

# Geometry of population activity in spiking networks with low-rank structure

Ljubica Cimesa, Lazar Ciric, Srdjan Ostojic

Laboratoire de Neurosciences Cognitives et Computationnelles, INSERM U960, Ecole Normale Supérieure - PSL University, 75005 Paris, France

## Abstract

Recurrent network models are instrumental in investigating how behaviorally-relevant computations emerge from collective neural dynamics. A recently developed class of models based on low-rank connectivity provides an analytically tractable framework for understanding of how connectivity structure determines the geometry of low-dimensional dynamics and the ensuing computations. Such models however lack some fundamental biological constraints, and in particular represent individual neurons in terms of abstract units that communicate through continuous firing rates rather than discrete action potentials. Here we examine how far the theoretical insights obtained from low-rank rate networks transfer to more biologically plausible networks of spiking neurons. Adding a low-rank structure on top of random excitatory-inhibitory connectivity, we systematically compare the geometry of activity in networks of integrate-and-fire neurons to rate networks with statistically equivalent low-rank connectivity. We show that the mean-field predictions of rate networks allow us to identify low-dimensional dynamics at constant population-average activity in spiking networks, as well as novel non-linear regimes of activity such as out-of-phase oscillations and slow manifolds. We finally exploit these results to directly build spiking networks that perform nonlinear computations.

## Author summary

Behaviorally relevant information processing is believed to emerge from interactions among neurons forming networks in the brain, and computational modeling is an important approach for understanding this process. Models of neuronal networks have been developed at different levels of detail, with typically a trade off between analytic tractability and biological realism. The relation between network connectivity, dynamics and computations is best understood in abstract models where individual neurons are represented as simplified units with continuous firing activity. Here we examine how far the results obtained in a specific, analytically-tractable class of rate models extend to more biologically realistic spiking networks where neurons interact through discrete action potentials. Our results show that abstract rate models provide accurate predictions for the collective dynamics and the resulting computations in more biologically faithful spiking networks.

## Introduction

Recurrent network models are an essential tool for understanding how the collective dynamics of activity in the brain give rise to computations that underlie behavior.

1  
2  
3

Network models at different levels of biological detail are typically used to describe different phenomena [Herz et al., 2006, Gerstner and Naud, 2009], but integrating findings across scales of abstraction remains challenging. Networks of excitatory and inhibitory spiking neurons [Brunel, 2000, Gerstner et al., 2014] are a popular class of models which incorporate the key biological fact that neurons interact through discrete action potentials, a.k.a. spikes. Balanced excitatory-inhibitory spiking networks in particular naturally lead to asynchronous irregular activity [Amit and Brunel, 1997, van Vreeswijk and Sompolinsky, 1996, Troyer and Miller, 1997, Renart et al., 2010, Boerlin et al., 2013a, Hennequin et al., 2017, Rosenbaum and Doiron, 2014, Sanzeni et al., 2022] that captures some of the main features of the spontaneous neural firing in vivo [Shadlen and Newsome, 2001, Destexhe et al., 2001, Shu et al., 2003, Haider et al., 2006, London et al., 2010, Sanzeni et al., 2020, Ahmadian and Miller, 2021]. Beyond spontaneous activity, how rich behavioral computations are implemented in spiking networks has been an open issue [Abbott et al., 2016, Ingrosso and Abbott, 2019, Tavanaei et al., 2019]. This question has so far been more easily tackled in more abstract models such as recurrent neural networks (RNNs) [Sussillo, 2014, Barak, 2017, Yang and Wang, 2020], where individual units are represented in terms of continuous firing rates rather than discrete spikes. A particularly fruitful approach has been to interpret the emerging computations in terms of the geometry of dynamics in the state space of joint activity of all neurons [Sussillo and Barak, 2013, Mante et al., 2013, Vyas et al., 2020, Chung and Abbott, 2021], as commonly done with experimental data [Churchland and Shenoy, 2007, Buonomano and Maass, 2009, Cunningham and Yu, 2014, Gallego et al., 2017, Saxena and Cunningham, 2019, Jazayeri and Ostojic, 2021]. In particular, in a large class of rate networks in which the connectivity contains a low-rank structure [Hopfield, 1982, Eliasmith and Anderson, 2004, Sussillo and Abbott, 2009, Boerlin et al., 2013b, Ahmadian et al., 2015, Pereira and Brunel, 2018, Landau and Sompolinsky, 2018, Beiran et al., 2021b, Landau and Sompolinsky, 2021, Schuessler et al., 2020b, Kadmon et al., 2020, Logiaco et al., 2021, Valente et al., 2022a], the geometry of activity and the resulting computations can be analytically predicted from the structure of connectivity [Mastrogiuseppe and Ostojic, 2018, Schuessler et al., 2020a, Beiran et al., 2021a, Dubreuil et al., 2022]. A comparable mechanistic picture has so far been missing in spiking networks.

A key question is therefore to which extent mechanistic insights from RNNs extend to more biologically plausible spiking models. In this regard, a central underlying issue is exactly how spiking models are related to abstract rate networks [Schaffer et al., 2013]. One common approach has been to interpret each unit in a rate network as an average over a sub-population of spiking neurons [Wilson and Cowan, 1972, Amit and Brunel, 1997, Shriki et al., 2003, Wong, 2006, Litwin-Kumar and Doiron, 2012, Hennequin et al., 2014, Baker et al., 2020, Timón et al., 2022, Zenke et al., 2015]. A possible alternative is instead to approximate each individual spiking neuron by a Poisson rate unit [Ostojic and Brunel, 2011], and therefore hypothesize that a full spiking network can be directly mapped onto a rate network with identical connectivity [Tetzlaff et al., 2012, Ostojic, 2014, Harish and Hansel, 2015, Nicola and Clopath, 2017, Kim and Chow, 2018, Kim et al., 2019]. If this hypothesis is correct, the analytic predictions for the geometry of activity in rate networks should directly translate to spiking networks with a low-rank structure in connectivity. This implies that the geometry of activity and range of dynamics in spiking networks may be much broader than apparent on the level of population-averaged spike trains.

To test this hypothesis, we consider a classical spiking network model consisting of excitatory-inhibitory integrate-and-fire neurons [Brunel, 2000], and add low-rank structure on top of the underlying random, sparse connectivity. Varying the statistics of the low-rank structure, we systematically compare the geometry of activity and

dynamical regimes in the spiking model with predictions of networks of rate units with statistically identical low-rank connectivity. We find that rate networks predict well the structure of activity in the spiking network even outside the asynchronous irregular regime, as long as spike-times are averaged over timescales longer than the synaptic and membrane time constants to estimate instantaneous firing rates. In particular, the predictions of the rate model allow us to identify low-dimensional dynamics at constant population-average activity in spiking networks, as well as novel non-linear regimes of activity such as out-of-phase oscillations and slow manifolds. We finally show that these results can be exploited to directly build spiking networks that perform nonlinear computations based on principles identified in rate networks.

## Results

### Geometry of the activity in the state space

We consider recurrent networks of  $N$  neurons, modeled either as rate units or leaky integrate-and-fire (LIF) neurons (Fig. 1 A, see Methods for details). We quantify the activity of each neuron  $i$  in terms of its time-dependent firing rate  $r_i(t)$ . In rate networks, each unit is described by the dynamics of its activation  $x_i(t)$ , an abstract variable usually interpreted as the total input or membrane potential [Miller and Fumarola, 2012], and the firing rate is obtained as  $r_i(t) = \phi(x_i(t))$ , where  $\phi(x)$  is the single-unit current-to-rate function that we here choose to be a positive sigmoid. In the LIF network, single-unit firing rates are instead estimated from a running average over spike times, computed using an exponential filter with timescale  $\tau_f$  (Fig. 1 C, Methods Eq. (17)).

Following a common approach for analyzing neural data [Harvey et al., 2012, Cunningham and Yu, 2014], we represent the collective activity at any time point as a vector  $\mathbf{r}(t) = \{r_i\}_{i=1\dots N}$  in the *activity state space* where each axis corresponds to the firing rate  $r_i$  of one neuron (Fig. 1 D). We then examine the geometry of the dynamical trajectories by projecting at each time point the activity vector  $\mathbf{r}(t)$  along different directions in that space. Each direction is specified by a vector  $\mathbf{w} = \{w_i\}_{i=1\dots N}$  in state space, so that projecting onto it is equivalent to assigning to every neuron a weight  $w_i$  and computing a weighted average of the activity (Methods Eq. (19)).

Analyses of experimental data and works on rate networks commonly examine how collective activity changes along arbitrary directions in state space, where the weight  $w_i$  of each neuron is chosen independently. In contrast, studies of spiking networks have often focused on firing rates averaged over the whole network or over specific sub-populations [Amit and Brunel, 1997, Shriki et al., 2003, Wong, 2006, Litwin-Kumar and Doiron, 2012, Hennequin et al., 2014, Baker et al., 2020, Timón et al., 2022]. Taking a population average over the whole network is equivalent to projecting activity along the direction  $(\mathbf{1}, \mathbf{1}, \dots, \mathbf{1})$ , which we call the *global axis* [Kobak et al., 2019]. Similarly, vectors with unit entries on a specific subset of neurons and zeros elsewhere define directions in state space that represent firing rates averaged over specific sub-populations. The goal of the present study is to instead examine how inputs and connectivity in spiking networks shape activity along arbitrary directions of state space, and in particular directions orthogonal to the global axis which correspond to changes in collective activity that modify the pattern of activity but keep the population-averaged activity constant. To this end, we compare the geometry of activity in spiking networks with rate networks based on an identical structure in the connectivity.

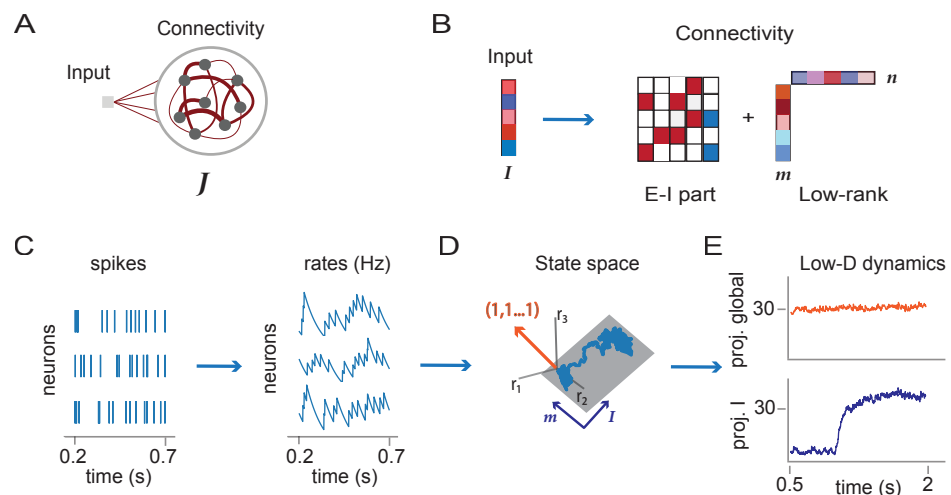
We specifically focus on networks for which the connectivity matrix contains a

low-rank structure parametrized as

$$P_{ij} = \frac{1}{N} \sum_{r=1}^R m_i^{(r)} n_j^{(r)} \quad (1)$$

where  $\mathbf{m}^{(r)} = \{m_i^{(r)}\}_{i=1\dots N}$  and  $\mathbf{n}^{(r)} = \{n_i^{(r)}\}_{i=1\dots N}$  for  $r = 1, \dots, R$  are *connectivity vectors* (Fig. 1 B). Rate networks with low-rank connectivity are analytically tractable, in the sense that the geometry of dynamics in state space can be directly predicted from the arrangement of connectivity vectors and inputs, as summarized below.

Our key hypothesis is that, for a spiking network in the asynchronous irregular state, each neuron can be directly mapped onto a unit in a rate network with statistically identical connectivity. To test this hypothesis, we start from an LIF network with sparse excitatory-inhibitory connectivity in the inhibition-dominated regime that leads to asynchronous irregular activity [Brunel, 2000] (Methods), and add to this random background an arbitrary low-rank connectivity structure of the type introduced in Eq. (1). We then compare the geometry of the resulting spiking activity to the predictions of a rate model with statistically identical connectivity.



**Fig 1. Low-rank connectivity and state space dynamics.** A: Illustration of recurrent neural network architecture, consisting of inputs and recurrent connectivity. B: Representation of inputs and connectivity in terms of vectors. The input weights form an input vector  $I$ . In spiking networks, the recurrent connectivity is composed of a sparse excitatory-inhibitory part (zero entries in white, excitatory connections in red, inhibitory in blue) and a low-rank structure defined by pairs of connectivity vectors  $m$  and  $n$ . The illustration shows a unit-rank example ( $R = 1$ ). C: Left: spike times of three neurons in the spiking network. Right: dynamics of instantaneous firing rates computed from spikes using an exponential filter with timescale  $\tau_f = 100\text{ms}$ . D: Three-dimensional illustration of low-dimensional dynamics in the activity state space where each axis represents the firing rate of one neuron. In a unit-rank network, the activity is expected to be confined to a two-dimensional plane spanned by the vectors,  $m$  and  $I$ . We refer to the direction  $(1, 1, \dots, 1)$  as the global axis (orange). E: Projections of activity on two axes: (top) global axis corresponding to the population-averaged firing rate; (bottom) axis defined by the input vector  $I$ .

## Geometry of responses to external inputs

We start by examining the geometry of transient dynamics in response to external inputs. We first summarize the predictions of low-rank rate models developed in previous studies [Mastrogiuseppe and Ostojic, 2018]. We then examine whether these predictions hold in networks of integrate-and-fire models with a statistically identical low-rank connectivity structure.

