Transferring State Representations in Hierarchical Spiking Neural Networks

Barna Zajzon[1,2,*], Renato Duarte[1], Abigail Morrison[1,3]
[1]Institute of Neuroscience and Medicine (INM-6), Institute for Advanced Simulation (IAS-6)
and JARA Institute Brain Structure-Function Relationships (JBI-1 / INM-10),
Jülich Research Centre, Jülich, Germany
[2] Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen University, Germany
[3]Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr-University Bochum, Germany
[*]E-mail: b.zajzon@fz-juelich.de

***Abstract*—Hierarchical modularity is a parsimonious design principle in many complex systems and underlies various key structural and functional aspects of neurobiological systems, whose modules are recurrent networks of spiking neurons. An essential requirement for such systems to adequately function is the ability to transfer information across multiple modules in a reliable and efficient manner. In this work, we study the characteristics of emergent stimulus representations in recurrent, spiking neural networks and the features that allow efficient information transfer among multiple, interacting sub-networks. We find that the specificity of structural mappings between the modules is strictly required for information to propagate to a sufficient depth, in a sequential setup. Conserved topography not only improves computational performance in all scenarios analyzed, but it proves to be more robust against noise and interference effects, results in less variability in the neural responses and increases memory capacity.**

***Index Terms*—stimulus representation, state transfer, modularity, spiking neural networks**

I. Introduction

To a first approximation, the mammalian neocortex can be seen as a large distributed hierarchy of interacting recurrent networks, whose rich high-dimensional dynamics subserves complex computational processes. Given this proof of principle, it is reasonable to expect that, when fed appropriate inputs and arranged in appropriate hierarchies, recurrent modules of spiking neural networks (SNN) may acquire tremendous processing capacity and efficiency, by hierarchically composing relevant stimulus features as structural rules at different temporal and spatial scales. Coherent conceptual accounts for spatiotemporal processing with SNNs already exist (e.g. [1]), but seldom consider the role of hierarchical modularity and focus entirely on local information processing in nonlinear dynamical systems. On the other hand, the computational benefits of complex hierarchical processing (hierarchical feature composition) have been largely demonstrated and exploited in the domain of artificial neural networks [2].

An essential first step towards more sophisticated processing across distributed hierarchies lies in the ability to retain and transfer adequate representations of the relevant stimulus features, in a usable manner that the subsequent processing stages can exploit. Propagation across distributed spiking networks has been studied, but typically devoid of functional context, i.e. while there have been a number of studies exploring the mechanisms underlying information transfer in spiking neural networks, they have mainly focused on the faithful propagation of spiking activity within one or across multiple neuronal populations without attributing functional meaning (information content) to such processes (but see, e.g. [3]). A systematic analysis of how meaningful information can be transferred in a computationally useful way in hierarchical spiking microcircuits remains relatively unexplored.

A fruitful way to address the issue while placing the focus on functional constraints lies in considering the spiking neural network to be a state-dependent processing reservoir whose high-dimensional transient dynamics supports information processing with fading memory. This framework, collectively termed reservoir computing (RC) and often referred to as Liquid State Machines [4] in the context of spiking networks, relies on the conceptual separation between a dynamic, excitable reservoir (a recurrent network seen as a nonlinear temporal expansion function) and a recurrence-free (memoryless) readout mechanism that linearly combines the high-dimensional population activity and extracts information from it, approximating a desired output. Through the effect of the nonlinear nodes and their recurrent interactions, the input is transformed in an expansive nonlinear embedding. Effectively, the reservoir projects a low dimensional input to a high dimensional feature space retaining time course information in the transient network responses. Importantly, apart from demonstrated computational proficiency, the RC framework relies on biologically plausible principles, consistent with the dynamics of information processing in real cortical microcircuits (see e.g. [5]).

In this work, we study the characteristics of emergent stimulus representations in recurrent, spiking neural networks and the features that allow efficient information transfer among multiple, interacting sub-networks. To exploit the complex, transient dynamics that such networks exhibit during active processing, we apply RC principles and probe the systems' computational performance with specific tasks. The goal of this work is to provide a relevant initial step towards modular, hierarchical processing with recurrent spiking networks.

978-1-5090-6014-6/18/$31.00 ©2018 IEEE

II. METHODS

Throughout this study, we explore the interactions between multiple recurrently coupled networks of spiking neurons, tuned to operate in a balanced, asynchronous irregular regime, under stationary conditions. We analyze the characteristics of stimulus representation and propagation and how the features of these sub-networks and the projections between them influence their representational capabilities.

A. Network Architecture

The systems analyzed consist of multiple sub-networks (layers) of sparsely and randomly coupled recurrent networks of excitatory (E) and inhibitory (I) neurons (see Figure 1A). Each sub-network is thus composed of $N = 10000$ leaky integrate-and-fire (LIF) neurons, sub-divided into $N^{\mathrm{E}} = 0.8N$ excitatory and $N^{\mathrm{I}} = 0.2N$ inhibitory populations (see subsection II-C). Recurrent connections (within layer) are randomly drawn with a fixed probability $\epsilon = 0.1$, common for all layers, i.e., on average, each neuron receives recurrent input from $K_{\mathrm{E}} = \epsilon N^{\mathrm{E}}$ excitatory and $K_{\mathrm{I}} = \epsilon N^{\mathrm{I}}$ inhibitory local synapses. The projections between the layers are considered, for simplicity, to be purely feed-forward and excitatory, i.e. the E_i population in layer L_i connects (with probability p_{ff}) to both populations E_j and I_j in a subsequent layer $L_j, \forall j > 1$, providing all neurons in layers L_j with an additional source of excitatory input, mediated via $K_{\mathrm{L_j}} = p_{\mathrm{ff}} N^{\mathrm{E}}$ synapses (see Figure 1). In addition, each neuron in each layer receives stochastic external input (background noise) from $K_{\mathrm{x}} = p_{\mathrm{x}} N^{\mathrm{x}}$ synapses, necessary to place the system in a responsive regime. We set $N^{\mathrm{x}} = N^{\mathrm{E}}$, as it is commonly assumed that the number of background input synapses modeling local and distant cortical input is in the same range as the number of recurrent excitatory connections.

