work, we have used networks of linear-nonlinear-Poisson neurons to give concrete examples of the concentration of measure phenomenon in networks of spiking neurons. This choice of neuron model was mainly motivated by the fact that results in this case take a particularly simple form. Namely, rate-based dynamics are described by standard RNN dynamics. However, concentration of measure does not only occur with this specific neuron model; the phenomenon also appears in networks with more general stochastic spiking neurons. To illustrate this, we briefly discuss below how Theorem 4 in Sec. V for feedforward networks of linear-nonlinear-Poisson neurons can be generalized to the case where diffusive membrane noise is added to the potential dynamics.

Using the same feedfoward setup as in Sec. V, we consider the case where diffusive membrane noise is added to the membrane potential dynamics of the neurons in the first layer, i.e. the system of stochastic differential equations (S22) is replaced by

$$\tau \, dh_j^{(1)}(t) = -h_j^{(1)}(t)dt + \tilde{\sigma} \, dW_j(t) + \frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^{p} \xi_{j\mu} dB_\mu(t), \tag{S24a}$$

$$Z_j^{(1)}(t) = \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \le \phi(h_j^{(1)}(s-))} \pi_j(ds, dz), \tag{S24b}$$

$$\tau \, dh_i^{(2)}(t) = -h_i^{(2)}(t)dt + \sum_{j=1}^{N/2} J_{ij} dZ_j^{(1)}(t), \tag{S24c}$$

where $W_1(t), \ldots, W_{N/2}(t)$ are $N/2$ independent standard Brownian motions that are also independent from the external input signals $B_1(t), \ldots, B_p(t)$, and $\tilde{\sigma} > 0$ is the intrinsic membrane noise parameter. The diffusive noise $\tilde{\sigma} \, dW_j(t)$ in (S24a) should be interpreted as intrinsic membrane noise due, for example, to the effect of channel noise on membrane potential dynamics. Note that, as in the original feedforward setup in Sec. V, neurons in the second layer, Eq. (S24c), simply follow the dynamics of passive membranes (without membrane noise).

We recall that, in Sec. V, Theorem 4 gives a bound for the deviation of the potentials $h_i^{(2)}(t)$ in the second layer from their expectations $x_i^{(2)}(t) = \mathbb{E}_{\{\pi_j\}_{j=1}^{N/2}} \left[h_i^{(2)}(t) \,|\, \{B_\mu\}_{\mu=1}^{p} \right]$ conditioned on the external input signals $\{B_\mu(t)\}_{\mu=1}^{p}$. In order to obtain an analog of Theorem 4 for the system with membrane noise (S24), our first task is to derive equations for the expectations

$$x_i^{(2)}(t) = \mathbb{E}_{\{W_j, \pi_j\}_{j=1}^{N/2}} \left[h_i^{(2)}(t) \,|\, \{B_\mu\}_{\mu=1}^{p} \right]$$

corresponding to the system (S24). By the linearity of the expectation, we have

$$\tau \frac{d}{dt} x_i^{(2)}(t) = -x_i^{(2)}(t) + \sum_{j=1}^{N/2} J_{ij} \, \mathbb{E}_{\{W_j, \pi_j\}_{j=1}^{N/2}} \left[\phi(h_j^{(1)}(t)) | \{B_\mu\}_{\mu=1}^p \right].$$

Thus, to derive a closed system of equations for the $x_i^{(2)}(t)$, it suffices to notice that the expected firing rates $r_j^{(1)}(t) := \mathbb{E}_{\{W_j, \pi_j\}_{j=1}^{N/2}} \left[\phi(h_j^{(1)}(t)) | \{B_\mu\}_{\mu=1}^p \right]$ of each individual neuron $j = 1, \ldots, N/2$ in the first layer are given by the solutions of Fokker-Planck equations (conditioned on the external inputs $I_j^{\text{ext}}(t) dt = \frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^p \xi_{j\mu} dB_\mu(t)$) describing the evolution of a time-varying probability densities $q_j^{(1)}(\cdot, t)$:

$$r_j^{(1)}(t) = \int_{\mathbb{R}} \phi(x) q_j^{(1)}(x, t) dx$$

$$\tau \frac{\partial}{\partial t} q_j^{(1)}(x, t) = \frac{\partial}{\partial x} \left([x - I_j^{\text{ext}}(t)] q_j^{(1)}(x, t) \right) + \frac{\tilde{\sigma}^2}{2\tau} \frac{\partial^2}{\partial x^2} q_j^{(1)}(x, t),$$

