Check for
updates

Passing the Message: Representation Transfer in Modular Balanced Networks

Barna Zajzon [1,2]*, Sepehr Mahmoudian [3,4], Abigail Morrison [1,5] and Renato Duarte [1]

[1] Jülich Research Centre, 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, Germany, [2] Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen University, Aachen, Germany, [3] Department of Data-Driven Analysis of Biological Networks, Campus Institute for Dynamics of Biological Networks, Georg August University Göttingen, Göttingen, Germany, [4] MEG Unit, Brain Imaging Center, Goethe University, Frankfurt, Germany, [5] Faculty of Psychology, Institute of Cognitive Neuroscience, Ruhr-University Bochum, Bochum, Germany

Neurobiological systems rely on hierarchical and modular architectures to carry out intricate computations using minimal resources. A prerequisite for such systems to operate adequately is the capability to reliably and efficiently transfer information across multiple modules. Here, we study the features enabling a robust transfer of stimulus representations in modular networks of spiking neurons, tuned to operate in a balanced regime. To capitalize on the complex, transient dynamics that such networks exhibit during active processing, we apply reservoir computing principles and probe the systems' computational efficacy with specific tasks. Focusing on the comparison of random feed-forward connectivity and biologically inspired topographic maps, we find that, in a sequential set-up, structured projections between the modules are strictly necessary for information to propagate accurately to deeper modules. Such mappings not only improve computational performance and efficiency, they also reduce response variability, increase robustness against interference effects, and boost memory capacity. We further investigate how information from two separate input streams is integrated and demonstrate that it is more advantageous to perform non-linear computations on the input locally, within a given module, and subsequently transfer the result downstream, rather than transferring intermediate information and performing the computation downstream. Depending on how information is integrated early on in the system, the networks achieve similar task-performance using different strategies, indicating that the dimensionality of the neural responses does not necessarily correlate with nonlinear integration, as predicted by previous studies. These findings highlight a key role of topographic maps in supporting fast, robust, and accurate neural communication over longer distances. Given the prevalence of such structural feature, particularly in the sensory systems, elucidating their functional purpose remains an important challenge toward which this work provides relevant, new insights. At the same time, these results shed new light on important requirements for designing functional hierarchical spiking networks.

Keywords: modularity, information transfer, reservoir computing, spiking neural networks, topographic maps

1. INTRODUCTION

Cortical information processing relies on a distributed functional architecture comprising multiple, specialized modules arranged in complex, but stereotyped networks (see, e.g., Felleman and Van Essen, 1991; Markov and Kennedy, 2013; Park and Friston, 2013). Structural organizational principles are noticeable at different scales and impose strong constraints on the systems' functionality (Duarte et al., 2017), while simultaneously suggest a certain degree of uniformity and a close relation between structure and function (Mountcastle, 1978, 1997).

On the lower levels of cortical processing, peripheral signals conveying sensory information need to be adequately routed, their content represented and integrated with internal, ongoing processes (Duarte, 2015) (based on both local and long-range interactions) as well as non-sensory signals such as attention (Macaluso et al., 2000), expectation (Keller et al., 2012), or reward (Shuler and Bear, 2006). A prerequisite for processing across such large distributed systems is therefore the ability to suitably represent relevant stimulus features, and transfer these representations in a reliable and efficient manner through various processing modules. Additionally, cortical areas are arranged in a functional hierarchy (Markov and Kennedy, 2013; Murray et al., 2014; Miller, 2016), whereby *higher*, more anterior, regions show sensitivity to increasingly complex and abstract features.

The computational benefits of such hierarchical feature aggregation and modular specialization have been consistently demonstrated in the domain of artificial neural networks (LeCun et al., 2015), with a primary focus on spatial and/or spectral features. However, given that, to a first approximation, cortical systems are recurrent networks of spiking neurons, temporal dynamics, and the ability to continuously represent and process spatio-temporal information are fundamental aspects of neural computation. The majority of previous studies on spatio-temporal processing with spiking neural networks have either focused on local information processing without considering the role of, or mechanisms for, modular specialization (e.g., Maass et al. 2004), or on the properties of signal transmission within one or across multiple neuronal populations regardless of their functional context (Diesmann et al. 1999; van Rossum et al. 2002; Kumar et al. 2008, 2010; Shadlen and Newsome 1998; Joglekar et al. 2018, but see, e.g., Vogels and Abbott 2005, 2009 for counter-examples).

In order to quantify transmission accuracy and, implicitly, information content, these studies generally look either at the stable propagation of synchronous spiking activity (Diesmann et al., 1999) or asynchronous firing rates (van Rossum et al., 2002). The former involves the temporally precise transmission of pulse packets (or spike volleys) aided by increasingly synchronous responses in multi-layered feed-forward networks (so-called "synfire chains"); the later refers to the propagation of asynchronous activity and assumes that information is contained and forwarded in the fidelity of the firing rates of individual neurons or certain sub-populations. An alternative approach was recently taken by Joglekar et al. (2018), in which signal propagation was analyzed in a large-scale cortical model and elevated firing rates across

areas were considered a signature of successful information transmission. However, no transformations on the input signals were carried out. Thus, a systematic analysis that considers both computation within a module and the transmission of computational results to downstream modules remains to be established.

In this study, we hypothesize that biophysically-based architectural features (modularity and topography) impose critical functional constraints on the reliability of information transmission, aggregation, and processing. To address some of the issues and limitations highlighted above, we consider a system composed of multiple interconnected modules, each of which is realized as a recurrently coupled network of spiking neurons, acting as a state-dependent processing reservoir whose high-dimensional transient dynamics supports online computation with fading memory, allowing simple readouts such as linear classifiers to learn a large set of input-output relations (Maass et al., 2002). Through the effect of the nonlinear nodes and their recurrent interactions, each module projects its inputs to a high dimensional feature space retaining time course information in the transient network responses. By connecting such spiking neural network modules, we uncover the architectural constraints necessary to enable a reliable transfer of stimulus representations from one module to the next. Using such a *reservoir computing* (RC) approach (Lukoševičius and Jaeger, 2009), the transmitted signals are conferred functional meaning and the circuits' information processing capabilities can be probed in various computational contexts. Preliminary results from this approach have been presented in a conference proceedings (Zajzon et al., 2018) and a preprint version of this manuscript has been released at Zajzon et al. (2019).

