Stochastic activity in low-rank recurrent neural networks

Francesca Mastrogiuseppe[*,1], Joana Carmona[1], Christian Machens[1]

[1] Champalimaud Foundation, Neuroscience Research Programme, Lisbon (Portugal)

[*] correspondence: francesca.mastrogiuseppe@research.fchampalimaud.org

The geometrical and statistical properties of brain activity depend on the way neurons connect together to form recurrent circuits. How the structure of connectivity shapes the emergent activity remains however not fully understood. We investigate this question in recurrent neural networks with linear additive stochastic dynamics. We assume that the synaptic connectivity can be expressed in a low-rank form, parameterized by a handful of connectivity vectors, and and examine how the geometry of emergent activity relates to these vectors. Our findings reveal that this relationship critically depends on the dimensionality of the external stochastic inputs. When inputs are low-dimensional, activity is confined to a low-dimensional subspace spanned by a subset of the connectivity vectors, whose dimensionality matches the rank of the connectivity matrix, along with the external inputs. Conversely, when inputs are high-dimensional, activity is generally high-dimensional; recurrent dynamics shape activity within a subspace spanned by all connectivity vectors, with a dimensionality equal to twice the rank of the connectivity matrix. Applying our formalism to excitatory-inhibitory circuits, we discuss how inputs geometry also play a crucial role in determining the amount of input amplification generated by non-normal dynamics. Our work provides a foundation for studying activity in structured brain circuits under realistic noise conditions, and offers a framework for interpreting stochastic models inferred from experimental data.

Introduction

Understanding the relationship between connectivity and activity in neural circuits is a central focus of theoretical neuroscience. A particularly active area of research investigates how large-scale circuits can give rise to low-dimensional activity patterns [27, 21, 46, 10]. This type of activity is commonly observed in neural recordings during behavioral tasks, particularly in cortical areas involved in high-level functions like motor planning and decision-making [15, 14, 62, 9].

Recurrent neural networks (RNNs) with low-rank connectivity have emerged as a flexible and mathematically-tractable framework for studying this phenomenon [12, 32, 43]. Low-rank matrices are characterized by a simple structure: they are defined by a finite number of non-zero singular values and can be constructed from a small set of generating connectivity vectors. This simplicity facilitates the study of the relationship between connectivity and activity, offering insights that are harder to extract from generic models [32, 42, 27, 4, 33].

Previous studies on low-rank RNNs have primarily focused on models with autonomous dynamics or specific types of external inputs. These inputs were assumed to have a simple and smooth temporal structure (e.g. constant [32] or sinusoidal [58]). Furthermore, they were low-dimensional: the inputs received by different neurons were highly correlated over time, as they originated from a small set of temporal signals broadcast across the entire network. Under these conditions, low-rank RNNs have been shown to generate low-dimensional activity, whose statistical properties can be directly traced back to the structure of the network's external inputs and recurrent connectivity [32, 2].

Far less is known about the behavior of low-rank RNNs in response to rapidly varying and stochastic inputs, which are routinely used to model trial-to-trial variability in circuit models [34, 17, 24, 20]. These inputs can

be either low-dimensional, when they are generated from a small set of stochastic signals and exhibit high correlations across units, or high-dimensional, when they arise from a large number of stochastic signals and remain uncorrelated across units.

Investigating the dynamics of low-rank RNNs in response to this type of inputs is important for several reasons. First, stochastic inputs naturally arise in cortical circuits from irregular spiking activity and the stochastic nature of biological processes like synaptic transmission and spike initiation [55, 53]. Understanding how these inputs influence the emergence of low-dimensional activity in recurrent circuits is therefore an important biological question. Second, stochastic inputs are routinely used during the optimization of RNNs on behavioural tasks [31, 64], where they help regularize the learning dynamics [22, 63] and improve the robustness of internal representations [50]. As optimization often results in approximately low-rank connectivity [49, 5], developing tools for describing stochastic activity in low-rank RNNs is essential for studying learning. Finally, statistical models for inferring circuit dynamics from recorded neural activity often explicitly incorporate stochastic terms [16, 54, 28, 41]. Understanding activity in stochastic models is therefore essential for interpreting the results of fitting procedures.

In this work, we study the emergent activity in linear, low-rank recurrent neural networks driven by fast stochastic inputs, exploring both low- and high-dimensional input regimes. We analyze the statistical and geometric properties of activity by examining its covariance matrix, which is expressed in terms of the vectors representing both the external inputs and the low-rank connectivity. Crucially, we find that the functional dependence between the covariance matrix and these vectors undergoes a qualitative shift depending on the dimensionality of the external stochastic inputs. For low-dimensional inputs, activity covariance is also low-dimensional, and depends only on a subset of the connectivity vectors, with dimensionality equal to the rank of the connectivity matrix. In this regime, activity qualitatively resembles that previously described in deterministic models [32, 2]. In contrast, for high-dimensional inputs, activity covariance is potentially high-dimensional and depends on the entire set of connectivity vectors, with dimensionality equal to twice the rank of the connectivity matrix. This regime qualitatively differs from previously reported ones, as the full geometrical structure of synaptic connectivity is reflected in the activity covariance. As a direct application of our formalism, we study stochastic activity in an excitatory-inhibitory circuit [36] and discuss how the variance amplification performed by the non-normal connectivity in this system depends on the input configuration. Our findings provide insights into how structured connectivity shapes neural activity in the presence of noise, laying the groundwork for a deeper understanding of brain circuit dynamics.

Results

1 Setup

We consider a stochastic recurrent neural network of N units as in Fig. 1A, governed by the linear dynamics:

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = -\boldsymbol{x}(t) + W\boldsymbol{x}(t) + U\boldsymbol{\chi}(t) \tag{1}$$

where, for simplicity, the activity evolution timescale is set to 1. On the right-hand side of Eq. 1, the first term represents the leak, the second term models the recurrent interactions mediated by the synaptic connectivity matrix W, and the third term corresponds to the external stochastic input. The stochastic input is expressed in terms of a Gaussian stochastic process $\boldsymbol{\chi}$, which has zero mean, is uncorrelated across units, and is temporally white. Formally: $\langle \boldsymbol{\chi}(t) \rangle = \mathbf{0}$ and $\langle \boldsymbol{\chi}(t)\, \boldsymbol{\chi}(s)^\top \rangle = I\delta(t-s)$, where $\langle\, . \,\rangle$ denotes an average over inputs realizations, $\mathbf{0}$ is a N-dimensional vector filled with zero entries, and I represents the $N \times N$ identity matrix. The covariance of the total stochastic inputs is therefore given by

$$\langle U\boldsymbol{\chi}(t)\, [U\boldsymbol{\chi}(s)]^\top \rangle = \bar{\Sigma}_{\mathrm{inp}}\delta(t-s), \tag{2}$$

where we use the notation $\bar{\Sigma}_{\mathrm{inp}} = UU^\top$ to indicate the static part of the covariance. Our analysis focuses on the activity generated by this model in the stationary regime, which is reached provided that the real part of all eigenvalues of W is less than 1. In this regime, the activity statistics can be equivalently computed over input realizations or time.

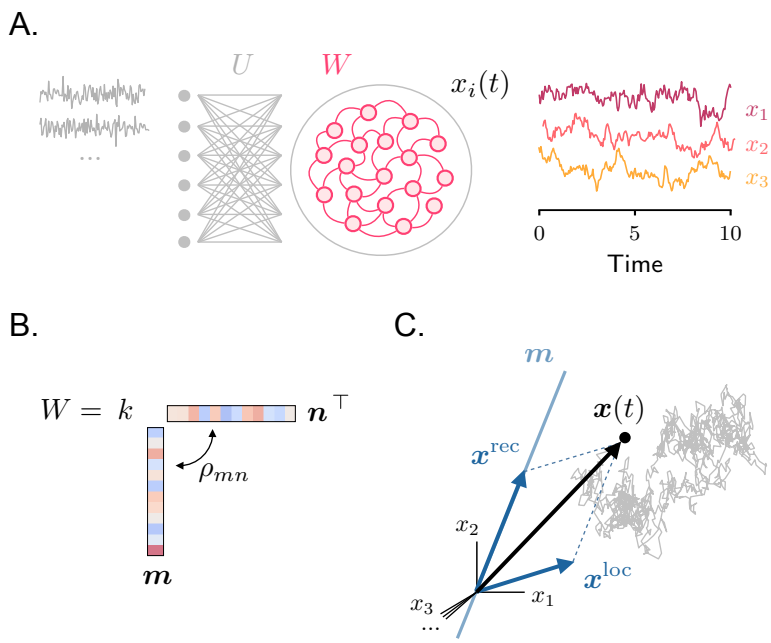

Figure 1: Setup. **A.** Left: model architecture. Right: sample activity traces from three randomly chosen neurons. **B.** Rank-one recurrent connectivity. **C.** Illustration of a sample activity trajectory in the high-dimensional space where each axis corresponds to the activity of a different neuron. Activity (black arrow) is given by the sum of two components (Eq. 8, blue arrows); the direction of the component generated from recurrent interactions is fixed, and is aligned with the connectivity vector $\boldsymbol{m}$.

We first consider rank-one connectivity matrices before turning to higher-rank cases. A rank-one connectivity matrix can be expressed as

$$W = k\,\boldsymbol{m}\boldsymbol{n}^\top, \tag{3}$$

where k is a scalar and $\boldsymbol{m}$ and $\boldsymbol{n}$ are unit-norm vectors (Fig. 1B). By defining the overlap function between two vectors $\boldsymbol{v}$ and $\boldsymbol{v}'$ as

$$\rho_{vv'} = \boldsymbol{v}^\top \boldsymbol{v}' \tag{4}$$

we therefore have $\rho_{mm} = \rho_{nn} = 1$. Vectors $\boldsymbol{m}$ and $\boldsymbol{n}$, which we refer to as *connectivity vectors*, characterize the structure of the synaptic connectivity. When $\boldsymbol{m}$ and $\boldsymbol{n}$ coincide, connectivity is symmetric. Conversely, when $\boldsymbol{m}$ and $\boldsymbol{n}$ are orthogonal, connectivity is maximally asymmetric. In general, the overlap between the two vectors is quantified by $\rho_{mn} = \boldsymbol{m}^\top \boldsymbol{n}$ (Eq. 4), which ranges from -1 to 1 (symmetric matrices) passing through 0 (maximally asymmetric). The only non-zero eigenvalue of W is given by $\lambda = k\rho_{mn}$, and is associated with the eigenvector $\boldsymbol{m}$. Stationary activity requires $\rho_{mn} < 1/k$. Without loss of generality, we fix $k > 0$.

Notably, due to normalization, the entries of connectivity vectors scale as $O(1/\sqrt{N})$ with the network size. We further assume that the scalar k is $O(1)$ in N, so that the eigenvalue λ takes finite values. This scaling is consistent with the scaling assumed for recurrent connectivity in several previous studies [19, 42, 32] (see [45, 27] for different choices). However, here we consider arbitrary network sizes, and do not restrict ourselves to the large-N limit.

2 General derivation

2.1 Computing activity

We start deriving closed-form expressions for the activity generated by the dynamics in Eq. 1. Integrating over time, we have [47]

$$\boldsymbol{x}(t) = \exp[(W-I)t]\,\boldsymbol{x}(0) + \int_0^t \exp[(W-I)(t-u)]U\boldsymbol{\chi}(u)\,\mathrm{d}u \tag{5}$$

from which, in the stationary state ($t \to \infty$):

$$\boldsymbol{x}(t) = \int_0^t \exp[(W-I)(t-u)]\,U\boldsymbol{\chi}(u)\,\mathrm{d}u. \tag{6}$$

We can now simplify this expression by computing the propagator operator $\exp[(W-1)t]$. For a rank-one connectivity matrix W, this has a simple expression (see Methods 1):

$$\exp[(W-I)t] = \exp(-t)\left[I + \frac{\exp(\lambda t)-1}{\lambda} k\,\boldsymbol{m}\boldsymbol{n}^\top\right]. \tag{7}$$

The first term in parentheses on the right-hand side corresponds to the propagator of an unconnected network ($W=0$), where the dynamics are driven solely by the local filtering, operated by the leak, of the external inputs. The second term captures instead the contribution to the propagator arising from the recurrent connectivity matrix W. By applying Eq. 7 to Eq. 6, we can therefore express activity as the sum of two components (Fig. 1C):

$$\boldsymbol{x}(t) = \boldsymbol{x}^{\mathrm{loc}}(t) + \boldsymbol{x}^{\mathrm{rec}}(t), \tag{8}$$

where we have defined

$$\boldsymbol{x}^{\mathrm{loc}}(t) = \int_0^t \exp[-(t-u)]\,U\boldsymbol{\chi}(u)\,\mathrm{d}u \tag{9}$$

$$\boldsymbol{x}^{\mathrm{rec}}(t) = k\,\boldsymbol{m}\boldsymbol{n}^\top \int_0^t \frac{\exp[-(t-u)]\{\exp[\lambda(t-u)]-1\}}{\lambda}\,U\boldsymbol{\chi}(u)\,\mathrm{d}u. \tag{10}$$

The first component, $\boldsymbol{x}^{\mathrm{loc}}$, reflects the behavior of an uncoupled RNN and represents the temporal integration of stochastic inputs performed independently by all units. Importantly, at each time point, the direction of vector $\boldsymbol{x}^{\mathrm{loc}}$ varies, depending on the direction of the most recent stochastic inputs. This activity component can therefore potentially be high-dimensional, depending on the structure of the input matrix U. The second component, $\boldsymbol{x}^{\mathrm{rec}}$, represents the contribution to activity arising from the temporal integration of recurrent inputs. Unlike $\boldsymbol{x}^{\mathrm{loc}}$, the direction of $\boldsymbol{x}^{\mathrm{rec}}$ is fixed and determined by the connectivity vector $\boldsymbol{m}$.

2.2 Computing activity covariance

We aim to study the statistical and geometrical properties of activity which, as detailed below, can be extracted from its covariance matrix. From Eq. 6, it follows that the mean activity, $\langle \boldsymbol{x}(t)\rangle$, vanishes in the stationary state. Consequently, the covariance matrix is defined as $\Sigma(t,s) = \langle \boldsymbol{x}(t)\boldsymbol{x}(s)^\top\rangle$, which can be expressed as:

$$\begin{aligned}\Sigma(t,s) &= \langle\left(\int_0^t \exp[(W-I)(t-u)]U\boldsymbol{\chi}(u)\,\mathrm{d}u\right)\left(\int_0^s \exp[(W-I)(s-v)]U\boldsymbol{\chi}(v)\,\mathrm{d}v\right)^\top\rangle \\ &= \int_0^t\int_0^s \exp[(W-I)(t-u)]\,\bar{\Sigma}_{\mathrm{inp}}\exp[(W-I)(s-v)]^\top\delta(u-v)\,\mathrm{d}u\,\mathrm{d}v.\end{aligned} \tag{11}$$

Substituting the expression for the propagator in Eq. 7, we see that covariance factorizes into four terms, which can be re-written as:

$$\Sigma(t,s) = c_1(t,s)\,\bar{\Sigma}_{\mathrm{inp}} + c_2(t,s)\,\bar{\Sigma}_{\mathrm{inp}}\boldsymbol{n}\boldsymbol{m}^\top + c_3(t,s)\,\boldsymbol{m}\boldsymbol{n}^\top\bar{\Sigma}_{\mathrm{inp}} + c_4(t,s)\,\boldsymbol{m}\boldsymbol{n}^\top\bar{\Sigma}_{\mathrm{inp}}\,\boldsymbol{n}\boldsymbol{m}^\top. \tag{12}$$

The expression for the time-varying coefficients in the right-hand side, as well as their derivation, are provided in Methods 3. In this work, we focus on the equal-time covariance. Evaluating those coefficients at $t=s$ yields the final expression:

$$\Sigma \equiv \Sigma(t,t) = \frac{1}{2}\left\{\bar{\Sigma}_{\mathrm{inp}} + \frac{k}{2-\lambda}\left[\bar{\Sigma}_{\mathrm{inp}}\boldsymbol{n}\boldsymbol{m}^\top + \boldsymbol{m}\boldsymbol{n}^\top\bar{\Sigma}_{\mathrm{inp}}\right] + \frac{k^2}{(2-\lambda)(1-\lambda)}\,\boldsymbol{m}\boldsymbol{n}^\top\bar{\Sigma}_{\mathrm{inp}}\,\boldsymbol{n}\boldsymbol{m}^\top\right\}. \tag{13}$$

It is straightforward to see that the four additive terms on the right-hand side correspond to the covariances of the two activity components defined in Eq. 8: $\boldsymbol{x}^{\text{loc}}$ (first term of the covariance) and $\boldsymbol{x}^{\text{rec}}$ (fourth term), as well as their cross-covariances (second and third terms). Eq. 13 further shows that the activity covariance is determined by matrix multiplications among the connectivity vectors and the static input covariance, indicating that the statistics of emergent activity are determined by the relative geometrical arrangement of the connectivity vectors and external inputs.