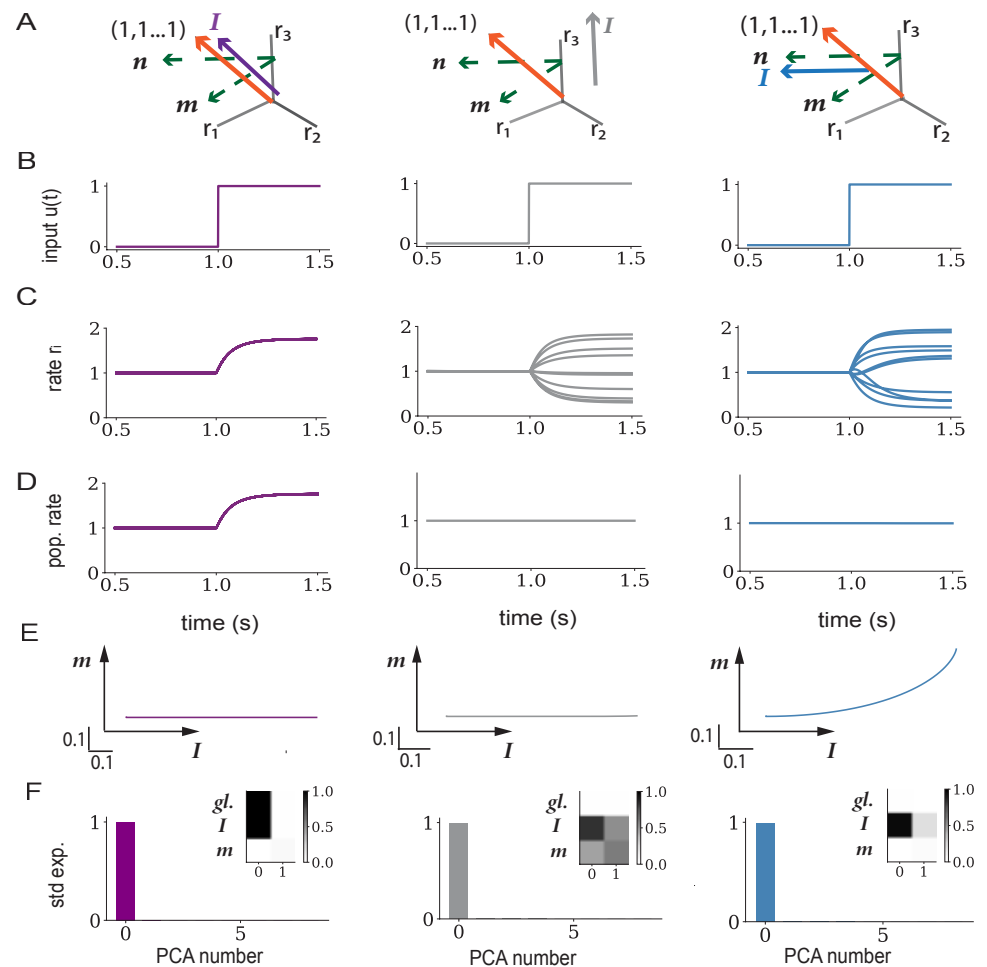
Each neuron  $i$  receives a step input  $u(t)$  (Fig. 2 B) multiplied by a weight  $I_i$ . The set of feed-forward weights  $I_i$  over neurons form an *input vector*  $\mathbf{I} = \{I_i\}_{i=1\dots N}$ . This input vector, as well as the connectivity vectors  $\mathbf{m}^{(r)}$  and  $\mathbf{n}^{(r)}$  introduced in Eq. 1, each define a specific direction in state space (Fig. 2 A). Previous work [Mastrogiuseppe and Ostojic, 2018, Beiran et al., 2021a] has shown that in rate networks with low-rank connectivity, the dynamics of the activations  $\mathbf{x}(t) = \{x_i(t)\}$  in response to an input are confined to a subspace of state space spanned by the input vector  $\mathbf{I}$  and connectivity vectors  $\mathbf{m}^{(r)}$  for  $r = 1 \dots R$ . Focusing on a unit rank network ( $R = 1$ ), this implies that the activation  $x_i(t)$  of unit  $i$  in the rate network can be expressed as

$$x_i(t) = \kappa(t) m_i + v(t) I_i \quad (2)$$

where  $\kappa(t)$  and  $v(t)$  are two scalar variables (Methods). The variable  $v(t)$  represents feed-forward activity propagated along the direction  $\mathbf{I}$ , while  $\kappa(t)$  quantifies activity that recurrent dynamics generate along the direction  $\mathbf{m}$ . An input along the direction  $\mathbf{I}$  will generate a non-zero recurrent response  $\kappa(t)$  only if  $\mathbf{I}$  has a non-zero overlap with the vector  $\mathbf{n}$ , i.e. if the scalar product  $\mathbf{n}^T \mathbf{I}$  is non-zero [Mastrogiuseppe and Ostojic, 2018].

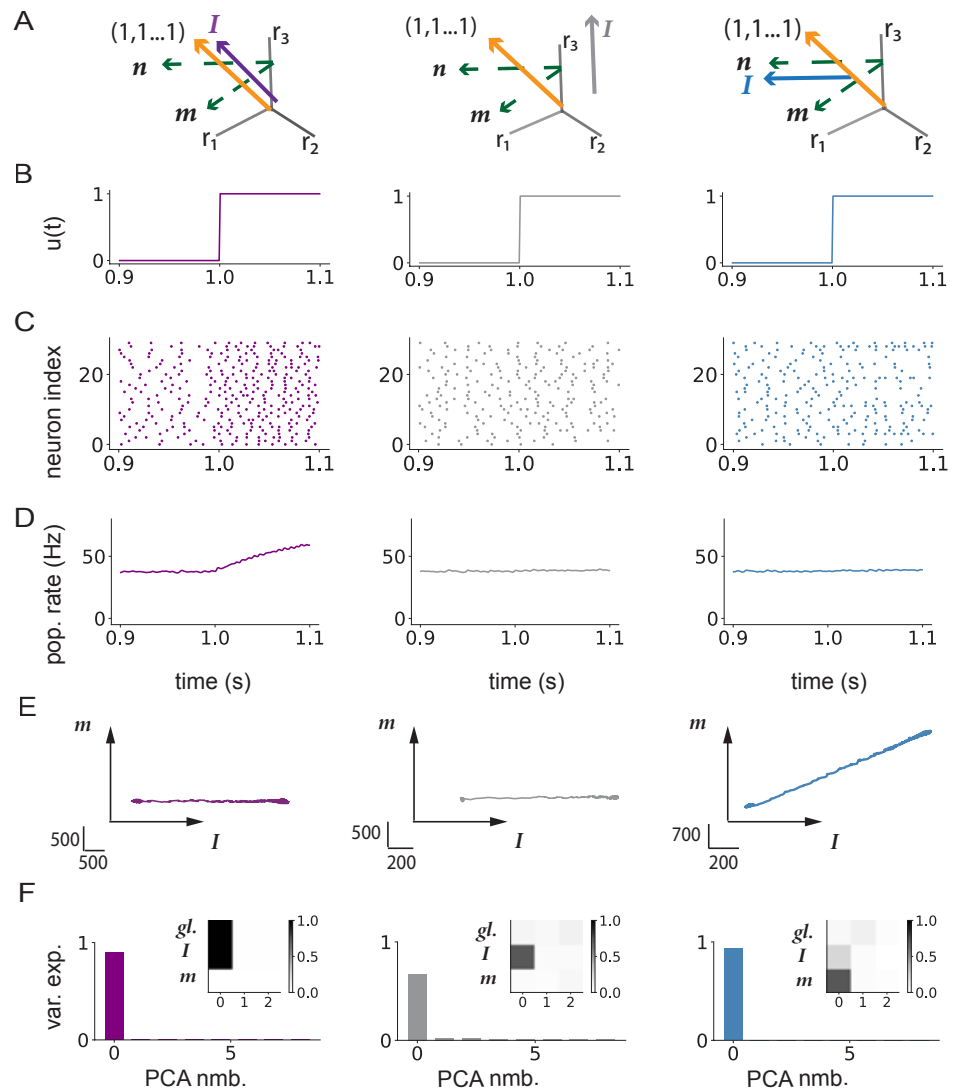
Additional analyses show that the low-dimensional geometry described by Eq. (2) at the level of activations  $\mathbf{x}(t)$  is largely preserved when applying the non-linear function  $\phi(x)$  to obtain rates (Methods). More specifically, the projection of the firing rates  $\mathbf{r}(t) = \{r_i(t)\}$  on an arbitrary axis  $\mathbf{w}$  is determined by the projection of  $\mathbf{x}(t)$  on that same axis if the network is in the linear regime, or if the entries of  $w_i$  of  $\mathbf{w}$  follow a Gaussian distribution (Eq. (33)), which we assume throughout this study. In low-rank rate networks, the dynamics of  $\mathbf{r}(t)$  in response to an input therefore dominantly lie in the subspace spanned by the input and connectivity vectors  $\mathbf{I}$  and  $\mathbf{m}$  (Fig. 2 E), with the non-linearity generating a potential additional component along the global axis (Methods Eq. (33)). These theoretical predictions were confirmed by a PCA analysis of simulated trajectories of firing rates  $\mathbf{r}(t)$  (Fig. 2 F).

Since the population-averaged firing rate is obtained by projecting  $\mathbf{r}(t)$  on the global axis  $(\mathbf{1}, \mathbf{1}, \dots, \mathbf{1})$ , the analysis of low-rank rate models predicts that a given input induces a strong change of population-averaged firing rates when the mean of inputs weights over neurons,  $\langle I \rangle = \frac{1}{N} \sum_{i=1}^N I_i$ , is non-zero (Fig. 2 D, left panel), or if both the average  $\langle m \rangle$  of elements of  $\mathbf{m}$  and the overlap  $\mathbf{n}^T \mathbf{I}$  are non-zero. Conversely, if  $\langle I \rangle = 0$  and  $\langle m \rangle = 0$ , inputs evoke changes in single-unit firing rate that essentially average-out on the population-average level (Fig. 2 D, middle and right panel), but instead explore the plane  $\mathbf{I} - \mathbf{m}$  that is orthogonal to the global axis in state space (Fig. 2 E, middle and right panel).



**Fig 2. Low-dimensional dynamics generated by external inputs in rate networks with low-rank connectivity.** A: Illustration of the geometry in the activity state-space. The input vector  $I$ , the connectivity vectors  $m$  and  $n$  (green), and the global axis  $(1, 1, \dots, 1)$  (orange) define a set of directions and a subspace within which the low-dimensional dynamics unfold. The overlaps of the vectors  $I$  and  $m$  with the global axis predict whether inputs give rise to a change in the population-averaged activity. The overlap of  $I$  with  $n$  instead determines whether an input engages recurrent activity along the direction  $m$ . The three columns display three different arrangements of the input vector (depicted in a different color in each column). Left:  $I$  aligned with the global axis; middle:  $I$  orthogonal to both  $n$  and the global axis; right:  $I$  aligned with  $n$ , but  $I$  and  $m$  orthogonal to the global axis. B: The input vector is multiplied by a scalar  $u(t)$  which is a step function from  $t = 1$ s. C: Individual firing rates  $r_i(t)$  for a subset of 10 neurons in each network. D: Population firing rate, averaged over all neurons in the network. E: Projections of the firing rate trajectory  $r(t)$  onto the  $(I, m)$  plane. F: PCA analysis of the firing rate dynamics  $r(t)$ . Variance explained by each of the first 8 PCs. Inserts: Projections of the first two PCs onto the global axis, the input vector  $I$  and the connectivity vector  $m$ . The connectivity vectors  $m$  and  $n$  have a zero mean and unit standard deviation, and are orthogonal to each other. Vectors  $n$  and  $I$  are orthogonal except in blue where the overlap is  $\sigma_{nI} = 1$ . Vectors  $I$  in gray and blue have a zero mean and unit standard deviation, while vector  $I$  in purple is along the global axis. Network parameters are given in Table 3.

We compared these predictions of low-rank rate models to the geometry of activity in spiking networks where a low-rank structure was added on top of sparse excitatory-inhibitory connectivity (Fig. 3). Starting from a network in the inhibition-dominated asynchronous irregular state [Brunel, 2000], as expected inputs evoked a change in population-averaged firing rates if the mean of the input vector  $\langle I \rangle$  was non-zero (Fig. 3 D, left panel). Input vectors of zero mean instead elicited patterns of responses across neurons that did not modify the population-averaged firing rate (Fig. 3 D, middle and right panel), but explored directions in state space orthogonal to the global axis. These directions were accurately predicted by low-rank rate models: input vectors  $\mathbf{I}$  orthogonal to the vector  $\mathbf{n}$  led to responses only along the direction  $\mathbf{I}$  (Fig. 3 E, left and middle panel), while inputs that overlapped with  $\mathbf{n}$  led to responses in the  $\mathbf{I} - \mathbf{m}$  plane (Fig. 3 E, right panel). A PCA analysis confirmed that these low-dimensional projections explained the dominant part of variance in the full trajectories (Fig. 3 F). Altogether, the predictions of low-rank rate models were therefore fully borne out when treating each individual spiking neuron as a rate unit.



**Fig 3. Low-dimensional dynamics generated by external inputs in spiking networks with low-rank structure.** A: Illustration of the geometry of input (varying color) and connectivity vectors (green) with respect to the global axis (orange). Left: input vector  $I$  along the global axis; middle: input vector  $I$  orthogonal to  $n$ ; right: input vector  $I$  along the vector  $n$ . B: The input vector is multiplied by a scalar  $u(t)$  which is a step function from  $t = 1$ s. C: Raster plot showing action potentials for a subset of 30 neurons in each network. D: Population firing rate obtained by averaging instantaneous firing rates of all neurons. E: Projections of the firing rate trajectory  $r(t)$  onto the  $(I, m)$  plane. F: PCA analysis of firing rate dynamics  $r(t)$ . Variance explained by each of the first 8 PCs. Inserts: Projections of the first 3 PCs onto the global axis (first row), and vectors  $I$  and  $m$ . The connectivity vectors  $m$  and  $n$  have a zero mean and unit standard deviation, and are orthogonal to each other. Vectors  $n$  and  $I$  are orthogonal except in blue where the overlap is  $n^T I / N = 0.4 \text{mV}^2$ . Vectors  $I$  in gray and blue have a zero mean and unit standard deviation, while vector  $I$  in purple is along the global axis. All analyses were performed on instantaneous firing rates computed using a filter timescale of  $\tau_f = 100 \text{ms}$ . Network parameters are given in Table 4

## Responses in spiking networks outside of the irregular asynchronous regime

Our initial hypothesis was that low-rank rate networks predict well the geometry of responses of spiking networks in the asynchronous irregular regime, where individual neurons can be approximated as independent Poisson processes [Brunel, 2000]. We next asked to which extent the predictions hold outside of this regime, when the spiking activity is either not asynchronous, i.e. exhibits some degree of synchronization and oscillations [Brunel, 2000], or is regular rather than irregular. Following Brunel 2000, we set the network to operate in a specific regime by varying the strength of the inhibition  $g$  in the random part of the connectivity, the external input  $\mu_{ext}$  and the synaptic delay  $\tau_{del}$  (Methods). We then examined how much the underlying regime influences the low-dimensional dynamics in response to external inputs in networks with a unit-rank structure. For this, we repeated the PCA analysis in spiking networks with zero-mean input and connectivity vectors, and  $\mathbf{I}$  overlapping with  $\mathbf{n}$  as in the right column of Fig. 2.