The total input from sources external to each layer is scaled to ensure that all neurons (regardless of their hierarchical position) receive, on average, the same amount of excitatory drive, in order to preserve the operating points of the sub-networks. Whereas in the first (input) layer L_1, $p_{\mathrm{x}} = \epsilon$, for deeper layers, the connection densities are chosen such that $p_{\mathrm{ff}} + p_{\mathrm{x}} = \epsilon$, with $p_{\mathrm{ff}} = 0.75\epsilon$ and $p_{\mathrm{x}} = 0.25\epsilon$, yielding a ratio of 3:1 between the number of feed-forward and external synapses.

B. Feed-forward Connectivity

Additionally, the feed-forward projections among the different layers can be completely random or topographically structured, following a ubiquitous design principle in neocortical hierarchies. As such, in addition to random feed-forward connections (Figure 1A), we also investigate the effects of structured projections (Figure 1B) on the ability to adequately propagate information across the hierarchy. To integrate such topographic projections, a network with random recurrent and feed-forward connectivity (subsection II-A) is modified by randomly assigning sub-groups of stimulus-specific neurons in each layer and conserving this map across the hierarchy. To be more concrete, each stimulus S_k projects onto a (randomly chosen) subset of excitatory and inhibitory neurons in L_1, which we will denote E_k^1 and I_k^1 (typically comprising 800 and 200 neurons, respectively). The connections from E_k^1 to layer L_2 are then rewired such that these input-specific neurons project only to populations E_k^2 and I_k^2 while maintaining the average connection density p_{ff}, unless otherwise stated. Similarly to the first layer, these sub-populations are randomly assigned to receive projections corresponding to stimulus S_k. By repeating these steps throughout the hierarchy, we assure that each stimulus is propagated through a specific pathway. This is illustrated for stimulus S_1 in Figure 1B. Note that the stimulus-specific sub-populations in each layer might overlap since they are selected randomly, and that the overlap increases with the number of stimuli.

C. Neuron and Synapse Model

All neurons are modeled with the common leaky integrate-and-fire (LIF) formalism, with fixed voltage threshold and conductance-based synapses. The time course of the sub-threshold membrane potential V_i of a neuron i is given by:

$$C_{\mathrm{m}} \frac{dV_i}{dt} = g_{\mathrm{leak}}(V_{\mathrm{rest}} - V_i(t)) + I_i^{\mathrm{E}}(t) + I_i^{\mathrm{I}}(t) + I_i^{\mathrm{x}}(t) \quad (1)$$

where I_i^{E} and I_i^{I} represent the total excitatory and inhibitory synaptic input current, respectively. I_i^{x} represents external background input assumed, for simplicity, to be excitatory (all parameters equal to recurrent excitatory synapses), unspecific and stochastic (Poisson process), coming from sources firing at a fixed rate $\nu_{\mathrm{x}} = 5$ spk/s. Whenever the membrane potential V_i crosses a fixed firing threshold of $V_{\mathrm{th}} = -50$ mV, an action potential is emitted and the membrane potential reset to $V_{\mathrm{reset}} = -60$ mV for a fixed refractory period of $t_{\mathrm{ref}} = 2$ ms, after which integration is resumed as above.

When a pre-synaptic neuron j from population $\beta \in \{\mathrm{E}, \mathrm{I}\}$ fires, the current it induces in post-synaptic neuron i is given by:

$$I_{ij}(t) = g_{ij}(t)(V_\beta - V_i(t)) \quad (2)$$

where V_β is the reversal potential of the corresponding synapse, set to the values of $V_{\mathrm{E}} = 0$ mV and $V_{\mathrm{I}} = -80$ mV. The total amount of synaptic current that flows into neuron i at time t is thus the sum of the individual currents from all pre-synaptic sources.

Spike-triggered synaptic conductances $g_{ij}(t)$ are modeled as an instantaneous rise, followed by an exponential decay with time constant $\tau_{\mathrm{E}} = 5$ ms and $\tau_{\mathrm{I}} = 10$ ms for excitatory and inhibitory synapses, respectively:

$$\frac{dg_{ij}(t)}{dt} = -\frac{g_{ij}(t)}{\tau_\beta} + \bar{g}^\beta \sum_{t_j} \delta(t - t_j - d) \quad (3)$$

with conduction delays $d = 1.5$ ms for all synapse types. Peak conductances are specific for each synapse type and chosen so that each layer operates in a balanced, low-rate asynchronous irregular regime when driven solely by background input (following [6]). This parameter tuning (data not shown) resulted in $\bar{g}^{\mathrm{E}} = 1$ and $\bar{g}^{\mathrm{I}} = 16$, giving rise to average firing rates of around 3 Hz, $\mathrm{CV_{ISI}} \in [1.0, 1.5]$ and $\mathrm{CC} \leq 0.01$.

D. Stimulus classification

To assess how accurately each layer is able to acquire, represent and transmit information, we drive the system with a stimulus sequence $\{S_1, S_2, ...\} \in S$, of finite total length T, comprising $|S|$ different stimuli. Each stimulus consists of a set of 800 Poisson processes at a fixed rate $\nu_{\mathrm{stim}} = \lambda * \nu_{\mathrm{x}}$ (with $\lambda = 3$, unless otherwise stated) and fixed duration of 200 ms. These input neurons project to a random, but stimulus-specific sub-population of E and I neurons in the first layer L_1, which we will commonly refer to as the *input layer*.

In each layer L_i, the responses of the excitatory population are sampled at fixed time points t^* relative to stimulus onset (typically $t^* = 200$ ms, unless otherwise stated) and these activity vectors are gathered in a state matrix $X_{\mathrm{L_i}} \in \mathbb{R}^{N^{E \times T}}$, comprising a sample of population responses for each stimulus in the sequence. Where applicable, the measured responses are quantified using the low-pass filtered spike trains of the individual neurons by convolving them with an exponential kernel with $\tau = 20$ ms and temporal resolution equal to the simulation resolution, 0.1 ms. Otherwise, most of the analyses are performed considering the membrane potential V_{m} of the neurons as the primary state variable, as it is parameter-free and constitutes a more natural choice (see [7]).