with initial conditions $q_j^{(1)}(x, 0) = \delta(x - h_j^{(1)}(0))$ (δ denotes the Dirac delta function). The Fokker-Planck equation above has an explicit solution

$$q_j^{(1)}(x, t) = \frac{1}{\sqrt{2\pi \Sigma(t)}} \exp \left(-\frac{(x - \bar{x}_j^{(1)}(t))^2}{2\Sigma(t)} \right),$$

where the mean potential $\bar{x}_j^{(1)}(t)$ follows the same dynamics as (S23a), i.e.

$$\tau \, d\bar{x}_j^{(1)}(t) = -\bar{x}_j^{(1)}(t) dt + \frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^p \xi_{j\mu} dB_\mu(t),$$

and the time-varying variance $\Sigma(t)$ is

$$\Sigma(t) = \frac{\tilde{\sigma}^2}{2\tau} (1 - e^{-2t/\tau}).$$

This gives us generalized rate-based equations for the system with membrane noise (S24):

$$\tau \, d\bar{x}_j^{(1)}(t) = -\bar{x}_j^{(1)}(t) dt + \frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^p \xi_{j\mu} dB_\mu(t) \tag{S25a}$$

$$\tau \frac{d}{dt} x_i^{(2)}(t) = -x_i^{(2)}(t) + \sum_{j=1}^{N/2} J_{ij} \Phi(\bar{x}_j^{(1)}(t), \Sigma(t)), \tag{S25b}$$

where the variance-dependent transfer function $\Phi(\bar{x}, \Sigma)$ is

$$\Phi(\bar{x}, \Sigma) = \int_{\mathbb{R}} \phi(x) \frac{1}{\sqrt{2\pi \Sigma}} \exp \left(-\frac{(x - \bar{x})^2}{2\Sigma} \right) dx.$$

24

Assuming that the transfer function ϕ is bounded, i.e. $\max \phi < \infty$, the statement and the proof of Theorem 4 can be readily adapted to the system (S24) if the original rate-based system (S23) is replaced by the generalized rate-based dynamics (S25).

To adapt the proof of Theorem 4, we simply need to notice that

$$
\mathbb{E}_{W_j, \pi_j}\left[\left(\int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \mathrm{d}Z_j^{(1)}(s) - \int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \int_{\mathbb{R}} \phi(x) q_j^{(1)}(x, s) \mathrm{d}x\, \mathrm{d}s\right)^2 \Big| \{B_\mu\}_{\mu=1}^p\right]
$$
$$
= \mathrm{Var}_{W_j, \pi_j}\left(\int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \mathrm{d}Z_j^{(1)}(s) \Big| \{B_\mu\}_{\mu=1}^p\right),
$$

and by the law of total variance,

$$
\mathrm{Var}_{W_j, \pi_j}\left(\int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \mathrm{d}Z_j^{(1)}(s) \Big| \{B_\mu\}_{\mu=1}^p\right)
$$
$$
= \mathbb{E}_{W_j}\left[\mathrm{Var}_{\pi_j}\left(\int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \mathrm{d}Z_j^{(1)}(s) \Big| W_j, \{B_\mu\}_{\mu=1}^p\right)\right]
$$
$$
+ \mathrm{Var}_{W_j}\left(\mathbb{E}_{\pi_j}\left[\int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \mathrm{d}Z_j^{(1)}(s) \Big| W_j, \{B_\mu\}_{\mu=1}^p\right]\right).
$$

But since

$$
\mathrm{Var}_{\pi_j}\left(\int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \mathrm{d}Z_j^{(1)}(s) \Big| W_j, \{B_\mu\}_{\mu=1}^p\right) = \int_0^T \left(\frac{1}{\tau} e^{-(T-s)/\tau}\right)^2 \int_{\mathbb{R}} \phi(h_j^{(1)}(s))\, \mathrm{d}s,
$$
$$
\mathbb{E}_{\pi_j}\left[\int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \mathrm{d}Z_j^{(1)}(s) \Big| W_j, \{B_\mu\}_{\mu=1}^p\right] = \int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \phi(h_j^{(1)}(s)) \mathrm{d}s,
$$