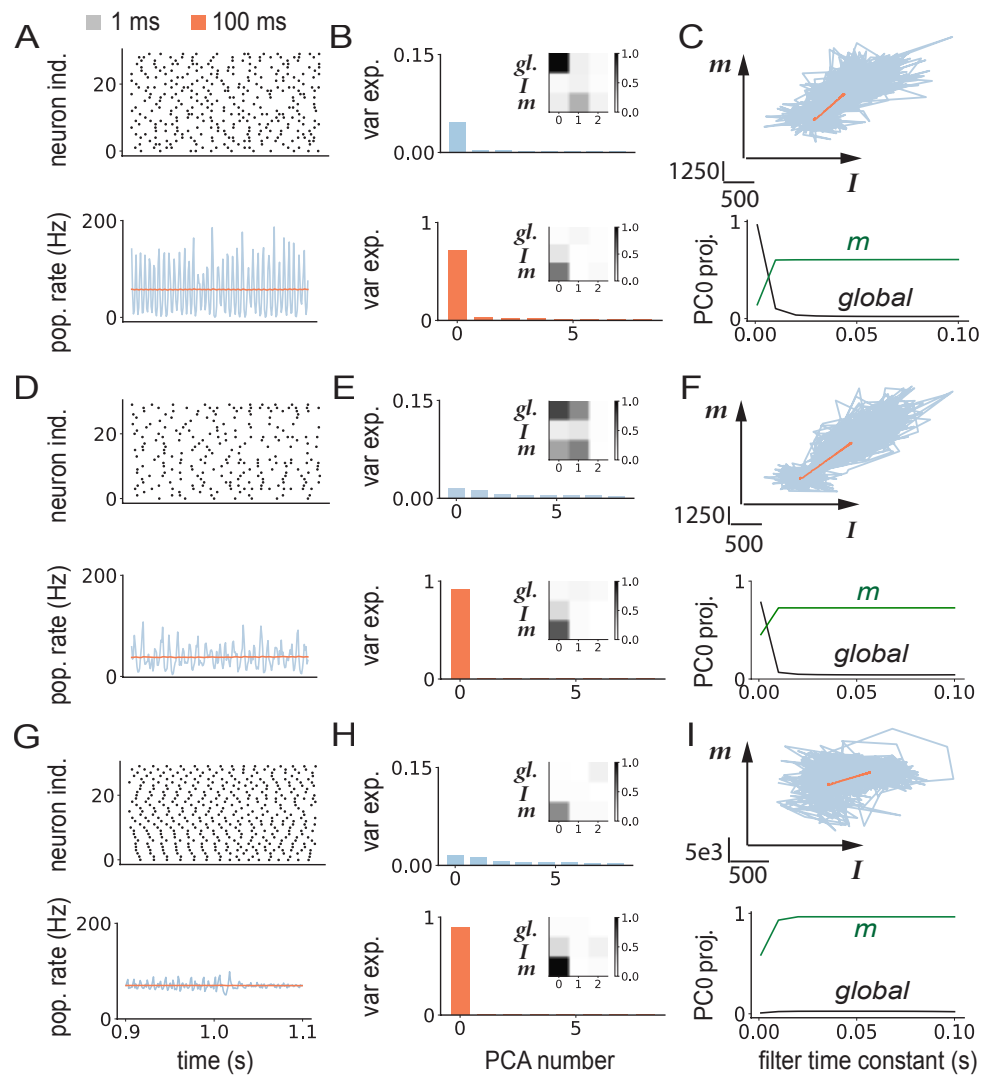
We first considered a network of integrate-and-fire neurons that operates in the synchronous irregular (SI) regime [Brunel, 2000] in which individual neurons fire irregularly (Fig. 4 A, top), but are sparsely synchronised, leading to oscillations in the population rate (Fig. 4 A, bottom). The frequency of these oscillations is set by the synaptic delay  $\tau_{del}$ , and is therefore high for physiologically realistic values of  $\tau_{del}$  [Brunel, 2000]. These oscillations can therefore only be observed in the firing rates when the filter timescale  $\tau_f$  used for averaging over spikes is comparable to the delays, i.e. of the order of milliseconds (Fig. 4 A, blue). Longer filter timescales instead totally average-out the oscillatory dynamics (Fig. 4 A, orange). We therefore found that the dimensionality and geometry of the responses in state space depend on the filter timescale used to determine single-unit firing rates (Fig. 4 B). Performing a PCA analysis on firing rate trajectories  $\mathbf{r}(t)$  obtained with a filter timescale of 1ms indicated that the activity was high-dimensional. Indeed, the explained variance was distributed along many principal components (Fig. 4 B, top), with the first PC capturing population-level oscillations along the global axis (insert in Fig. 4 B, top panel), while strong fluctuations were present in other directions (Fig. 4 C). In contrast, for a filter timescale of 100ms the first PC explained a much larger fraction of variance (Fig. 4 B, bottom), and corresponds instead to activity along a combination between the connectivity vector  $\mathbf{m}$  and the input vector  $\mathbf{I}$  (insert in Fig. 4 B, bottom), as predicted by the rate model (insert in Fig. 2 F, blue). In between these two extremes, progressively increasing the filter timescale (Fig. 4 C, bottom) shows that for timescales below 10ms, the geometry of activity is dominated by fluctuations along the global axis, while for longer timescales the dynamics are lower-dimensional and reside dominantly in the  $(\mathbf{m}, \mathbf{I})$  plane as expected from the rate network (Fig. 2 E and F, blue).

Given the strong influence of the filter timescale on the results, we repeated the same analysis in the asynchronous irregular (AI) regime, which in Fig. 3 was investigated only using a long timescale of 100ms. We found that the results of the PCA were similar to the SI regime: fluctuations along the global axis dominated at timescales below 10ms, and low-dimensional dynamics predicted by the rate model emerged at longer timescales (Fig. 4, D-F). The main difference between the AI and SI regimes is that the global fluctuations at short timescales are weaker in the AI regime (with an amplitude that decays as the network size is increased), and do not show the periodic structure found in the SI regime (Fig. 4 D).

We then examined the role of irregular activity, by turning to networks in which the connectivity consisted only of a low-rank structure without the random E-I part. In such networks, individual neurons fired almost periodically, in contrast to Poisson-like activity in the AI regime. The action potentials of the different neurons were however

highly asynchronous (Fig. 4 G top), and the fluctuations in the population activity were 223  
weak even for filter timescales of 1ms (Fig. 4 G bottom). Similarly to SI and AI regime, 224  
the dynamics in this network became low-dimensional for long filter timescales (Fig. 4 225  
H and I), but the projection along  $m$  was higher for all filter timescales, and saturated 226  
above 10ms (Fig. 4 I bottom). 227

In summary, our analyses indicate that the predictions of the rate networks for the 228  
geometry of responses hold independently of the activity regime in the spiking network 229  
if the single neuron firing rates are determined by averaging action potentials on a 230  
timescale longer than the synaptic delays. At shorter timescales, the activity is 231  
dominated by spiking synchronization that leads to prominent fluctuations along the 232  
global axis which corresponds to the population-averaged firing rate. 233



**Fig 4. Influence of firing regime and filter timescales on low-dimensional dynamics in spiking networks.** A-C: Synchronous irregular (SI) regime. A: Top: raster plot showing action potentials for a subset of 30 neurons in the network. Bottom: population-averaged firing rate computed using filter time constants of 1ms (blue) and 100ms (orange). B: PCA analysis of trajectories of instantaneous firing rates computed from spike trains using two different filter time constants (top: 1ms, bottom: 100ms). Main panels: variance explained by each of the first 8 PCs; inserts: projections of the first 3 principal components on the global vector,  $I$  and  $m$ . C: Top: projections of the firing rate trajectories on the plane defined by vectors  $m$  and  $I$ . Bottom: Projection of the first principal component on the global axis (black) and on the vector  $m$  (green) as a function of the filter time constant. D-F: Similar to A-C, for the network in asynchronous irregular regime shown in the right column of Fig. 3. G-I Similar to A-C, for a network without the background E-I connectivity. The firing regime was controlled by varying the inhibition strength in the random EI connectivity, the baseline input and synaptic delays (see Table 4). The unit-rank connectivity structure was identical to Fig. 3 right column, with zero-mean input and connectivity vectors. At time  $t = 1s$ , a step input was given along the input vector  $I$  that was aligned with  $n$ . Network parameters are given in Table 4.

## Nonlinear autonomous activity in networks with unit-rank structure

In previous sections we studied the geometry of dynamics in response to external inputs. We next turned to autonomous dynamics generated by the recurrent connectivity in the absence of inputs. As before, our goal was to determine whether the dynamics in a spiking network with low-rank connectivity are well predicted by a rate network with an analogous low-rank structure in the connectivity. We first summarize the results for rate networks developed in earlier studies, and then compare dynamics in spiking networks with these predictions.

In a rate network with unit rank structure, in absence of time-varying external inputs, the low-dimensional dynamics in Eq. (2) are described only by the recurrent variable  $\kappa(t)$ . The temporal evolution of  $\kappa(t)$  obeys (Methods):

$$\frac{d\kappa}{dt} = -\kappa + \frac{1}{N} \sum_{i=1}^N \phi(x_i) n_i. \quad (3)$$

The steady state value of  $\kappa$  is therefore given by:

$$\kappa = \sum_{i=1}^N n_i \phi(x_i) / N. \quad (4)$$

Assuming as previously a Gaussian distribution of the entries  $(m_i, n_i)$  of the connectivity vectors, and using a mean-field analysis in the large  $N$  limit, Eq. 4 can be further expressed as (Methods Eq. (39))

$$\kappa = \langle n \rangle \langle \phi(\mu, \Delta) \rangle + \sigma_{mn} \kappa \langle \phi'(\mu, \Delta) \rangle \quad (5)$$

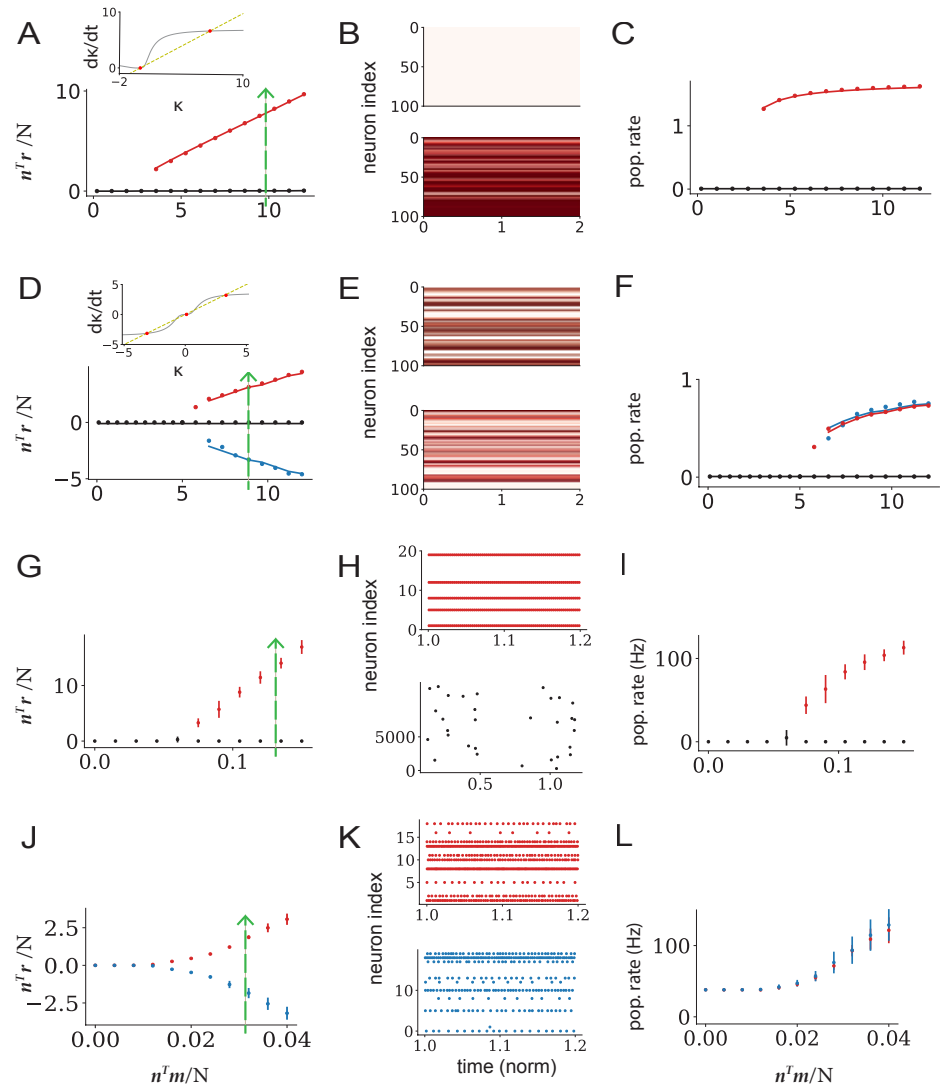
where  $\langle m \rangle$ ,  $\langle n \rangle$  and  $\sigma_{mn}$  are the mean values and covariance of connectivity vectors  $\mathbf{m}$  and  $\mathbf{n}$ , while  $\langle \phi(\mu, \Delta) \rangle$  and  $\langle \phi'(\mu, \Delta) \rangle$  are the mean firing rate and mean gain obtained by averaging the transfer function and its derivative over a Gaussian distribution of mean  $\mu = \langle m \rangle \kappa$  and variance  $\Delta = \sigma_m^2 \kappa^2$  (Methods Eq. (39), Appendix Eq. (50)). Eq. (5) provides a self-consistent equation for the steady state value of  $\kappa$ , which enters implicitly in the r.h.s. through  $\mu$  and  $\Delta$ . The two terms in the r.h.s can therefore be interpreted as two different sources of feedback, a first one controlled by the mean values  $\langle m \rangle$ ,  $\langle n \rangle$ , and a second one controlled by the covariance  $\sigma_{mn}$  between  $\mathbf{m}$  and  $\mathbf{n}$ . Previous works analyzed the bifurcations in networks with a symmetric transfer function [Mastrogiuseppe and Ostojic, 2018], or positive transfer function with non-zero  $\langle m \rangle$  and  $\langle n \rangle$  [Mastrogiuseppe and Ostojic, 2018, Shao and Ostojic, 2022]. The respective contributions of the two sources of feedback in networks with a positive transfer function have so far not been examined.

To extend previous studies, we therefore analyzed the bifurcations obtained by separately increasing each source of feedback in Eq. (5) in networks with a positive transfer function. For  $\sigma_{mn} = 0$ , the feedback is generated only by the first term, and we controlled it by changing  $\langle n \rangle$  while keeping  $\langle m \rangle$  fixed. As the non-linearity in that term is given by  $\langle \phi \rangle(\kappa)$ , which is a positive sigmoid (Fig. 5 A, insert), increasing  $\langle n \rangle$  beyond a critical value leads to a bifurcation to two *asymmetric* states with low and high values of  $\kappa$  (Fig. 5 A). Since the mean  $\langle m \rangle$  of the vector  $\mathbf{m}$  is non-zero, these two values of  $\kappa$  correspond to two states with a low and a high population-averaged firing rate (Fig. 5 C), as usually found when positive feedback is high [Lerchner and Latham, 2015, Mastrogiuseppe and Ostojic, 2018, Shao and Ostojic, 2022].

In contrast, when  $\langle m \rangle = \langle n \rangle = 0$  and  $\sigma_{mn} \neq 0$ , the recurrent feedback is generated only by the second term in Eq. (5), for which the non-linearity is given by  $\kappa \langle \phi'(0, \Delta) \rangle$ .

Independently of the precise form of  $\phi$ ,  $\kappa\langle\phi'(0, \Delta)\rangle$  as function of  $\kappa$  is in general symmetric around zero (Fig. 5 D, insert). In consequence, increasing  $\sigma_{mn}$  beyond a critical value leads to the emergence of two *symmetric* fixed points for  $\kappa$  (Fig. 5 D), which correspond to two activity states with different patterns of activity (Fig. 5 E), but identical population-averaged firing rates (Fig. 5 F).

In summary, a mean-field analysis of rate networks with unit-rank connectivity predicts two qualitatively different types of bifurcations and bistable states depending on whether the connectivity vectors  $\mathbf{m}$  and  $\mathbf{n}$  have zero or non-zero mean. We therefore examined whether these two types of bifurcations appeared when increasing the overlap between  $\mathbf{n}$  and  $\mathbf{m}$  in spiking networks with unit-rank connectivity added on top of a random EI background. Increasing  $\langle n \rangle$  with non-zero  $\langle m \rangle$  and zero  $\sigma_{mn}$  is in fact equivalent to increasing the mean excitation in the underlying EI connectivity [Shao and Ostojic, 2022]. In agreement with previous studies [Brunel, 2000], we found that this could lead to the emergence of an asymmetric bistability between a low and a high average activity state (Fig. 5 G-I). Increasing  $\sigma_{mn}$  in networks with  $\langle m \rangle = \langle n \rangle = 0$  instead gives rise to a bifurcation to two symmetric activity patterns with equal population-averaged firing rates (Fig. 5 J-L). The predictions of the mean-field analysis in low-rank rate networks were therefore directly verified in spiking networks, and allowed us to identify a novel bifurcation to two symmetric states of activity.