To assess the classification accuracy in each layer, we determine the capacity to linearly combine the input-driven population responses to approximate the desired output [8]:

$$\hat{Y} = W^{\mathrm{out}} X \tag{4}$$

where $\hat{Y} \in \mathbb{R}^{r \times T}$ and $X \in \mathbb{R}^{N \times T}$ are the collection of all outputs and corresponding states over all time steps T, W^{out} is a $N^{E \times \mathrm{r}}$ matrix of output weights from the excitatory populations in each layer to their dedicated readout units, chosen to minimize the quadratic error between the readout's output $\hat{\mathbf{y}}$ and the desired, binary target values $\mathbf{y}$ (*training* the readout):

$$W^{\mathrm{out}} = Y X^T \left(X X^T + \beta I \right)^{-1} \tag{5}$$

The regularization parameter β penalizes solutions with large norm, essentially forcing the readout to avoid overweighting specific dimensions but instead gather information in the distributed population activity. The input sequence of length T is divided into a train ($0.8T$) and a test ($0.2T$) set. The former is used to train a set of r linear readouts to correctly classify the sequence of stimulus patterns. For each layer, the dimensionality of the readout layer matched the number of different stimuli to be classified, $r = |S|$, and the target output y is a binary representation of the stimulus sequence. The regularization parameter β is chosen by Leave-One-Out cross-validation on the training dataset.

Predicted stimulus labels are obtained by applying the winner-takes-all (WTA) operation on the readout outputs $\hat{y}$ for each stimulus in the test set. Average classification performance is then measured as the fraction of correctly classified patterns.

Numerical Simulations and Analyses

All numerical simulations were run using NMSAT[1] [9], a high-level wrapper for NEST customized for the application of reservoir computing principles to complex microcircuits. NEST version 2.12.0 [10] was used for all the numerical simulations.

III. Results

As a necessary first step towards more specialized processing, information about stimulus identity ought to be easily maintained in the input layer and propagated through the hierarchy. To assess the quality of emergent representations and the adequacy of information propagation, we employ a simple classification task and carefully analyze the responses to stimuli observed in the different layers.

A. Stimulus representation in sequential hierarchies

Fully random projections (Figure 1A) allowed maximum classification accuracy to be achieved in the first two layers, L_1 and L_2, with little or no variance across trials (Figure 1C). From layer L_3 onwards, however, there is a significant drop in classification accuracy, with a mean (over 10 trials) of ≈ 0.55. These results imply that despite the random connectivity between L_1 and L_2, the emerging population responses in L_2 are sufficiently discernible and unique to propagate to L_3 in a distinguishable manner, but not enough for an adequate classification performance or for further propagation downstream. This drop in representational accuracy in L_3 then hinders the ability to propagate stimulus information further and L_4 is entirely unable to distinctly represent the different stimuli (accuracy remains at chance level).

Incorporating topographic projections into the hierarchy (Figure 1B), whereby the neurons that receive direct stimulation at L_i connect exclusively to another set of stimulus-specific neurons in the subsequent layer, counteracts the effects observed with random mappings and leads to a considerable improvement in classification accuracy in the last two layers, i.e. stimulus information was accurately propagated to the deeper layers. This improvement suggests that stimulus-specific topographic maps play an essential role in successful propagation of signals across multiple interacting sub-populations.

Since computing the accuracy scores relies on a non-linear post-processing step (WTA, see subsection III-A), we additionally verify whether this operation significantly affects the results, by evaluating the mean squared error of the raw readout output $\hat{y}$ (MSE, Figure 1D). As expected, the error values are close to zero in the first layer and then increase with depth for both network setups, being significantly higher in L_3 and L_4 for random connections. The magnitude of the impact of topographic mappings is very noticeable and significant in the deeper layers, consistent with the corresponding accuracy scores.