Our results demonstrate that the connectivity structure between the modules strongly affects the transmission efficacy. We contrast random projections with biologically-inspired topographic maps, which are particularly prominent in sensory systems (Kaas, 1997) and have been associated with a variety of important functional roles, ranging from information segregation and transmission along sensory pathways (Silver et al., 2005; Harris and Shepherd, 2015) to spatio-temporal feature aggregation (Hagler and Sereno, 2006). Additionally, conserved topography was shown to support the development of stable one-to-one mappings between abstract cognitive representations in higher cortical regions (Thivierge and Marcus, 2007). We show that incorporating such structured projections between the modules facilitates the reliable transmission of information, improving the overall computational performance. Such ordered mappings lead to lower-dimensional neural responses, allowing a more stable and efficient propagation of the input throughout the modules while enabling a computationally favorable dynamic regime. These results suggest that, while random connectivity can be applied for local processing within a module or between a few populations, accurate and robust information transmission over longer distances benefits from spatially segregated pathways and thus offers a potential functional interpretation for the existence of conserved topographic maps patterning cortical mesoscopic connectivity.

2. MATERIALS AND METHODS

2.1. Network Architecture

We model systems composed of multiple sub-networks (modules). Each module is a balanced random network (see, e.g., Brunel 2000), i.e., a sparsely and randomly connected recurrent network containing $N = 10,000$ leaky integrate-and-fire neurons (described below), sub-divided into $N^E = 0.8N$ excitatory and $N^I = 0.2N$ inhibitory populations. Neurons make random recurrent connections within a module with a fixed probability common for all modules, $\epsilon = 0.1$, such that on average each neuron in every module receives recurrent input from $K_E = \epsilon N^E$ excitatory and $K_I = \epsilon N^I$ inhibitory local synapses.

For simplicity, all projections between the modules are considered to be purely feed-forward and excitatory. Specifically, population E_i in module M_i connects, with probability p_{ff}, to both populations E_{i+1} and I_{i+1} in subsequent module M_{i+1}. This way, every neuron in M_{i+1} receives an additional source of excitatory input, mediated via $K_{M_{i+1}} = p_{ff}N^E$ synapses (see **Figure 1**).

To place the system in a responsive regime, all neurons in each module further receive stochastic external input (background noise) from $K_x = p_x N^x$ synapses. We set $N^x = N^E$, 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 (Brunel, 2000; Kumar et al., 2008; Kremkow et al., 2010).

In order to preserve the operating point of the different sub-networks, we scale the total input from sources external to each module to ensure that all neurons (regardless of their position in the setup) receive, on average, the same amount of excitatory drive. Whereas, $p_x = \epsilon$ holds in the first (input) module, M_0, the connection densities for deeper modules are chosen such that $p_{ff} + p_x = \epsilon$, with $p_{ff} = 0.75\epsilon$ and $p_x = 0.25\epsilon$, yielding a ratio of 3:1 between the number of feed-forward and background synapses.

For a complete, tabular description of the models and model parameters used throughout this study (see **Tables S1, S2**).

2.2. Structured Feed-Forward Connectivity

We explore the functional role of long-range connectivity profiles by investigating and comparing networks with random (**Figure 1A**) and topographically structured feed-forward projections (**Figure 1B**).

To build systems with topographic projections in a principled, but simple, manner, a network with random recurrent and feed-forward connectivity (as described in the previous section) is modified by systematically assigning sub-groups of stimulus-specific neurons in each module. Each of these then connects only to the corresponding sub-group across the different modules. More specifically, each stimulus S_k projects onto a randomly chosen subset of 800 excitatory and 200 inhibitory neurons in M_0 (input module), denoted E_0^k and I_0^k. The connections from E_0^k to module M_1 are then rewired such that neurons in E_0^k project, with probability p_{ff}, exclusively to similarly chosen stimulus-specific neurons E_1^k and I_1^k. These sub-populations in M_1 thus extend the topographic map associated with stimulus S_k. By repeating

these steps throughout the system, we ensure that each stimulus is propagated through a specific pathway while inter-module projections from neurons not belonging to any topographic map remain unchanged (random). This connectivity scheme is illustrated for stimulus S_1 in **Figure 1B**.

It is worth noting that, as the stimulus-specific sub-populations are randomly chosen, overlaps occur (depending on the total number of stimuli). By allowing multiple feed-forward synaptic connections between neurons that are part of different clusters, the effective connection density along the topographic maps (p_{ff}) is slightly increased compared with the random case (from 0.075 to 0.081). Any given neuron belongs to at most three different maps, ensuring that information transmission is not heavily biased by only a few strong connections. The average overlap between maps, measured as the mean fraction of neurons shared between any two maps, was 0.61. These values are representative for all sequential setups, unless stated otherwise.

2.3. Neuron and Synapse Model

The networks are composed of leaky integrate-and-fire (LIF) neurons, with fixed voltage threshold and conductance-based, static synapses. The dynamics of the membrane potential V_i for neuron i follows:

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

where the leak-conductance is given by g_{leak}, and I_i^E and I_i^I represent the total excitatory and inhibitory synaptic input currents, respectively. We assume the external background input, denoted by I_i^x, to be excitatory (all parameters equal to recurrent excitatory synapses), unspecific and stochastic, modeled as a homogeneous Poisson process with constant intensity $\nu_x = 5$ Hz. Spike-triggered synaptic conductances are modeled as exponential functions, with fixed and equal conduction delays for all synapse types. The equations of the model dynamics, along with the numerical values for all parameters are summarized in **Tables S1, S2**.

Following Duarte and Morrison (2014), the peak conductances were chosen such that the populations operate in a balanced, low-rate asynchronous irregular regime when driven solely by background input. For this purpose, we set $\bar{g}^E = 1$ nS and $\bar{g}^I = 16$ nS, giving rise to average firing rates of ~ 3 Hz, $CV_{ISI} \in [1.0, 1.5]$ and $CC \leq 0.01$ in the first two modules of the networks, as described in the previous sections.