2.3 Covariance structure

To clarify the geometrical structure of the covariance, we rewrite it as

$$\Sigma = \frac{1}{2}\left\{\bar{\Sigma}_{\text{inp}} + \frac{k}{2-\lambda}\left[\boldsymbol{x}\boldsymbol{m}^\top + \boldsymbol{m}\boldsymbol{x}^\top\right] + \frac{\sigma k^2}{(2-\lambda)(1-\lambda)}\,\boldsymbol{m}\boldsymbol{m}^\top\right\} \tag{14}$$

where we have defined $\boldsymbol{x} = \bar{\Sigma}_{\text{inp}}\boldsymbol{n}$ and $\sigma = \boldsymbol{n}^\top\bar{\Sigma}_{\text{inp}}\,\boldsymbol{n}$. The first term on the right-hand side represents the input covariance, which can be either a low-rank or full-rank matrix, depending on whether the inputs are low- or high-dimensional. The remaining terms are rank-one matrices, spanned by the connectivity vectors $\boldsymbol{m}$ and $\boldsymbol{x}$. The direction of $\boldsymbol{x}$ is jointly determined by the connectivity vector $\boldsymbol{n}$ and the input covariance.

In the rest of the paper, we focus our analysis on two limit cases. First, we consider the case where the inputs are one-dimensional, which can be achieved by setting $U = [\boldsymbol{u}, \boldsymbol{0}, \ldots, \boldsymbol{0}]$, where $\boldsymbol{u}$ is an N-dimensional input vector (not necessarily normalized), and $\boldsymbol{0}$ is a vector of zeros. In this case, we have that $\bar{\Sigma}_{\text{inp}} = \boldsymbol{u}\boldsymbol{u}^\top$, and thus, the vector $\boldsymbol{x}$ is fully aligned with the input vector: $\boldsymbol{x} = \rho_{nu}\boldsymbol{u}$ (we have used Eq. 4 to define the overlap between the connectivity vector $\boldsymbol{n}$ and the input vector $\boldsymbol{u}$, denoted as ρ_{nu}). We then consider the case where inputs have maximal dimensionality, which corresponds to an input matrix U that is proportional to an orthogonal matrix. Without loss of generality, we take $U = u\,I$, where u is a scalar, implying $\bar{\Sigma}_{\text{inp}} = u^2\,I$. In this case, we have that the vector $\boldsymbol{x}$ is fully aligned with the connectivity vector $\boldsymbol{n}$, as $\boldsymbol{x} = u^2\boldsymbol{n}$. Intermediate cases, which correspond to inputs of intermediate dimensionality, are characterized by activity properties interpolating between these two limit cases.

We conclude this section by remarking that, in the general case, the geometrical structure of the covariance depends on both connectivity vectors, $\boldsymbol{m}$ and $\boldsymbol{n}$. While the dependence of activity covariance on the connectivity vector $\boldsymbol{m}$ – which corresponds to the right eigenvector of the connectivity matrix – has been reported in previous work on low-rank RNNs [32, 11, 61], the dependence on the connectivity vector $\boldsymbol{n}$ is perhaps more surprising. We will explore this dependence further in the subsequent sections.

3 Low-dimensional stochastic inputs

We begin with the case of one-dimensional stochastic inputs, where a single stochastic signal is broadcast across the entire network, leading to temporally correlated inputs for individual neurons (Fig. 2A). This case closely relates to previous work that examined the impact of deterministic low-dimensional inputs in low-rank RNNs [32, 11].

Setting $\bar{\Sigma}_{\text{inp}} = \boldsymbol{u}\boldsymbol{u}^\top$ in Eq. 14, the activity covariance becomes:

$$\Sigma = \frac{1}{2}\left\{\boldsymbol{u}\boldsymbol{u}^\top + \frac{\rho_{nu}k}{(2-\lambda)}\left[\boldsymbol{u}\boldsymbol{m}^\top + \boldsymbol{m}\boldsymbol{u}^\top\right] + \frac{\rho_{nu}^2 k^2}{(2-\lambda)(1-\lambda)}\,\boldsymbol{m}\boldsymbol{m}^\top\right\}. \tag{15}$$

As in the general case, the first and fourth terms on the right-hand side represent the covariance of the activity components $\boldsymbol{x}^{\text{loc}}$ and $\boldsymbol{x}^{\text{rec}}$, while the second and third terms represent the cross-covariances between them. All terms, except the first one, are proportional to the overlap parameter ρ_{nu}. This implies that, when the input vector $\boldsymbol{u}$ has zero overlap with the connectivity vector $\boldsymbol{n}$, the network behaves as if it were unconnected. Therefore, recurrent connectivity contributes to shaping activity in the case of low-dimensional inputs only when the vectors $\boldsymbol{u}$ and $\boldsymbol{n}$ are characterized by a non-vanishing overlap.

3.1 Covariance eigenvalues and eigenvectors

Covariance matrices provide important information on the statistical and geometrical properties of activity. One effective way to analyze these properties is by computing the covariance matrix eigenvectors and eigenvalues,

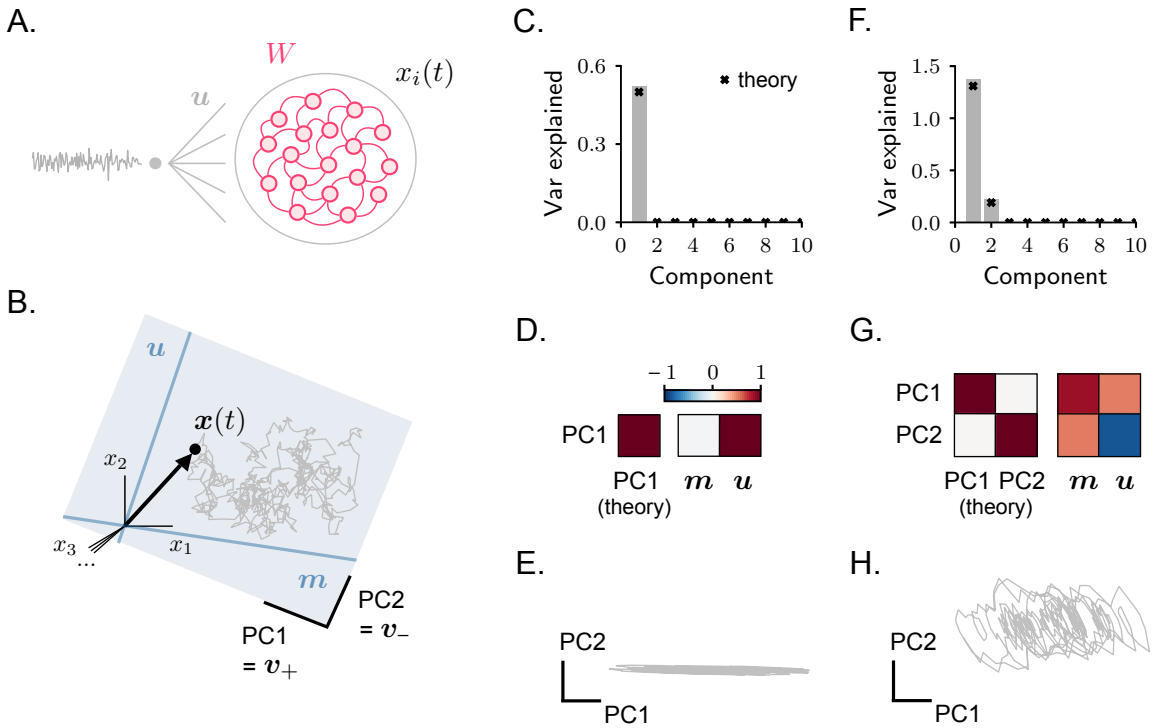

Figure 2: Rank-one RNN with one-dimensional stochastic inputs. **A.** Model architecture. **B.** Activity covariance is low-dimensional, and is spanned by connectivity vector $\boldsymbol{m}$ together with the external input vector $\boldsymbol{u}$. As a consequence, activity is contained within the plane collinear with these two vectors. **C–D–E.** Example of a simulated network with $\rho_{nu} = 0$. In C: covariance spectrum. Components larger than 10 are not displayed (they are all close to zero). In D: overlap between the dominant principal components estimated from simulated activity and the theoretically-estimated PCs (left), or the vectors $\boldsymbol{m}$ and $\boldsymbol{u}$ (right). Overlaps are quantified via Eq. 4, with input vectors $\boldsymbol{u}$ chosen to be normalized. Note that here, but not in G, only one principal component can be identified. In E: simulated activity projected on the two dominant PCs. **F–G–H.** Same as in C–D–E, example with $\rho_{nu} > 0$.

which provide information on the directions and magnitude of co-variation of activity over time. The eigenvectors of the covariance matrix, denoted as $\{\boldsymbol{v}_i\}_i$, are also known as the Principal Components (PCs) of activity. The corresponding eigenvalues, denoted as $\{\mu_i\}_i$, quantify the variance of activity along each principal component. From the covariance eigenvalues, the embedding dimensionality of activity is commonly estimated using the participation ratio [29, 46]:

$$\mathcal{D} = \frac{\left(\sum_{i=1}^{N} \mu_i\right)^2}{\sum_{i=1}^{N} \mu_i^2}, \tag{16}$$

which quantifies the concentration of the eigenvalue distribution.

The covariance in Eq. 15 is a rank-two matrix. Its eigenvalues can be estimated through the method of reduced matrices [37] (see Methods 4), yielding:

$$\mu_\pm = \frac{1}{4}\left[\rho_{uu} + 2\alpha\rho_{mu} + \beta \pm \sqrt{(\rho_{uu} + 2\alpha\rho_{mu} + \beta)^2 - 4(\rho_{uu} - \rho_{mu}^2)(\beta - \alpha^2)}\right] \tag{17}$$

where we have defined $\rho_{mu} = \boldsymbol{m}^\top \boldsymbol{u}$ (Eq. 4) and

$$\begin{aligned} \alpha &= \frac{\rho_{nu} k}{2-\lambda} \\ \beta &= \frac{\rho_{nu}^2 k^2}{(2-\lambda)(1-\lambda)}. \end{aligned} \tag{18}$$

To find the eigenvectors associated with these eigenvalues, we leverage the low-dimensional structure of Eq. 15 to formulate the ansatz:

$$\boldsymbol{v}_\pm = a_\pm \boldsymbol{m} + \boldsymbol{u}, \tag{19}$$

and compute the coefficients $a_\pm$ through straightforward algebra (see Methods 4).

The fact that the covariance matrix has only two non-zero eigenvalues indicates that, in the stationary regime, the activity $\boldsymbol{x}(t)$ is confined to a two-dimensional plane, spanned by the connectivity vector $\boldsymbol{m}$ and the input vector $\boldsymbol{u}$ (Fig. 2B). While the component of activity that is aligned with $\boldsymbol{u}$ reflects the integration of external inputs that is performed locally and independently by single neurons, the component aligned with vector $\boldsymbol{m}$ represents the contribution to activity from recurrent interactions. Within the $\boldsymbol{m}$-$\boldsymbol{u}$ plane, the principal components $\boldsymbol{v}_\pm$ define the orthogonal basis that is most collinear with the activity fluctuations. When the variance of activity fluctuations is primarily concentrated along the first PC, the activity dimensionality is close to 1; conversely, when the variance is distributed similarly along the two PCs, the dimensionality approaches 2.

We have seen that when the input vector $\boldsymbol{u}$ and the connectivity vector $\boldsymbol{n}$ are characterized by zero overlap ($\rho_{nu} = 0$), recurrent interactions do not contribute to the activity. This corresponds to setting $\alpha = \beta = 0$, from which the eigenvalues (Eq. 17) become:

$$\mu_\pm = \frac{1}{4}(1 \pm 1) = \frac{1}{2}, 0. \tag{20}$$

The eigenvalue μ_- vanishes, and therefore activity is one-dimensional (Fig. 2C, E). The only non-trivial PC of activity, $\boldsymbol{v}_+$, is fully aligned with the input vector $\boldsymbol{u}$, as the value of a_+ vanishes (Fig. 2D, see Methods 4). As expected, activity reduces to a local filtering of the stochastic inputs, completely operating along the input direction.

When the input vector $\boldsymbol{u}$ and the connectivity vector $\boldsymbol{n}$ are characterized by non-zero overlap ($\rho_{nu} \neq 0$), activity is characterized by a non-vanishing component along $\boldsymbol{m}$ (Fig. 2F, G, H). The exact orientation of the two principal components with respect to the vectors $\boldsymbol{m}$ and $\boldsymbol{u}$, as well as and the variance they capture, depend on the overlap parameters ρ_{mn} and ρ_{mu}, specifying the alignment among the three vectors $\boldsymbol{m}$, $\boldsymbol{n}$ and $\boldsymbol{u}$.

To summarize, we found that rank-one RNNs receiving one-dimensional stochastic inputs generate activity that is confined to the plane spanned by the connectivity vector $\boldsymbol{m}$ and the input vector $\boldsymbol{u}$. The contribution of recurrent connectivity to the activity covariance is, in this scenario, one-dimensional. These findings are in qualitative agreement with previous results on low-rank RNNs receiving deterministic low-dimensional inputs [32, 11]. Next, we consider the opposite case of high-dimensional stochastic inputs, and reveal that major qualitative differences emerge in that setting.

4 High-dimensional stochastic inputs

We now consider high-dimensional and fully unstructured stochastic inputs, for which the input to each neuron in the network is uncorrelated and has identical variance (Fig. 3A). Using $\bar{\Sigma}_{\text{inp}} = u^2\, I$ in Eq. 14, the covariance becomes

$$\Sigma = \frac{u^2}{2}\left\{ I + \frac{k}{2-\lambda}\left[\boldsymbol{n}\boldsymbol{m}^\top + \boldsymbol{m}\boldsymbol{n}^\top + \frac{k}{1-\lambda}\,\boldsymbol{m}\boldsymbol{m}^\top\right]\right\}. \tag{21}$$

As expected for high-dimensional inputs, the covariance matrix is now full-rank. However, it exhibits a particularly simple structure: an additive low-rank perturbation, generated from the connectivity vectors $\boldsymbol{m}$ and $\boldsymbol{n}$, on the identity matrix. This structure enables us to derive the behavior of its eigenvalues and eigenvectors in closed form. Before delving into the derivations, we explore the covariance matrix from a different perspective.

4.1 Dependence on connectivity vector n

The covariance in Eq. 21 differs from that observed in response to low-dimensional inputs in several ways, one of which is its dependence not only on the connectivity vector $\boldsymbol{m}$, but also on $\boldsymbol{n}$. This dependence is particularly interesting, as it contrasts with previous studies on low-rank RNNs which have primarily emphasized the dominant role of connectivity vector $\boldsymbol{m}$ over $\boldsymbol{n}$ in shaping the geometrical structure of activity [32, 2, 11]. In this section, we examine this dependence in detail to better understand its origin. In the next section, we characterize the impact of the connectivity vector $\boldsymbol{n}$ on the covariance structure.

The covariance terms that depend on $\boldsymbol{n}$ correspond to cross-covariances among the two terms constituting activity $\boldsymbol{x}^{\mathrm{loc}}$ and $\boldsymbol{x}^{\mathrm{rec}}$ (Eq. 8). To elucidate the origin of the dependence on $\boldsymbol{n}$, we re-derive those covariance terms starting from the explicit expressions for activity contributions $\boldsymbol{x}^{\mathrm{loc}}$ and $\boldsymbol{x}^{\mathrm{rec}}$ (Eqs. 9, 10). (While equivalent to the general derivation provided in the sections above, this calculation offers a more direct perspective on the origin of the dependence on $\boldsymbol{n}$.) We focus on the second covariance term in Eq. 21; similar algebra applies to third one. The element ij of the covariance term we seek to re-compute is given by:

$$\langle x_i^{\mathrm{loc}}\, x_j^{\mathrm{rec}}\rangle = \frac{u^2 k}{\lambda} m_j \,\langle \int_0^t \exp[-(t-v)]\chi_i(v)\,\mathrm{d}v \cdot \sum_{k=1}^{N} n_k \int_0^t \exp[-(t-v)]\{\exp[\lambda(t-v)]-1\}\chi_k(v)\,\mathrm{d}v\rangle. \tag{22}$$

where we used Eqs. 9 and 10 together with $U = u\,I$. We now introduce a compact notation to represent the signal obtained by filtering the stochastic process $\chi_i(t)$ with an exponential kernel of timescale Λ:

$$\hat{\chi}_i(\Lambda) = \int_0^t \exp[\Lambda(t-v)]\chi_i(v)\,\mathrm{d}v. \tag{23}$$

Eq. 22 can therefore be rewritten as:

$$\begin{aligned}\langle x_i^{\mathrm{loc}}\, x_j^{\mathrm{rec}}\rangle &= \frac{u^2 k}{\lambda} m_j \sum_{k=1}^{N} n_k \;[\langle\, \hat{\chi}_i(-1)\hat{\chi}_k(\lambda-1)\rangle - \langle \hat{\chi}_i(-1)\hat{\chi}_k(-1)\rangle] \\ &= \frac{u^2 k}{\lambda} m_j n_i \;\left[\langle\, \hat{\chi}_i(-1)\hat{\chi}_i(\lambda-1)\rangle - \langle [\hat{\chi}_i(-1)]^2\rangle\right]\end{aligned} \tag{24}$$

where we have used the fact that filtered processes $\hat{\chi}_i$ and $\hat{\chi}_k$ are uncorrelated: $\langle \hat{\chi}_i(\Lambda)\hat{\chi}_k(\Lambda')\rangle = 0$ for $i \neq k$ and every value of Λ, Λ'. In the last equation, the dependence on $\boldsymbol{n}$ appears in the same form as in Eq. 21.