[1] github.com/rcfduarte/nmsat

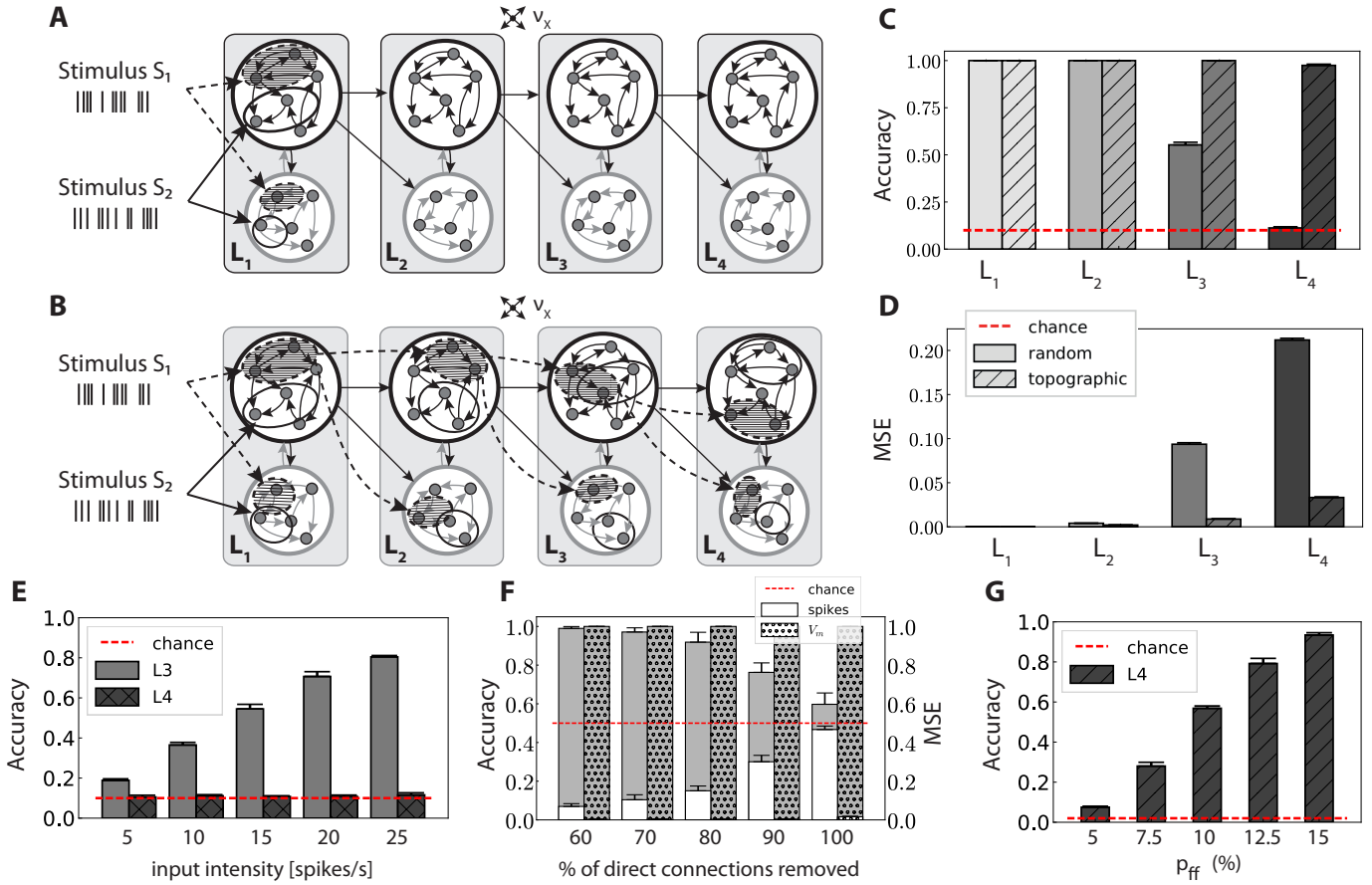

Fig. 1. Stimulus classification in sequential hierarchies. Hierarchies are composed of four identical layers, with random (**A**) or topographically structured (**B**) feed-forward connections from each E population (larger, black circles) to both E and I (smaller, gray circles) populations in the next layer (subsection II-A). Structured stimuli were used to drive specific sub-populations in L_1 as described in subsection II-D. For stimulus S_1, the topographic projections (**B**, dashed lines) are represented explicitly in addition to the corresponding stimulus-specific sub-populations (shaded ellipses), as detailed in subsection II-B. For S_2, only the sub-populations are depicted (solid lines, no shading). **C, D**: Mean classification accuracy and corresponding mean error for a stimulus set of $|S| = 10$ in each of the four sequentially connected layers averaged over ten trials , each using $T = 5000$ samples. Error bars show the standard deviations. **E**: Classification accuracy in L_3 and L_4 based on V_m as a function of the intensity of the input stimuli for ten different stimuli and ten trials per condition. **F**: Classification accuracy (gray) and MSE (white) for two stimuli as a function of the percentage of connections removed between neurons receiving direct stimulus input in L_1 and neurons in L_2. **G**: Classification accuracy in L_4 for 50 stimuli as a function of the connection probabilities within the topographic projections.

1) Modulating stimulus propagation: In this section, we study the impact of stimulus intensity and different connectivity properties on information transmission. For random networks, stimuli with larger input rates are more likely to be successfully propagated through the hierarchy, as reflected by the classification performance in Figure 1E. The stimulus intensity was varied between 5 and 25 $\mathrm{spk/s}$. For these values, the accuracy is always 1 in the first two layers (not shown). In L_3, the performance increases linearly with the stimulus intensity, up to a value of 0.8, but this improvement does not transfer to the last layer, where the results remain at chance level. This means that unstructured feed-forward connectivity does not allow signal propagation to reach a fourth layer, regardless of the intensity of the input stimulus and, surprisingly, regardless of the classification accuracy obtained in layer L_3. Note that due to the large scaling factor, for 20 $\mathrm{spk/s}$ and above, the activity already enters an unrealistically bursty regime with very high firing rates (data not shown).

Randomly connected networks provide no structured pathways to enable information transfer between the layers, yet stimuli can be read out as far as L_3. However, by construction, some neurons in the first layer that receive input stimulus directly also project (randomly) to the second layer. To assess the importance of these direct projections for signal propagation, we gradually remove them and measure the impact on the classification accuracy in L_2 (Figure 1F). The performance begins to drop only after half of the direct synapses are removed, and the decrease exclusively affects the low-pass filtered responses. This means that the system is fairly robust with respect to the loss of direct connections between the layers. Moreover, the populations in L_1 are able to create an internal representation of the input through their recurrent connections, and transfer this state representation to the next layer in a useful manner, at the level of membrane potentials