2.4. Stimulus Input and Computational Tasks

We evaluate the information processing capabilities of the different networks on simple linear and nonlinear computational tasks. For this purpose, the systems are presented with a sequence of stimuli $\{S_1, S_2, ...\} \in S$, of finite total length T and comprising $|S|$ different stimuli.

Each stimulus consists of a set of 800 Poisson processes at a fixed rate $\nu_{stim} = \lambda * \nu_x$ and fixed duration of 200 ms, mimicking sparse input from an external population of size
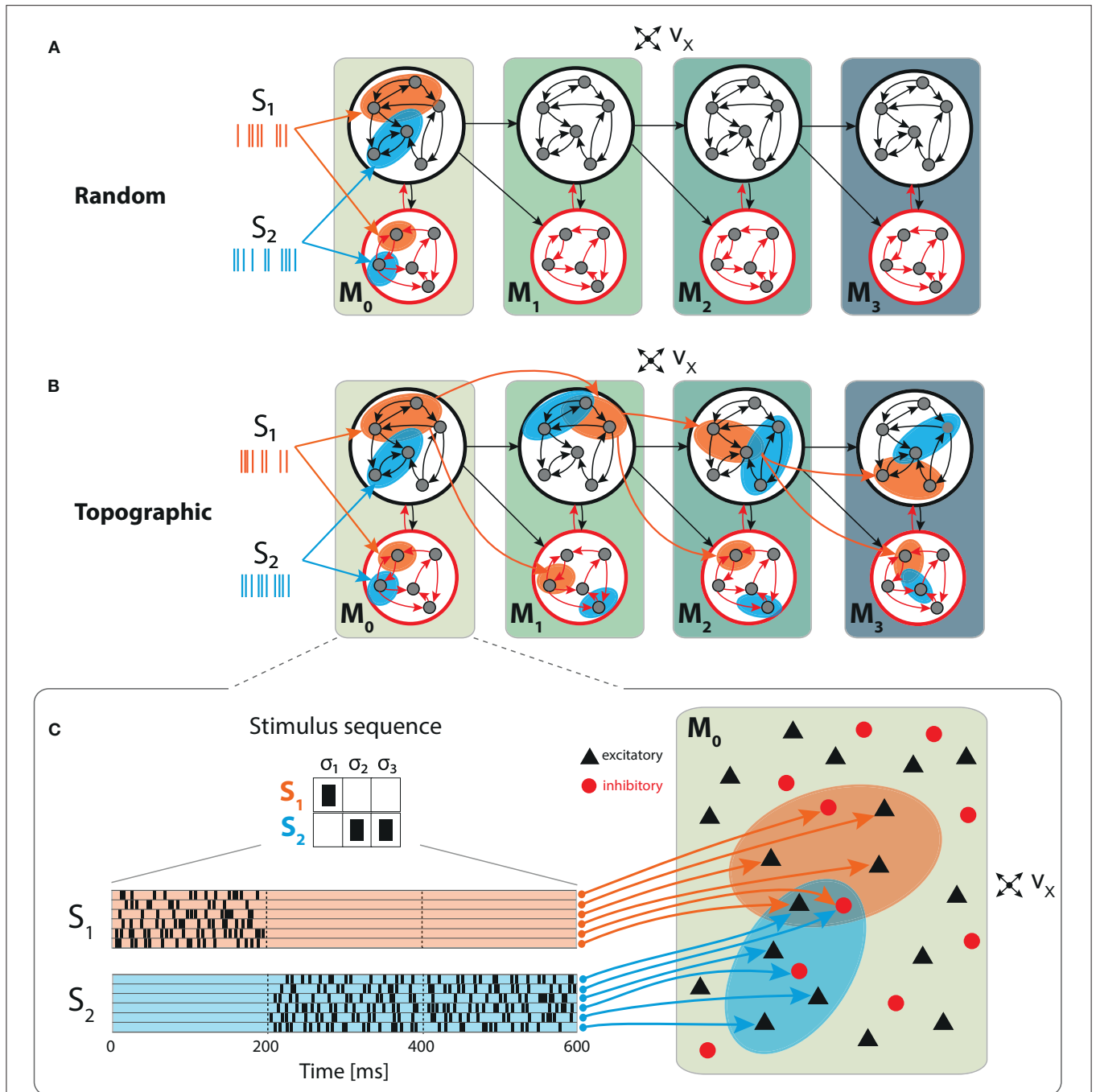
FIGURE 1 | Schematic overview of the sequential setup and input stimuli. Networks are composed of four modules with identical internal structure, with random **(A)** or topographically structured **(B)** feed-forward projections. Structured stimuli drive specific, randomly selected sub-populations in M_0. For stimulus S_1, the topographic projections (**B**, orange arrows) between the modules are represented explicitly in addition to the corresponding stimulus-specific sub-populations (orange ellipses), whereas for S_2 only the sub-populations are depicted (blue ellipses). The black feed-forward arrows depict the remaining sparse random connections from neurons that are not part of any stimulus-specific cluster. **(C)** Illustrative example of the input encoding scheme: a symbolic input sequence of length T (3 in this example), containing $|S|$ different, randomly ordered stimuli ($S = \{S_1, S_2\}$), is encoded into a binary matrix of dimensions $|S| \times T$. Each stimulus is then converted into a set of 800 Poissonian spike trains of fixed duration (200 ms) and rate ν_{stim} and delivered to a subset of ϵN^E excitatory and ϵN^I inhibitory neurons.

N^E (**Figure 1C**). These input neurons are mapped to randomly chosen, but stimulus-specific sub-populations of ϵN^E excitatory and ϵN^I inhibitory neurons in the first module M_0, which we denote the *input module*. Unless otherwise stated, we set $\lambda = 3$, resulting in mean firing rates ranging between 2 and 8 spikes/s across the modules.

To sample the population responses for each stimulus in the sequence, we collect the responses of the excitatory population in each module M_i at fixed time points t^*, relative to stimulus onset (with $t^* = 200$ ms, unless otherwise stated). These activity vectors are then gathered in a state matrix $X_{M_i} \in \mathbb{R}^{N^E \times T}$. In some cases, the measured responses are quantified using the low-pass filtered spike trains of the individual neurons, obtained by convolving them with an exponential kernel with $\tau = 20$ ms and temporal resolution equal to the simulation resolution, 0.1 ms. However, for most of the analyses, we consider the membrane potential V_m as the primary state variable, as it is parameter-free and constitutes a more natural choice (van den Broek et al., 2017; Duarte et al., 2018).