We now paraphrase this derivation in words. For neuron i, the local activity term, x_i^{loc}, is given by a filtered version of the local stochastic input, χ_i. In contrast, for neuron j, the recurrent activity term, x_j^{rec}, is expressed as a sum over *all* stochastic inputs χ_k, each filtered and weighted by $m_j n_k$. When computing the cross-covariance between these two terms, only the stochastic input $k = i$ within x_j^{rec} contributes a non-zero term. Consequently, the covariance between units i and j is proportional to $m_j n_i$.

This derivation highlights how the dependence of covariance on the connectivity vector $\boldsymbol{n}$ arises from activity comprising two components: one representing the integrated local stochastic input and the other reflecting integrated inputs to all neurons, aggregated through recurrent connectivity (Eq. 8). In Methods 2, we discuss how these two activity components robustly emerge in continuous- and discrete-time RNNs. We also show that the local component takes a much simplified form in discrete-time RNNs characterized by extremely fast updates [34, 6]. For these models, which are widely used in machine learning [59] and statistical applications [16], the activity covariance is expected to take a different functional form.

4.2 Covariance eigenvalues and eigenvectors

To quantify the statistical properties of activity and its dependence on connectivity vectors, we now compute the eigenvalues and eigenvectors of Σ (Eq. 21). Because of the properties of the identity matrix, the eigenvalues of Σ, $\{\mu_i\}_i$, are given by:

$$\mu_i = \frac{u^2}{2}\left(1 + \mu_i^{\mathrm{lr}}\right), \tag{25}$$

where $\{\mu_i^{\mathrm{lr}}\}_i$ are the eigenvalues of the low-rank covariance component

$$\Sigma^{\mathrm{lr}} = \frac{k}{2-\lambda}\left[\boldsymbol{n}\boldsymbol{m}^\top + \boldsymbol{m}\boldsymbol{n}^\top + \frac{k}{1-\lambda}\,\boldsymbol{m}\boldsymbol{m}^\top\right]. \tag{26}$$

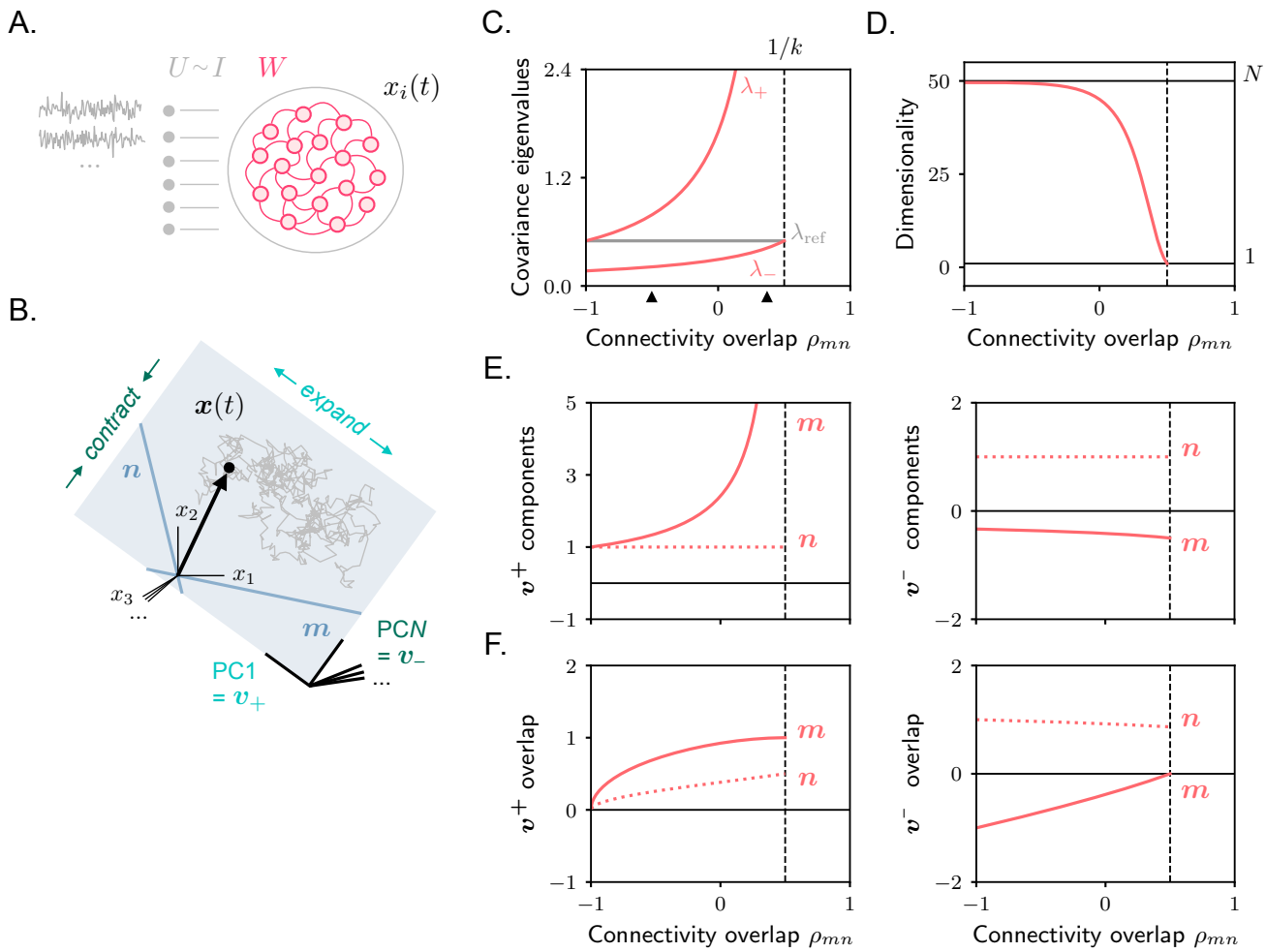

Figure 3: Rank-one RNN receiving high-dimensional stochastic inputs. **A.** Model architecture. **B.** The activity covariance is high-dimensional, with all eigenvalues taking identical values except for two – one larger and one smaller. The principal components associated with these two eigenvalues lie within the plane spanned by the connectivity vectors $\boldsymbol{m}$ and $\boldsymbol{n}$. **C.** Covariance eigenvalues as a function of overlap between connectivity vectors. The dashed vertical line indicates the value of ρ_{mn} for which dynamics becomes unstable. Black arrows indicate the value of ρ_{mn} that is used for simulations in Fig. 4. **D.** Dimensionality. Horizontal black lines indicate the maximum (N) and the minimum (1) possible values. **E.** Components of $\boldsymbol{v}_+$ (or PC1 vector, left) and $\boldsymbol{v}_-$ (or PCN vector, right) along connectivity vectors $\boldsymbol{m}$ and $\boldsymbol{n}$, as from Eq. 28. **F.** Overlap (Eq. 4) between the principal components $\boldsymbol{v}_+$ and $\boldsymbol{v}_-$ (after normalization) and the connectivity vectors $\boldsymbol{m}$ and $\boldsymbol{n}$.

Σ^{lr} has $N-2$ vanishing eigenvalues, resulting in $N-2$ eigenvalues for Σ that are equal to $\mu_{\text{ref}} = u^2/2$. These eigenvalues correspond to eigenvectors, or PCs, along which activity is dominated by external inputs and receives no contribution from recurrent interactions. To compute the non-vanishing eigenvalues, we again follow [37]. A little algebra gives (see Methods 5):

$$\mu_{\pm}^{\text{lr}} = \frac{k}{2(2-\lambda)}\left[2\rho_{mn} + \frac{k}{1-\lambda} \pm \sqrt{\left(2\rho_{mn} + \frac{k}{1-\lambda}\right)^2 + 4(1-\rho_{mn}^2)}\right] \tag{27}$$

from which $\mu_{\pm}$ can be computed via Eq. 25. These eigenvalues correspond to eigenvectors, or PCs, along which activity is shaped by recurrent interactions. One can see from Eq. 27 that μ_+^{lr} and μ_-^{lr} take values that are, respectively, always positive and negative. Therefore, μ_+ and μ_- take values that are, respectively, larger and

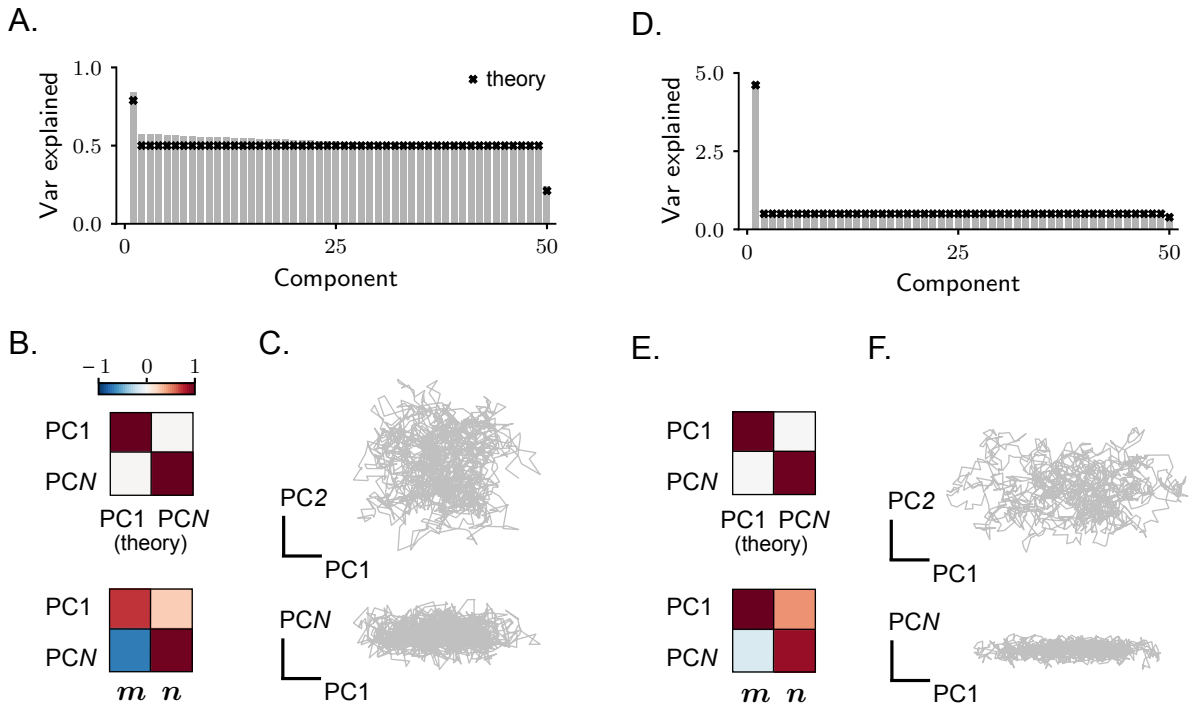

Figure 4: Rank-one RNN receiving high-dimensional stochastic inputs. **A–B–C.** Example of a simulated network with $\rho_{mn} = -0.5$. In A: covariance spectrum. In B: overlap between two principal components (the strongest and the weakest) estimated from simulated activity and the theoretically-estimated vectors $\boldsymbol{v}_+$ and $\boldsymbol{v}_-$ (top), or vectors $\boldsymbol{m}$ and $\boldsymbol{n}$ (bottom). Overlaps are quantified via Eq. 4. In C: simulated activity projected on two different pairs of PCs. **D–E–F.** Same as in A–B–C, example with $\rho_{mn} = 0.3$. Note that, although the qualitative behaviour of activity in the two examples is similar, activity in the example network in A–B–C is overall higher dimensional.

smaller than the reference value $\mu_{\text{ref}} = u^2/2$. This implies that the principal components associated with μ_+ and μ_- identify directions along which activity is maximally amplified and maximally shrunk (Fig. 3B).

These eigenvectors can be computed as the eigenvectors of Σ^{lr} associated with $\mu_\pm^{\text{lr}}$. Similarly to the previous section, we set the ansatz:

$$\boldsymbol{v}_\pm = a_\pm \boldsymbol{m} + \boldsymbol{n}. \tag{28}$$

A little algebra gives (see Methods 5):

$$a_\pm = \frac{1}{2}\left[\frac{k}{1-\lambda} \pm \sqrt{\left(2\rho_{mn} + \frac{k}{1-\lambda}\right)^2 + 4(1-\rho_{mn}^2)}\right]. \tag{29}$$

Combining these results, we conclude that the activity is isotropic, with amplitude $\mu_{\text{ref}} = u^2/2$, across all dimensions orthogonal to the plane defined by the connectivity vectors $\boldsymbol{m}$ and $\boldsymbol{n}$. Within this plane, activity exhibits greater variance along one dimension (corresponding to the principal component associated with the largest covariance eigenvalue, PC1) and lower variance along the other (corresponding to the principal component associated with the smallest covariance eigenvalue, PCN). Notably, unlike the case of low-dimensional inputs considered earlier, the recurrent connectivity shapes activity along two dimensions (both of which collapse into one in the case of symmetric connectivity, for $\rho_{mn} \in \{-1, 1\}$). This occurs because, in addition to the connectivity vector $\boldsymbol{m}$, also $\boldsymbol{n}$ has an impact on the activity covariance structure.

4.3 Dependence on connectivity vectors overlap

We analyze in detail the behaviour of the covariance eigenvalues and eigenvectors as a function of ρ_{mn}, the parameter that controls the level of symmetry of the recurrent connectivity matrix by specifying the overlap between the connectivity vectors $\boldsymbol{m}$ and $\boldsymbol{n}$. The network activity is stable for ρ_{mn} values ranging between -1 and $1/k$ (where $\lambda = 1$). At one extreme, when $\boldsymbol{m}$ and $\boldsymbol{n}$ are strongly anti-correlated ($\rho_{mn} \to -1$), the network functions as a one-dimensional auto-encoder [3, 23], with recurrent connectivity generating strong negative feedback. At the other extreme, when $\boldsymbol{m}$ and $\boldsymbol{n}$ are strongly correlated ($\rho_{mn} \to 1/k$), the network behaves as a line-attractor [52], where recurrent connectivity mediates the integration of the external inputs by generating positive feedback that counteracts the leak dynamics.

We start by examining the case of strong positive overlap, with $\rho_{mn} \to 1/k$ and $\lambda \to 1$. Keeping only the dominant terms, we have

$$\mu_{\pm}^{\mathrm{lr}} \simeq \frac{k}{2}\left[\frac{k}{1-\lambda} \pm \left|\frac{k}{1-\lambda}\right|\right], \tag{30}$$

implying that $\mu_{+}^{\mathrm{lr}} \to +\infty$ and $\mu_{-}^{\mathrm{lr}} \to 0$ or, equivalently, $\mu_{+} \to +\infty$ and $\mu_{-} \to u^2/2$ (Fig. 3C). Therefore, activity is characterized by a diverging component along $\boldsymbol{v}_{+}$, resulting from noise integration. Along of all other directions, including $\boldsymbol{v}_{-}$, the amplitude of activity is fixed to $u^2/2$. Therefore, the dimensionality of activity approaches one (Fig. 3D). The principal component $\boldsymbol{v}_{+}$ is primarily aligned with the connectivity vector $\boldsymbol{m}$, as indicated by the fact that $a_{+} \to \infty$ (Fig. 3E). However, $\boldsymbol{v}_{+}$ still retains a nonzero overlap with $\boldsymbol{n}$, since $\boldsymbol{m}$ and $\boldsymbol{n}$ are not orthogonal. Simulated activity for networks with positive overlap, operating close to this regime, is showed in Fig. 4D–E–F.

We now examine the opposite scenario of strong negative overlap, corresponding to $\rho_{mn} \to -1$, for which $\lambda \to -k$. We show in Methods 5 that, in this case, the value of μ_{+} is equal to the reference value $u^2/2$. The value of μ_{-}, instead, is significantly smaller than the reference value, and approaches zero as the strength of recurrent connectivity k is increased (Fig. 3C). Therefore, activity is characterized by a strongly reduced component along $\boldsymbol{v}_{-}$. Along of all other directions, including $\boldsymbol{v}_{+}$, the amplitude of activity is fixed to $u^2/2$. As the suppression along $\boldsymbol{v}_{-}$ has a limited impact on activity (in contrast to the previous case with positive overlap, where the expansion along $\boldsymbol{v}_{+}$ is unbounded), the dimensionality of activity is high (Fig. 3D). The principal component $\boldsymbol{v}_{-}$ is mostly aligned with the connectivity vector $\boldsymbol{n}$, as it can be showed that the value of a_{-} approaches zero for strong recurrent connectivity (Fig. 3E). Also in this case, $\boldsymbol{v}_{-}$ retains a nonzero overlap with $\boldsymbol{m}$, since $\boldsymbol{m}$ and $\boldsymbol{n}$ are strongly aligned. Simulated activity for networks with negative overlap, operating close to this regime, is showed in Fig. 4A–B–C.