we obtain

$$
\mathrm{Var}_{W_j, \pi_j}\left(\int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \mathrm{d}Z_j^{(1)}(s) \Big| \{B_\mu\}_{\mu=1}^p\right)
$$
$$
= \int_0^T \left(\frac{1}{\tau} e^{-(T-s)/\tau}\right)^2 \int_{\mathbb{R}} \phi(x) q_j^{(1)}(x, s) \mathrm{d}x\, \mathrm{d}s
$$
$$
+ \mathrm{Var}_{W_j}\left(\int_0^T \frac{1}{\tau} e^{-(T-s)/\tau} \phi(h_j^{(1)}(s)) \mathrm{d}s \Big| \{B_\mu\}_{\mu=1}^p\right). \quad \text{(S26)}
$$

The second term on the right-hand side of (S26) implies that the final bound for the network with membrane noise (S24) has to be looser than the bound (S21) for the network with linear-nonlinear-Poisson neurons. From (S26), we can easily derive the final bound

$$
\lim_{T \to \infty} \frac{1}{T} \int_0^T \left|h_i^{(2)}(t) - x_i^{(2)}(t)\right| \mathrm{d}t \leq \left(\sqrt{\frac{\max \phi}{2\tau}} + \max \phi\right) \|\mathbf{J}_i\|_2.
$$

Note that this bound might not be optimal and a tighter bound may be obtainable.

The model treated in this section is instructive as it illustrates an interesting fact: the potentials in the second layer $h_j^{(2)}(t)$ can converge to $x_j^{(2)}(t)$ in large networks (if the norms $\|\mathbf{J}_i\|_2$ converge to 0) even if the the potentials in the first layer $h_j^{(1)}(t)$ do not converge to the mean potentials $\bar{x}_j^{(1)}(t)$. Indeed, in this model, we know that the variance of $h_j^{(1)}(t) - \bar{x}_j^{(1)}(t)$ is $\Sigma(t)$, which does not depend on the network size. This shows that concentration of measure can occur even in cases where membrane potentials dynamics are intrinsically noisy.

---

[1] S. Méléard, Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models, in *Probabilistic Models for Nonlinear Partial Differential Equations*, Lecture Notes in Mathematics (Springer, 2006) p. 42.

[2] A. Gut, *Probability: A Graduate Course*, Vol. 200 (Springer, 2006).

[3] F. Dai and Y. Xu, Approximation theory and harmonic analysis on spheres and balls, Springer Monographs in Mathematics (2013).

[4] U. Pereira and N. Brunel, Attractor dynamics in networks with learning rules inferred from in vivo data, Neuron **99**, 227 (2018).

[5] F. Mastrogiuseppe and S. Ostojic, Linking connectivity, dynamics, and computations in low-rank recurrent neural networks, Neuron **99**, 609 (2018).

[6] F. Schuessler, A. Dubreuil, F. Mastrogiuseppe, S. Ostojic, and O. Barak, Dynamics of random recurrent networks with correlated low-rank structure, Phys.Rev. Research **2**, 013111 (2020).

[7] W. Gerstner, Time structure of the activity in neural network models, Phys. Rev. E **51**, 738 (1995).

[8] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal dynamics: From single neurons to networks and models of cognition* (Cambridge University Press, 2014).

[9] P. C. Bressloff, Spatiotemporal dynamics of continuum neural fields, J. Phys. A **45**, 033001 (2011).

[10] W. Gerstner and W. M. Kistler, *Spiking neuron models: Single neurons, populations, plasticity* (Cambridge University Press, 2002).

[11] J. Chevallier, A. Duarte, E. Löcherbach, and G. Ost, Mean field limits for nonlinear spatially extended hawkes processes with exponential memory kernels, Stochastic Process. Appl. **129**, 1

(2019).

[12] J. F. C. Kingman, *Poisson processes*, Vol. 3 (Clarendon Press, 1992).

[13] E. Löcherbach, Spiking neurons: interacting hawkes processes, mean field limits and oscillations, ESAIM Proc. **60**, 90 (2017).

[14] V. Schmutz, *Taming neuronal noise with large networks*, Ph.D. thesis, École Polytechnique Fédérale de Lausanne (2022).

[15] D. Applebaum, *Lévy processes and stochastic calculus* (Cambridge University Press, 2009).