Unless otherwise stated, all results are averaged over multiple trials. Each trial consists of a single simulation of a particular network realization, driven by the relevant input stream(s). For each trial, the input-driven network responses are used to evaluate performance on a given task. In the case of the classification and XOR tasks described below, the performances within a single trial are always averaged over all stimuli.

2.4.1. Classification of Stimulus Identity
In the simplest task, the population responses are used to decode the identity of the input stimuli. The classification accuracy is determined by the capacity to linearly combine the input-driven population responses to approximate a target output (Lukoševičius and Jaeger, 2009):

$$\hat{Y} = W_{out}^{\mathsf{T}} X \qquad (2)$$

where $\hat{Y} \in \mathbb{R}^{r \times T}$ and $X \in \mathbb{R}^{N^E \times T}$ are the collection of all readout outputs and corresponding states over all time steps T, respectively, and W_{out} is the $N^E \times r$ matrix of output weights from the excitatory populations in each module to their dedicated readout units. We use 80% of the input data for training a set of r linear readouts to correctly classify the sequence of stimulus patterns in each module, where $r = |S|$ is the number of different stimuli to be classified. Training is performed using ridge regression (L_2 regularization), with the regularization parameter chosen by leave-one-out cross-validation on the training dataset. In the test phase, we obtain the predicted stimulus labels for the remaining 20% of the input sequence by applying the winner-takes-all (WTA) operation on the readout outputs $\hat{Y}$. Average classification performance is then measured as the fraction of correctly classified patterns.

2.4.2. Non-linear Exclusive-or (XOR)
We also investigate the more complex XOR task, involving two parallel stimulus sources S and S' injected into either the same or two separate input modules. Given stimulus sets $S = \{S_0, S_1\}$ and $S' = \{S'_0, S'_1\}$, the task is to compute the XOR on the stimulus labels, i.e., the target output is 1 for input combinations $\{S_0, S'_1\}$ and $\{S_1, S'_0\}$, and 0 otherwise. In this case, computational performance is quantified using the point-biserial correlation coefficient (PBCC), which is suitable for determining the correlation between a binary and a continuous variable (Haeusler and Maass, 2007; Klampfl and Maass, 2013; Duarte and

Morrison, 2014). The coefficient is computed between the binary target variable and the analog (raw) readout output $\hat{Y}(t)$, taking values in the $[-1, 1]$ interval, with any significantly positive value reflecting a performance above chance.

2.5. State Space Analysis
For a compact visualization and interpretation of the geometric arrangement of the population response vectors in the network's state-space, we analyze the characteristics of a low-dimensional projection of the population state vectors (membrane potentials) obtained through principal component analysis (PCA). More specifically, each N^E-dimensional state vector $x_i \in X_{M_i}$ is first mapped onto the sub-space spanned by the first three principal components (PCs), yielding a cloud of data points i which we label by their corresponding stimulus id.

In this lower-dimensional representation of the neuronal activity, we then evaluate how similar each data point in one stimulus-specific cluster is to its own cluster compared to neighboring clusters. This is done by assigning a silhouette coefficient $s(i)$ (Rousseeuw, 1987) to each sample i, computed during a single trial as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}. \qquad (3)$$

$a(i)$ represents the average distance between i and all other data points in the same cluster (same stimulus label), while $b(i)$ is mean distance of i to all points in the nearest cluster, i.e., corresponding to a different stimulus label. The coefficients $s(i)$ take values between $[-1, 1]$, with a value close to 1 indicating that the data point lies well within its assigned cluster (correct stimulus label), whereas values close to -1 imply an incorrect cluster assignment and therefore indicate overlapping stimulus representations in the network activity.

To get a single value that is representative of the overall clustering quality in one specific trial, we computed the silhouette score by averaging over all the silhouette coefficients $s(i)$. Note that for the results presented in **Figure 4B**, the silhouette scores were computed using projections onto the first 10 PCs, and were further averaged across ten different trials.

In addition to the cluster separation, we also quantify the dimensionality of the subspace where the neuronal activity predominantly lies, using the method introduced in Abbott et al. (2009) and Mazzucato et al. (2016). After performing a standard Principal Component Analysis on the firing rate vectors (average neuronal activity during a single stimulus presentation), we calculated the effective dimensionality as:

$$d = \left(\sum_{i=1}^{N} \widetilde{\lambda}_i^2 \right)^{-1}, \qquad (4)$$

where N is the real dimensionality of the network's state-space, i.e., the total number of neurons, and $\widetilde{\lambda}_i$ represents the fraction of the variance explained by the corresponding principal component, i.e., the normalized eigenvalues of the covariance

FIGURE 2 | Stimulus classification in sequentially connected modular networks. **(A,B)** Mean classification accuracy over $|S| = 10$ stimuli and corresponding mean squared error in each of the four modules in the random (plain bars) and topographic (hatched bars) conditions. **(C,D)** Mean classification accuracy and corresponding mean squared error in M_1 as a function of the number of direct projections (from neurons receiving direct stimulus input in M_0 to neurons in M_1) when decoding stimulus information from the low-pass filtered spike trains (stippled bars) and the membrane potential (plain bars). **(E)** Classification accuracy in M_2 and M_3 decoded from the membrane potential as a function of the input intensity. **(F)** Classification accuracy over $|S| = 50$ stimuli in M_3 as a function of the connection density within the topographic projections. All panels show the mean and standard deviations obtained from ten simulations per condition.

matrix of the firing rates. For the analysis in **Figures 4C,D**, **8**, the number of PCs considered was limited to 500.