Finally, in the case of maximally asymmetric connectivity ($\rho_{mn} \to 0$), activity is characterized by intermediate properties: the values of μ_{+} and μ_{-} are respectively larger and smaller than the reference values (Fig. 3C). Therefore, activity is characterized by an enhanced component along $\boldsymbol{v}_{+}$ and a reduced one along $\boldsymbol{v}_{-}$; dimensionality takes intermediate values (Fig. 3D). The principal components $\boldsymbol{v}_{+}$ and $\boldsymbol{v}_{-}$ have non-vanishing components along both connectivity vectors $\boldsymbol{m}$ and $\boldsymbol{n}$; $\boldsymbol{v}_{+}$ retains a stronger overlap with $\boldsymbol{m}$, while $\boldsymbol{v}_{-}$ a stronger overlap with $\boldsymbol{n}$ (Fig. 3E).

To summarize, in the limit of strongly positive or negative overlap – corresponding to the well-established cases of autoencoder and line-attractor models – the effect of recurrent interactions on activity, which is generally two-dimensional, reduces to a single dominant dimension. This effect differs significantly in the two cases: for positive overlap, activity is strongly amplified along the direction most correlated with $\boldsymbol{m}$, leading to low dimensionality; for negative overlap, activity is suppressed along the direction most correlated with $\boldsymbol{n}$, resulting in rather high dimensionality.

5 Higher-rank networks

Our analysis so far focused on connectivity matrices of unit rank. In this section, we discuss how results generalize to matrices of higher rank. Those matrices can be written as

$$W = k \sum_{r=1}^{R} \boldsymbol{m}^r \boldsymbol{n}^{r\top} \tag{31}$$

where once again k is a finite scalar and connectivity vectors $\boldsymbol{m}^r$ and $\boldsymbol{n}^r$ for $r = 1, \ldots, R < N$ are unit norm. In particular, we focus on a subset of low-rank matrices: those for which the connectivity vectors $\boldsymbol{m}^r$ and $\boldsymbol{n}^r$ associated with different unit-rank components are orthogonal ($\boldsymbol{m}^{r\top}\boldsymbol{n}^{r'} = 0$ for $r \neq r'$). These matrices have a set of R non-zero eigenvalues given by $\lambda^r = k\rho_{m^r n^r}$, each corresponding to an eigenvector $\boldsymbol{m}^r$ (therefore, for example, low-rank matrices with complex eigenvalues are not included within this ensemble).

For these matrices, the propagator can again be computed in closed form (Eq. 93, see Methods 6), and therefore activity can be computed analytically. We have:

$$\boldsymbol{x}(t) = \boldsymbol{x}^{\text{loc}}(t) + \sum_{r=1}^{R} \boldsymbol{x}^{\text{rec},r}(t) \tag{32}$$

where the component $\boldsymbol{x}^{\text{loc}}$ is given by Eq. 9, and each component $\boldsymbol{x}^{\text{rec},r}$ is given by Eq. 10, with $\boldsymbol{m}$, $\boldsymbol{n}$, and λ replaced, respectively, by $\boldsymbol{m}^r$, $\boldsymbol{n}^r$, and λ^r. Thus, increasing the rank of the synaptic connectivity increases the dimensionality of the activity component generated by recurrent connectivity.

From activity, we can compute the equal-time covariance, which is given by (see Methods 7):

$$\Sigma = \frac{1}{2}\left\{\bar{\Sigma}_{\text{inp}} + \sum_{r=1}^{R} \frac{k}{2-\lambda^r}\left[\boldsymbol{x}^r\boldsymbol{m}^{r\top} + \boldsymbol{m}^r\boldsymbol{x}^{r\top}\right] + \sum_{r=1}^{R}\sum_{r'=1}^{R} \frac{\sigma^{rr'}k^2(4-\lambda^r-\lambda^{r'})}{(2-\lambda^r-\lambda^{r'})(2-\lambda^r)(2-\lambda^{r'})}\boldsymbol{m}^r\boldsymbol{m}^{r'\top}\right\} \tag{33}$$

where we defined $\boldsymbol{x}^r = \bar{\Sigma}_{\text{inp}}\boldsymbol{n}^r$ and $\sigma^{rr'} = \boldsymbol{n}^{r\top}\bar{\Sigma}_{\text{inp}}\,\boldsymbol{n}^{r'}$. To inspect the properties of this covariance matrix, we once again focus on the two limit cases, corresponding to low- and high-dimensional stochastic inputs.

For one-dimensional inputs, Eq. 33 becomes:

$$\Sigma = \frac{1}{2}\left\{\boldsymbol{u}\boldsymbol{u}^\top + \sum_{r=1}^{R} \frac{\rho_{n^r u}\,k}{2-\lambda^r}\left[\boldsymbol{u}\boldsymbol{m}^{r\top} + \boldsymbol{m}^r\boldsymbol{u}^\top\right] + \sum_{r=1}^{R}\sum_{r'=1}^{R} \frac{\rho_{n^r u}\,\rho_{n^{r'} u}\,k^2(4-\lambda^r-\lambda^{r'})}{(2-\lambda^r-\lambda^{r'})(2-\lambda^r)(2-\lambda^{r'})}\boldsymbol{m}^r\boldsymbol{m}^{r'\top}\right\}. \tag{34}$$

The maximum rank of this covariance matrix is $R+1$, which is reached when external inputs have a nonzero overlap with one of the connectivity vectors $\boldsymbol{n}^r$. Activity is therefore low-dimensional, and occupies the hyperplane spanned by the input vector $\boldsymbol{u}$ and the connectivity vectors $\boldsymbol{m}$. As for rank-one networks, activity in this input regime qualitatively resembles the activity described in previous work on deterministic low-rank RNNs [2].

For high-dimensional inputs, we have instead

$$\Sigma = \frac{1}{2}\left\{I + \sum_{r=1}^{R} \frac{k}{2-\lambda^r}\left[\boldsymbol{n}^r\boldsymbol{m}^{r\top} + \boldsymbol{m}^r\boldsymbol{n}^{r\top}\right] + \sum_{r=1}^{R}\sum_{r'=1}^{R} \frac{\rho_{n^r n^{r'}}\,k^2(4-\lambda^r-\lambda^{r'})}{(2-\lambda^r-\lambda^{r'})(2-\lambda^r)(2-\lambda^{r'})}\boldsymbol{m}^r\boldsymbol{m}^{r'\top}\right\}. \tag{35}$$

This covariance matrix consists of the sum of a full-rank term (first term on the right-hand side) and a low-rank term arising from recurrent interactions. The maximum rank of the low-rank component is $2R$. Generalizing the results from rank-one networks, we conclude that in the presence of high-dimensional stochastic inputs, low-rank recurrent connectivity shapes activity covariance along a number of dimensions equal to twice the rank of the connectivity. This occurs because activity is structured along both sets of connectivity vectors, $\boldsymbol{m}$ and $\boldsymbol{n}$.

In Fig. 5, we analyze in detail rank-two networks by focusing on two specific connectivity parametrizations. In Fig. 5B–C, we fix the overlap $\rho_{m^1m^2}$ and $\rho_{n^1n^2}$ to zero, while varying $\rho_{m^1n^1}$ and $\rho_{m^2n^2}$. This case corresponds to rank-one components in the connectivity (Eq. 31) that operate on orthogonal subspaces. The eigenvalues of the recurrent connectivity in this parametrization vary together with $\rho_{m^1m^2}$ and $\rho_{n^1n^2}$. In Fig. 5D–E, instead, we fix $\rho_{m^1n^1}$ and $\rho_{m^2n^2}$ to zero, while varying $\rho_{m^1m^2}$ and $\rho_{n^1n^2}$. This parametrization results in a connectivity matrix with vanishing eigenvalues, and arbitrary correlations among rank-one components.

As expected, for matrices obeying both the first (Fig. 5B) and the second (Fig. 5D) parametrizations, recurrent connectivity impacts the covariance by modifying the value of four eigenvalues. Among these four eigenvalues, two of them are lowered with respect to the reference value μ_{ref}, while two are increased. Similarly to the rank-one case, activity is therefore compressed along two directions and stretched along two others (Fig. 5F). The activity PCs that correspond to these transformations are linear combinations of the four connectivity vectors (Fig. 5G).

In Fig. 5C–E, we finally leverage the predictions for the covariance eigenvalues to compute dimensionality. We find that activity generally exhibits relatively high dimensionality across most of the parameter space.

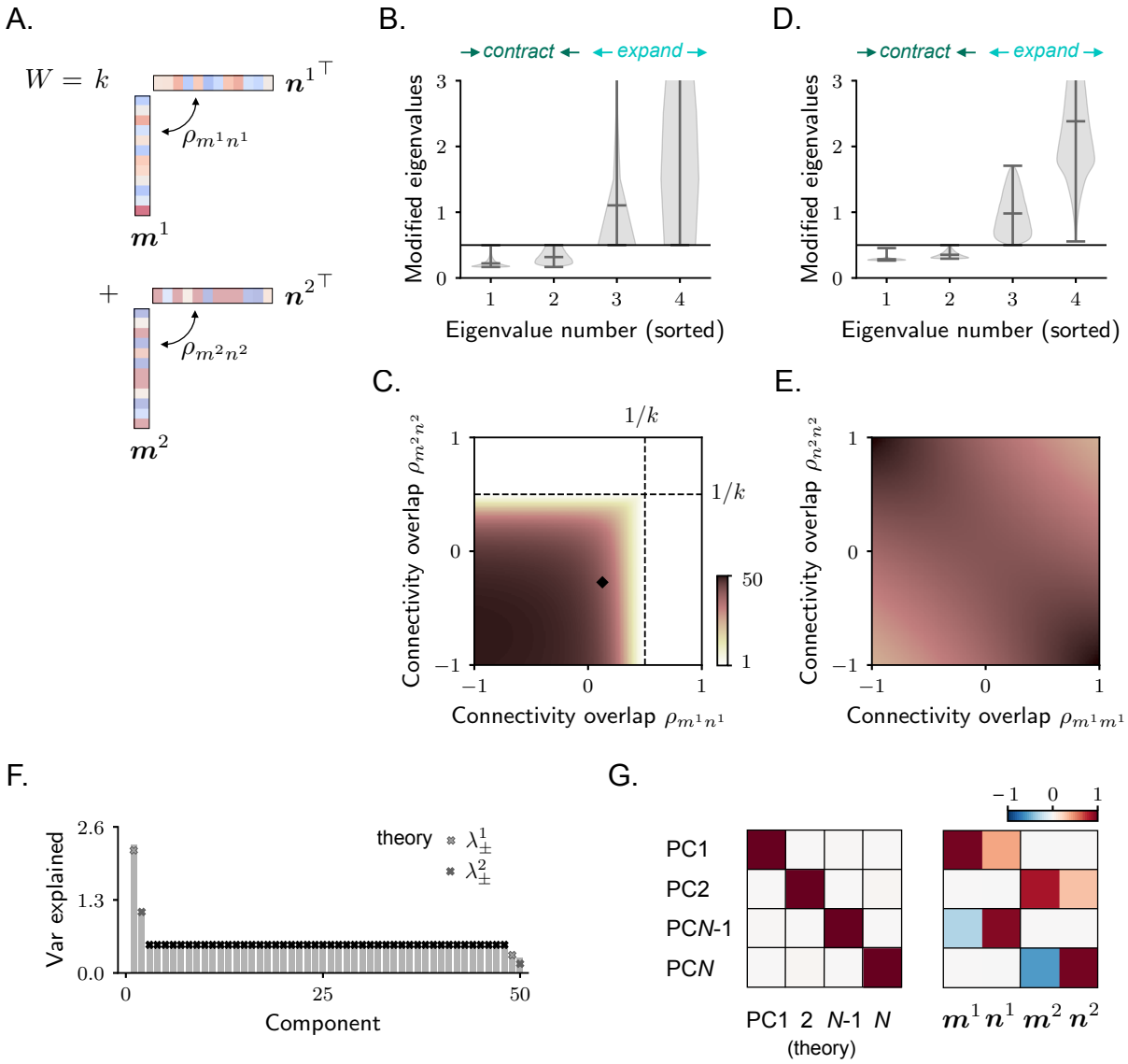

Figure 5: Stochastic activity in rank-two recurrent neural networks. **A.** Rank-two connectivity. **B.** Eigenvalues of the covariance matrix that are different than the reference value μ_{ref}. As connectivity is rank-two, four eigenvalues are perturbed; we sort them in ascending order. Violin plots show the distribution of perturbed eigenvalues for different values of the parameters $\rho_{m^1 n^1}$ and $\rho_{m^2 n^2}$. Note that, for all sets of parameters, two eigenvalues are increased and two decreased with respect to μ_{ref}. **C.** Dimensionality as a function of $\rho_{m^1 n^1}$ and $\rho_{m^2 n^2}$. The black dashed lines indicate the parameter values for which dynamics becomes unstable. The tiny black square indicates the parameter values that are used for simulations in F–G. In both B and C, we keep the values of $\rho_{m^1 m^2}$ and $\rho_{n^1 n^2}$ fixed to zero (see Methods 7). **D–E.** Same as for B–C, but for a different parametrization, where we keep $\rho_{m^1 n^1}$ and $\rho_{m^2 n^2}$ fixed to zero. **F–G.** Example of a simulated network, parameters indicated in C. In F: covariance spectrum. In G: overlap between four selected principal components (the strongest and the weakest) estimated from simulated activity and the the theoretically-estimated covariance eigenvectors (left) and the connectivity vectors (right). Overlaps are quantified via Eq. 4. The theoretical predictions for this case are reported in Methods 7.

Low-dimensional activity emerges only in regions where one or both rank-one components of the connectivity are characterized by a large internal overlap $\rho_{m^r n^r}$, leading to large and positive eigenvalues. In these cases,

recurrent connectivity generates positive feedback resulting in a temporal integration of the external inputs, similarly to rank-one case discussed in Fig. 4D–F.

6 Amplification in excitatory-inhibitory networks

As a final application of major biological relevance, we consider a circuit consisting of one excitatory (E) and one inhibitory (I) unit. The connectivity is given by

$$W = w \begin{pmatrix} 1 & -g \\ 1 & -g \end{pmatrix} \tag{36}$$

where w and g are two non-negative scalars. This matrix is rank-one and can be rewritten in the form of Eq. 3 by setting

$$\begin{aligned} \boldsymbol{m} &= \frac{1}{\sqrt{2}}(1,1) \\ \boldsymbol{n} &= \frac{1}{\sqrt{g^2+1}}(1,-g) \end{aligned} \tag{37}$$

as well as $k = w\sqrt{2(g^2+1)}$. The connectivity vector $\boldsymbol{m}$ therefore corresponds to the *sum* direction, expressing co-modulation of the E-I units, while the connectivity vector $\boldsymbol{n}$ is approximately aligned with the *difference* direction, capturing opposite-sign modulations of E and I [36]. The non-zero eigenvalue of the connectivity is given by $\lambda = w(1-g)$, which is negative in the inhibition-dominated regime ($g > 1$).

A key property of E-I connectivity in cortical circuits is its ability to amplify inputs while preserving fast dynamics [36]. This is achieved through a mechanism known as non-normal amplification, arising in circuits with non-symmetric connectivity matrices and non-orthogonal eigenvectors. The stochastic setting we consider here provides a natural framework for studying inputs amplification [13, 17]. We therefore leverage our mathematical framework to analyze stochastic dynamics in the rank-one E-I circuit (Eq. 36), with a particular focus on amplification and its dependence on the geometry of external inputs.

We begin by considering the case in which E and I units receive uncorrelated stochastic inputs (high-dimensional inputs, Fig. 6A). The covariance matrix (Eq. 21) has two eigenvalues, illustrated in Fig. 6B. The second eigenvalue is significantly smaller than the first (note the different color scales), indicating that activity is approximately one-dimensional, with nearly all variance concentrated along the first principal component. To characterize PC1 (Eq. 28), we compute its overlap with the sum and difference directions (Fig. 6C). Consistent with previous work, we find that PC1 is much more strongly aligned with the sum mode than with the difference mode.

The amplitude of the dominant covariance eigenvalue provides a measure of input amplification. Fig. 6B shows that the degree of input amplification performed by the recurrent connectivity depends on the parameters w and g, quantifying the total connectivity strength and the relative strength of inhibition over excitation, respectively. In particular, amplification is large in a parameter region adjacent to the unstable region, where the connectivity eigenvalue λ exceeds the stability threshold. At the threshold, amplification is likely dominated by a Hebbian-like mechanism, similar to Fig. 4D, with external inputs integrated via slow dynamics [17]. However, relatively large amplification can also occur in an adjacent region where the connectivity eigenvalue λ is negative (Fig. 6D), and non-normal amplification mechanisms are at play. This region corresponds to values of g slightly greater than one (weak inhibition dominance) and large w (strong excitation and inhibition).