**Fig 5. Nonlinear autonomous activity in networks with unit-rank**

**connectivity structure.** A-F: Rate networks. A-C: connectivity vectors  $\mathbf{m}$  and  $\mathbf{n}$  with non-zero means  $\langle \mathbf{m} \rangle$ ,  $\langle \mathbf{n} \rangle$ , and zero covariance  $\sigma_{mn}$ . A: Fixed points of the collective variable  $\kappa$  as a function of the overlap  $n^T \mathbf{m} / N$ , low (black) and high (red) activity state. Insert: RHS of the equation  $d\kappa/dt$  (Eq. (5)),  $\kappa$  (yellow) and  $\langle \mathbf{n} \rangle \phi(\kappa)$  (gray), shown for the overlap  $n^T \mathbf{m} / N = 10$ . Fixed points (red dots) correspond to the intersections of  $\kappa$  and  $\langle \mathbf{n} \rangle \phi(\kappa)$  which is a positive function. The bifurcation therefore leads to a low and a high state. B: Illustration of the single-unit firing rates in the two states when  $n^T \mathbf{m} / N = 10$  (dashed line in A, green) for 100 units. Top: low activity state. Bottom: high activity state. C: Population-averaged firing rate as a function of  $n^T \mathbf{m} / N$ . D-F: same as A-C, for connectivity vectors  $\mathbf{m}$  and  $\mathbf{n}$  with zero means  $\langle \mathbf{m} \rangle$ ,  $\langle \mathbf{n} \rangle$ , and non-zero covariance  $\sigma_{mn}$ . D: Fixed points of the collective variable  $\kappa$  as a function of the overlap  $n^T \mathbf{m} / N$ . Insert: RHS of the equation  $d\kappa/dt$  (Eq. (5)),  $\kappa$  (yellow) and  $\kappa \phi'(\kappa)$  (gray), shown for the overlap  $n^T \mathbf{m} / N = 11.2$ . Fixed points (red dots) correspond to the intersection of  $\kappa$  and  $\kappa \phi'(\kappa)$ , which is symmetric around the  $y$  axis. The bifurcation therefore leads to two symmetric states (red and blue) on top of the low activity state. E: Illustration of the single-unit firing rates in the two symmetric states. F: Population-averaged firing rate as a function of  $n^T \mathbf{m} / N$ . G-I: same as A-C, for connectivity vectors  $\mathbf{m}$  and  $\mathbf{n}$  with non-zero means  $\langle \mathbf{m} \rangle$ ,  $\langle \mathbf{n} \rangle$ , and zero covariance  $\sigma_{mn}$ . J-L: same as A-C, for connectivity vectors  $\mathbf{m}$  and  $\mathbf{n}$  with zero means  $\langle \mathbf{m} \rangle$ ,  $\langle \mathbf{n} \rangle$ , and non-zero covariance  $\sigma_{mn}$ .

**Fig 5 (previous page).** G-L: Simulations of the spiking network. G-I: connectivity vectors  $\mathbf{m}$  and  $\mathbf{n}$  with non-zero means  $\langle m \rangle$ ,  $\langle n \rangle$  and zero covariance  $\sigma_{mn}$ . G: bifurcation to low and high states as  $\langle n \rangle$  is increased. H: raster plots of the spiking activity in the two states when  $n^T \mathbf{m}/N = 1.35\text{mV}$  (dashed line in J, green) for 20 neurons. Top: activity of 20 neurons in the high state. Bottom: activity of all (12500) neurons in the low state. The activity in the low state is highly sparse [Brunel, 2000]. I: population-averaged firing rate in the two states. J-L: same as G-I connectivity vectors  $\mathbf{m}$  and  $\mathbf{n}$  with zero means  $\langle m \rangle$ ,  $\langle n \rangle$  and non-zero covariance  $\sigma_{mn}$ . J: bifurcation to two symmetric states as  $\sigma_{mn}$  is increased. K: raster plots of the spiking activity in the two states when  $n^T \mathbf{m}/N = 32\text{mV}$  (dashed line in J, green) for 20 neurons. L: population-averaged firing rate in the two states. Dots: simulations, lines: Monte Carlo integration predictions. Network parameters are shown in Tables 5, 6.

## Geometry of nonlinear autonomous activity in rank-two networks

Going beyond unit-rank connectivity, we next examined non-linear autonomous dynamics in network with a rank-two structure. As before, we first summarize the analyses of rate networks performed in previous studies, and then compare the dynamics in spiking networks with those predictions.

A rank-two connectivity structure is defined by two pairs of vectors  $(\mathbf{m}^{(1)}, \mathbf{n}^{(1)})$  and  $(\mathbf{m}^{(2)}, \mathbf{n}^{(2)})$ :

$$J_{ij} = \frac{1}{N}(m_i^{(1)}n_j^{(1)} + m_i^{(2)}n_j^{(2)}). \quad (6)$$

In absence of external inputs, the activation dynamics  $\mathbf{x}(t)$  are confined to the two-dimensional plane spanned by the vectors  $\mathbf{m}^{(1)}$  and  $\mathbf{m}^{(2)}$ , so that, in analogy to Eq. (2) the activation  $x_i$  of unit  $i$  can be expressed as:

$$x_i(t) = \kappa_1(t) m_i^{(1)} + \kappa_2(t) m_i^{(2)}. \quad (7)$$

Here  $\kappa_1(t)$  and  $\kappa_2(t)$  are two collective variables that describe the projection of  $\mathbf{x}$  on the connectivity vectors  $\mathbf{m}^{(1)}$  and  $\mathbf{m}^{(2)}$ .

Previous works [Schuessler et al., 2020a, Beiran et al., 2021a] have shown that in low-rank rate networks with Gaussian connectivity vectors, non-linear dynamics are fully determined by the eigenspectrum of the connectivity matrix. A rank- $R$  matrix defined as in Eq. (1) has in general  $R$  non-zero eigenvalues, that coincide with the eigenvalues of the  $R \times R$  overlap matrix  $\mathbf{J}^{ov}$  obtained from scalar products between pairs of connectivity patterns [Schuessler et al., 2020a]:

$$J_{rs}^{(ov)} = \mathbf{n}^{(r)T} \mathbf{m}^{(s)} / N. \quad (8)$$

For rank-one networks, the overlap matrix reduces to a single parameter given in Eq. (38), while for rank-two networks it is a  $2 \times 2$  matrix. In the following, we focus on connectivity vectors with zero-mean entries, in which case for large  $N$  the overlap matrix converges to

$$\mathbf{J}^{ov} = \begin{pmatrix} \sigma_{n_1 m_1} & \sigma_{n_1 m_2} \\ \sigma_{n_2 m_1} & \sigma_{n_2 m_2} \end{pmatrix} \quad (9)$$

where  $\sigma_{n_r m_s}$  is the covariance between the entries of vectors  $\mathbf{n}^{(r)}$  and  $\mathbf{m}^{(s)}$ . A mean-field analysis then predicts that such networks have a fixed point at  $(0, 0)$ , the stability of which is determined by the eigenvalues of  $\phi' \mathbf{J}^{ov}$  (Methods).

We specifically examined connectivity structures with two different forms of the overlap matrix, that lead to different configurations of eigenvalues and thereby generate

qualitatively different types of nonlinear dynamics in rate networks [Beiran et al., 2021a].

We first consider rank-two networks with overlap matrices of the form:

$$\mathbf{J}^{(ov)} = \begin{pmatrix} \sigma & -\sigma_w \\ \sigma_w & \sigma \end{pmatrix}. \quad (10)$$

Such matrices have two complex conjugate eigenvalues  $\sigma \pm i\sigma_w$ . A mean-field analysis predicts spiral dynamics around the origin that decays to zero if  $\phi'(0)\sigma < 1$ , or generate limit cycle in the  $\mathbf{m}^{(1)} - \mathbf{m}^{(2)}$  plane if  $\phi'(0)\sigma > 1$ . On the other hand,  $\phi'(0)\sigma_w$  determines the frequency of these oscillations. Simulations of rate networks for  $\phi'(0)\sigma > 1$  show that the firing rates of individual units oscillate strongly (Fig. 6 A, top), but out of phase, so that oscillations are not visible at the level of the population average (Fig. 6 A, bottom). Projecting  $\mathbf{r}(t)$  on the  $\mathbf{m}^{(1)} - \mathbf{m}^{(2)}$  plane however uncovers a clear limit cycle (Fig. 6 B) that corresponds to oscillations of  $\kappa_1(t)$  and  $\kappa_2(t)$  (Fig. 6 C).

To check whether qualitatively similar dynamics occur in spiking networks, we added a rank-two structure with complex eigenvalues on top of random excitatory-inhibitory connectivity. As the two parts of the connectivity are independent, the spectrum of the full connectivity matrix consists of a continuous bulk corresponding to the random part and discrete outliers given by the eigenvalues of the rank-two structure [Rajan and Abbott, 2006, Tao, 2011, Mastrogiuseppe and Ostojic, 2018]. For large values of  $\sigma$ , simulations of the resulting spiking network show that the firing rates of individual neurons oscillate strongly (Fig. 6 D, top), but out of phase with each other, so that oscillations on the population-averaged level are weak (Fig. 6 D, bottom). Projections of the population rate  $\mathbf{r}$  on the plane  $\mathbf{m}^{(1)} - \mathbf{m}^{(2)}$  however identified clear limit cycles (Fig. 6 E, F).

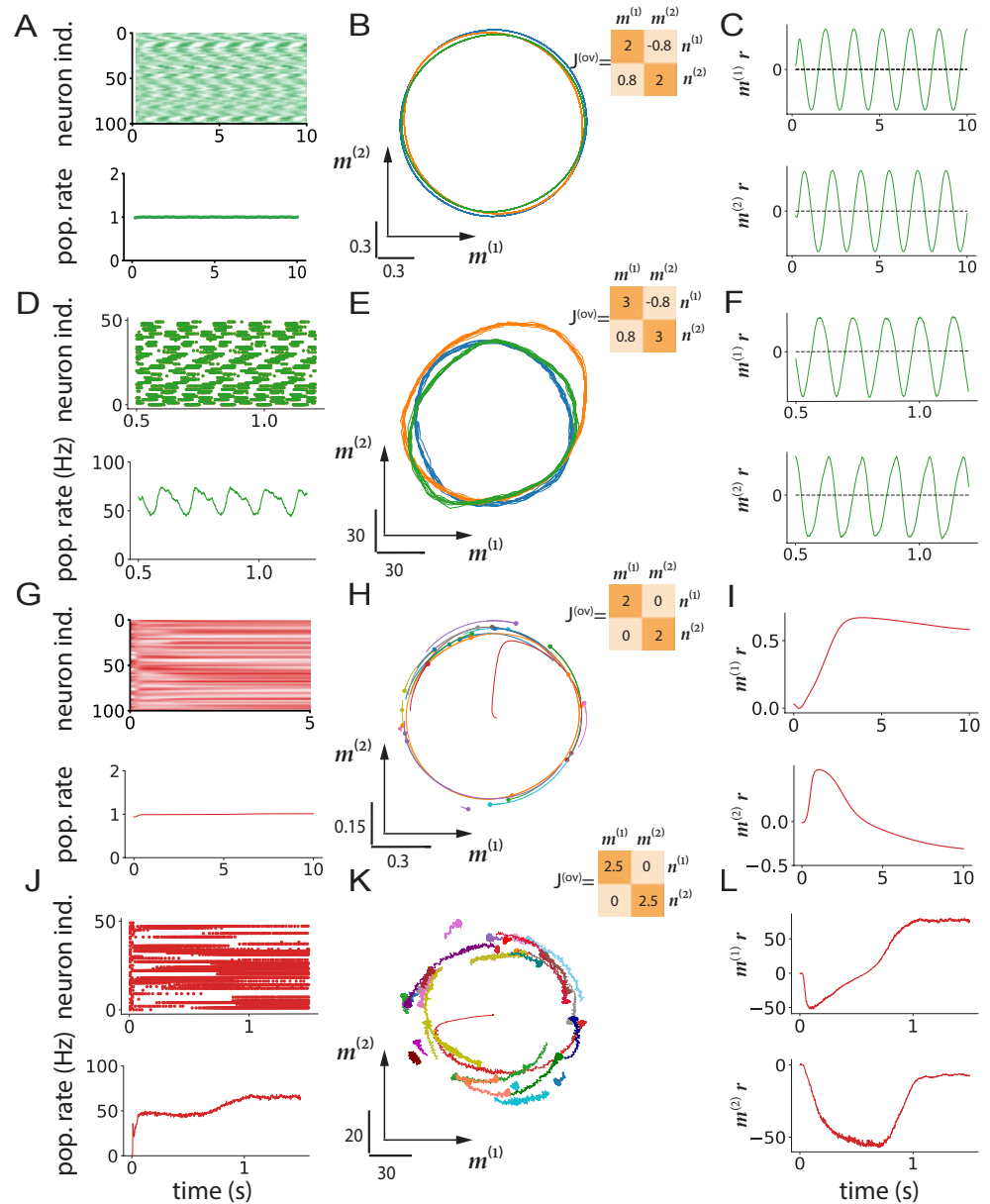
We next turned to rank-two structure with overlap matrices of the form:

$$\mathbf{J}^{ov} = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}. \quad (11)$$

The resulting connectivity matrices have two degenerate real eigenvalues  $\sigma$ , and mean-field analyses of rate networks have shown that in the limit  $N \rightarrow \infty$ , as  $\sigma$  is increased this leads to a continuum of fixed points arranged on a ring in the  $\mathbf{m}^{(1)} - \mathbf{m}^{(2)}$  plane [Mastrogiuseppe and Ostojic, 2018, Beiran et al., 2021a, Beiran et al., 2021b]. In finite-size networks, sampling fluctuations of random connectivity vectors breaks the exact degeneracy, so that only a small number of points on the ring attractor remain actual stable fixed points while the rest form a slow manifold: dynamics quickly converge to the ring, after which they slowly evolve on it until reaching a fixed point (Fig. 6 H, I).

We verified that analogous dynamics emerge in spiking networks with a degenerate rank-two structure added on top of the random excitatory-inhibitory connectivity matrix. As in rate networks, dynamics quickly converge to a ring in the  $\mathbf{m}^{(1)} - \mathbf{m}^{(2)}$  plane, after which they evolve along the ring towards stable fixed points (Fig. 6 K, L). Different instances of the rank-two structure generated with identical statistics lead to different fixed points that are all located on the same ring (Fig. 6 K). In spiking networks, the fluctuations in activity are stronger than in rate networks because of a combination of spiking noise, random excitatory-inhibitory connectivity and fluctuations in low-rank connectivity, but the low-dimensional dynamics are qualitatively similar.

In summary, mean-field analyses of rate networks with low-rank connectivity allow us to identify analogous non-trivial dynamical regimes in networks of spiking neurons.



**Fig 6. Nonlinear dynamics in networks with rank-two structure.** A-F: Connectivity structure with two complex-conjugate eigenvalues. A-C: Rate networks. A: Top: illustration of the single-unit firing rates for the first 100 neurons. Bottom: population-averaged firing rate. B: Projections of the firing rates  $\mathbf{r}(t)$  on the  $\mathbf{m}^{(1)} - \mathbf{m}^{(2)}$  plane. Insert: overlap matrix. C: Projections of the firing rates  $\mathbf{r}(t)$  on vectors  $\mathbf{m}^{(1)}$  and  $\mathbf{m}^{(2)}$  as a function of time. D-F: Analogous to (A-C), spiking network. D: Top: raster plots of the spiking activity for first 50 neurons. Bottom panel: population firing rate. E: Projections of the firing rates  $\mathbf{r}(t)$  on the  $\mathbf{m}^{(1)} - \mathbf{m}^{(2)}$  plane. Insert: overlap matrix. F: Projections of the firing rates  $\mathbf{r}(t)$  on vectors  $\mathbf{m}^{(1)}$  and  $\mathbf{m}^{(2)}$  as a function of time.