2.6. Numerical Simulations and Analysis

All numerical simulations were conducted using the Neural Microcircuit Simulation and Analysis Toolkit (NMSAT) v0.2 (Duarte et al., 2017), a high-level Python framework for creating, simulating and evaluating complex, spiking neural microcircuits in a modular fashion. It builds on the PyNEST interface for NEST (Gewaltig and Diesmann, 2007), which provides the core simulation engine. To ensure the reproduction of all the numerical experiments and figures presented in this study, and abide by the recommendations proposed in Pauli et al. (2018), we provide a complete code package that implements project-specific functionality within NMSAT (see **Supplementary Materials**) using a modified version of NEST 2.12.0 (Kunkel et al., 2017).
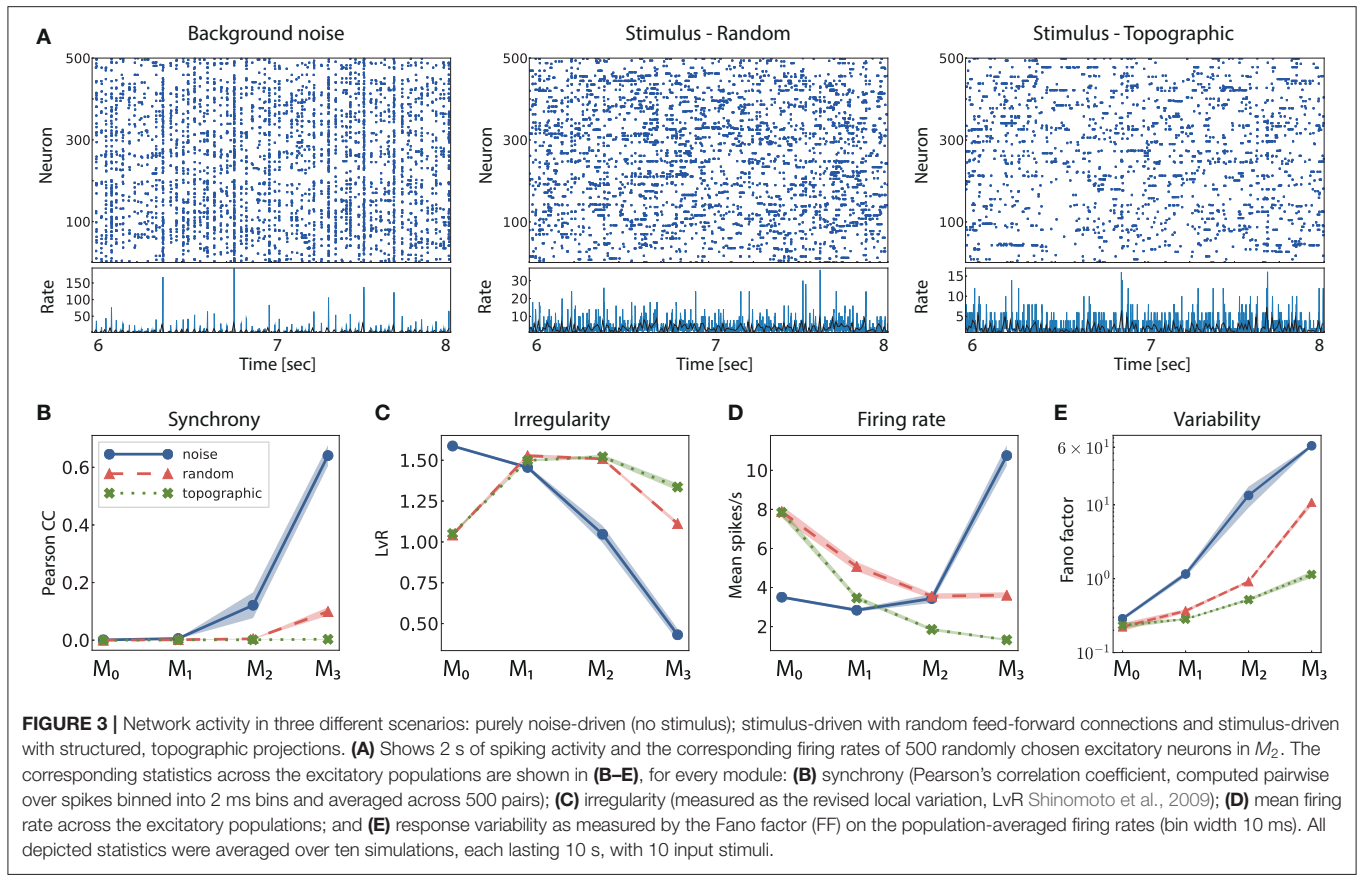
3. RESULTS

Distributed information processing across multiple neural circuits requires, in a first instance, an accurate representation of the stimulus identity and a reliable propagation of this information throughout the different modules. In the following section, we assess these capabilities using a linear classification task in a sequential setup (illustrated in **Figure 1**), and analyze the characteristics of population responses in the different modules.

Subsequently, we look at how different network setups handle information from two concurrent input streams by examining their ability to perform nonlinear transformations on the inputs.

3.1. Sequential Transmission of Stimulus Representation

In networks with fully random projections (**Figure 1A**), stimulus information can be accurately decoded up to a maximum depth of 3, i.e., the first three modules in the sequential setup contain sufficient information to classify (significantly beyond chance level) which of the ten stimuli had been presented to the input module (see section 2 for details of the stimulus generation and classification assessment). Whereas, the first two modules, M_0 and M_1, achieve maximum classification performance with virtually no variance across trials (**Figure 2A**, plain bars), the accuracy of ≈ 0.55 observed in M_2 indicates that the stimulus representations have become degraded. These results suggest that while random connectivity between the modules allows the input signal to reach M_2, the population responses at this depth are already insufficiently discernible to propagate further downstream, with M_3 entirely unable to distinctly represent the different stimuli.