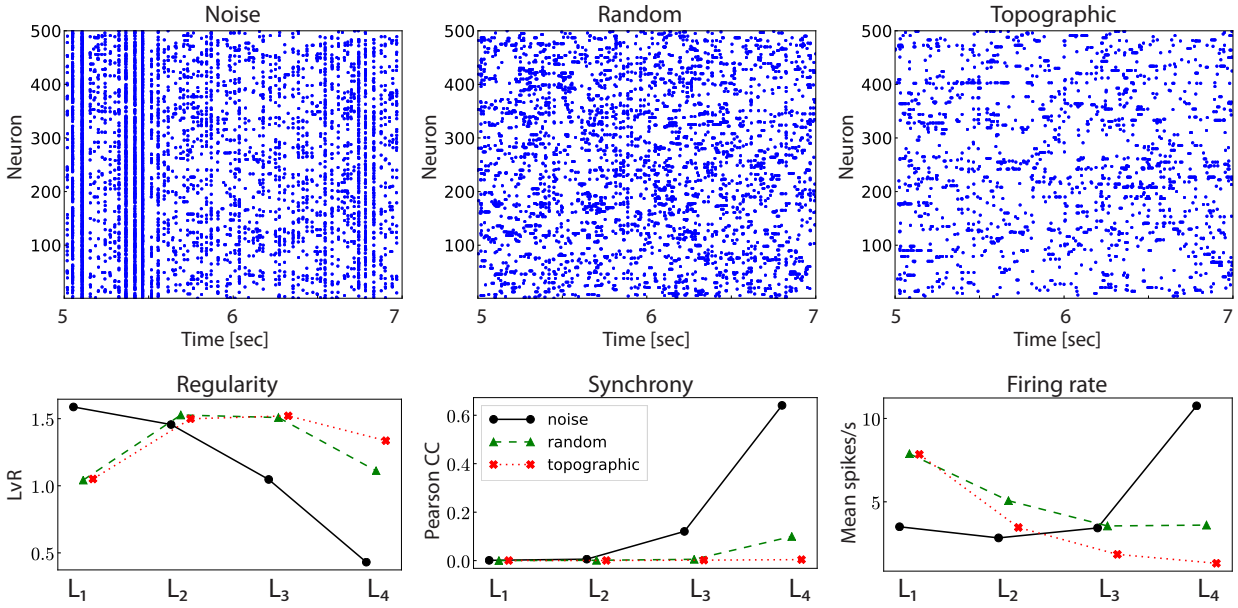

Fig. 2. Population spiking activity in three different scenarios: purely noise-driven (no stimulus), random feed-forward connections and topographic mappings. Top row shows the spiking activity of 500 randomly chosen excitatory neurons in L_3, over a period of two seconds. The corresponding statistics across the whole population are shown in the bottom row, for all the different layers: regularity (measured as the revised local variation, LvR, left), synchrony (Pearson's correlation coefficient, computed pairwise and averaged across 500 pairs, center) and mean firing rate across the excitatory populations (right).

(for which accuracy remains 1). For this state variable, the MSE increases only marginally, signaling robust responses. The different results obtained when considering spiking activity and sub-threshold dynamics indicate that the functional impact of recurrence is much more evident in the population membrane potentials.

The results discussed above involve at most ten stimuli, and the connection probability along the topographic maps was always fixed to $p_{\text{ff}} = 0.75\epsilon$, or 7.5%. However, in addition to stimulus intensity and directed connectivity, the density of these projections is bound to play a significant role in successful propagation. To quantify this effect, we test the ability of the last layer, L_4, to discriminate 50 different stimuli. As Figure 1G shows, this is possible, but the performance depends strongly on the density of topographic projections. Increasing the number of stimuli leads to a significantly lower performance for the same projection density , from ≈ 1 with ten stimuli (Figure 1C) to ≈ 0.3 with 50. Since the projections between the layers can overlap, more stimulus-specific pathways naturally lead to more overlap between these regions, causing less discriminable responses. However, this seems to be compensated by increasing connection density, and thus strengthening the connectivity along the topographic maps.

2) Population activity: The ability to internally represent and transfer information in a usable manner is determined by the macroscopic features that characterize population activity in the different scenarios (Figure 2), which in turn, is a direct consequence of the manner in which excitatory inputs drive the different sub-networks. When driven only by the Poissonian background noise, the activity in the first two layers is asynchronous and irregular, but evolves into a more synchronous regime in L_3 (see noise condition in Figure 2, bottom center), and in L_4 the system's operating point is entirely switched to a synchronous regular regime, which is known to hinder information processing [6]. The excessive synchronization in the last layer is also responsible for the higher firing rate of ≈ 10 spk/s and is mainly a consequence of an increase in shared pre-synaptic inputs. As we move to deeper layers, the feed-forward projections increase the convergence and the post-synaptic responses. This moves the network activity away from Poissonian statistics, an effect that accumulates from layer to layer and gradually skews the population activity towards states of increased synchrony.

Compared to baseline activity, the presence of a patterned stimulus increases the irregularity in all layers except the very first one, and reduces the synchrony in the last two layers significantly, allowing the system to globally maintain the asynchronous irregular regime. In line with the performance results, the underlying population activity demonstrates the computational benefits of operating in such states. Furthermore, random projections are not enough to entirely overcome the shared-input effect in L_4 (chararacterized by increased synchrony, CC ≈ 0.12), but topographic maps are, maintaining an asynchronous firing profile even in the deeper layers. The lower firing rates in networks with topography can be explained by the fact that neurons receiving direct stimulus input in L_1, having high firing rates, only project to a restricted group of neurons in the subsequent layers, thus having a smaller impact on the average population activity downstream. This implies that topography enables both more accurate and more resource-efficient information transmission, achieving