To understand how variance amplification is linked to the geometry of external inputs, we next consider the case of one-dimensional inputs (Fig. 6E). We construct a family of input vectors, $\boldsymbol{u}$, each characterized by a different relative strength of input to E versus I. As before, we compute the covariance eigenvalue and eigenvectors as functions of the parameters w and g. As in the high-dimensional input configuration, we find that activity is often concentrated along PC1, expressing co-modulatory variance in the E and I units (sum direction, Fig. 6H). Crucially, however, the amplitude of input amplification and its dependence on the connectivity parameters vary with the input vector direction (Fig. 6F–G). For input vectors strongly aligned with the difference direction, input amplification can be large; in particular, it reaches a maximum for large values of w and values of g slightly greater than one, as in the previous case. Instead, for inputs strongly aligned with the sum direction,

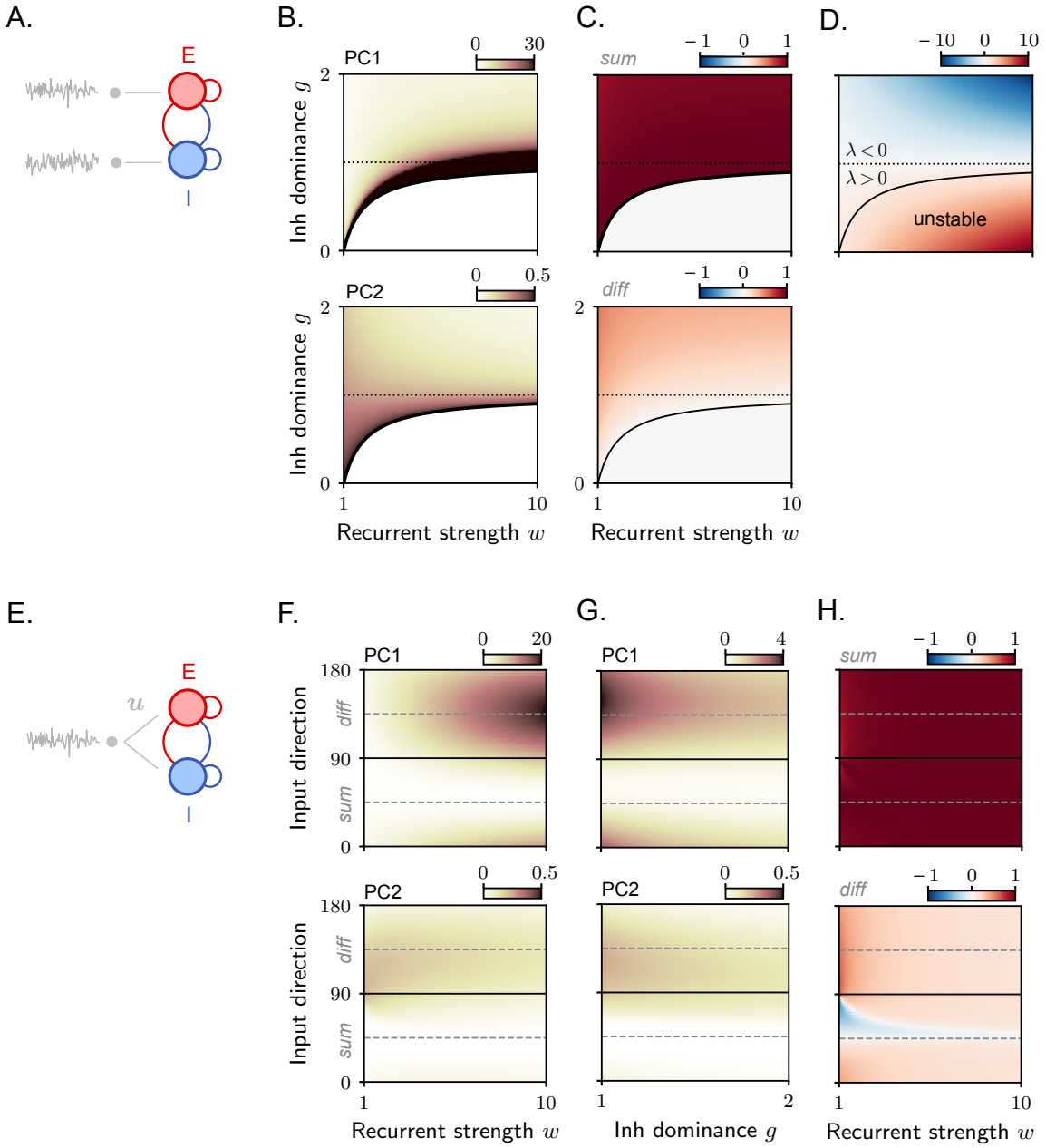

Figure 6: Stochastic activity in a rank-one excitatory-inhibitory circuit. **A.** E-I circuit with high-dimensional inputs. **B.** Variance explained by PC1 (top) and PC2 (bottom) as a function of the overall recurrent connectivity strength w and the relative dominance of inhibition g. In B–C–D, the black solid line separates the regions for which the non-zero eigenvalue λ is larger or smaller than one. The black dotted line separates the regions for which the non-zero eigenvalue λ is larger or smaller than zero. Note the different color scales in the top and bottom plots. **C.** Overlap between PC1 and the sum (top) and diff (bottom) direction. The sum and difference directions are defined as $(1, \pm 1)/\sqrt{2}$. **D.** Non-zero eigenvalue of the synaptic connectivity matrix $\lambda = w(1-g)$. **E.** E-I circuit with one-dimensional inputs. **F.** Variance explained by PC1 (top) and PC2 (bottom) as a function of the overall recurrent connectivity strength w and the direction of the input vector $\boldsymbol{u}$. The input direction is parametrized by an angle θ (see Methods 8), so that $\theta = 0°$ (resp. 90°) correspond to inputs entering only E (resp. I), while $\theta = 45°$ (resp. 135°) corresponds to inputs aligned with the sum (resp. diff) direction. **G.** Variance explained by PC1 (top) and PC2 (bottom) as a function of the relative dominance of inhibition g and the direction of the input vector $\boldsymbol{u}$. **H.** Overlap between PC1 and the sum (top) and diff (bottom) direction.

input amplification is small and largely independent of the connectivity parameters. This can be understood from Eq. 15, as this configuration corresponds to very small values of ρ_{nu}, leading to minimal engagement of recurrent connectivity in processing the inputs. Note that along the vertical slices of the plots in Fig. 6F–G, the eigenvalues of the synaptic connectivity remain fixed. Thus, the observed modulation in variance as a function of input direction must be attributed to the non-normal properties of the connectivity.

The dependence of variance on input direction described here is broadly consistent with previous findings in deterministic networks [36, 4]. In particular, for fixed connectivity parameters, the input direction that leads to maximal amplification is close to the difference direction [36], but is not identical – consistent with the results from [4].

Discussion

The rapid advancement of experimental techniques over the past decade has led to extensive datasets characterizing single-neuron activity across large populations and brain regions [57, 56, 30], as well as increasingly detailed maps of their synaptic connectivity [39, 18, 35]. A key challenge is to understand how these two levels of organization relate to each other during brain function. Theoretical models play a fundamental role in this effort, by helping synthesizing hypotheses about the relationship between connectivity and activity.

In this work, we investigated this relationship in linear RNNs with low-rank connectivity under the influence of stochastic external inputs. We analyzed activity across various choices of external input dimensionality, ranging from low to high. Our findings show that the dimensionality of emergent activity is strongly influenced by input dimensionality, with low-dimensional activity arising systematically only when the inputs themselves are low-dimensional. For high-dimensional inputs, the dimensionality of activity is instead shaped by the statistical properties of the connectivity. In particular, low-dimensional activity emerges in networks with connectivity matrices that have large positive eigenvalues (Fig. 4D) or exhibit specific non-normal structure (Fig. 6B, F). Our analysis uncovered a second fundamental difference between the low- and high-dimensional input regimes. In networks receiving low-dimensional inputs, the structure of the activity covariance depends on connectivity solely through the set of vectors $\boldsymbol{m}$. In contrast, in networks receiving high-dimensional inputs, covariance depends on both sets of connectivity vectors, $\boldsymbol{m}$ and $\boldsymbol{n}$. These two sets of vectors define a $2R$-dimensional hyperplane in the activity space where activity distribution is stretched along some directions and shrunk along others.

The dependence of activity covariance on the connectivity vectors $\boldsymbol{m}$ has already been reported in previous work on low-rank RNNs [32, 11, 61]. This dependence arises because, due to the low-rank structure of recurrent connectivity, the contribution to activity that stems from recurrent interactions (here, Eq. 10) is always constrained to align with $\boldsymbol{m}$. In contrast, the dependence of activity covariance on the connectivity vector $\boldsymbol{n}$ has not been previously emphasized. We show that this dependence emerges only when the inputs are high-dimensional – a setup that is relevant for both biological and artificial networks, but has not been mathematically analyzed in detail before.

Previous work has demonstrated that connectivity vectors $\boldsymbol{m}$ can be robustly inferred from activity traces – whether obtained from simulations or experimental data – using simple statistical techniques, such as principal component analysis [32, 61]. In contrast, reliably estimating the direction of vectors $\boldsymbol{n}$ has generally been considered a more challenging problem [61]. In this work, we show that for the activity covariance matrix to explicitly encode information about $\boldsymbol{n}$, the RNN dynamics must be probed by external inputs along a large number of directions. This observation suggests that in high-dimensional input regimes, inference algorithms may more easily and accurately reconstruct the directions of these vectors. This idea is supported by a simple intuition: uniquely identifying the input direction that generates recurrent activity along $\boldsymbol{m}$ (which is precisely the role of $\boldsymbol{n}$) requires observing the system's response across multiple input directions. This hypothesis is further consistent with recent work by [44], which demonstrated that connectivity vectors can be reliably estimated in a teacher-student setup involving high-dimensional stochastic inputs as long as activity traces from all network units is available. That study focused on a specific class of low-rank RNNs, where the eigenvalues of the connectivity matrix vanish, and mostly examined the case of very strong connectivity weights, which lead to approximately low-dimensional activity even in response to high-dimensional inputs. In contrast, here we considered generic low-rank matrices, with a particular focus on rank-one structures, and we provided a detailed analysis of how arbitrary input and connectivity vector geometries affect the emergent activity.

The work by [44] also highlights an important challenge: available algorithms for inferring low-rank dynamics

from activity datasets often struggle when applied to subsampled neural data. Subsampling leads to an overestimation of symmetry in the connectivity matrix and, in general, poor performance in reconstructing non-normal dynamics [40]. Developing algorithms that can handle such challenges, and provide an effective and scalable estimate of both connectivity vectors regardless of their overlap, represents an important avenue for future research. Recently, novel algorithms for inferring RNN models from neural data have been proposed, which leverage the probabilistic structure of activity that emerges from stochastic inputs [25, 41, 26]. Investigating the performance of these algorithms under subsampling and other aversive conditions represents an important step for future research. In the context of those studies, our analysis provides a solid mathematical framework on which these statistical techniques can be benchmarked.

As a straightforward application of our framework, we considered a two-dimensional circuit consisting of one excitatory and one inhibitory unit, with rank-one connectivity. Although this system and its input amplification properties have been the focus of several studies [36, 17, 4, 7], to our knowledge, a systematic analysis for different input structures in the stochastic setting has not been conducted before. Our analysis highlights that, in these systems, large non-normal amplification can be observed provided that both excitation and inhibition are strong and connectivity is weakly inhibition-dominated. This range of parameters maximizes amplification for both low- [36] and high-dimensional [17] external inputs. However, as expected from Eq. 15, amplification occurs for low-dimensional inputs only when input vectors are specifically oriented with respect to the connectivity ones. In particular, non-normal amplification is maximized when the input vector modulates the excitatory and inhibitory units in opposition (difference direction) [36]. Non-normal amplification takes intermediate values for inputs targeting only the excitatory or inhibitory units – a scenario relevant to the study of optogenetic perturbations in cortical circuits [48, 40]. In particular, amplification is greater for inputs targeting only the excitatory unit than for those targeting only the inhibitory unit (Fig. 6F–G; compare the horizontal black lines at 90° and 180°).

Finally, our framework opens new avenues for studying how trial-to-trial variability in brain circuits relates to the underlying anatomical structure. Stochastic inputs have long been used in recurrent neural networks to model fluctuations in brain activity [17, 24, 38, 20]. However, most studies have focused on models with simple connectivity structures, such as fully random networks or those incorporating local connectivity motifs. Our work extends this approach to low-rank connectivity structures, which are global and more naturally linked to network computations [11, 49]. In particular, low-rank RNNs have recently been used to model mesoscopic cortical circuits spanning multiple brain areas [8, 1]. Experimental studies indicate that these circuits exhibit structured trial-to-trial variability, with cross-covariances differing in dimensionality when computed within versus across areas [51]. Our work provides the mathematical tools needed to analyze variability in these models, offering a foundation for refining them through direct comparison with experimental data.

Acknowledgments

We would like to thank Giulio Bondanelli for discussion and feedback on the manuscript. F.M. is supported by a Transition To Independence grant from Simons Foundation (code 00002313). J.C. is supported by a PhD studentship from Fundação para a Ciencia e a Tecnologia (DOI: 10.54499/2020.05021.BD). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Bibliography

[1] J. Barbosa, R. Proville, C. Rodgers, M. DeWeese, S. Ostojic, and Y. Boubenec. Early selection of task-relevant features through population gating. *Nature Communications*, 14:6837, 2023.

[2] M. Beiran, A. Dubreuil, A. Valente, F. Mastrogiuseppe, and S. Ostojic. Shaping dynamics with multiple populations in low-rank recurrent networks. *Neural Comput.*, 33(6):1572–1615, 2021.

[3] M. Boerlin, C. K. Machens, and S. Deneve. Predictive coding of dynamical variables in balanced spiking networks. *PLOS Comput. Biol.*, 9(11):1–16, 11 2013.

[4] G. Bondanelli and S. Ostojic. Coding with transient trajectories in recurrent neural networks. *PLOS Comput Biol*, 16(2):e1007655, 2020.

[5] B. Bordelon, J. Cotler, C. Pehlevan, and J. Zavatone-Veth. Dynamically learning to integrate in recurrent neural networks. *arXiv*, 2503.18754, 2025.

[6] B. Cessac, B. Doyon, M. Quoy, and M. Samuelides. Mean-field equations, bifurcation map and route to chaos in discrete time neural networks. *Physica D*, 74(1):24 – 44, 1994.

[7] G. Christodoulou, T. Vogels, and E. Agnes. Regimes and mechanisms of transient amplification in abstract and biological neural networks. *PLOS Comput. Biol.*, 18(8):1–32, 2022.

[8] D. Clark and M. Beiran. Structure of activity in multiregion recurrent neural networks. *Proc. Natl. Acad. Sci. USA*, 122(10):e2404039122, 2025.

[9] C. Cueva, A. Saez, E. Marcos, A. Genovesio, M. Jazayeri, R. Romo, D. Salzman, M. Shadlen, and S. Fusi. Low-dimensional dynamics for working memory and time encoding. *Proc. Natl. Acad. Sci. USA*, 117(37): 23021–23032, 2020.

[10] D. Dahmen, M. Layer, L. Deutz, P. A. Dąbrowska, N. Voges, M. von Papen, T. Brochier, A. Riehle, M. Diesmann, S. Grün, and M. Helias. Global organization of neuronal activity only requires unstructured local connectivity. *eLife*, 11:e68422, 2023.

[11] A. Dubreuil, A. Valente, M. Beiran, F. Mastrogiuseppe, and S. Ostojic. The role of population structure in computations through neural dynamics. *Nat. Neurosci.*, 25:783–794, 2022.

[12] C. Eliasmith and C. Anderson. *Neural Engineering - Computation, Representation, and Dynamics in Neurobiological Systems*. MIT press, 2004.

[13] B. Farrell and P. Ioannou. Generalized stability theory. Part I: autonomous operators. *J. Atmos. Sci.*, 53: 2025–2040, 1996.

[14] J. Gallego, M. Perich, L. Miller, and S. Solla. Neural manifolds for the control of movement. *Neuron*, 94: 978–984, 2017.

[15] P. Gao and S. Ganguli. On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.*, 32:148–55, 2015.

[16] Z. Ghahramani and G. Hinton. Parameter estimation for linear dynamical systems. *Technical Report CRG-TR-96-2*, 1996.

[17] G. Hennequin, T. Vogels, and W. Gerstner. Non-normal amplification in random balanced neuronal networks. *Phys. Rev. E*, 86:011909, 2012.

[18] H. Hintiryan, N. Foster, and I. Bowman. The mouse cortico-striatal projectome. *Nat. Neurosci.*, 19: 1100–1114, 2016.

[19] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, 79(8):2554–2558, 1982.

[20] Y. Hu and H. Sompolinsky. The spectrum of covariance matrices of randomly connected recurrent neuronal networks with linear dynamics. *PLOS Computat. Biol.*, 18:1–27, 2022.