Including structured projections in the system (**Figure 1B**) counteracts these effects, allowing stimulus information to be accurately transferred to the deeper modules (**Figure 2A**, hatched bars). This indicates that stimulus-specific topographic maps, whereby the neurons receiving direct stimulation at M_i

FIGURE 3 | Network activity in three different scenarios: purely noise-driven (no stimulus); stimulus-driven with random feed-forward connections and stimulus-driven with structured, topographic projections. **(A)** Shows 2 s of spiking activity and the corresponding firing rates of 500 randomly chosen excitatory neurons in M_2. The corresponding statistics across the excitatory populations are shown in **(B–E)**, for every module: **(B)** synchrony (Pearson's correlation coefficient, computed pairwise over spikes binned into 2 ms bins and averaged across 500 pairs); **(C)** irregularity (measured as the revised local variation, LvR Shinomoto et al., 2009); **(D)** mean firing rate across the excitatory populations; and **(E)** response variability as measured by the Fano factor (FF) on the population-averaged firing rates (bin width 10 ms). All depicted statistics were averaged over ten simulations, each lasting 10 s, with 10 input stimuli.

connect exclusively to another set of stimulus-specific neurons in the subsequent module (see section 2), play a critical role in the successful propagation of signals across multiple interacting sub-networks.
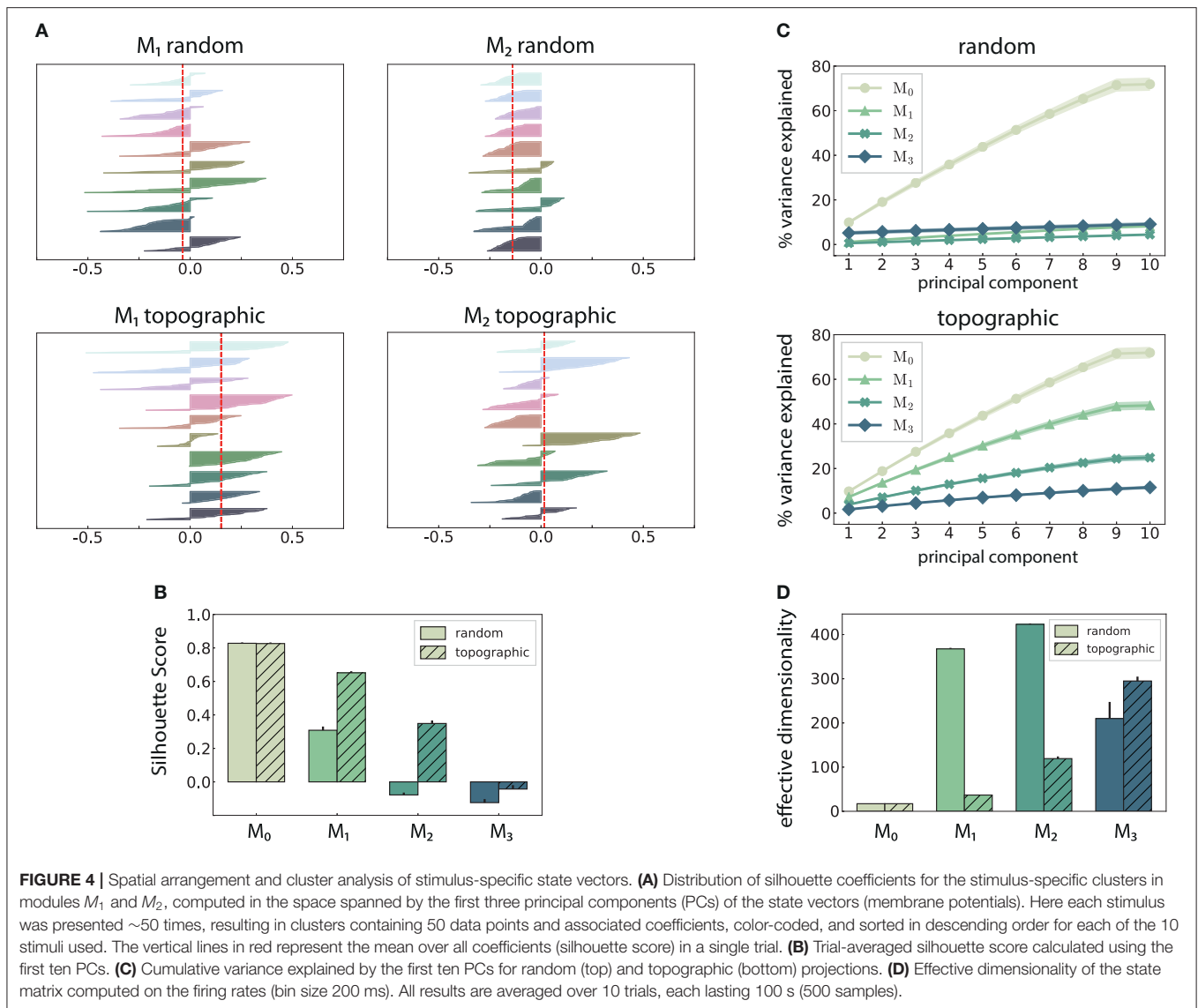
As computing the accuracy scores involves a nonlinear post-processing step (winner-takes-all, see section 2), we additionally verify whether this operation significantly biases the results by evaluating the mean squared error (MSE) between the raw readout outputs $\hat{Y}$ and the binary targets Y. These MSE values, depicted in **Figure 2B**, are consistent: performance decays with depth for both network setups, with topography leading to significant computational benefits for all modules beyond the input module. In the following two sections, we investigate the factors influencing stimulus propagation and uncover the relationships between the underlying population dynamics and the system's task performance.

3.1.1. Modulating Stimulus Propagation

Since random networks provide no clearly structured feed-forward pathways to facilitate signal propagation, it is unclear how stimulus information can be read out as far as M_2 (**Figure 2A**), considering the nonlinear transformations at each processing stage. However, by construction, some neurons in M_0 that receive input stimulus directly also project (randomly) to M_1. To assess the importance of these directed projections for information transmission, we gradually remove them and

measure the impact on the performance in M_1 (**Figures 2C,D**). The system shows substantial robustness with respect to the loss of such direct feed-forward projections, as the onset of the decline in performance only occurs after removing half of the direct synapses. Furthermore, this decay is observed almost exclusively in the low-pass filtered responses, while the accuracy of state representations at the level of membrane potentials remains maximal. This suggests that the populations in the input module are not only able to create internal representations of the stimuli through their recurrent connections, but also transfer these to the next module in an suitable manner. 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.

It is reasonable to assume that the transmission quality in the two networks, as presented above, is susceptible to variations in the input intensity. For random networks, one might expect that increasing the stimulus intensity would enable its decoding in all four modules. Although stronger input does improve the classification performance in M_2 (**Figure 2E**), this improvement is not visible in the last module. When varying the input rates between 5 and 25 spk/s, the accuracy increases linearly with the stimulus intensity in M_2. However, the signal does not propagate to the last module in a decipherable manner (results remain at chance level), regardless of the input rate and, surprisingly, regardless of the representational accuracy in M_2.