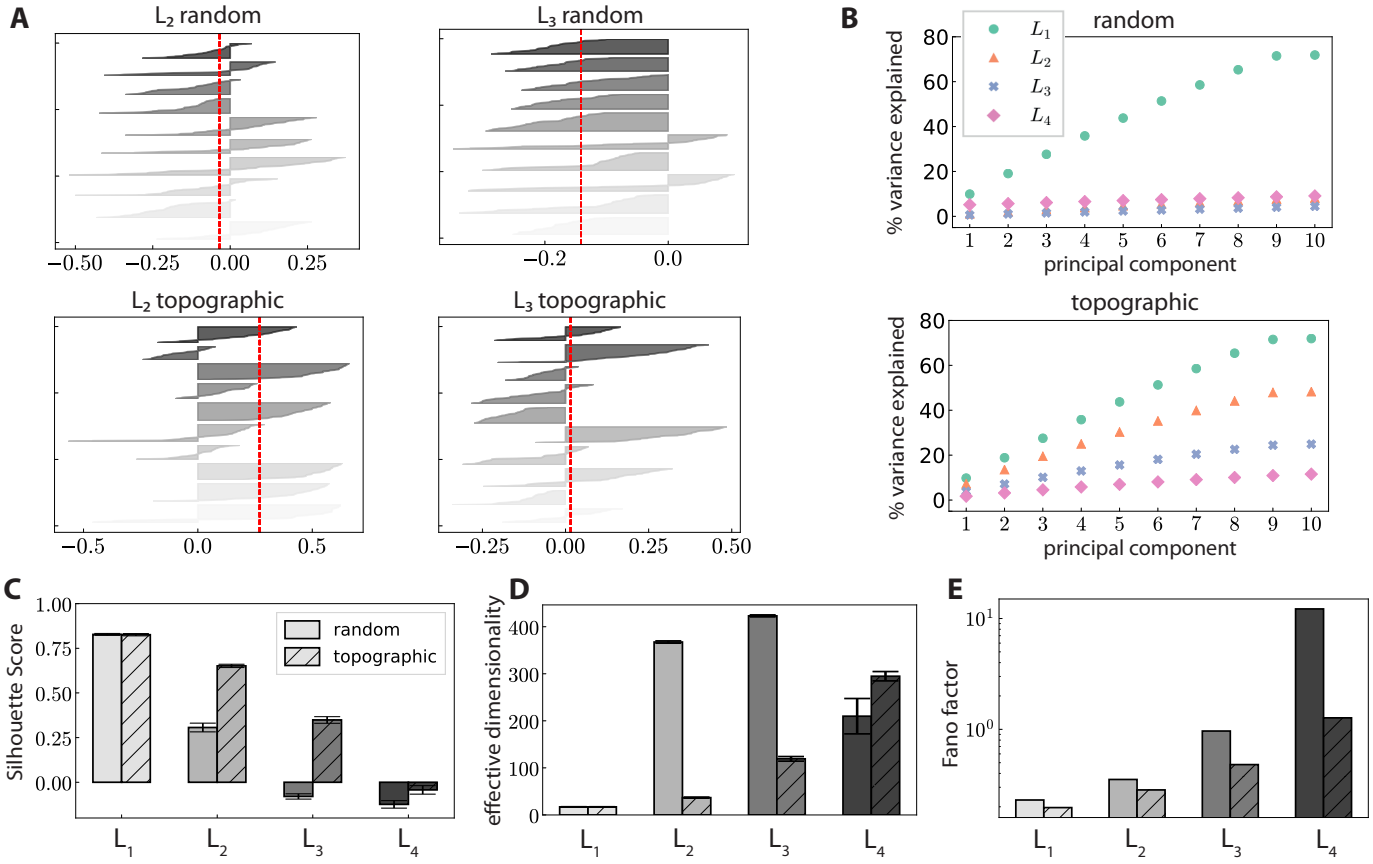

Fig. 3. Spatial clustering of stimulus-specific state vectors. **A**: Distribution of silhouette coefficients for the stimulus-specific clusters in layers L_2 and L_3, color-coded for the 10 different stimuli used (500 samples). The coefficients are sorted in descending order for each stimulus. The vertical lines in red represent the mean over all coefficients (silhouette score) in one trial. **B**: Cumulative variance explained by the first ten principal components for random (top) and topographic (bottom) maps. **C-E**: For each layer in random and topographic maps: **C** Mean silhouette score (and standard deviation) **D** Effective dimensionality of the state matrix. **E** Variability of neural responses measured by the Fano factor (FF) on the population-averaged firing rates during a single trial of 10 seconds (bin size 10 ms). With the exception of **E**, all other statistics are averaged over ten trials, each lasting 100 seconds, with ten different stimuli as input.

better performance for lower overall network activity.

3) State separability: Adequate input-state mappings are reflected in the geometric organization of the circuit's state-space, whereby different stimuli are mapped to different, stimulus-specific sub-spaces (the *separation property*, see [4]). In this section, we evaluate the quality of these mappings as the representations are transferred from layer to layer.

The differences in the representational performance of the two connectivity profiles explored (random versus topographic) can be understood by the degree of uniqueness in the elicited stimulus-specific responses. Ideally, patterned external stimuli would allow the state-space to be partitioned into clearly segregated stimulus-specific state-vector clusters, i.e. population activity in response to different stimuli flows along well segregated sub-spaces. We assess the degree of clustering by analyzing the characteristics of a low-dimensional projection (obtained through principal component analysis) and evaluating how close each data point (state vector projected onto the space spanned by the first 3 PCs) in one stimulus-specific cluster is to points in neighboring clusters. Each labeled sample of population activity is thus given a silhouette coefficient (SC, Figure 3A) ranging between $[-1, 1]$. A coefficient value close to 1 indicates that the data point is close to the mean of its assigned cluster (stimulus label), near 0 points to partially overlapping clusters, while negative values imply wrong cluster assignment. This measurement thus allows us to quantify the compactness of stimulus-dependent state vector clusters, and observe a decay in the clustering quality throughout the hierarchy. Starting with L_2, there is a consistent difference of about 0.15 in the silhouette score between the random and topographic setups. The decay is not reflected uniformly for all stimuli, but the trend is consistent.

It is worth noting that the silhouette plots were computed for a single trial and rely only on clustering the space spanned by the first three PCs. To obtain a more representative result, we repeat the analysis for multiple trials (and considering the first ten principal components) and summarize the results using the silhouette score (Figure 3C). These results reveal a clear disparity in the clustering quality for random and topographic networks, in accordance with the classification performances (Figure 1).

To further quantify how efficiently the networks use their

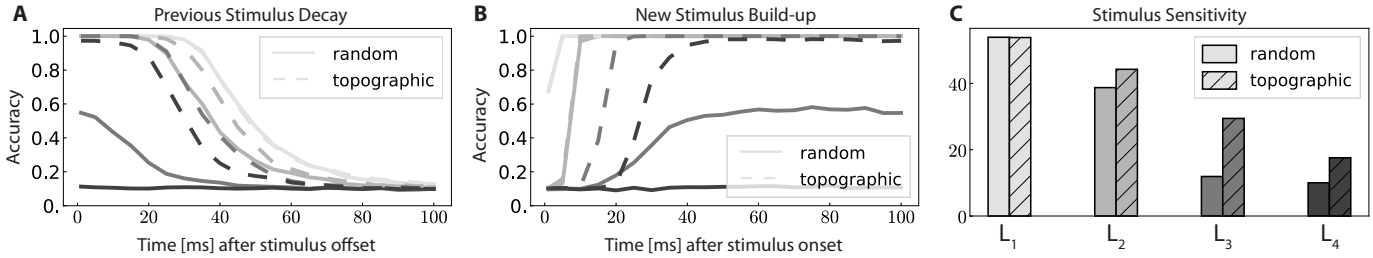

Fig. 4. Changes in the state representations over time, as reflected by classification accuracy. For an offset of the previous stimulus and onset of the new stimulus at $t = 0$, readout accuracy is shown for: (**A**) previous stimulus (**B**) new stimulus. Shown is the mean accuracy score over 5 trials, with linear interpolation of sampling offsets $t_{\text{samp}} = 1, 5, 10, 15, ..., 100$ ms, with solid and dashed curves representing networks with random connectivity and topography respectively. Curves coloured (in grayscale) according to layers, key in **C**. The stimulus sensitivity (**C**) is defined as the area below the intersection of corresponding curves from **A** and **B**.

high-dimensional state space, we evaluate how many PCs are necessary to account for most of the variance in the data (Figure 3B, D). In the first layer, the variance captured by each subsequent PC is almost constant ($\approx 10\%$), and reaches around 70% by the 9th PC (Figure 3B), indicating that population activity lies in a very low-dimensional sub-space (Figure 3D) where the stimulus impact is strongest, "enslaving" the activity and creating narrower, stimulus-specific trajectories. For random networks, however, this trend is not reflected in the subsequent layers (where the first 10 PCs account for less than 10% of the total variance), indicating a significant increase in effective dimensionality with hierarchical depth (Figure 3D), a pattern which also exists, to a much lesser extent, in the topographic case.