[21] C. Huang, D. Ruff, R. Pyle, R. Rosenbaum, M. Cohen, and B. Doiron. Circuit models of low-dimensional shared variability in cortical networks. *Neuron*, 101:337–348, 2019.

[22] H. Jaeger. Adaptive nonlinear system identification with echo state networks. *NIPS*, pages 609–616, 2002.

[23] J. Kadmon, J. Timcheck, and S. Ganguli. Predictive coding in balanced neural networks with noise, chaos and delays. *Adv. Neural Inf. Process. Syst.*, 33, 2020.

[24] T. Kanashiro, G. Ocker, M. Cohen, and B. Doiron. Attentional modulation of neuronal variability in circuit models of cortex. *eLife*, 6:e23978, 2017.

[25] T. Kim, T. Luo, T. Can, K. Krishnamurthy, J. Pillow, and C. Brody. Flow-field inference from neural data using deep recurrent networks. *bioRxiv*, 2023.11.14.567136, 2023.

[26] A. Kumar, L. Frank, and K. Bouchard. Identifying feedforward and feedback controllable subspaces of neural population dynamics. *arXiv:2408.05875*, 2024.

[27] I. Landau and H. Sompolinsky. Coherent chaos in a recurrent neural network with structured connectivity. *PLOS Comput. Biol.*, 14(12):1–27, 12 2018.

[28] S. Linderman, M. Johnson, A. Miller, R. Adams, D. Blei, and L. Paninski. Bayesian learning and inference in recurrent switching linear dynamical systems. *Proceedings of Machine Learning Research*, 54:914–922, 2017.

[29] A. Litwin-Kumar, K. Harris, R. Axel, H. Sompolinsky, and L. Abbott. Optimal degrees of synaptic connectivity. *Neuron*, 93(5):1153 – 1164.e7, 2017.

[30] T. Machado, I. Kauvar, and K. Deisseroth. Multiregion neuronal activity: the forest and the trees. *Nat. Rev. Neurosci.*, 23:683–704, 2022.

[31] V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78–84, 2013.

[32] F. Mastrogiuseppe and S. Ostojic. Linking connectivity, dynamics and computations in low-rank recurent neural networks. *Neuron*, 99(3):609–623, 2018.

[33] F. Mastrogiuseppe, N. Hiratani, and P. Latham. Evolution of neural activity in circuits bridging sensory and abstract knowledge. *eLife*, 12:e79908, 2023.

[34] L. Molgedey, J. Schuchhardt, and H. G. Schuster. Suppressing chaos in neural networks by noise. *Phys. Rev. Lett.*, 69:3717–3719, 1992.

[35] R. Munoz-Castaneda, R. Zingg, K. Matho, and al. Cellular anatomy of the mouse primary motor cortex. *Nature*, 598:159–166, 2021.

[36] B. Murphy and K. Miller. Balanced amplification: A new mechanism of selective amplification of neural activity patterns. *Neuron*, 61:635–648, 2009.

[37] Y. Nakatsukasa. The low-rank eigenvalue problem. *arXiv*, 1905.11490, 2019.

[38] G. K. Ocker, Y. Hu, M. A. Buice, B. Doiron, K. Josić, R. Rosenbaum, and E. Shea-Brown. From the statistics of connectivity to the statistics of spike times in neuronal networks. *Curr. Opin. Neurobiol.*, 46: 109–119, 2017.

[39] S. Oh, J. Harris, L. Ng, and al. A mesoscale connectome of the mouse brain. *Nature*, 508:207–214, 2013.

[40] D. O'Shea, L. Duncker, W. Goo, X. Sun, S. V. E. Trautmann, I. Diester, C. Ramakrishnan, K. Deisseroth, M. Sahani, and K. Shenoy. Direct neural perturbations reveal a dynamical mechanism for robust computation. *bioRxiv*, 2022.12.16.520768, 2022.

[41] M. Pals, E. Sagtekin, F. Pei, M. Gloeckler, and J. Macke. Inferring stochastic low-rank recurrent neural networks from neural data. *Adv. Neural Inf. Process. Syst.*, 37, 2024.

[42] U. Pereira and N. Brunel. Attractor dynamics in networks with learning rules inferred from in vivo data. *Neuron*, 99(1):227–238.e4, 2018.

[43] W. Podlaski and C. Machens. Approximating nonlinear functions with latent boundaries in low-rank excitatory-inhibitory spiking networks. *Neural Comput*, 36:803–857, 2024.

[44] W. Quian, J. Zavatone-Veth, B. Ruben, and C. Pehlevan. Partial observation can induce mechanistic mismatches in data-constrained models of neural dynamics. *Adv. Neural Inf. Process. Syst.*, 38, 2024.

[45] K. Rajan and L. F. Abbott. Eigenvalue spectra of random matrices for neural networks. *Phys. Rev. Lett.*, 97:188104, 2006.

[46] S. Recanatesi, G. Ocker, M. Buice, and E. Shea-Brown. Dimensionality in recurrent spiking networks: Global trends in activity and local origins in connectivity. *PLOS Computat. Biol.*, 15:e1006446, 2019.

[47] H. Risken. *The Fokker-Planck Equation*. Springer, 2012.

[48] A. Sanzeni, B. Akitake, H. Goldbach, C. Leedy, N. Brunel, and M. Histed. Inhibition stabilization is a widespread property of cortical networks. *eLife*, 9:e54875, 2020.

[49] F. Schuessler, F. Mastrogiuseppe, A. Dubreuil, S. Ostojic, and O. Barak. The interplay between randomness and structure during learning in rnns. *Adv. Neural Inf. Process. Syst.*, 33:13352–13362, 2020.

[50] F. Schuessler, F. Mastrogiuseppe, S. Ostojic, and O. Barak. Aligned and oblique dynamics in recurrent neural networks. *eLife*, 2024.

[51] J. Semedo, A. Zandvakili, C. Machens, B. Yu, and A. Kohn. Cortical areas interact through a communication subspace. *Neuron*, 102(1):249–259, 2019.

[52] H. Seung. How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. USA*, 93(23):13339–13344, 1996.

[53] M. N. Shadlen and W. T. Newsome. The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J. Neurosci.*, 18(10):3870–3896, 1998.

[54] A. Smith and E. Brown. Estimating a state-space model from point process observations. *Neural Comput*, 15:965–991, 2003.

[55] W. R. Softky and C. Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *J. Neurosci.*, 13(1):334–350, 1993.

[56] N. Steinmetz and al. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science*, 372:eabf4588, 2021.

[57] C. Stringer, M. Pachitariu, N. Steinmetz, C. Reddy, M. Carandini, and K. Harris. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364(6437):255, 2019.

[58] L. Susman, F. Mastrogiuseppe, N. Brenner, and O. Barak. Quality of internal representation shapes learning performance in feedback neural networks. *Phys. Rev. Research*, 3:013176, 2021.

[59] I. Sutskever, O. Vinyals, and Q. Le. Sequence to sequence learning with neural networks. *Adv. Neural Inf. Process. Syst.*, 27, 2014.

[60] T. Toyoizumi and L. Abbott. Beyond the edge of chaos: Amplification and temporal integration by recurrent networks in the chaotic regime. *Phys. Rev. E*, 84, 2011.

[61] A. Valente, J. Pillow, and S. Ostojic. Extracting computational mechanisms from neural data using low-rank rnns. *Adv. Neural Inf. Process. Syst.*, 35, 2022.

[62] J. Wang, D. Narain, E. Hosseini, and M. Jazayeri. Flexible timing by temporal scaling of cortical responses. *Nat. Neurosci.*, 21(1):102–110, 2018.

[63] A. Wikner, J. Harvey, M. Girvan, B. Hunt, A. Pomerance, T. Antonsen, and E. Ott. Stabilizing machine learning prediction of dynamics: Novel noise-inspired regularization tested with reservoir computing. *Neural Networks*, 170:94–110, 2024.

[64] G. Yang, M. Joglekar, H. Song, W. Newsome, and X.-J. Wang. Task representations in neural networks trained to perform many cognitive tasks. *Nat. Neurosci.*, 22:297–306, 2019.

Methods

1 Propagator of rank-one matrices

We derive here the expression (Eq. 7) for the propagator operator $\exp[(W-I)t]$ for a connectivity matrix W of rank one as defined in the main text. We follow [4] and start observing that, since W and I commute,

$$\exp[(W-I)t] = \exp(Wt)\exp(-t). \tag{38}$$

Moreover, we have:

$$\begin{aligned}\exp(Wt) &= I + kt\,\boldsymbol{m}\boldsymbol{n}^\top + \frac{(kt)^2}{2!}\,\boldsymbol{m}\boldsymbol{n}^\top\boldsymbol{m}\boldsymbol{n}^\top + \frac{(kt)^3}{3!}\,\boldsymbol{m}\boldsymbol{n}^\top\boldsymbol{m}\boldsymbol{n}^\top\boldsymbol{m}\boldsymbol{n}^\top\ldots \\ &= I + \frac{1}{\rho_{mn}}\left[(k\rho_{mn}t)\,\boldsymbol{m}\boldsymbol{n}^\top + \frac{(k\rho_{mn}t)^2}{2!}\,\boldsymbol{m}\boldsymbol{n}^\top + \frac{(k\rho_{mn}t)^3}{3!}\,\boldsymbol{m}\boldsymbol{n}^\top\ldots\right] \\ &= I + \frac{\exp(\lambda t)-1}{\lambda}k\,\boldsymbol{m}\boldsymbol{n}^\top\end{aligned} \tag{39}$$

Finally:

$$\exp[(W-I)t] = \exp(-t)\left[I + \frac{\exp(\lambda t)-1}{\lambda}k\,\boldsymbol{m}\boldsymbol{n}^\top\right]. \tag{40}$$

Note that this expression is only valid for $\lambda \neq 0$. When $\lambda = 0$, we have that

$$\exp(Wt) = I + k\,\boldsymbol{m}\boldsymbol{n}^\top t \tag{41}$$

(all the higher-order terms in the power expansion vanish, being multiplied by λ). Therefore:

$$\exp[(W-I)t] = \exp(-t)\left[I + k\,\boldsymbol{m}\boldsymbol{n}^\top t\,\right]. \tag{42}$$

2 Origin of local and recurrent activity terms

To clarify the origin and implications of the separation of activity in two terms (Eq. 8), we consider the discrete-time approximation of Eq. 1:

$$\boldsymbol{x}(t+\tau) = (1-\tau)\,\boldsymbol{x}(t) + \tau\,W\boldsymbol{x}(t) + \tau^{1/2}\,\boldsymbol{\xi}(t+\tau) \tag{43}$$

which is accurate in the $\tau \to 0$ limit. We have used the short-hand notation $\boldsymbol{\xi}(t) = U\boldsymbol{\chi}(t)$.

Starting from a simple initial condition $\boldsymbol{x}(0) = \boldsymbol{0}$, we have $\boldsymbol{x}(\tau) = \tau^{1/2}\,\boldsymbol{\xi}(\tau)$, and therefore

$$\begin{aligned}\boldsymbol{x}(2\tau) &= (1-\tau)\,\boldsymbol{x}(\tau) + \tau\,W\boldsymbol{x}(\tau) + \tau^{1/2}\,\boldsymbol{\xi}(2\tau) \\ &= (1-\tau)\tau^{1/2}\,\boldsymbol{\xi}(\tau) + \tau^{3/2}W\boldsymbol{\xi}(\tau) + \tau^{1/2}\,\boldsymbol{\xi}(2\tau).\end{aligned} \tag{44}$$

In analogy with Eq. 8, we can rewrite this as $\boldsymbol{x}(2\tau) = \boldsymbol{x}^{\text{loc}}(2\tau) + \boldsymbol{x}^{\text{rec}}(2\tau)$ by defining

$$\begin{aligned}\boldsymbol{x}^{\text{loc}}(2\tau) &= (1-\tau)\tau^{1/2}\,\boldsymbol{\xi}(\tau) + \tau^{1/2}\,\boldsymbol{\xi}(2\tau) \\ \boldsymbol{x}^{\text{rec}}(2\tau) &= \tau^{3/2}\,W\boldsymbol{\xi}(\tau).\end{aligned} \tag{45}$$

Taking one extra step, one can easily show that:

$$\begin{aligned}\boldsymbol{x}^{\text{loc}}(3\tau) &= (1-\tau)^2\tau^{1/2}\,\boldsymbol{\xi}(\tau) + (1-\tau)\tau^{1/2}\,\boldsymbol{\xi}(2\tau) + \tau^{1/2}\,\boldsymbol{\xi}(3\tau) \\ \boldsymbol{x}^{\text{rec}}(3\tau) &= \tau^{5/2}\,W^2\boldsymbol{\xi}(\tau) + \tau^{3/2}\,W\boldsymbol{\xi}(2\tau) + (1-\tau)\tau^{3/2}W\boldsymbol{\xi}(\tau).\end{aligned} \tag{46}$$

We highlight the following points. First, both activity terms $\boldsymbol{x}^{\text{loc}}$ and $\boldsymbol{x}^{\text{rec}}$ are expressed as sums over terms that depend on the stochastic input at different time points. Unlike $\boldsymbol{x}^{\text{loc}}$, $\boldsymbol{x}^{\text{rec}}$ depends only on inputs from previous time steps (after passing through the recurrent connectivity). Second, the leak in the dynamics

(Eq. 1), which contributes to all the terms proportional to $(1-\tau)$, feeds into both activity terms, $\boldsymbol{x}^{\text{loc}}$ and $\boldsymbol{x}^{\text{rec}}$. Third, removing the leak from the dynamics modifies the temporal properties of activity, but does not affect its two-contribution structure (Eq. 8). This can be easily verified by noting that removing the leak is equivalent to transforming all $(1-\tau)$ terms into 1 in the equations above.

Finally, we consider one specific type of discrete-time dynamics, obtained by setting $\tau = 1$ (very large updates, corresponding to very fast leak and dynamics). This type of model is of particular interest because it has been widely studied in the context of RNNs with random connectivity [34, 6, 60] and is broadly used in machine learning [59] and statistics [16] applications. For these models, one gets:

$$\boldsymbol{x}^{\text{loc}}(n\tau) = \tau^{1/2}\boldsymbol{\xi}(n\tau), \tag{47}$$

implying that the local activity component loses memory of inputs from previous time steps. However, this is not the case for the recurrent component, for which

$$\boldsymbol{x}^{\text{rec}}(n\tau) = \sum_{m=1}^{n-1} \tau^{(n-m)+\frac{1}{2}} W^{n-m}\boldsymbol{\xi}(m\tau). \tag{48}$$

We conclude that, in discrete-time models with $\tau = 1$, the network's memory of previous inputs is maintained solely through recurrent interactions. The two activity contributions, $\boldsymbol{x}^{\text{loc}}$ and $\boldsymbol{x}^{\text{rec}}$, reflect stochastic inputs at different time points, and are therefore uncorrelated. We therefore expect the activity covariance for these models to take a simpler form than Eq. 13, where two of the four terms arise from the cross-covariance between $\boldsymbol{x}^{\text{loc}}$ and $\boldsymbol{x}^{\text{rec}}$.