FIGURE 4 | Spatial arrangement and cluster analysis of stimulus-specific state vectors. **(A)** Distribution of silhouette coefficients for the stimulus-specific clusters in modules M_1 and M_2, computed in the space spanned by the first three principal components (PCs) of the state vectors (membrane potentials). Here each stimulus was presented ~50 times, resulting in clusters containing 50 data points and associated coefficients, color-coded, and sorted in descending order for each of the 10 stimuli used. The vertical lines in red represent the mean over all coefficients (silhouette score) in a single trial. **(B)** Trial-averaged silhouette score calculated using the first ten PCs. **(C)** Cumulative variance explained by the first ten PCs for random (top) and topographic (bottom) projections. **(D)** Effective dimensionality of the state matrix computed on the firing rates (bin size 200 ms). All results are averaged over 10 trials, each lasting 100 s (500 samples).

Previous studies have shown that, when structured feed-forward connections are introduced, the spiking activity propagation generally depends on both the synaptic strength and connection density along the structures, with higher values increasing the transmission success (Vogels and Abbott, 2005; Kumar et al., 2010). To evaluate this in our model without altering the synaptic parameters, we increase the task difficulty and test the ability of the last module, M_3, to discriminate 50 different stimuli. The results, shown in **Figure 2F**, exhibit a significantly lower performance for the initial topographic density of (7.5%), from $\approx$ 1 for ten stimuli (**Figure 2A**) to $\approx$ 0.3. This drop can be likely attributed to overlapping projections between the modules, since more stimulus-specific pathways naturally lead to more overlap between these regions, causing less discriminable responses. However, this seems to be compensated for by increasing the projection density, with stronger connectivity significantly improving the performance.

Thus, our simulations corroborate these previous experiments: increasing the connection density within topographic maps increases the network's computational capacity.

3.1.2. Population Activity and State Separability

To ensure a perfect linear decoding of the input, population responses elicited by different stimuli must flow along well segregated, stimulus-specific regions in the network's state-space (separation property, see Maass et al., 2002). In this section, we evaluate the quality of these input-state mappings as the representations are transferred from module to module, and identify population activity features that influence the networks' computational capabilities in various scenarios.

When a random network is driven only by background noise, the activity in the first two modules is asynchronous and irregular, but evolves into a more synchronous regime in M_2 (see example activity in **Figure 3A** left, and noise condition in **Figure 3B**). In

the last module, the system enters a synchronous regime, which has been previously shown to negatively impact information processing by increasing redundancy in the population activity (Duarte and Morrison, 2014). This excessive synchronization explains the increased firing rates, reaching $\approx$ 10 spk/s in M_3 (**Figure 3D**). Previous works have shown that even weak correlations within an input population can induce correlations and fast oscillations in the network (Brunel, 2000). This phenomenon arises in networks with sequentially connected populations, and is primarily a consequence of an increase in shared pre-synaptic inputs between successive populations (Shadlen and Newsome, 1998; Tetzlaff et al., 2003; Kumar et al., 2008). As the feed-forward projections gradually increase the convergence of the inter-module connections, the corresponding magnitude of post-synaptic responses also increase toward the deeper modules. Effectively stronger synapses then shift the network's operating point away from the desired Poissonian statistics. This effect accumulates from module to module and gradually skews the population activity toward states of increased synchrony.

Compared to baseline activity, the presence of a patterned stimulus increases the irregularity in all modules except the very first one. This is visualized in the example activity plots in **Figure 3A** (center and right). Furthermore, active input substantially reduces the synchrony in the last two modules, allowing the system to globally maintain the asynchronous irregular regime (see random and topographic conditions in **Figures 3B,C**). Such alterations in the population response statistics during active processing have also been confirmed experimentally: *in vivo* recordings show that neuronal activity in awake, behaving animals is characterized by weak correlations and low firing rates in the presence of external stimuli (Vaadia et al., 1995; Ecker et al., 2010).

Despite the beneficial influence of targeted stimulation, it appears that random projections are not sufficient to entirely overcome the effects of shared input and excessive synchronization in the deeper modules (e.g., in M_3, CC $\approx$ 0.12, with a correspondingly high firing rate). The existence of structured connectivity, through conserved topographic maps, on the other hand, allows the system to retain an asynchronous firing profile throughout the network. Whereas the more synchronous activity in random networks, coupled with a larger variability in the population responses (**Figure 3E**), contributes to their inability to represent the input in the deeper modules, topographic projections lead to more stable and reliable neuronal responses that enable the maintenance of distinguishable stimulus mappings, in line with the performance results observed in **Figure 2A**.

Furthermore, networks with structured connectivity are also more resource-efficient, achieving better performance with lower overall activity (**Figure 3D**). This can be explained by the fact that neurons receiving direct stimulus input in M_0, firing at higher rates, project only to a restricted sub-population in the subsequent modules, thereby having a smaller impact on the average population activity downstream.

The above observations are also reflected in the geometric arrangement of the population response vectors, as visualized by the silhouette coefficients of a low-dimensional projection