As we traverse the hierarchy, population activity becomes less constrained, leading to the exploration of a larger region of the state-space. Whereas this tendency is consistent and more gradual for topographic networks, it is faster in random networks, suggesting a quicker dispersion of the stimulus representations. Moreover, since the stimulus does not reach the last layer with random mappings, the activity tends to become more synchronous, resulting in a reduced effective dimensionality. These observations are in line with the classification performance: the effective dimensionality (Figure 3D) and variability (Figure 3E) of the neural responses are inversely related to the classification accuracy, and both increase with hierarchical depth whereas performance drops.

B. Memory capacity and stimulus sensitivity

While the ability to discriminate the identity of the current stimulus and adequately propagate this information is critically important for distributed processing across multiple modules, it is equally important to understand how information is retained within these modules and how the systems deal with interference effects from sequentially presented stimuli.

In this section, we study the time course of stimulus representation, measuring the memory capacity, i.e. how long stimulus information is retained in each sub-network through reverberations of the current state; also how quickly the network reacts in the presence of a new stimulus as well as the potential interference between past and present inputs.

We quantify these properties by evaluating classification accuracy across time: Let $t_i^0 = i * 200$ ms represent the offset of stimulus i, with $1 \leq i \leq T$. The network responses are sampled at intervals of $t^* = t_i^0 + t_{\text{samp}}$ and collected in the state matrix (see subsection III-A), a process repeated for different sampling offsets, t_{samp}. For each t_{samp}, the corresponding state matrix is used to train two readouts: one on the labels of the previous stimuli, and a second one on the current stimulus label. This procedure leads to the memory decay and build-up curves displayed in Figure 4. The degree of overlap between the two curves for a given network configuration indicates how long the system is able to retain useful information about both the previous and the present stimuli (Figure 4C). With this analysis we can measure three important properties of the hierarchical system: how long stimulus information is retained in the network activity; how long the network requires to accumulate sufficient evidence to classify the present input; and what are the interference effects of multiple stimuli.

Stimulus information and input representations gradually disappear after stimulus offset (*fading memory*), a characteristic illustrated by the decay in the classification accuracy measured at increasing delays after their offset (Figure 4A). For computational reasons, only the first 100 ms are plotted, but the decreasing trend in the accuracy continued and invariably reached the chance level within the first 150 ms, indicating that these networks have a rather short memory capacity, unable to span multiple input elements. In any case, it is evident that the ability to memorize stimulus information decays with hierarchical depth (Figure 4A). The higher overall accuracy achieved by topographic networks is again reflected in the memory curves, demonstrating its important functional consequences.

In addition, stimulus representations also appear to build up over exposure time (Figures 4B). A common trend is the slower build-up of information with depth, more pronounced for random mappings but clearly observable in the topographic case as well, reflecting an unsurprising cumulative delay. Furthermore, topography enables a faster information build-up beginning with L_3.

We subsequently determine the *stimulus sensitivity* of a

population, as the extent of time where useful non-interfering representations are retained in each sub-network. This is calculated as the area below the intersection of its memory and build-up curves, representing the previous and current stimulus, respectively. Following a similar trend to performance and memory, sensitivity to stimulus decreases through the hierarchy and the existence of structured propagation pathways leads to clear benefits, particularly pronounced in the deeper layers.

In summary, deeper layers forget faster and take longer (than the inter-layer delays) to build up stimulus representations. No population is able to represent two stimuli accurately for a significant amount of time (longer than 100 ms), although topographic maps improve memory capacity and stimulus sensitivity.

IV. Discussion

The hierarchical and modular nature of the cortex requires efficient and reliable transmission of information among the various cortical regions to allow for real-time interaction with the external environment. In this study, we investigated the properties of emerging state representations in multiple, interacting populations of spiking neurons, and examined the temporal dynamics of the information transferred between them. We tested the computational capabilities of the system with linear classification tasks in the context of reservoir computing. A critical aspect of this work was the comparison of random connectivity and biologically inspired topographic projections.

Our results suggest that while random connectivity might be applied successfully at a local population level, signal propagation over longer distances and across several cortical modules requires topographic precision for accurate, robust and reliable transmission. Such might be the case for the early sensory systems, where real-time computation is crucial and where the existence of topographic maps is well supported by anatomical studies. An important aspect of topographic maps in the brain, outside the scope of our current study, concerns the decrease in topographic specificity along the hierarchy, due to increasingly overlapping projections stemming partly from larger map sizes, which is assumed to have an important computational role (see e.g. [11]).

With random feed-forward connectivity, information could be decoded up to the third layer but not beyond (Figure 1), due to a decrease in the stimulus-specific tuning with hierarchical depth (Figure 3). This effect is much less prominent when topographic projections are used; in that condition, accuracy remains close to maximum in all four layers. This suggests, that at least in our simple model, topographic precision is necessary to accurately transmit stimulus information. Topography also prevents the development of synchronous regimes (Figure 2) in the deeper layers due to the shared-input effect, since the stimulus propagates more robustly through the layers.

The sequences of network states are trajectories in an N-dimensional Euclidean space. However, effectively, the system's dynamics (particularly when driven by an external stimulus) flows along compact sub-regions, and is constrained to operate in a low-dimensional sub-space. Intuitively, a reduction in dimensionality means that the trajectory induced by a stimulus explores only a stereotypical sub-space and reflects the spatiotemporal features of the input. The effective dimensionality of the explored state space was lowest in the first layer where the input is strongest, and increased with hierarchical depth as the stimulus information fades (Figure 3). The rate of increase is significantly higher for random connections, indicating that topography entrains the neural responses to lower-dimensional stereotypical trajectories. This behavior is also reflected by the variability of the neuronal activity, with random connectivity leading to less consistent responses after the second layer.