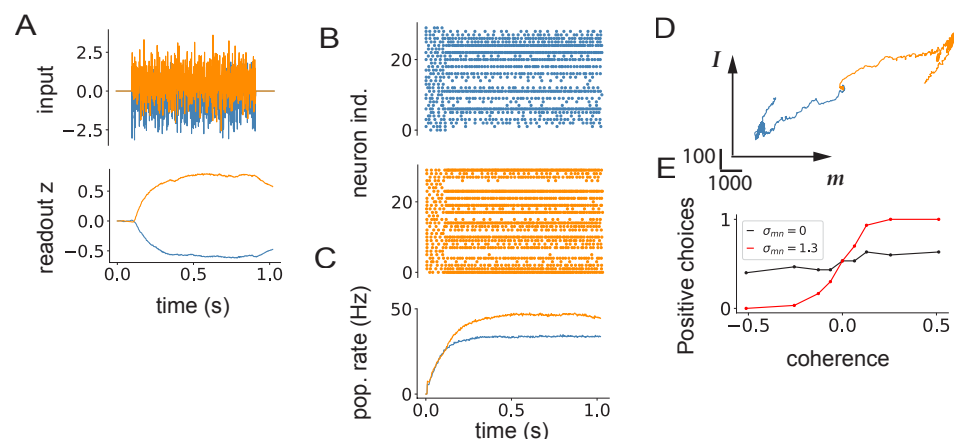
**Fig 6 (previous page).** (G-I) Rate network dynamics for an overlap matrix that has two real, degenerate eigenvalues. (G) Top panel: illustration of the single-unit firing rates for the first 100 neurons. Bottom panel: population firing rate. (H) Projections of the firing rates  $\mathbf{r}(t)$  on the  $\mathbf{m}^{(1)} - \mathbf{m}^{(2)}$  plane. Insert: overlap matrix. (I) Projections of the firing rate  $\mathbf{r}(t)$  on vectors  $\mathbf{m}^{(1)}$  and  $\mathbf{m}^{(2)}$  as a function of time. (J-L) same analysis as in (G-I) for a spiking model. (J) Top panel: raster plots of the spiking activity for first 50 neurons. Bottom panel: population firing rate. (K) Projections of the firing rates  $\mathbf{r}(t)$  on the  $\mathbf{m}^{(1)} - \mathbf{m}^{(2)}$  plane. Insert: overlap matrix. (L) Projections of the firing rate  $\mathbf{r}(t)$  on vectors  $\mathbf{m}^{(1)}$  and  $\mathbf{m}^{(2)}$  as a function of time. Different colors in the middle column (B,E,H,K) corresponds to network instances with different connectivity vectors but identical statistics. Network parameters are shown in Tables 7, 8.

## Perceptual decision making task

Our results so far show that the geometry and firing regimes in networks of spiking neurons can be predicted from the statistics of low-rank connectivity by following the principles identified in rate networks. In a final step, here we illustrate how this finding can be exploited to directly implement computations in spiking networks. We consider the perceptual decision-making task [Gold and Shadlen, 2007] in which a network receives a noisy scalar stimulus along a random input vector  $\mathbf{I}$ , and needs to report the sign of its temporal average along a random readout vector  $\mathbf{w}$ .

Previous works have identified requirements on rank-one network to perform this task [Dubreuil et al., 2022]. They showed that a unit-rank network was sufficient to implement the task, with connectivity statistics requiring a strong overlap  $\sigma_{nI}$  to integrate inputs, and an overlap  $\sigma_{mn} \approx 1$  to generate a long integration timescale via positive feedback. We built a spiking network based on an analogous connectivity configuration.

Fig. 7 illustrates the dynamics in the network in response to two inputs with positive and negative means. The two inputs lead to different patterns of activity with opposite readout values (Fig. 7 A-B), but similar population averaged firing rates (Fig. 7 C). As expected from the theory of low-rank networks, the dynamics evolve in a two-dimensional plane spanned by the input pattern  $\mathbf{I}$  and the output connectivity pattern  $\mathbf{m}$  (Fig. 7 D), as observed in experimental data [Mante et al., 2013]. The psychometric curve generated by the network strongly depends on the values of the connectivity overlaps (Fig. 7 E).



**Fig 7. Spiking network implementation of the perceptual decision-making task.** A: Top panel: two instances of the fluctuating input signal with a positive (orange) and a negative (blue) mean. Bottom panel: network readout of the activity generated by the two inputs. B: Raster plots for the first 50 neurons. C: Population firing rate. D: Dynamics projected onto the  $I - m$  plane. E: Psychometric function showing the fraction of positive outputs at different values of the overlap  $\sigma_{nm}$ . Orange color corresponds to positive ( $\bar{u} = 0.512$ ), while blue to negative mean-input ( $\bar{u} = -0.512$ ). Parameters:  $N = 12500$ ,  $\sigma_u = 1$ ,  $\sigma_{nI} = 0.26$ ,  $\sigma_{mw} = 2.1$ ,  $\sigma_{m^2} = 0.02$ ,  $\tau_f = 100\text{ms}$

## Discussion

In this study, we set out to examine how far theoretical predictions for the relation between connectivity and dynamics in recurrent networks of rate units translate to networks of spiking neurons. To this end, we compared the population activity in rate networks with low-rank connectivity to networks of integrate-and-fire neurons in which a low-rank structure was added on top of random, sparse excitatory-inhibitory connectivity. Altogether, we found the geometry of low-pass filtered activity in spiking networks is largely identical to rate networks when the low rank structure in connectivity is statistically identical. In particular, this allowed us to identify novel regimes of linear and non-linear dynamics in spiking networks, and construct networks that implement specific computations.

A widespread experimental observation across cortical areas is that sensory inputs lead to both increases and decreases of activity in individual neurons, so that different stimuli are often indistinguishable at the population-average level albeit they induce distinct patterns of responses [Wohrer et al., 2013, Bagur et al., 2018, Kobak et al., 2019]. Within the state-space picture, this implies that the responses take place primarily along directions orthogonal to the global axis [Kobak et al., 2019], suggesting that behaviorally-relevant computations may rely on dynamics along these dimensions complementary to the population-average. So far, most studies on spiking networks have however focused on averaging spiking activity either over the whole network or over sub-populations. Here we instead show that, when a low-rank connectivity structure is included in the connectivity, spiking networks naturally lead to rich dynamics along dimensions orthogonal to the global axis. Our results therefore open the door to a closer match between spiking models and analyses of experimental data.

Our starting hypothesis was that spiking networks in the asynchronous irregular regime can be directly mapped onto rate networks with identical connectivity, by identifying each integrate-and-fire neuron with a rate unit. Here we tested a restricted

version of this hypothesis by focusing exclusively on low-rank structure in the connectivity. We found that the population dynamics in spiking networks with a superposition of random and low-rank connectivity match well the predictions of rate networks with connectivity given by an identical low-rank part. To which extent these results extend to more general types of connectivity remains to be determined. A key feature of a low-rank connectivity structure is that it leads to discrete, isolated eigenvalues in the complex plane [Rajan and Abbott, 2006, Tao, 2011, Mastrogiuseppe and Ostojic, 2018, Logiaco et al., 2021, Herbert and Ostojic, 2022] (or singular values on the real line [Bondanelli and Ostojic, 2020, Benaych-Georges and Nadakuditi, 2011], while purely random connectivity in general generates a continuously distributed bulk of eigenvalues [Sompolinsky et al., 1988a, Sommers et al., 1988]. We expect that our findings hold as long as the eigenspectrum (or singular value distribution) consists of discrete outliers well separated from the random bulk, in which case the connectivity can be accurately approximated by a low-rank structure [Shao and Ostojic, 2022]. Networks performing specific computations typically rely on such outliers in the connectivity spectrum [Ingrosso and Abbott, 2019, Schuessler et al., 2020b, Susman et al., 2021], so that our results may help explain in which case functional spiking networks can be directly built from trained rate networks [Nicola and Clopath, 2017, Kim and Chow, 2018, Kim et al., 2019].

A surprising result of our analyses is that rate networks predict well the activity in spiking networks even outside of the asynchronous irregular regime, i.e. when neurons spike regularly, or with some degree of synchrony. Indeed we found that our results hold as long as spike-trains are averaged over timescales longer than the synaptic or membrane time constants. When do spiking networks then qualitatively differ from their rate-based counter-parts? Do spikes have a potential advantage over rate-based computations? One regime we have not explored here is ultra-sparse activity, in which each neuron emits only a handful of spikes in response to a stimulus. In this regime, information can be encoded in the precise timing of isolated spikes of individual neurons [Masquelier and Thorpe, 2007, Masquelier et al., 2009, Kheradpisheh et al., 2018], and a comparison with state-space trajectories predicted by rate-based models may be less useful. The ultra-sparse firing regime provides a fruitful framework for energy-efficient neuromorphic computing [Zenke et al., 2021], and suggests a potential computational role for spikes distinct from rate-based coding. It remains to be determined which parts of the brain may function in this regime. An alternative possibility is that action potentials play mainly an implementational role, as a biological mechanism for transmitting information across long distances over myelinated axons, and therefore act as a discretization of a fundamentally continuous underlying signal. Ultimately, the computational and implementational interpretations of action potentials are not mutually exclusive, and it is likely that spikes play different functional roles in different brain structures or species.

## Acknowledgments

The project was supported by the CRCNS project PIND (ANR-19-NEUC-0001-01), the program “Ecoles Universitaires de Recherche” launched by the French Government and implemented by the ANR, with the reference ANR-17-EURE-0017. There are no competing interests.

# References

- Abbott et al., 2016. Abbott, L. F., DePasquale, B., and Memmesheimer, R.-M. (2016). Building functional networks of spiking model neurons. *Nature Neuroscience*, 19(3):350–355.
- Ahmadian et al., 2015. Ahmadian, Y., Fumarola, F., and Miller, K. D. (2015). Properties of networks with partially structured and partially random connectivity. *Phys. Rev. E*, 91:012820.
- Ahmadian and Miller, 2021. Ahmadian, Y. and Miller, K. D. (2021). What is the dynamical regime of cerebral cortex? *Neuron*, 109(21):3373–3391.
- Amit and Brunel, 1997. Amit, D. J. and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, 7(3):237–252.
- Bagur et al., 2018. Bagur, S., Averseng, M., Elgueda, D., David, S., Fritz, J., Yin, P., Shamma, S., Boubenec, Y., and Ostojic, S. (2018). Go/no-go task engagement enhances population representation of target stimuli in primary auditory cortex. *Nature Communications*, 9(1).
- Baker et al., 2020. Baker, C., Zhu, V., and Rosenbaum, R. (2020). Nonlinear stimulus representations in neural circuits with approximate excitatory-inhibitory balance. *PLOS Computational Biology*, 16(9):e1008192.
- Barak, 2017. Barak, O. (2017). Recurrent neural networks as versatile tools of neuroscience research. *Current Opinion in Neurobiology*, 46:1–6.
- Beiran et al., 2021a. Beiran, M., Dubreuil, A., Valente, A., Mastrogiuseppe, F., and Ostojic, S. (2021a). Shaping Dynamics With Multiple Populations in Low-Rank Recurrent Networks. *Neural Computation*, 33(6):1572–1615.
- Beiran et al., 2021b. Beiran, M., Meirhaeghe, N., Sohn, H., Jazayeri, M., and Ostojic, S. (2021b). Parametric control of flexible timing through low-dimensional neural manifolds. *bioRxiv*.
- Benaych-Georges and Nadakuditi, 2011. Benaych-Georges, F. and Nadakuditi, R. R. (2011). The singular values and vectors of low rank perturbations of large rectangular random matrices.
- Boerlin et al., 2013a. Boerlin, M., Machens, C. K., and Denève, S. (2013a). Predictive coding of dynamical variables in balanced spiking networks. *PLOS Computational Biology*, 9(11):1–16.
- Boerlin et al., 2013b. Boerlin, M., Machens, C. K., and Denève, S. (2013b). Predictive coding of dynamical variables in balanced spiking networks. *PLOS Computational Biology*, 9(11):1–16.
- Bondanelli and Ostojic, 2020. Bondanelli, G. and Ostojic, S. (2020). Coding with transient trajectories in recurrent neural networks. *PLOS Computational Biology*, 16(2):1–36.
- Brunel, 2000. Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, 8.

- Buonomano and Maass, 2009. Buonomano, D. and Maass, W. (2009). State-dependent computations: Spatiotemporal processing in cortical networks. *Nature reviews. Neuroscience*, 10:113–25.
- Chung and Abbott, 2021. Chung, S. and Abbott, L. (2021). Neural population geometry: An approach for understanding biological and artificial neural networks. *Current Opinion in Neurobiology*, 70:137–144. Computational Neuroscience.
- Churchland and Shenoy, 2007. Churchland, M. M. and Shenoy, K. V. (2007). Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex. *Journal of Neurophysiology*, 97(6):4235–4257.
- Cunningham and Yu, 2014. Cunningham, J. and Yu, B. (2014). Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17.
- Destexhe et al., 2001. Destexhe, A., Rudolph, M., Fellous, J.-M., and Sejnowski, T. (2001). Fluctuating synaptic conductances recreate in vivo-like activity in neocortical neurons. *Neuroscience*, 107(1):13–24.
- Dubreuil et al., 2022. Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F., and Ostojic, S. (2022). The role of population structure in computations through neural dynamics. *Nature Neuroscience*, 25:1–12.
- Eliasmith and Anderson, 2004. Eliasmith, C. and Anderson, C. H. (2004). Neural engineering: Computation, representation, and dynamics in neurobiological systems. *IEEE Transactions on Neural Networks*, 15:528–529.
- Gallego et al., 2017. Gallego, J. A., Perich, M. G., Miller, L. E., and Solla, S. A. (2017). Neural manifolds for the control of movement. *Neuron*, 94(5):978–984.
- Gerstner et al., 2014. Gerstner, W., Kistler, W. M., Naud, R., and Paninski, L. (2014). *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. Cambridge University Press, USA.
- Gerstner and Naud, 2009. Gerstner, W. and Naud, R. (2009). How good are neuron models? *Science*, 326(5951):379–380.
- Gold and Shadlen, 2007. Gold, J. I. and Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, 30(1):535–574. PMID: 17600525.
- Haider et al., 2006. Haider, B., Duque, A., Hasenstaub, A. R., and McCormick, D. A. (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *Journal of Neuroscience*, 26(17):4535–4545.
- Harish and Hansel, 2015. Harish, O. and Hansel, D. (2015). Asynchronous rate chaos in spiking neuronal circuits. *PLOS Computational Biology*, 11(7):1–38.
- Harvey et al., 2012. Harvey, C., Coen, P., and Tank, D. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, 484:62–8.
- Hennequin et al., 2017. Hennequin, G., Agnes, E. J., and Vogels, T. P. (2017). Inhibitory plasticity: Balance, control, and codependence. *Annual Review of Neuroscience*, 40(1):557–579. PMID: 28598717.
- Hennequin et al., 2014. Hennequin, G., Vogels, T. P., and Gerstner, W. (2014). Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron*, 82(6):1394–1406.