3 Covariance derivation

We start from Eq. 11, and we use the expression for the propagator (Eq. 7), yielding

$$\begin{aligned}
\Sigma(t,s) = {} & \bar{\Sigma}_{\text{inp}} \int_0^t \int_0^s \exp[-(t+s-u-v)]\,\delta(u-v)\,\mathrm{d}u\,\mathrm{d}v \\
& + \frac{k}{\lambda}\bar{\Sigma}_{\text{inp}}\boldsymbol{n}\boldsymbol{m}^\top \int_0^t \int_0^s \exp[-(t+s-u-v)]\,\{\exp[\lambda(s-v)]-1\}\delta(u-v)\,\mathrm{d}u\,\mathrm{d}v \\
& + \frac{k}{\lambda}\boldsymbol{m}\boldsymbol{n}^\top\bar{\Sigma}_{\text{inp}} \int_0^t \int_0^s \exp[-(t+s-u-v)]\,\{\exp[\lambda(t-u)]-1\}\delta(u-v)\,\mathrm{d}u\,\mathrm{d}v \\
& + \frac{k^2}{\lambda^2}\boldsymbol{m}\boldsymbol{n}^\top\bar{\Sigma}_{\text{inp}}\,\boldsymbol{n}\boldsymbol{m}^\top \int_0^t \int_0^s \exp[-(t+s-u-v)]\,\{\exp[\lambda(t-u)]-1\}\,\{\exp[\lambda(s-v)]-1\}\delta(u-v)\,\mathrm{d}u\,\mathrm{d}v.
\end{aligned} \tag{49}$$

This expression can be re-written as in Eq. 12 by defining:

$$\begin{aligned}
c_1 &= \int_0^t \int_0^s \exp[-(t+s-u-v)]\,\delta(u-v)\,\mathrm{d}u\,\mathrm{d}v \\
c_2 &= \frac{k}{\lambda}\int_0^t \int_0^s \exp[-(t+s-u-v)]\,\{\exp[\lambda(s-v)]-1\}\delta(u-v)\,\mathrm{d}u\,\mathrm{d}v \\
c_3 &= \frac{k}{\lambda}\int_0^t \int_0^s \exp[-(t+s-u-v)]\,\{\exp[\lambda(t-u)]-1\}\delta(u-v)\,\mathrm{d}u\,\mathrm{d}v \\
c_4 &= \frac{k^2}{\lambda^2}\int_0^t \int_0^s \exp[-(t+s-u-v)]\,\{\exp[\lambda(t-u)]-1\}\,\{\exp[\lambda(s-v)]-1\}\delta(u-v)\,\mathrm{d}u\,\mathrm{d}v.
\end{aligned} \tag{50}$$

Working out the integrals in the right-hand side yields the final expression for the coefficients (the complete derivation is provided below):

$$\begin{aligned}
c_1 &= \frac{\exp(t-s)}{2} \qquad t < s \\
&= \frac{\exp(s-t)}{2} \qquad t > s
\end{aligned} \tag{51}$$

$$
\begin{aligned}
c_2 &= \frac{k}{\lambda}\left[\frac{\exp[(t-s)(1-\lambda)]}{2-\lambda} - \frac{\exp(t-s)}{2}\right] && t<s \\
&= \frac{k}{2}\frac{\exp(s-t)}{2-\lambda} && t>s
\end{aligned}
\tag{52}
$$

$$
\begin{aligned}
c_3 &= \frac{k}{2}\frac{\exp(t-s)}{2-\lambda} && t<s \\
&= \frac{k}{\lambda}\left[\frac{\exp[(s-t)(1-\lambda)]}{2-\lambda} - \frac{\exp(s-t)}{2}\right] && t>s
\end{aligned}
\tag{53}
$$

$$
\begin{aligned}
c_4 &= \frac{k^2}{\lambda}\left[\frac{\exp[(1-\lambda)(t-s)]}{2(1-\lambda)(2-\lambda)} - \frac{\exp(t-s)}{2(2-\lambda)}\right] && t<s \\
&= \frac{k^2}{\lambda}\left[\frac{\exp[(1-\lambda)(s-t)]}{2(1-\lambda)(2-\lambda)} - \frac{\exp(s-t)}{2(2-\lambda)}\right] && t>s.
\end{aligned}
\tag{54}
$$

It is interesting to see that the only covariance terms that can give rise to temporal asymmetries, where $\Sigma(t,s) \neq \Sigma(s,t)$, are those multiplying c_2 and c_3, expressing cross-correlations between the two activity components $\boldsymbol{x}^{\text{loc}}$ and $\boldsymbol{x}^{\text{rec}}$. The temporal asymmetries arise because, as observed in the main text, these two activity components encode stochastic inputs filtered with two different timescales.

The derivation above returns the activity covariance for any pair of time points t, s, from which the equal-time covariance can be obtained by setting $t = s$. The equal-time covariance could have been equivalently derived by solving the Lyapunov equation associated with Eq. 1, which reads:

$$
(k\,\boldsymbol{m}\boldsymbol{n}^\top - I)\Sigma + \Sigma(k\,\boldsymbol{n}\boldsymbol{m}^\top - I) + \bar{\Sigma}_{\text{inp}} = 0. \tag{55}
$$

It is easy to verify that the expression for the equal-time covariance provided in Eq. 13 satisfies the Lyapunov equation above.

Coefficient c_1 For $t < s$:

$$
c_1 = \int_0^t \exp[-(t+s-2u)]\ \mathrm{d}u = \frac{\exp(t-s)+\exp[-(t+s)]}{2} \to \frac{\exp(t-s)}{2} \tag{56}
$$

as the system approaches the stationary state (t, $s \to \infty$). For $t > s$, similarly:

$$
c_1 = \int_0^s \exp[-(t+s-2v)]\ \mathrm{d}v \to \frac{\exp(s-t)}{2}. \tag{57}
$$

Coefficient c_2 For $t < s$:

$$
\begin{aligned}
c_2 &= \frac{k}{\lambda}\int_0^t \exp[-(t+s-2u)]\,\{\exp[\lambda(s-u)]-1\}\,\mathrm{d}u \\
&= \frac{k}{\lambda}\exp[-(t+s)]\left\{\exp(\lambda s)\int_0^t \exp[u(2-\lambda)]\,\mathrm{d}u - \int_0^t \exp(2u)\,\mathrm{d}u\right\} \\
&= \frac{k}{\lambda}\exp[-(t+s)]\left\{\frac{\exp[2t+\lambda(s-t)-\exp(\lambda s)}{2-\lambda} - \frac{\exp(2t)-1}{2}\right\} \\
&= \frac{k}{\lambda}\left\{\frac{\exp[(t-s)(1-\lambda)]-\exp[-t-s(1-\lambda)]}{2-\lambda} - \frac{\exp(t-s)+\exp[-(t+s)]}{2}\right\} \\
&\to \frac{k}{\lambda}\left\{\frac{\exp[(t-s)(1-\lambda)]}{2-\lambda} - \frac{\exp(t-s)}{2}\right\}.
\end{aligned}
\tag{58}
$$

For $t > s$, instead:

$$
\begin{aligned}
c_2 &= \frac{k}{\lambda}\int_0^s \exp[-(t+s-2v)]\left\{\exp[\lambda(s-v)]-1\right\}\mathrm{d}v \\
&= \frac{k}{\lambda}\exp[-(t+s)]\left\{\exp(\lambda s)\int_0^s \exp[v(2-\lambda)]\,\mathrm{d}v - \int_0^s \exp(2v)\,\mathrm{d}v\right\} \\
&= \frac{k}{\lambda}\exp[-(t+s)]\left\{\frac{\exp(2s)-\exp(\lambda s)}{2-\lambda} - \frac{\exp(2s)-1}{2}\right\} \\
&= \frac{k}{\lambda}\left\{\frac{\exp(s-t)-\exp[-t-s(1-\lambda)]}{2-\lambda} - \frac{\exp(s-t)+\exp[-(t+s)]}{2}\right\} \\
&\to \frac{k}{\lambda}\left\{\frac{\exp(s-t)}{2-\lambda} - \frac{\exp(s-t)}{2}\right\} = \frac{k}{2}\frac{\exp(s-t)}{2-\lambda}.
\end{aligned}
\tag{59}
$$

Coefficient c_3 It follows from the same math as above that, for $t < s$:

$$
c_3 = \frac{k}{\lambda}\int_0^t \exp[-(t+s-2u)]\left\{\exp[\lambda(t-u)]-1\right\}\mathrm{d}u \to \frac{k}{2}\frac{\exp(t-s)}{2-\lambda}, \tag{60}
$$

while for $t > s$:

$$
c_3 = \frac{k}{\lambda}\int_0^s \exp[-(t+s-2v)]\left\{\exp[\lambda(s-v)]-1\right\}\mathrm{d}v \to \frac{k}{\lambda}\left\{\frac{\exp[(s-t)(1-\lambda)]}{2-\lambda} - \frac{\exp(s-t)}{2}\right\}. \tag{61}
$$

Coefficient c_4 For $t < s$:

$$
\begin{aligned}
c_4 &= \frac{k^2}{\lambda^2}\int_0^t \exp[-(t+s-2u)]\left\{\exp[\lambda(t-u)]-1\right\}\left\{\exp[\lambda(s-u)]-1\right\}\mathrm{d}u \\
&= \frac{k^2}{\lambda^2}\exp[-(t+s)]\left\{\exp[\lambda(t+s)]\int_0^t \exp[2u(1-\lambda)]\,\mathrm{d}u - [\exp(\lambda t)+\exp(\lambda s)]\int_0^t \exp[u(2-\lambda)]\,\mathrm{d}u + \int_0^t \exp(2u)\,\mathrm{d}u\right\} \\
&= \frac{k^2}{\lambda^2}\exp[-(t+s)]\left\{\frac{\exp[t(2-\lambda)+\lambda s]-\exp[\lambda(t+s)]}{2(1-\lambda)} - \frac{\exp(2t)+\exp[2t+\lambda(s-t)]-\exp(\lambda t)-\exp(\lambda s)}{2-\lambda} + \frac{1}{2}[\exp(2t)-1]\right\} \\
&\to \frac{k^2}{\lambda^2}\left\{\frac{\exp[(1-\lambda)(t-s)]}{2(1-\lambda)} - \frac{\exp(t-s)+\exp[(t-s)(1-\lambda)]}{2-\lambda} + \frac{\exp(t-s)}{2}\right\} = \frac{k^2}{\lambda}\left\{\frac{\exp[(1-\lambda)(t-s)]}{2(1-\lambda)(2-\lambda)} - \frac{\exp(t-s)}{2(2-\lambda)}\right\},
\end{aligned}
$$

while, using similar algebra, for $t > s$

$$
c_4 = \frac{k^2}{\lambda^2}\int_0^s \exp[-(t+s-2v)]\left\{\exp[\lambda(t-v)]-1\right\}\left\{\exp[\lambda(s-v)]-1\right\}\mathrm{d}v \to \frac{k^2}{\lambda}\left\{\frac{\exp[(1-\lambda)(s-t)]}{2(1-\lambda)(2-\lambda)} - \frac{\exp(s-t)}{2(2-\lambda)}\right\}. \tag{62}
$$

4 Low-dimensional inputs

Scaling analysis What is the scaling of the different terms in the covariance (Eq. 15) with respect to the network size N? A quick scaling analysis of Eq. 15 yields:

$$
O(\Sigma_{ij}) = O(u_i)^2 + O(\rho_{nu})\frac{O(u_i)}{\sqrt{N}} + O(\rho_{nu})\frac{O(u_i)}{\sqrt{N}} + O(\rho_{nu})^2\frac{1}{N} \tag{63}
$$

where $O(\Sigma_{ij})$ and $O(u_i)$ represent the scaling of a typical entry for the matrix Σ and the vector $\boldsymbol{u}$, respectively.

We can rewrite the scaling of the overlap ρ_{nu} in terms of the normalized input vector $\bar{\boldsymbol{u}}$, for which $\rho_{\bar{u}\bar{u}} = 1$ and therefore $O(\bar{u}_i) = 1/\sqrt{N}$. This gives:

$$
O(\rho_{nu}) = \frac{O(u_i)}{O(\bar{u}_i)}O(\rho_{n\bar{u}}) = \sqrt{N}O(u_i)O(\rho_{n\bar{u}}). \tag{64}
$$

Substituting this expression into Eq. 63, we obtain:

$$O(\Sigma_{ij}) = O(u_i)^2 \left[1 + O(\rho_{n\bar{u}}) + O(\rho_{n\bar{u}}) + O(\rho_{n\bar{u}})^2\right]. \tag{65}$$

This shows that the four covariance components display identical scaling with N, provided that $O(\rho_{n\bar{u}}) = O(1)$. This condition holds when the normalized connectivity vector $\boldsymbol{n}$ and the input vector $\boldsymbol{u}$ retain a finite overlap for all values of N. If the entries of $\boldsymbol{n}$ and $\boldsymbol{u}$ are drawn randomly, this condition is satisfied when the entries are correlated across the two vectors. Importantly, this condition is independent of the scaling chosen for the input vector $\boldsymbol{u}$.

Covariance eigenvalues and eigenvectors We compute the eigenvalues of the covariance matrix in Eq. 15, which has low rank. From [37], we have that the non-zero eigenvalues of a $N \times N$ low-rank matrix expressed in terms of R couples of $\boldsymbol{m}^r$ and $\boldsymbol{n}^r$ vectors as

$$M_{ij} = \sum_{r=1}^{R} m_i^r n_j^r \tag{66}$$

are identical to the eigenvalues of the $R \times R$ reduced matrix defined by

$$M_{rs}^{\text{red}} = \boldsymbol{m}^{r\top} \boldsymbol{n}^s. \tag{67}$$

We apply this rule to Eq. 15, yielding the reduced matrix:

$$\frac{1}{2}\begin{pmatrix} \rho_{uu} & \rho_{mu} & \rho_{uu} & \rho_{mu} \\ \alpha\rho_{uu} & \alpha\rho_{mu} & \alpha\rho_{uu} & \alpha\rho_{mu} \\ \alpha\rho_{mu} & \alpha & \alpha\rho_{mu} & \alpha \\ \beta\rho_{mu} & \beta & \beta\rho_{mu} & \beta \end{pmatrix}. \tag{68}$$

with α and β defined in Eq. 18. This matrix has clearly only two non-zero eigenvalues. We therefore use the reduction scheme once again, yielding the further reduced 2×2 matrix:

$$\frac{1}{2}\begin{pmatrix} \rho_{uu} + \alpha\rho_{mu} & \rho_{mu} + \alpha \\ \alpha\rho_{uu} + \beta\rho_{mu} & \alpha\rho_{mu} + \beta \end{pmatrix}, \tag{69}$$

From this matrix, eigenvalues can easily be computed, resulting in Eq. 17. Dimensionality (Eq. 16) takes a simple form:

$$\mathcal{D} = \left[1 - 2\frac{(\rho_{uu} - \rho_{mu}^2)(\beta - \alpha^2)}{(\rho_{uu} + 2\alpha\rho_{mu} + \beta)^2}\right]^{-1}, \tag{70}$$

which is bounded between 1 and 2.

We now compute the eigenvectors associated with those eigenvalues. We formulate the ansatz:

$$\boldsymbol{v}_\pm = a_\pm \boldsymbol{m} + \boldsymbol{u}. \tag{71}$$

(Note that those vectors are not normalized, but can be normalized by dividing by $\sqrt{a_\pm^2 + 2\rho_{mu}a_\pm + \rho_{uu}}$). We then impose:

$$\Sigma \boldsymbol{v}^\pm = \mu_\pm \boldsymbol{v}^\pm. \tag{72}$$

With a little algebra, we obtain

$$\Sigma \boldsymbol{v}^\pm = \frac{1}{2}\left[\left(a_\pm\alpha\rho_{mu} + a_\pm\beta + \alpha\rho_{uu} + \beta\rho_{mu}\right)\boldsymbol{m} + \left(a_\pm\rho_{mu} + a_\pm\alpha + \rho_{uu} + \alpha\rho_{mu}\right)\boldsymbol{u}\right]. \tag{73}$$

Combining this with Eq. 72 yields a system of two equations for $a_\pm$ and $\mu_\pm$:

$$\begin{aligned} a_\pm\mu_\pm &= \frac{1}{2}\left(a_\pm\alpha\rho_{mu} + a_\pm\beta + \alpha\rho_{uu} + \beta\rho_{mu}\right) \\ \mu_\pm &= \frac{1}{2}\left(a_\pm\rho_{mu} + a_\pm\alpha + \rho_{uu} + \alpha\rho_{mu}\right). \end{aligned} \tag{74}$$

Using the second equation together with Eq. 17, we find:

$$a_\pm = \frac{2\mu_\pm - \rho_{uu} - \alpha\rho_{mu}}{\rho_{mu} + \alpha} = \frac{1}{2(\rho_{mu} + \alpha)} \left[-\rho_{uu} + \beta \pm \sqrt{(\rho_{uu} + 2\alpha\rho_{mu} + \beta)^2 - 4(\rho_{uu} - \rho_{mu}^2)(\beta - \alpha^2)} \right]. \quad (75)$$

Note that this equation is ill-defined for a_+ when $\rho_{mu} = \rho_{nu} = 0$, for which $\alpha = \beta = 0$. In that case, one can use the first equation in Eq. 74 to see that $a_\pm = 0$.

Simulations detail In Fig. 2, we simulated RNNs of size $N = 100$. To construct external inputs and recurrent connectivity, we first generated a set of three normalized orthogonal vectors $\{\boldsymbol{z}^i\}_{i=1,2,3}$. To this end, we independently sampled vector entries from a Gaussian distribution of mean 0 and variance $1/N$, and then fine-tuned orthonormality through a Gram-Schimdt process. We then set: $\boldsymbol{m} = \boldsymbol{z}^1$, $\boldsymbol{n} = \boldsymbol{z}^2$, implying that $\rho_{mn} = 0$. In Fig. 2C–D–E, we then set $\boldsymbol{u} = \boldsymbol{z}^3$, implying that $\rho_{uu} = 1$ and $\rho_{nu} = \rho_{mu} = 0$. In Fig. 2F–G–H, we set $\boldsymbol{u} = \boldsymbol{n}$, implying that $\rho_{nu} = \rho_{uu} = 1$ and $\rho_{mu} = 0$.

5 High-dimensional inputs

Scaling analysis The scaling with respect to the network size N of Eq. 21 is given by:

$$O(\Sigma_{ij}) = O(u)^2 \left\{ \delta_{ij} + \frac{1}{N} + \frac{1}{N} + \frac{1}{N} \right\}. \quad (76)$$

The equation above shows that, in the case of high-dimensional stochastic inputs, the diagonal entries of the covariance matrix, which quantify single-neuron variances, are significantly larger in amplitude than the off-diagonal entries, which quantify cross-covariances among neurons. This is due to the first term of the covariance (Eq. 21), that is generated from local activity (Eq. 8), being diagonal and having much larger amplitude than the other three. Eq. 76 also indicates that the remaining terms, which express different combinations of the $\boldsymbol{m}$ and $\boldsymbol{n}$ vectors, are characterized by a similar scaling. Importantly, as in the case of low-dimensional inputs, the scaling properties of the covariance are independent of the scaling assumed for the input u.