Overall, the insights gained from this work could provide useful constraints for building hierarchical systems composed of spiking balanced networks that enable accurate information transmission. Whereas random connectivity is sufficient for information transfer between two layers, more complex systems might benefit from some form of structured connectivity. Topography was found not only to improve computational performance in all scenarios analyzed, but it proved to be more robust against noise (data not shown) and interference effects (Figure 4), due to the spatial segregation of the projections. Additionally, it led to a reduction in the variability of the neural responses (Figure 3), prevented the development of synchronous regimes in the deeper layers (Figure 2), and also slightly increased memory capacity (Figure 4).

The work presented here is preliminary and exploratory in nature and thus suffers from a number of limitations. For simplicity reasons, only relatively simple and homogeneous networks were considered. However, there is a whole host of relevant phenomena that is bound to significantly impact the network's behavior and computational performance [12], particularly when considered in a hierarchical setting. Heterogeneity and variability within and across the different modules may have a critical role to play in differentially modulating the circuits' operating points. For example, the variations in synaptic strengths, transmission delays and/or response kinetics have been hypothesized to underlie temporal specialization in neocortical hierarchies [13] and may thus modulate the circuit's memory capacity and intrinsic time constants.

Thus, the lack of timescale diversity and the short memory spans we observed in the deeper layers could be improved upon by, for example, incorporating synaptic mechanisms that operate over longer time scales. Interestingly, our results stand in contrast with the findings in [14], where authors report the development of longer timescales in the deeper layers of hierarchical Echo State Networks (ESNs) [15], despite using fixed parameters throughout the hierarchy. This difference suggests that the hierarchical organization of processing timescales may not stem merely from recurrence (the main commonality between the two approaches), but from the dynamical properties of the reservoirs.

Thus, it is worth investigating the components that give

rise to multiple timescales in hierarchical spiking networks, as known from experimental data, along with their influence on the network's functional capabilities. Future work should also focus on including multiple and different layers of biological realism, and at the same time explore computationally more complex tasks. From a computer science perspective, hybrid deep architectures composed of convolutional (spiking) layers for pre-processing intertwined with randomly connected reservoirs for feature discovery may represent a promising research direction.

ACKNOWLEDGMENTS

We acknowledge partial support by the Erasmus Mundus Joint Doctoral Program EuroSPIN, the German Ministry for Education and Research (BMBF Grant 01GQ0420 to BCCN Freiburg), the Initiative and Networking Fund of the Helmholtz Association and the Helmholtz Portfolio theme "Supercomputing and Modeling for the Human Brain".

REFERENCES

[1] W. Maass, T. Natschläger, and H. Markram, "Fading memory and kernel properties of generic cortical microcircuit models," *Journal of Physiology Paris*, vol. 98, no. 4-6 SPEC. ISS., pp. 315–330, 2004.

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[3] T. P. Vogels and L. F. Abbott, "Gating multiple signals through detailed balance of excitation and inhibition in spiking networks," *Nature Neuroscience*, vol. 12, no. 4, pp. 483–491, 2009.

[4] W. Maass, T. Natschläger, and H. Markram, "Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations," *Neural Computation*, vol. 14, no. 11, pp. 2531–2560, 2002.

[5] P. Enel, E. Procyk, R. Quilodran, and P. F. Dominey, "Reservoir Computing Properties of Neural Dynamics in Prefrontal Cortex," *PLoS Computational Biology*, vol. 12, no. 6, p. e1004967, 2016.

[6] R. C. Duarte and A. Morrison, "Dynamic stability of sequential stimulus representations in adapting neuronal networks," *Front Comput Neurosci*, vol. 8, no. October, p. 124, 2014.

[7] D. van den Broek, M. Uhlmann, H. Fitz, R. Duarte, P. Hagoort, and K. M. Petersson, "The best spike filter kernel is a neuron," 2017.

[8] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Computer Science Review*, vol. 3, no. 3, pp. 127–149, 2009.

[9] R. Duarte, B. Zajzon, and A. Morrison, "Neural Microcircuit Simulation And Analysis Toolkit," 2017.

[10] S. Kunkel, A. Morrison, P. Weidel, J. M. Eppler, A. Sinha, W. Schenck, M. Schmidt, S. B. Vennemo, J. Jordan, A. Peyser, D. Plotnikov, S. Graber, T. Fardet, D. Terhorst, H. Mørk, G. Trensch, A. Seeholzer, R. Deepu, J. Hahne, I. Blundell, T. Ippen, J. Schuecker, H. Bos, S. Diaz, E. Hagen, S. Mahmoudian, C. Bachmann, M. E. Lepperød, O. Breitwieser, B. Golosio, H. Rothe, H. Setareh, M. Djurfeldt, T. Schumann, A. Shusharin, J. Garrido, E. B. Muller, A. Rao, J. H. Vieites, and H. E. Plesser, "NEST 2.12.0," 2017.

[11] M. Rigotti, O. Barak, M. R. Warden, X. J. Wang, N. D. Daw, E. K. Miller, and S. Fusi, "The importance of mixed selectivity in complex cognitive tasks," *Nature*, vol. 497, no. 7451, pp. 585–590, 2013.

[12] R. Duarte and A. Morrison, "Leveraging heterogeneity for neural computation with fading memory in layer 2/3 cortical microcircuits," *bioRxiv*, p. 230821, 2017.

[13] R. Duarte, A. Seeholzer, K. Zilles, and A. Morrison, "Synaptic patterning and the timescales of cortical dynamics," *Current Opinion in Neurobiology*, vol. 43, pp. 156–165, 2017.

[14] C. Gallicchio, A. Micheli, and L. Pedrelli, "Deep reservoir computing: A critical experimental analysis," *Neurocomputing*, vol. 268, pp. 87–99, 2017.

[15] H. Jaeger, "Adaptive Nonlinear System Identification with Echo State Networks," in *Proceedings of the 15th International Conference on Neural Information Processing Systems*, ser. NIPS'02. Cambridge, MA, USA: MIT Press, 2002, pp. 609–616.