- Herbert and Ostojic, 2022. Herbert, E. and Ostojic, S. (2022). The impact of sparsity in low-rank recurrent neural networks. *PLOS Computational Biology*, 18(8):1–21.
- Herz et al., 2006. Herz, A. V. M., Gollisch, T., Machens, C. K., and Jaeger, D. (2006). Modeling single-neuron dynamics and computations: A balance of detail and abstraction. *Science*, 314(5796):80–85.
- Hopfield, 1982. Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558.
- Ingrosso and Abbott, 2019. Ingrosso, A. and Abbott, L. F. (2019). Training dynamically balanced excitatory-inhibitory networks. *PLOS ONE*, 14(8):e0220547.
- Jazayeri and Ostojic, 2021. Jazayeri, M. and Ostojic, S. (2021). Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity.
- Kadmon et al., 2020. Kadmon, J., Timcheck, J., and Ganguli, S. (2020). Predictive coding in balanced neural networks with noise, chaos and delays. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 16677–16688. Curran Associates, Inc.
- Kheradpisheh et al., 2018. Kheradpisheh, S. R., Ganjtabesh, M., Thorpe, S. J., and Masquelier, T. (2018). STDP-based spiking deep convolutional neural networks for object recognition. *Neural Networks*, 99:56–67.
- Kim and Chow, 2018. Kim, C. M. and Chow, C. C. (2018). Learning recurrent dynamics in spiking networks. *eLife*, 7:e37124.
- Kim et al., 2019. Kim, R., Li, Y., and Sejnowski, T. J. (2019). Simple framework for constructing functional spiking recurrent neural networks. *Proceedings of the National Academy of Sciences*, 116(45):22811–22820.
- Kobak et al., 2019. Kobak, D., Pardo-Vazquez, J. L., Valente, M., Machens, C. K., and Renart, A. (2019). State-dependent geometry of population activity in rat auditory cortex. *eLife*, 8:e44526.
- Landau and Sompolinsky, 2018. Landau, I. D. and Sompolinsky, H. (2018). Coherent chaos in a recurrent neural network with structured connectivity. *PLOS Computational Biology*, 14(12):e1006309.
- Landau and Sompolinsky, 2021. Landau, I. D. and Sompolinsky, H. (2021). Macroscopic fluctuations emerge in balanced networks with incomplete recurrent alignment. *Phys. Rev. Research*, 3:023171.
- Lerchner and Latham, 2015. Lerchner, A. and Latham, P. E. (2015). A unifying framework for understanding state-dependent network dynamics in cortex.
- Litwin-Kumar and Doiron, 2012. Litwin-Kumar, A. and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature Neuroscience*, 15(11):1498–1505.
- Logiaco et al., 2021. Logiaco, L., Abbott, L., and Escola, S. (2021). Thalamic control of cortical dynamics in a model of flexible motor sequencing. *Cell Reports*, 35(9):109090.

- London et al., 2010. London, M., Roth, A., Beeren, L., Häusser, M., and Latham, P. E. (2010). Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature*, 466(7302):123–127.
- Mante et al., 2013. Mante, V., Sussillo, D., Shenoy, K., and Newsome, W. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503:78–84.
- Masquelier et al., 2009. Masquelier, T., Guyonneau, R., and Thorpe, S. J. (2009). Competitive STDP-based spike pattern learning. *Neural Computation*, 21(5):1259–1276.
- Masquelier and Thorpe, 2007. Masquelier, T. and Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLOS Computational Biology*, 3(2):1–11.
- Mastrogiuseppe and Ostojic, 2018. Mastrogiuseppe, F. and Ostojic, S. (2018). Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, 99(3):609–623.e29.
- Miller and Fumarola, 2012. Miller, K. D. and Fumarola, F. (2012). Mathematical equivalence of two common forms of firing rate models of neural networks. *Neural Computation*, 24(1):25–31.
- Nicola and Clopath, 2017. Nicola, W. and Clopath, C. (2017). Supervised learning in spiking neural networks with FORCE training. *Nature Communications*, 8(1).
- Ostojic, 2014. Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nature neuroscience*, 17.
- Ostojic and Brunel, 2011. Ostojic, S. and Brunel, N. (2011). From spiking neuron models to linear-nonlinear models. *PLOS Computational Biology*, 7(1):1–16.
- Pereira and Brunel, 2018. Pereira, U. and Brunel, N. (2018). Attractor dynamics in networks with learning rules inferred from in vivo data. *Neuron*, 99(1):227–238.e4.
- Rajan and Abbott, 2006. Rajan, K. and Abbott, L. F. (2006). Eigenvalue spectra of random matrices for neural networks. *Phys. Rev. Lett.*, 97:188104.
- Renart et al., 2010. Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science*, 327(5965):587–590.
- Rosenbaum and Doiron, 2014. Rosenbaum, R. and Doiron, B. (2014). Balanced networks of spiking neurons with spatially dependent recurrent connections. *Phys. Rev. X*, 4:021039.
- Sanzeni et al., 2020. Sanzeni, A., Akitake, B., Goldbach, H. C., Leedy, C. E., Brunel, N., and Histed, M. H. (2020). Inhibition stabilization is a widespread property of cortical networks. *eLife*, 9:e54875.
- Sanzeni et al., 2022. Sanzeni, A., Histed, M. H., and Brunel, N. (2022). Emergence of irregular activity in networks of strongly coupled conductance-based neurons. *Phys. Rev. X*, 12:011044.
- Saxena and Cunningham, 2019. Saxena, S. and Cunningham, J. (2019). Towards the neural population doctrine. *Current Opinion in Neurobiology*, 55:103–111.

- Schaffer et al., 2013. Schaffer, E. S., Ostojic, S., and Abbott, L. F. (2013). A complex-valued firing-rate model that approximates the dynamics of spiking networks. *PLOS Computational Biology*, 9(10):1–11.
- Schuessler et al., 2020a. Schuessler, F., Dubreuil, A., Mastrogiuseppe, F., Ostojic, S., and Barak, O. (2020a). Dynamics of random recurrent networks with correlated low-rank structure. *Phys. Rev. Research*, 2:013111.
- Schuessler et al., 2020b. Schuessler, F., Mastrogiuseppe, F., Dubreuil, A., Ostojic, S., and Barak, O. (2020b). The interplay between randomness and structure during learning in rnns. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 13352–13362. Curran Associates, Inc.
- Shadlen and Newsome, 2001. Shadlen, M. N. and Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area lip) of the rhesus monkey. *Journal of Neurophysiology*, 86(4):1916–1936. PMID: 11600651.
- Shao and Ostojic, 2022. Shao, Y. and Ostojic, S. (2022). Relating local connectivity and global dynamics in recurrent excitatory-inhibitory networks. *bioRxiv*.
- Shriki et al., 2003. Shriki, O., Hansel, D., and Sompolinsky, H. (2003). Rate models for conductance-based cortical neuronal networks. *Neural Computation*, 15(8):1809–1841.
- Shu et al., 2003. Shu, Y., Hasenstaub, A., and McCormick, D. A. (2003). Turning on and off recurrent balanced cortical activity. *Nature*, 423(6937):288–293.
- Sommers et al., 1988. Sommers, H. J., Crisanti, A., Sompolinsky, H., and Stein, Y. (1988). Spectrum of large random asymmetric matrices. *Phys. Rev. Lett.*, 60:1895–1898.
- Sompolinsky et al., 1988a. Sompolinsky, H., Crisanti, A., and Sommers, H. J. (1988a). Chaos in random neural networks. *Phys. Rev. Lett.*, 61:259–262.
- Sompolinsky et al., 1988b. Sompolinsky, H., Crisanti, A., and Sommers, H. J. (1988b). Chaos in random neural networks. *Phys. Rev. Lett.*, 61:259–262.
- Stimberg et al., 2019. Stimberg, M., Brette, R., and Goodman, D. F. (2019). Brian 2, an intuitive and efficient neural simulator. *eLife*, 8:e47314.
- Susman et al., 2021. Susman, L., Mastrogiuseppe, F., Brenner, N., and Barak, O. (2021). Quality of internal representation shapes learning performance in feedback neural networks. *Phys. Rev. Research*, 3:013176.
- Sussillo, 2014. Sussillo, D. (2014). Neural circuits as computational dynamical systems. *Current Opinion in Neurobiology*, 25:156–163.
- Sussillo and Abbott, 2009. Sussillo, D. and Abbott, L. (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron*, 63(4):544–557.
- Sussillo and Barak, 2013. Sussillo, D. and Barak, O. (2013). Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Computation*, 25(3):626–649.
- Tao, 2011. Tao, T. (2011). Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probability Theory and Related Fields*, 155(1-2):231–263.

- Tavanaei et al., 2019. Tavanaei, A., Ghodrati, M., Kheradpisheh, S. R., Masquelier, T., and Maida, A. (2019). Deep learning in spiking neural networks. *Neural Networks*, 111:47–63.
- Tetzlaff et al., 2012. Tetzlaff, T., Helias, M., Einevoll, G. T., and Diesmann, M. (2012). Decorrelation of neural-network activity by inhibitory feedback. *PLOS Computational Biology*, 8(8):1–29.
- Timón et al., 2022. Timón, L. B., Ekermans, P., Konrad, S., Nold, A., and Tchumatchenko, T. (2022). Synaptic plasticity controls the emergence of population-wide invariant representations in balanced network models. *Physical Review Research*, 4(1).
- Troyer and Miller, 1997. Troyer, T. W. and Miller, K. D. (1997). Physiological Gain Leads to High ISI Variability in a Simple Model of a Cortical Regular Spiking Cell. *Neural Computation*, 9(5):971–983.
- Valente et al., 2022a. Valente, A., Ostojic, S., and Pillow, J. W. (2022a). Probing the Relationship Between Latent Linear Dynamical Systems and Low-Rank Recurrent Neural Network Models. *Neural Computation*, 34(9):1871–1892.
- Valente et al., 2022b. Valente, A., Ostojic, S., and Pillow, J. W. (2022b). Probing the Relationship Between Latent Linear Dynamical Systems and Low-Rank Recurrent Neural Network Models. *Neural Computation*, 34(9):1871–1892.
- van Vreeswijk and Sompolinsky, 1996. van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726.
- Vyas et al., 2020. Vyas, S., Golub, M., Sussillo, D., and Shenoy, K. (2020). Computation through neural population dynamics. *Annual Review of Neuroscience*, 43:249–275.
- Wilson and Cowan, 1972. Wilson, H. R. and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12(1):1–24.
- Wohrer et al., 2013. Wohrer, A., Humphries, M. D., and Machens, C. K. (2013). Population-wide distributions of neural activity during perceptual decision-making. *Progress in Neurobiology*, 103:156–193.
- Wong, 2006. Wong, K.-F. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4):1314–1328.
- Yang and Wang, 2020. Yang, G. R. and Wang, X.-J. (2020). Artificial neural networks for neuroscientists: A primer. *Neuron*, 107(6):1048–1070.
- Zenke et al., 2015. Zenke, F., Agnes, E. J., and Gerstner, W. (2015). Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications*, 6(1).
- Zenke et al., 2021. Zenke, F., Bohtë, S. M., Clopath, C., Comşa, I. M., Göltz, J., Maass, W., Masquelier, T., Naud, R., Neftci, E. O., Petrovici, M. A., Scherr, F., and Goodman, D. F. (2021). Visualizing a joint future of neuroscience and neuromorphic engineering. *Neuron*, 109(4):571–575.

## Materials and methods

### Rate network model

We consider rate networks of  $N$  units. Each unit is described by its activation  $x_i(t)$ , with dynamics evolving according to [Sompolsky et al., 1988b]:

$$\tau \dot{x}_i(t) = -x_i(t) + \sum_{j=1}^N J_{ij} \phi(x_j) + I_i u(t). \quad (12)$$

Here  $u(t)$  is the input amplitude shared by all units,  $I_i$  is the weight of the external input on unit  $i$ , and  $\phi(x) = 1 + \tanh(x - x_{off})$  is the firing rate transfer function. The firing rate of unit  $i$  is therefore  $r_i = \phi(x_i)$ .

The recurrent connectivity matrix  $\mathbf{J}$  consists of a rank  $R$  structure  $\mathbf{P}$ , represented as a sum of  $R$  unit-rank terms, where the  $r$ -th term is given by the outer product of two vectors  $\mathbf{m}^{(r)}$ ,  $\mathbf{n}^{(r)}$ :

$$P_{ij} = \frac{1}{N} \sum_{r=1}^R m_i^{(r)} n_j^{(r)}. \quad (13)$$

We refer to vectors  $\mathbf{m}^{(r)} = \{m_i^{(r)}\}_{i=1\dots N}$ ,  $\mathbf{n}^{(r)} = \{n_i^{(r)}\}_{i=1\dots N}$  as the right and left connectivity vectors, and to  $\mathbf{I} = \{I_i\}_{i=1\dots N}$  as the input vector.

In this study, we focus on the case where the entries  $m_i^{(r)}$ ,  $n_i^{(r)}$ ,  $I_i$  of connectivity and input vectors are generated independently for each unit from a Gaussian distribution with means  $\langle m_r \rangle$ ,  $\langle n_r \rangle$ ,  $\langle I \rangle$ , standard deviations  $\sigma_{m_r}$ ,  $\sigma_{n_r}$ ,  $\sigma_I$  and covariances  $\sigma_{xy}$  ( $x, y \in \{n_r, m_r, I\}$ ).

To simulate network activity, Eq. (12) was discretised using Euler's method with time step  $dt$ , for a total simulation time  $t_{run}$ . Network parameters are shown in Tables 3, 5, 7.

### Spiking network model

We consider networks of  $N$  leaky-integrate and fire neurons [Brunel, 2000], where the membrane potential of neuron  $i$  evolves according to:

$$\tau_m \frac{dV_i}{dt} = -V_i + \mu_0 + \sqrt{\tau_m} \sigma_0 \xi_i(t) + \mu_i^{rec}(t) + I_i u(t). \quad (14)$$

Here  $\tau_m$  is the membrane time constant,  $\mu_0$  a constant baseline input,  $\xi_i(t)$  a white noise independent for each neuron,  $\sigma_0$  the amplitude of the noise,  $\mu_i^{rec}$  total recurrent input defined below, and  $I_i$ ,  $u(t)$  the weights and the amplitude of the external input.

An action potential, or "spike", is generated when the membrane potential crosses the threshold  $V_{thr}$ . The membrane potential is then reset to the value of  $V_r$ , and maintained at that value during a refractory period  $t_{ref}$ .

The total recurrent input to neuron  $i$  is given by

$$\mu_i^{rec}(t) = \tau_m \sum_{j=1}^N J_{ij} \sum_k \delta(t - t_j^{(k)} - \tau_{del}) \quad (15)$$

where  $J_{ij}$  is strength of the synaptic connection from neuron  $j$  to neuron  $i$ ,  $t_j^{(k)}$  is the time of the  $k^{th}$  spike of the presynaptic neuron  $j$ ,  $\tau_{del}$  is the synaptic delay and  $\delta(t)$  is the delta function.