Covariance eigenvalues and eigenvectors We compute the non-zero eigenvalues of the low-rank component of the covariance matrix, Σ^{lr} (Eq. 26), following again [37], and arriving at the reduced matrix:

$$\frac{k}{2-\lambda} \begin{pmatrix} \rho_{mn} & 1 & \rho_{mn} \\ 1 & \rho_{mn} & 1 \\ k/(1-\lambda) & \rho_{mn}k/(1-\lambda) & k/(1-\lambda) \end{pmatrix}. \quad (77)$$

The last two rows are linearly dependent, implying that the original matrix is rank-two. By reducing the matrix further, we get to:

$$\frac{k}{2-\lambda} \begin{pmatrix} \rho_{mn} & 1 \\ 1 + \rho_{mn}k/(1-\lambda) & \rho_{mn} + k/(1-\lambda) \end{pmatrix} \quad (78)$$

whose eigenvalues are given by Eq. 27.

To compute the eigenvectors associated with those eigenvalues, we formulate the ansatz:

$$\boldsymbol{v}_\pm = a_\pm \boldsymbol{m} + \boldsymbol{n}, \quad (79)$$

where the normalization factor is in this case given by $\sqrt{a_\pm^2 + 2\rho_{mn}a_\pm + 1}$. We then impose:

$$\Sigma^{\text{lr}} \boldsymbol{v}^\pm = \mu_\pm^{\text{lr}} \boldsymbol{v}^\pm. \quad (80)$$

With a little algebra, we obtain

$$\Sigma \boldsymbol{v}^\pm = \frac{k}{2-\lambda} \left[\left(a_\pm \rho_{mn} + 1 + \frac{k}{1-\lambda} a_\pm + \frac{k}{1-\lambda} \rho_{mn} \right) \boldsymbol{m} + (a_\pm + \rho_{mn}) \boldsymbol{n} \right]. \quad (81)$$

Combining it with Eq. 80, we see that

$$a_\pm = \frac{2-\lambda}{k} \mu_\pm^{\text{lr}} - \rho_{mn} = \frac{1}{2} \left[\frac{k}{1-\lambda} \pm \sqrt{\left(2\rho_{mn} + \frac{k}{1-\lambda} \right)^2 + 4(1 - \rho_{mn}^2)} \right]. \quad (82)$$

Limit cases We start considering the case of maximally symmetric recurrent connectivity. To derive Eq. 30, we use $\rho_{mn} \to 1/k$ and $\lambda \to 1$, and then only keep the diverging terms. Similarly, one gets

$$a_{\pm} \simeq \frac{1}{2}\left[\frac{k}{1-\lambda} \pm \left|\frac{k}{1-\lambda}\right|\right], \tag{83}$$

implying that a_+ diverges as $\lambda \to 1$.

We then consider the opposite scenario: using $\rho_{mn} \to -1$ and $\lambda \to -k$ in Eq. 27, we have

$$\mu_{\pm}^{\mathrm{lr}} = \frac{k}{2(2+k)}\left[-2 + \frac{k}{1+k} \pm \left| -2 + \frac{k}{1+k}\right|\right] \tag{84}$$

The argument of the absolute value on the right-hand side is negative. Therefore we have: $\mu_+^{\mathrm{lr}} = 0$, implying that μ_+^{lr} is again equal to $u^2/2$. Instead:

$$\mu_-^{\mathrm{lr}} = \frac{k}{2+k}\left[-2 + \frac{k}{1+k}\right] = -\frac{k}{1+k}. \tag{85}$$

which is a negative number smaller than -1. When $k \to \infty$ (strong recurrent connectivity), $\mu_- \to -1$. Correspondingly,

$$\mu_- = \frac{u^2}{2}\frac{1}{1+k}, \tag{86}$$

which is a positive number that is always smaller than $u^2/2$, and converges to zero for very large k. We also have:

$$a_{\pm} = \frac{1}{2}\left[\frac{k}{1+k} \pm \left| -2 + \frac{k}{1+k}\right|\right] \tag{87}$$

from which $a_+ = 1$, and

$$a_- = \frac{1}{2}\left[\frac{2k}{1+k} - 2\right] = -\frac{1}{1+k}. \tag{88}$$

Note than the latter is negative, and vanishes for $k \to \infty$. This implies that the eigenvector $\boldsymbol{v}_-$ has a small and negative component along connectivity vector $\boldsymbol{m}$, and a larger and positive component along connectivity vector $\boldsymbol{n}$.

Simulations detail In Fig. 4, we simulated RNNs of size $N = 50$. To construct external inputs and recurrent connectivity, we first generated a set of two normalized orthogonal vectors $\{\boldsymbol{z}^i\}_{i=1,2}$ as detailed above. We then set: $\boldsymbol{m} = \boldsymbol{z}^1$, and

$$\boldsymbol{n} = \rho_{mn}\boldsymbol{z}^1 + \sqrt{1-\rho_{mn}^2}\boldsymbol{z}^2. \tag{89}$$

In Fig. 4A–B–C, we then set $\rho_{mn} = -0.5$ In Fig. 4D–E–F, we then set $\rho_{mn} = 0.3$

6 Propagator of higher-rank matrices

For a connectivity matrix of rank greater than one (Eq. 31), we have:

$$\exp[(W-I)t] = \exp\left[\left(k\sum_{r=1}^{R}\boldsymbol{m}^r\boldsymbol{n}^{r\top}\right)t\right]\exp(-t). \tag{90}$$

Evaluating this matrix exponential is generally challenging for a generic low-rank matrix W. However, we can consider a specific subclass of these matrices, where the connectivity vectors $\boldsymbol{m}^r$ and $\boldsymbol{n}^r$ associated with different rank-one components are mutually orthogonal:

$$\boldsymbol{n}^{r\top}\boldsymbol{m}^{r'} = 0 \tag{91}$$

for $r \neq r'$. For these matrices, the nonzero eigenvalues are given by $\lambda^r = k\rho_{m^r n^r}$ for $r = 1, \ldots, R$, with corresponding eigenvectors $\boldsymbol{m}^r$. This class includes, as a special case, the low-rank matrices studied in [4] and [44], for which all eigenvalues vanish.

Eq. 91 implies that all rank-one components commute, leading to:

$$\exp\left[\left(k\sum_{r=1}^{R}\boldsymbol{m}^r\boldsymbol{n}^{r\top}\right)t\right]=\prod_{r=1}^{R}\exp\left(k\,\boldsymbol{m}^r\boldsymbol{n}^{r\top}t\right)=\prod_{r=1}^{R}\left[I+\frac{\exp(\lambda^r t)-1}{\lambda^r}k\,\boldsymbol{m}^r\boldsymbol{n}^{r\top}\right]=I+\sum_{r=1}^{R}\frac{\exp(\lambda^r t)-1}{\lambda^r}k\boldsymbol{m}^r\boldsymbol{n}^{r\top}. \tag{92}$$

Thus, the propagator takes the form:

$$\exp[(W-I)t]=\exp\left[I+\sum_{r=1}^{R}\frac{\exp(\lambda^r t)-1}{\lambda^r}k\,\boldsymbol{m}^r\boldsymbol{n}^{r\top}\right]\exp(-t). \tag{93}$$

This expression assumes that $\lambda^r \neq 0$ for all r. If the eigenvalue associated with a given rank-one component vanishes, the corresponding term in the sum in Eq. 93 must be replaced with the form given in Eq. 41.

7 Covariance of higher-rank networks

To compute activity covariance, we start from Eq. 11, use again Eq. 93, and we arrive at

$$\Sigma(t,s)=c_1(t,s)\,\bar{\Sigma}_{\text{inp}}+\sum_{r=1}^{R}\left[c_2^r(t,s)\,\bar{\Sigma}_{\text{inp}}\boldsymbol{n}^r\boldsymbol{m}^{r\top}+c_3^r(t,s)\,\boldsymbol{m}^r\boldsymbol{n}^{r\top}\bar{\Sigma}_{\text{inp}}\right]+\sum_{r=1}^{R}\sum_{r'=1}^{R}c_4^{rr'}(t,s)\,\boldsymbol{m}^r\boldsymbol{n}^{r\top}\bar{\Sigma}_{\text{inp}}\,\boldsymbol{n}^{r'}\boldsymbol{m}^{r'\top}. \tag{94}$$

Using algebra similar to that in 3, one can show that $c_1(t,s)$ is given by Eq. 51, while $c_2^r(t,s)$ and $c_3^r(t,s)$ obey Eqs. 52 and 53 (with λ replaced by λ^r). Finally, $c_4^{rr'}(t,s)$ is given by:

$$\begin{aligned} c_4^{rr'} &= \frac{k^2}{\lambda^{r'}}\left[\frac{\exp[(1-\lambda^r)(t-s)]}{(2-\lambda^2-\lambda^{r'})(2-\lambda^{r'})}-\frac{\exp(t-s)}{2(2-\lambda^r)}\right] \qquad t<s \\ &= \frac{k^2}{\lambda^{r}}\left[\frac{\exp[(1-\lambda^{r'})(s-t)]}{(2-\lambda^2-\lambda^{r'})(2-\lambda^{r'})}-\frac{\exp(s-t)}{2(2-\lambda^{r'})}\right] \qquad t>s \end{aligned} \tag{95}$$

which becomes equal to Eq. 54 for $\lambda^r=\lambda^{r'}$. Evaluating the covariance at $t=s$ yields:

$$\Sigma=\frac{1}{2}\left\{\bar{\Sigma}_{\text{inp}}+\sum_{r=1}^{R}\frac{k}{2-\lambda^r}\left[\bar{\Sigma}_{\text{inp}}\boldsymbol{n}^r\boldsymbol{m}^{r\top}+\boldsymbol{m}^r\boldsymbol{n}^{r\top}\bar{\Sigma}_{\text{inp}}\right]+\sum_{r=1}^{R}\sum_{r'=1}^{R}\frac{k^2(4-\lambda^r-\lambda^{r'})}{(2-\lambda^r-\lambda^{r'})(2-\lambda^r)(2-\lambda^{r'})}\boldsymbol{m}^r\boldsymbol{n}^{r\top}\bar{\Sigma}_{\text{inp}}\,\boldsymbol{n}^{r'}\boldsymbol{m}^{r'\top}\right\} \tag{96}$$

from which Eq. 33 can be derived.

We focus on the case of high-dimensional stochastic inputs (Eq. 35). The eigenvalues of the covariance matrix can be computed following Eq. 25, with μ_i^{lr} being the eigenvalues of the low-rank component

$$\Sigma^{\text{lr}}=\sum_{r=1}^{R}\frac{k}{2-\lambda^r}\left[\boldsymbol{n}^r\boldsymbol{m}^{r\top}+\boldsymbol{m}^r\boldsymbol{n}^{r\top}\right]+\sum_{r=1}^{R}\sum_{r'=1}^{R}\frac{\rho_{n^rn^{r'}}\,k^2(4-\lambda^r-\lambda^{r'})}{(2-\lambda^r-\lambda^{r'})(2-\lambda^r)(2-\lambda^{r'})}\boldsymbol{m}^r\boldsymbol{m}^{r'\top}. \tag{97}$$

To compute the eigenvalues of this matrix, we first observe that we can re-write it as

$$\Sigma^{\text{lr}}=\sum_{r=1}^{R}\left[\boldsymbol{n}^r(\alpha^r\boldsymbol{m}^r)^\top+\boldsymbol{m}^r\left(\alpha^r\boldsymbol{n}^{r\top}+\sum_{r'=1}^{R}\rho_{n^rn^{r'}}\beta^{rr'}\boldsymbol{m}^{r'\top}\right)\right] \tag{98}$$

where we have defined

$$\alpha^r=\frac{k}{2-\lambda^r} \tag{99}$$

$$\beta^{rs}=\frac{k^2(4-\lambda^r-\lambda^s)}{(2-\lambda^r-\lambda^s)(2-\lambda^r)(2-\lambda^s)} \tag{100}$$

From this formulation, we can derive a $2R \times 2R$ reduced matrix [37], constructed by concatenating 2×2 blocks horizontally and vertically. These blocks take the values

$$B^{rs} = \begin{pmatrix} \delta^{rs}\rho_{m^r n^r}\alpha^r & \rho_{n^r n^s}\alpha^s + \rho_{m^r n^r}\rho_{n^r n^s}\beta^{sr} \\ \rho_{m^r m^s}\alpha^s & \delta^{rs}\rho_{m^r n^r}\alpha^r + \sum_{s'}\rho_{m^r m^{s'}}\rho_{n^s n^{s'}}\beta^{ss'} \end{pmatrix} \tag{101}$$

where δ^{rs} denotes the Kronecker delta. Computing the eigenvalues of this reduced matrix yields the eigenvalues of Σ^{lr}.

These eigenvalues take particularly simple values for a subclass of low-rank matrices in which $\rho_{m^r n^r}$ can take arbitrary values, ranging between -1 and $1/k$, while $\rho_{m^r m^s}$ and $\rho_{n^r n^s}$ vanish for all $r \neq s$ (and are equal to 1 otherwise). This case corresponds to different rank-one components operating in orthogonal subspaces. For these matrices, all entries of off-diagonal blocks B^{rs} for $r \neq s$ vanish, while the diagonal blocks take the form:

$$B^{rr} = \begin{pmatrix} \rho_{m^r n^r}\alpha^r & \alpha^s + \rho_{m^r n^r}\beta^{rr} \\ \alpha^r & \rho_{m^r n^r}\alpha^r + \beta^{rr} \end{pmatrix}. \tag{102}$$

for $r = 1, 2$. Thus, the eigenvalues of the total reduced matrix coincide with those of the 2×2 diagonal blocks. Using the identity

$$\beta^{rr} = \frac{k^2}{(1-\lambda^r)(2-\lambda^r)}, \tag{103}$$

one can show that each diagonal block is associated with a pair of eigenvalues $\mu^r_\pm$, obeying Eq. 27, where λ is replaced by λ^r and ρ_{mn} by $\rho_{m^r n^r}$. The eigenvectors corresponding to $\mu^r_\pm$ are linear combinations of the connectivity vectors $\boldsymbol{m}^r$ and $\boldsymbol{n}^r$, similar to Eq. 28.

In Fig. 5, we consider rank-2 connectivity matrices with two specific parametrizations. The first corresponds to the simplified case described above, where $\rho_{m^1 n^1}$ and $\rho_{m^2 n^2}$ take arbitrary values, while $\rho_{m^1 m^2}$ and $\rho_{n^1 n^2}$ are set to zero. The second parametrization, in contrast, involves setting $\rho_{m^1 n^1}$ and $\rho_{m^2 n^2}$ to zero while allowing $\rho_{m^1 m^2}$ and $\rho_{n^1 n^2}$ to take arbitrary values within the range $[-1, 1]$. Unlike the first case, this choice leads to a connectivity matrix with vanishing eigenvalues. Moreover, in this regime, the covariance eigenvectors are linear combinations of all four connectivity vectors: $\boldsymbol{m}^1$, $\boldsymbol{n}^1$, $\boldsymbol{m}^2$, and $\boldsymbol{n}^2$.

Simulations detail In Fig. 5, we simulated RNNs of size $N = 50$. To construct external inputs and recurrent connectivity, we first generated a set of four normalized orthogonal vectors $\{\boldsymbol{z}^i\}_{i=1,\dots,4}$ as detailed above. We then set: $\boldsymbol{m}^1 = \boldsymbol{z}^1$, $\boldsymbol{m}^2 = \boldsymbol{z}^2$, and

$$\begin{aligned} \boldsymbol{n}^1 &= \rho_{m^1 n^1}\boldsymbol{z}^1 + \sqrt{1-\rho^2_{m^1 n^1}}\boldsymbol{z}^3 \\ \boldsymbol{n}^2 &= \rho_{m^2 n^2}\boldsymbol{z}^2 + \sqrt{1-\rho^2_{m^2 n^2}}\boldsymbol{z}^4 \end{aligned} \tag{104}$$

We fix $\rho_{m^1 n^1} = 0.1$, $\rho_{m^2 n^2} = -0.3$.

8 Excitatory-inhibitory network

In the case of one-dimensional inputs, we parametrize the input vector as $\boldsymbol{u} = \big(u_{\mathrm{E}}, u_{\mathrm{I}}\big) = (\cos(\theta), \sin(\theta))$, where θ varies between 0 and 2π. Special cases of interest are: $\theta = 0$ (input only to E), $\theta = \pi/4$ (equal inputs to E and I), $\theta = \pi/2$ (input only to I), $\theta = 3\pi/4$ (opposite-sign, otherwise equal input to E and I). This results in the following overlaps:

$$\begin{aligned} \rho_{mn} &= \frac{1-g}{\sqrt{2(g^2+1)}} \\ \rho_{nu} &= \frac{1}{\sqrt{g^2+1}}[\cos(\theta) - g\sin(\theta)] \\ \rho_{mu} &= \frac{1}{\sqrt{2}}[\cos(\theta) + \sin(\theta)] \end{aligned} \tag{105}$$

while $\rho_{uu} = \rho_{mm} = 1$.