The connectivity matrix  $\mathbf{J}$  consists of a sum of a full-rank excitatory-inhibitory part  $\mathbf{J}^{EI}$  and a low-rank structure  $\mathbf{P}$ :

Notation	Description
$N$	number of neurons
$C$	number of EI connections each neuron receives
$J$	excitatory synaptic strength
$g$	relative inhibition strength
$\tau_m$	membrane time constant
$\mu_{rec}$	total recurrent input
$\mu_0$	baseline input
$\sigma_0$	amplitude of the noise
$V_{thr}$	threshold potential
$V_r$	reset potential
$\tau_{ref}$	refractory period
$\tau_{del}$	synaptic delay
$\tau_f$	filter time constant
$ I $	amplitude of the vector $I$
$t_{run}$	simulation run time
$\delta(t)$	delta function

**Table 1.** List of notations for spiking network models.

$$\mathbf{J} = \mathbf{J}^{EI} + \mathbf{P}. \quad (16)$$

The low-rank structure is identical to Eq. (13), while  $\mathbf{J}^{EI}$  is a sparse, random excitatory-inhibitory matrix identical to [Brunel, 2000]. Each neuron receives inputs from  $C$  neurons,  $C$  being much smaller of the total number of neurons  $N$ . The fraction of non-zero connections is  $f_c = C/N = 0.1$ , where 80% of incoming connections are excitatory and the rest are inhibitory. All non-zero excitatory synapses have the same strength  $J$ , while non-zero inhibitory synapses have the strength  $-gJ$ .

The network was simulated using the Euler method implemented in Brian2 package [Stimberg et al., 2019] with simulation step  $dt$  and simulation time  $t_{run}$ .

Single-neuron firing rates were computed from spikes using an exponential filter with a time constant  $\tau_f$ . The instantaneous rate of  $i$ -th neuron at time  $t$  is given by

$$\tau_f \frac{dr_i(t)}{dt} = -r_i(t) + \sum \delta(t - t_k) \quad (17)$$

where  $\delta(t - t_k)$  is the delta function centered at  $t_k$ . In the case of multiple trials, rates  $r_i$  are averaged over trials:

$$\langle r_i \rangle = \frac{1}{N_{tr}} \sum_{k=1}^{N_{tr}} r_i(k) \quad (18)$$

The parameter notations for spiking models are summarized in Table 1. Parameters whose value do not change over different simulations/figures are given in Table 2. If different values are used in a specific figure, these value are specified in a dedicated table (Tables 4, 6, 8).

## Geometry of responses to external inputs

To characterize the geometry of activity in the high-dimensional state space, we examined the projections of the firing rate trajectories  $\mathbf{r}(t) = \{r_i(t)\}_{i=1\dots N}$  on an arbitrary direction  $\mathbf{w} = \{w_i\}_{i=1\dots N}$ , defined as:

$$\langle \mathbf{w}^T \mathbf{r}(t) \rangle = \frac{1}{N} \sum_{i=1}^N w_i r_i(t). \quad (19)$$

Variable	Value
$N$	12500
$C$	1250
$\tau_m$	20ms
$\tau_{ref}$	0.5ms
$dt$	1ms
$\sigma_0$	0.71mV
$V_{thr}$	20mV
$V_r$	10mV
$\tau_{del}$	1.5ms

**Table 2.** Common parameters in Fig. 3-7

In particular, taking  $\mathbf{w}$  to be the global axis where  $w_i = 1$  for all  $i$ , the projection gives the population-averaged firing rate:

$$\langle r \rangle(t) = \frac{1}{N} \sum_{i=1}^N r_i(t). \quad (20)$$

Based on previous works [Mastrogiuseppe and Ostojic, 2018, Schuessler et al., 2020a, Beiran et al., 2021a, Dubreuil et al., 2022], below we summarize the predictions of low-rank rate models for the geometry of activity, and then describe a method for verifying these predictions using principal components analysis.

### Rate networks

In rate networks with a low-rank connectivity matrix, the dynamics of the activations  $\mathbf{x}(t) = \{x_i(t)\}_{i=1\dots N}$  are explicitly confined to a low-dimensional subspace of state space [Beiran et al., 2021a, Dubreuil et al., 2022, Valente et al., 2022b], meaning that projections of  $\mathbf{x}(t)$  are non-zero only on vectors  $\mathbf{w}$  belonging to this subspace. Here we first reproduce the derivation of the geometry of the activations  $\mathbf{x}(t)$  [Beiran et al., 2021a, Dubreuil et al., 2022]. We then explore the implications for the geometry of firing rates  $\mathbf{r}(t)$  where  $r_i(t) = \phi(x_i(t))$ .

For low-rank connectivity, the dynamics of  $x_i(t)$  are given by

$$\tau \dot{x}_i(t) = -x_i(t) + \frac{1}{N} \sum_{j=1}^N \sum_{k=1}^R m_i^{(k)} n_j^{(k)} \phi(x_j) + \sum_{s=1}^{N_{in}} I_i^{(s)} u_s(t), \quad (21)$$

where for completeness we included  $N_{in}$  scalar inputs  $u_s(t)$  along input vectors  $\mathbf{I}^{(s)}$  with  $s = 1 \dots N_{in}$ .

We start by assuming that at time 0, the initial state  $\mathbf{x}(0)$  lies in the subspace spanned by  $\mathbf{m}^{(r)}$  and  $\mathbf{I}^{(s)}$ , i.e. that  $\langle \mathbf{w}^T \mathbf{x}(0) \rangle \neq 0$  if and only if  $\mathbf{w}$  is a linear combination of  $\mathbf{m}^{(r)}$  for  $r = 1 \dots R$  and  $\mathbf{I}^{(s)}$  for  $s = 1 \dots N_{in}$ . This assumption can be made without loss of generality. Indeed, if it is not full-filled, the initial state  $\mathbf{x}(0)$  can be included as an additional input with  $u_s(t) = \delta(t)$  and  $\mathbf{I}^{(s)} = \mathbf{x}(0)$  in Eq. (21).

It is then straightforward to show by induction from Eq. (21) that for any  $t$ ,  $\langle \mathbf{w}^T \mathbf{x}(t) \rangle \neq 0$  if and only if  $\mathbf{w}$  is a linear combination of  $\mathbf{m}^{(r)}$  and  $\mathbf{I}^{(s)}$ . The activations  $\mathbf{x}(t)$  therefore lie for any  $t$  in the subspace spanned by  $\mathbf{m}^{(r)}$  and  $\mathbf{I}^{(s)}$ . Assuming for simplicity that these vectors form an orthogonal set, the activation of the  $i$ -th neuron  $x_i$ , can be written as:

$$x_i(t) = \sum_{r=1}^R \kappa_r(t) m_i^{(r)} + \sum_{s=1}^{N_{in}} v_s(t) I_i^{(s)}. \quad (22)$$

Here  $\kappa_r$  and  $v_s$  are scalar latent variables that correspond to the coordinates of  $x(t)$  along the vectors  $\mathbf{m}^{(r)}$  and  $\mathbf{I}^{(s)}$ , and can be computed by projecting  $\mathbf{x}(t)$  on normalized directions  $\mathbf{m}^{(r)}/\|\mathbf{m}^{(r)}\|^2$  and  $\mathbf{I}^{(s)}/\|\mathbf{I}^{(s)}\|^2$ :

$$\mathbf{x}^T \mathbf{m}^{(r)} / \|\mathbf{m}^{(r)}\|^2 = \kappa_r \quad (23)$$

$$\mathbf{x}^T \cdot \mathbf{I}^{(s)} / \|\mathbf{I}^{(s)}\|^2 = v_s. \quad (24)$$

Projecting Eq. (21) on the vector  $\mathbf{I}^{(s)}/\|\mathbf{I}^{(s)}\|^2$ , we then obtain

$$\tau \frac{dv_s}{dt} = -v_s + u_s(t), \quad (25)$$

while the projection on  $\mathbf{m}^{(r)}/\|\mathbf{m}^{(r)}\|^2$  gives

$$\tau \frac{d\kappa_r}{dt} = -\kappa_r + \frac{1}{N} \sum_{j=1}^N n_j^{(r)} \phi(x_j). \quad (26)$$

Inserting Eq. (22) into Eq. (26) then leads to the following dynamical system:

$$\tau \frac{d\kappa_r}{dt} = -\kappa_r + \kappa_r^{rec} \quad (27)$$

$$\kappa_r^{rec} = \frac{1}{N} \sum_{i=1}^N n_i^{(r)} \phi \left( \sum_{l=1}^R \kappa_l m_i^{(l)} + \sum_{s=1}^{N_{in}} v_s I_i^{(s)} \right). \quad (28)$$

To simplify notations, from here on we consider unit-rank networks with a single input ( $R = 1$  and  $N_{in} = 1$ , we therefore drop the indices  $r$  and  $s$ ), where the entries of  $\mathbf{m}$ ,  $\mathbf{n}$  and  $\mathbf{I}$  are generated from a joint Gaussian distribution with means  $\langle m \rangle$ ,  $\langle n \rangle$ ,  $\langle I \rangle$ , standard deviations  $\sigma_m$ ,  $\sigma_n$ ,  $\sigma_I$  and covariances  $\sigma_{xy}$  for  $x, y \in \{m, n, I\}$ . In the limit  $N \rightarrow \infty$ , the sum over  $j$  in Eq. (28) can then be replaced by an integral over the joint Gaussian distribution, which can be computed using Stein's Lemma for Gaussian integrals [Schuessler et al., 2020a, Beiran et al., 2021a, Dubreuil et al., 2022]. The dynamics for  $\kappa$  then become (Methods):

$$\tau \frac{d\kappa}{dt} = -\kappa + \langle n \rangle \langle \phi(\mu, \Delta) \rangle + (\sigma_{nm}\kappa + \sigma_{nI}v) \langle \phi'(\mu, \Delta) \rangle \quad (29)$$

where the brackets denote the following Gaussian integral

$$\langle f(\mu, \Delta) \rangle = \int dx (2\pi)^{-\frac{1}{2}} \exp^{-x^2/2} f(\mu + \sqrt{\Delta}x) \quad (30)$$

and

$$\begin{aligned} \mu &= \langle I \rangle v + \langle m \rangle \kappa \\ \Delta &= (\kappa \sigma_m)^2 + (v \sigma_I)^2. \end{aligned} \quad (31)$$

We next turn to the geometry of firing rates  $\mathbf{r}(t)$  where  $r_i(t) = \phi(x_i(t))$ , and examine the projection of  $\mathbf{r}(t)$  on an arbitrary direction  $\mathbf{w}$  in the activity state space:

$$\begin{aligned} \langle \mathbf{w}^T \phi(\mathbf{x}) \rangle &= \frac{1}{N} \sum_{j=1}^N w_j \phi(x_j) \\ &= \frac{1}{N} \sum_{j=1}^N w_j \phi(\kappa(t) m_j + v(t) I_j). \end{aligned} \quad (32)$$

In linear networks (i.e. when  $\phi(x) = x$ ), firing rates are equivalent to activations  $\mathbf{x}(t)$ , and therefore their dynamics are confined to the subspace spanned by  $\mathbf{m}$  and  $\mathbf{I}$ . The projection of  $\mathbf{r}(t)$  on any direction orthogonal to  $\mathbf{m}$  and  $\mathbf{I}$  is therefore zero. In particular, the projection on the global axis is non-zero only if  $\langle I \rangle \neq 0$ , or if  $\langle m \rangle \neq 0$  and  $\kappa \neq 0$ .

Here we focus on non-linear networks, and directions  $\mathbf{w}$  with entries generated from a joint Gaussian distribution with entries of  $\mathbf{m}$  and  $\mathbf{I}$ , specified by a mean  $\langle w \rangle$ , variance  $\sigma_w^2$ , and covariances  $\sigma_{wm}$  and  $\sigma_{wI}$ . As for Eq. (28), the r.h.s. of Eq. (32) can be replaced by a Gaussian integral, and, using Stein's Lemma, be expressed as (see Appendix):

$$\langle w^T \mathbf{r} \rangle = \langle w \rangle \langle \phi(\mu, \Delta) \rangle + \sigma_{wm} \kappa(t) \langle \phi'(\mu, \Delta) \rangle + \sigma_{wI} v(t) \langle \phi'(\mu, \Delta) \rangle. \quad (33)$$

In Eq. (33), the first term in the r.h.s. represents the population-averaged firing rate, i.e. the projection of  $\mathbf{r}(t)$  on the global axis. Indeed, taking  $w_i = 1/N$  for  $i = 1 \dots N$ , only the first term is non-zero. Moreover, Eqs. (30),(31) show that changes in the population averaged firing-rate  $\langle \phi(\mu, \Delta) \rangle$  can be induced either through the mean input  $\mu$  by non-zero means  $\langle I \rangle$  and  $\langle m \rangle$ , or through the variance of the input  $\Delta$  by non-zero variances  $\sigma_I$  and  $\sigma_m$  of the input and connectivity vectors.

The last two terms in Eq. (33) respectively represent the projection of firing rates on the zero-mean parts of  $\mathbf{m}$  and  $\mathbf{I}$ , i.e. changes in  $\mathbf{r}(t)$  along directions orthogonal to the global axis. Altogether, Eq. (33) therefore predicts that the projection of the firing rate vector  $\mathbf{r}(t)$  is zero on any direction  $\mathbf{w}$  orthogonal to the global axis,  $\mathbf{m}$  and  $\mathbf{I}$ . Interestingly, for Gaussian connectivity vectors considered here, the geometry of firing rate dynamics is therefore largely equivalent to linear networks (i.e. to the geometry of  $\mathbf{x}(t)$ ). The main difference is that in the non-linear case, the heterogeneity across neurons quantified by the input variance  $\Delta$  can induce a non-zero component along the global axis even when  $\langle I \rangle = 0$  and  $\langle m \rangle = 0$ .

These theoretical predictions are verified through simulations in Fig. 2. The corresponding network parameters are given in Table 3.

## Principal Component Analysis

In order to extract the low-dimensional subspace of the population activity from simulations, we performed dimensional reduction via a standard Principal Component Analysis (PCA). First, we construct the matrix  $X$  in which every column corresponds to the time trace of firing rates  $X[:, i] = r_i(t)$ . The matrix  $X$  is then normalized by subtracting the mean in every column. We compute the principal components (PCs) as the normalized eigenvectors of the correlation matrix  $C = X^T X$ , sorted in decreasing order of their eigenvalues  $\lambda_i$ . The activity matrix  $X$  is then projected on the orthonormal basis generated by the PC vectors, yielding  $X' = XE$  where  $E$  is the  $N \times N$  matrix with columns formed by PC components. The variance explained by each component is the corresponding entry on the diagonal of the rotated correlation matrix  $C' = X'^T X'$ . For rate networks, we run PCA on individual trial, as we did not include noise in the dynamics. For spiking networks we run the PCA on firing rates averaged over trials  $N_{tr}$  (see Tables 6, 8).

## Geometry of nonlinear autonomous activity in unit-rank networks

### Rate network

We now turn to the autonomous activity in unit-rank networks without external inputs. The autonomous dynamics of the collective variable  $\kappa$  are described by Eqs. (27), (28)

Notation	Description
$N$	1000
$\sigma_m$	1
$\sigma_n$	1
$\sigma_I$	1 (gray, blue), 0 (purple)
$\sigma_{nI}$	1
$\langle I \rangle$	0 (gray, blue), 1 (purple)
$\tau_m$	100ms
$t_{run}$	5s
$dt$	1ms
$u(t)$	input amplitude

**Table 3.** Parameters Fig. 2

Variable	Value			
$C$	1250			
$\sigma_m$	2			
$\sigma_n$	20mV			
$\tau_{ref}$	0.5ms			
$\tau_{del}$	1.5ms			
$\tau_f$	1 – 100ms			
$t_{run}$	2s			
$\sigma_{nI}$	0.4mV <sup>2</sup>			
	Fig. 3	Fig. 4 A-C	Fig. 4 D-F	Fig. 4 G-I
$J$	0.1mV	0.1mV	0.1mV	0mV
$g$	5	6	5	5
$\mu_0$	40mV	80mV	40mV	30mV
$ I $ , global	22.5mV	-/-	-/-	-/-
$ I $ , orthogonal	125mV	-/-	-/-	-/-
$ I $ , along	125mV	50mV	50mV	50mV

**Table 4.** Parameters Fig. 3 and 4

Variable	Value	
$\sigma_m$	2	
$\sigma_n$	6	
$\tau_m$	100ms	
$t_{run}$	20s	
$dt$	10ms	
$x_{off}$	2.9	
	Panels A-C	Panels D-F
$\mu_m$	2	0
$\mu_n$	0.1 – 6	0
$\sigma_{mn}$	0	0.1 – 10
$N_{nets}$	1	1
$N_{tr}$	4	8

**Table 5.** Parameters Fig. 5, rate

in which the external input is zero:

$$\frac{d\kappa}{dt} = -\kappa + \frac{1}{N} \sum_{i=1}^N n_i \phi(\kappa m_i). \quad (34)$$

Any fixed point  $\kappa_0$  obeys:

$$\kappa_0 = F(\kappa_0). \quad (35)$$

where

$$F(\kappa) = \frac{1}{N} \sum_{i=1}^N n_i \phi(\kappa m_i), \quad (36)$$

The stability of  $\kappa_0$  is determined by linearizing Eq. 34, yielding:

$$\left. \frac{d\kappa}{dt} \right|_{\kappa=\kappa_0} = -1 + \frac{1}{N} \sum n_i m_i \phi'(\kappa_0 m_i). \quad (37)$$

The stability of  $\kappa_0$  is therefore controlled by the overlap

$$\langle \mathbf{n}^T \phi' \mathbf{m} \rangle = \frac{1}{N} \sum_{i=1}^N n_i m_i \phi'(\kappa_0 m_i). \quad (38)$$

In the large  $N$  limit, replacing the sum with a Gaussian integral and applying Stein's lemma, the r.h.s in Eq. (36) can be further expressed as

$$F(\kappa) = \langle n \rangle \langle \phi(\mu, \Delta) \rangle + \sigma_{mn} \kappa \langle \phi'(\mu, \Delta) \rangle. \quad (39)$$

To examine the effects of the two terms in  $F(\kappa)$ , in the results we vary the overlap either by setting  $\sigma_{mn} = 0$  and changing  $\langle n \rangle$  or by setting  $\langle m \rangle, \langle n \rangle = 0$  and changing  $\sigma_{mn}$ . To compute  $F(\kappa)$ , we approximate the Gaussian integrals  $\langle \phi' \rangle, \langle \phi \rangle$  in Eq. (39) using the Monte-Carlo method. Specifically, we choose an array of values for  $\kappa$ , and for each element compute the corresponding  $F(\kappa)$  (Eq. (36)) by averaging over 50 different realisation of vectors  $\mathbf{m}$  and  $\mathbf{n}$ . We then determine the fixed point by solving for  $\kappa = F(\kappa)$ . The predicted population-averaged firing rate can then be computed as  $\frac{1}{N} \sum_i \phi(\kappa m_i)$ .

The corresponding results are shown in Fig. 5. Network parameters are given in Table 5.

Variable	Value	
$\sigma_m$	2	
$\sigma_n$	20mV	
$t_{run}$	1.2s	
$\tau_f$	100ms	
	Panels G-I	Panels J-L
$C$	4000	1250
$\mu_m$	0.01	0
$\mu_n$	0 – 150mV	0
$\sigma_{mn}$	0	0 – 40mV
$J$	0.2mV	0.1mV
$g$	4.8	5
$\mu_0$	17.7mV	40mV
$\tau_{ref}$	2ms	0.5ms
$\tau_{del}$	2.5ms	1.5ms
$N_{nets}$	5	7
$N_{tr}$	3	2

**Table 6.** Parameters Fig. 5, SNN

### Spiking network

In Fig. 5 the overlap is varied as in the rate network, either through  $\langle n \rangle$  or the covariance  $\sigma_{mn}$ . We run the dynamics for  $N_{nets}$  different network instances keeping the overlap  $\mathbf{n}^T \mathbf{m} / N$  fixed, while resampling connectivity vectors  $\mathbf{m}$  and  $\mathbf{n}$  from Gaussians with mean  $\langle m \rangle$ ,  $\langle n \rangle$  and variance  $\sigma_m$ ,  $\sigma_n$  respectively. The dynamics for each network instance is run for  $N_{tr}$  number of trials. In every trial, we resample the initial membrane potential  $V(0)$  from a Gaussian distribution. At a fixed overlap, for each network configuration and at each trial, the collective variable is computed as  $\kappa_{curr} = \frac{1}{N} \sum_i r_i n_i$ . To get the plot in Fig. 5 we first set a threshold  $\kappa_{thr}$  that is a boundary between zero state and the the high state (Fig. 5 G) or two symmetric states (Fig. 5 J). Then we average over all collective variables  $\kappa_{curr}$  that have  $|\kappa_{curr}| < \kappa_{thr}$  to compute low states. For those  $|\kappa_{curr}| > \kappa_{thr}$ , we average over all positive or over all negative  $\kappa_{curr}$  values to get the high states or the two symmetric states. The parameters used for simulating spiking model in Fig. 5 are presented in Table 6

### Geometry of nonlinear autonomous activity in rank-two networks

We extend the previous results to rank-two networks where the autonomous dynamics are described by two collective variables  $\kappa_1$  and  $\kappa_2$  (Eq. (22)). Projecting Eq. (21) onto  $\mathbf{m}^{(1)}$  and  $\mathbf{m}^{(2)}$  gives

$$\begin{aligned} \tau \frac{d\kappa_1}{dt} &= -\kappa_1 + \frac{1}{N} \sum_{i=1}^N n_i^{(1)} \phi(\kappa_1 m_i^{(1)} + \kappa_2 m_i^{(2)}) = G_1(\kappa_1, \kappa_2) \\ \tau \frac{d\kappa_2}{dt} &= -\kappa_2 + \frac{1}{N} \sum_{i=1}^N n_i^{(2)} \phi(\kappa_1 m_i^{(1)} + \kappa_2 m_i^{(2)}) = G_2(\kappa_1, \kappa_2) \end{aligned} \quad (40)$$

Assuming zero-mean Gaussian connectivity vectors, replacing sums by integrals in

$\tau_m$	100ms	
Variable	Value	
$t_{run}$	10s	
$dt$	1ms	
	Panels A-C	Panels G-I
$\sigma_{m_1^2}$	1	1
$\sigma_{n_1^2}$	7.24	7.24
$\sigma_{m_2^2}$	1	1
$\sigma_{n_2^2}$	3.63	3.63
$\sigma_{m_1 n_1}$	2	2
$\sigma_{m_1 n_2}$	0.8	0
$\sigma_{m_2 n_1}$	-0.8	0
$\sigma_{m_2 n_2}$	2	2
$N_{nets}$	3	15
$N_{tr}$	5	3

**Table 7.** Parameters Fig. 6, rate

the  $N \rightarrow \infty$  limit, and applying Stein's lemma leads to:

$$\begin{aligned}\frac{d\kappa_1}{dt} &= -\kappa_1 + (\sigma_{n_1 m_1} \kappa_1 + \sigma_{n_1 m_2} \kappa_2) \langle \phi'(\mu, \Delta) \rangle \\ \frac{d\kappa_2}{dt} &= -\kappa_2 + (\sigma_{n_2 m_1} \kappa_1 + \sigma_{n_2 m_2} \kappa_2) \langle \phi'(\mu, \Delta) \rangle\end{aligned}\quad (41)$$

where

$$\begin{aligned}\mu &= 0 \\ \Delta &= \sigma_{m_1^2} \kappa_1^2 + \sigma_{m_2^2} \kappa_2^2.\end{aligned}\quad (42)$$

For zero-mean connectivity vectors,  $(\kappa_1, \kappa_2) = (0, 0)$  is always a fixed point. A linear analysis shows that its stability is given by the eigenvalues of  $\phi'(0) \mathbf{J}^{ov}$  [Schuessler et al., 2020a, Beiran et al., 2021a], where  $\phi'(0)$  is the gain at zero, and  $\mathbf{J}^{ov}$  the overlap matrix :

$$\mathbf{J}^{ov} = \begin{pmatrix} \sigma_{n_1 m_1} & \sigma_{n_1 m_2} \\ \sigma_{n_2 m_1} & \sigma_{n_2 m_2} \end{pmatrix}.\quad (43)$$

For Fig. 6, we ran simulations for  $N_{nets}$  network instances and  $N_{tr}$  trials for each instance, and plot the projections without averaging over trials. The parameters used for simulating rate and spiking model in Fig. 6 are presented in Tables 7 and 8.

## Perceptual decision-making task

We start from a network in AI regime as in Section Geometry of responses to external inputs, and add a unit rank structure on top of the random part.

In each trial, the model was run for  $t_{run} = 1020\text{ms}$ : a fixation epoch of duration  $T_{fix} = 100\text{ms}$  was followed by a stimulation epoch of  $T_{stim} = 800\text{ms}$ , delay epoch of  $T_{del} = 100\text{ms}$  and a decision epoch  $T_{dec} = 20\text{ms}$ . The feed-forward input to neuron  $i$  on trial  $k$  was

$$I_i^{FF}(t) = I_i u^{(k)}(t) \quad (44)$$

where during the stimulation,  $u^{(k)} = \bar{u}^{(k)}(t) + \psi^{(k)}(t)$ , with  $\psi^{(k)}(t)$  a zero-mean Gaussian white noise of standard deviation  $\sigma_u = 1$ . Connectivity vectors and the input vector were generated from a Gaussian distribution with zero mean. The standard deviation of vector  $\mathbf{m}$  was  $\sigma_{m^2} = 0.02$ , and the covariance between pairs of vectors

Variable	Value	
$C$	1250	
$J$	0.1mV	
$g$	5	
$t_{ref}$	0.5ms	
$\tau_{del}$	1.5ms	
$\tau_f$	20ms	
	Panels D-F	Panels J-L
$\sigma_{m_1^2}$	1	1
$\sigma_{n_1^2}$	82.4mV	26mV
$\sigma_{m_2^2}$	1	1
$\sigma_{n_2^2}$	46mV	26mV
$\sigma_{m_1 n_1}$	30mV	25mV
$\sigma_{m_1 n_2}$	8mV	0
$\sigma_{m_2 n_1}$	-8mV	0
$\sigma_{m_2 n_2}$	26mV	25mV
$N_{nets}$	3	35
$N_{tr}$	1	1
$t_{run}$	1.2s	3s

**Table 8.** Parameters Fig. 6, SNN

$\sigma_{mn} = 0.016$ ,  $\sigma_{nI} = 0.26$ ,  $\sigma_{mw} = 2.1$ . During the decision epoch, a single readout was evaluated by output of the network is defined by readout value:

$$z = \frac{1}{N} \sum_{j=1}^N w_j r_j \quad (45)$$

where  $\mathbf{w}$  is a readout vector generated from a Gaussian with zero mean and

On trial  $k$ , if the mean of the readout is above zero, we label the output as 1, and as 0 otherwise. At every value of the overlap, psychometric curve is computed by plotting the fraction of trials that had an output 1. The network was run for 30 trials at each overlap.

## Appendix A: Mean-field theory and gaussian integrals

Using the mean-field theory, we derive in detail the projection in Eq. (32) for the rank-one case, which can then be extended to higher ranks. Vectors  $\mathbf{m}$ ,  $\mathbf{I}$  and  $\mathbf{w}$  are generated as

$$\mathbf{m} = \sigma_m \mathbf{X} \quad (46)$$

$$\mathbf{I} = \sigma_I \mathbf{Y}, \quad (47)$$

$$\mathbf{w} = \sigma_{mw}/\sigma_m \mathbf{X} + \sigma_{Iw}/\sigma_I \mathbf{Y} + \sqrt{\sigma_w^2 - (\sigma_{mw}^2/\sigma_m^2 + \sigma_{Iw}^2/\sigma_I^2)} \mathbf{Z} \quad (48)$$

where  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$  are independent vectors generated from a Gaussian distribution with zero mean and unit standard deviation,  $\sigma_m, \sigma_I, \sigma_w$  standard deviations of vectors  $\mathbf{m}$ ,  $\mathbf{I}$ ,  $\mathbf{w}$  and  $\sigma_{mw}, \sigma_{Iw}$  overlaps of vectors  $\mathbf{w}$  and  $\mathbf{m}$ ,  $\mathbf{I}$  respectively.

The dynamics in Eq. (32) consist of a sum over the  $N$  units in the network. In the limit of large networks with defined statistics, the sum over  $N$  elements corresponds to

the empirical average over the distribution of its elements. Therefore, we can replace the sum by an integral over the distribution  $P(m, n, I)$ .

$$\begin{aligned} \mathbf{w} \cdot \phi(x) &= \frac{1}{N} \sum_{j=1}^N w_j \phi(\kappa m_i + v_s I_i) = \\ &= \int dm dI dw P(m, I, w) w \phi(\kappa m + v_s I) \\ &= \int dX dY dZ P(X, Y, Z) w(X, Y, Z) \phi(\kappa m(X, Y, Z) + v_s I(X, Y, Z)) \end{aligned} \quad (49)$$

We represented the integral in Eq. (49) as a function of variables  $X$ ,  $Y$  and  $Z$ , which are independent, so that the joint distribution obeys  $P(X, Y, Z) = P(X)P(Y)P(Z)$ . Eq. (49) then becomes:

$$\begin{aligned} \mathbf{w} \cdot \phi(x) &= \\ &= \iiint P(X)P(Y)P(Z) dX dY dZ \left( \frac{\sigma_{mw}}{\sigma_m} X + \frac{\sigma_{Iw}}{\sigma_I} Y + \sqrt{\sigma_w^2 - \frac{\sigma_{mw}^2}{\sigma_m^2} - \frac{\sigma_{Iw}^2}{\sigma_I^2}} Z \right) \cdot \\ &\cdot \phi(\kappa_r \sigma_m X + v_s \sigma_I Y) = \\ &= \frac{\sigma_{mw}}{\sigma_m} \int X \phi(\kappa_r \sigma_m X + v_s \sigma_I Y) P(X) dX \int P(Y) dY \int P(Z) d(Z) + \\ &+ \frac{\sigma_{Iw}}{\sigma_I} \int Y \phi(\kappa_r \sigma_m X + v_s \sigma_I Y) P(Y) dY \int P(X) dX \int P(Z) d(Z) + \\ &+ \sqrt{\sigma_w^2 - \frac{\sigma_{mw}^2}{\sigma_m^2} - \frac{\sigma_{Iw}^2}{\sigma_I^2}} \iint \phi(\kappa_r \sigma_m X + v_s \sigma_I Y) P(X) P(Y) dX dY \int Z P(Z) dZ = \\ &= \sigma_{mw} \kappa_r(t) \langle \phi'(\mu_r, \Delta) \rangle + \sigma_{Iw} v_s(t) \langle \phi'(\mu_s, \Delta) \rangle \end{aligned} \quad (50)$$

where  $\mu_r = \kappa_r \sigma_m$ ,  $\mu_s = v_s \sigma_I$  and  $\Delta = (\kappa_r \sigma_m)^2 + (v_s \sigma_I)^2$ . In the last line we use the Gaussian integral notation:

$$\langle f(\mu, \Delta) \rangle = \int dx (2\pi)^{-\frac{1}{2}} \exp^{-x^2/2} f(\mu + \sqrt{\Delta}x) \quad (51)$$