of their firing rates in **Figure 4A** (see Methods). As stimulus responses become less distinguishable with network depth, the coefficients decrease, indicating more overlapping representations. This demonstrates a reduction in the compactness of stimulus-dependent state vector clusters, which, although not uniformly reflected for all stimuli, is consistent across modules (only M_1 and M_2 shown). However, these coefficients are computed using only the first three principal components (PCs) of the firing rate vectors and are trial-specific. We can obtain a more representative result by repeating the analysis over multiple trials and taking into account the first ten PCs (**Figure 4B**). The silhouette scores computed in this way reveal a clear disparity between random and topographic network for the spatial segregation of the clusters, beginning with M_1, in accordance with the classification performances (**Figure 2A**).
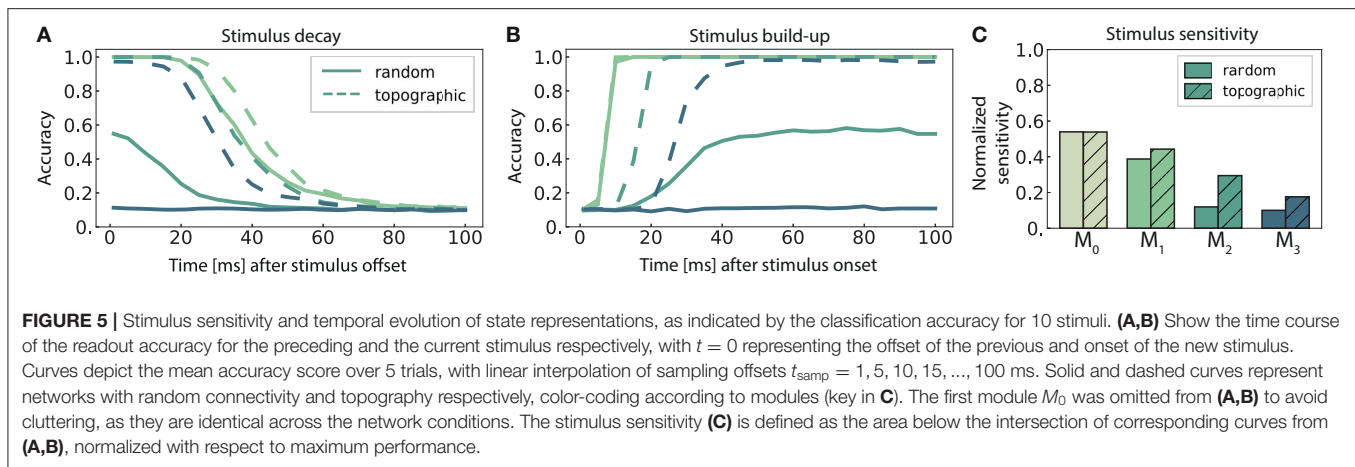
We can further assess the effectiveness with which the networks utilize their high-dimensional state-space by evaluating how many PCs are required to capture the majority of the variance in the data (**Figure 4C**). In the input module, where the stimulus impact is strongest, the variance captured by each subsequent PC is fairly constant ($\approx$ 10%), reaching around 75% by the ninth PC. This indicates that population activity can represent the input in a very low-dimensional subspace through narrow, stimulus-specific trajectories. In random networks, however, this trend is not reflected in the subsequent modules, where the first ten PCs account for <10% of the total variance.

There is thus a significant increase in the effective dimensionality (see section 2) in the deeper modules (**Figure 4D**), a pattern which is also exhibited, to a lesser extent, in the topographic case. As the population activity becomes less entrained by the input, the deeper modules explore a larger region of the state-space. Whereas, this tendency is consistent and more gradual for topographic networks, it is considerably faster in networks with unstructured projections, suggesting a quicker dispersion of the stimulus representations. Since in these networks the stimulus does not effectively reach the last module (**Figure 2A**), there is no de-correlation of the responses, and the elevated synchrony (**Figure 3B**) leads to a reduced effective dimensionality.

Overall, these results demonstrate that patterned stimuli push the population activity toward an asynchronous-irregular regime across the network, but purely random systems cannot sustain this state in the deeper modules. Networks with structured connectivity, on the other hand, display a more stable activity profile throughout the system, allowing the stimuli to propagate more efficiently and more accurately to all modules. Accordingly, the state representations are more compact and distinguishable, and these representations decay significantly slower with module depth than in random networks, in line with the observed classification results (**Figure 2A**).

3.2. Memory Capacity and Stimulus Sensitivity

As demonstrated above, both random and topographic networks are able to create unique representations of single stimuli in their internal dynamics and transfer these across multiple recurrent

FIGURE 5 | Stimulus sensitivity and temporal evolution of state representations, as indicated by the classification accuracy for 10 stimuli. **(A,B)** Show the time course of the readout accuracy for the preceding and the current stimulus respectively, with $t = 0$ representing the offset of the previous and onset of the new stimulus. Curves depict the mean accuracy score over 5 trials, with linear interpolation of sampling offsets $t_{samp} = 1, 5, 10, 15, ..., 100$ ms. Solid and dashed curves represent networks with random connectivity and topography respectively, color-coding according to modules (key in **C**). The first module M_0 was omitted from **(A,B)** to avoid cluttering, as they are identical across the network conditions. The stimulus sensitivity **(C)** is defined as the area below the intersection of corresponding curves from **(A,B)**, normalized with respect to maximum performance.

modules. In order to better understand the nature of distributed processing in these systems, it is critical to investigate how they retain information over time and whether representations of multiple, sequentially presented, stimuli can coexist in a superimposed manner, a property exhibited by cortical circuits as demonstrated by *in vivo* recordings (Nikolic et al., 2009).

To quantify these properties, we use the classification accuracy to evaluate how, for consecutive stimuli, the first stimulus decays and the second stimulus builds up (**Figures 5A,B**). For a given network configuration, the degree of overlap between the two curves indicates how long the system is able to retain useful information about both the previous and the present stimuli (**Figure 5C**). This analysis allows us to measure three important properties of the system: how long stimulus information is retained in each sub-network through reverberations of the current state; how long the network requires to accumulate sufficient evidence to classify the present input; and what are the potential interference effects between multiple stimuli. Note that the procedure used in the following experiments is virtually identical to that in section 3.1, the only difference being the time at which the network's responses are sampled. In **Figure 5A**, the readout is trained to classify the stimulus identity at increasing time lags after its offset, whereas in **Figure 5B**, the classification accuracy is evaluated at various time points after the stimulus onset.

The decay in performance measured at increasing delays after stimulus offset (**Figure 5A**) shows how input representations gradually disappear over time (the *fading memory* property, see Maass et al., 2004). For computational reasons, only the first 100 ms are plotted, but the decreasing trend in the accuracy continues and invariably reaches chance level within the first 150 ms. This demonstrates that the networks have a rather short memory capacity which is unable to span multiple input elements, and that the ability to memorize stimulus information decays with network depth. Adding to the functional benefits of topographic maps, the memory curves reflect the higher overall accuracy achieved by these networks.

We further observe that the networks require exposure time to acquire discernible stimulus representations (**Figure 5B**). The time for classification accuracy to reach its maximum

increases with depth, resulting in an unsurprising cumulative delay. Notably, topography enables a faster information build-up beginning with M_2.
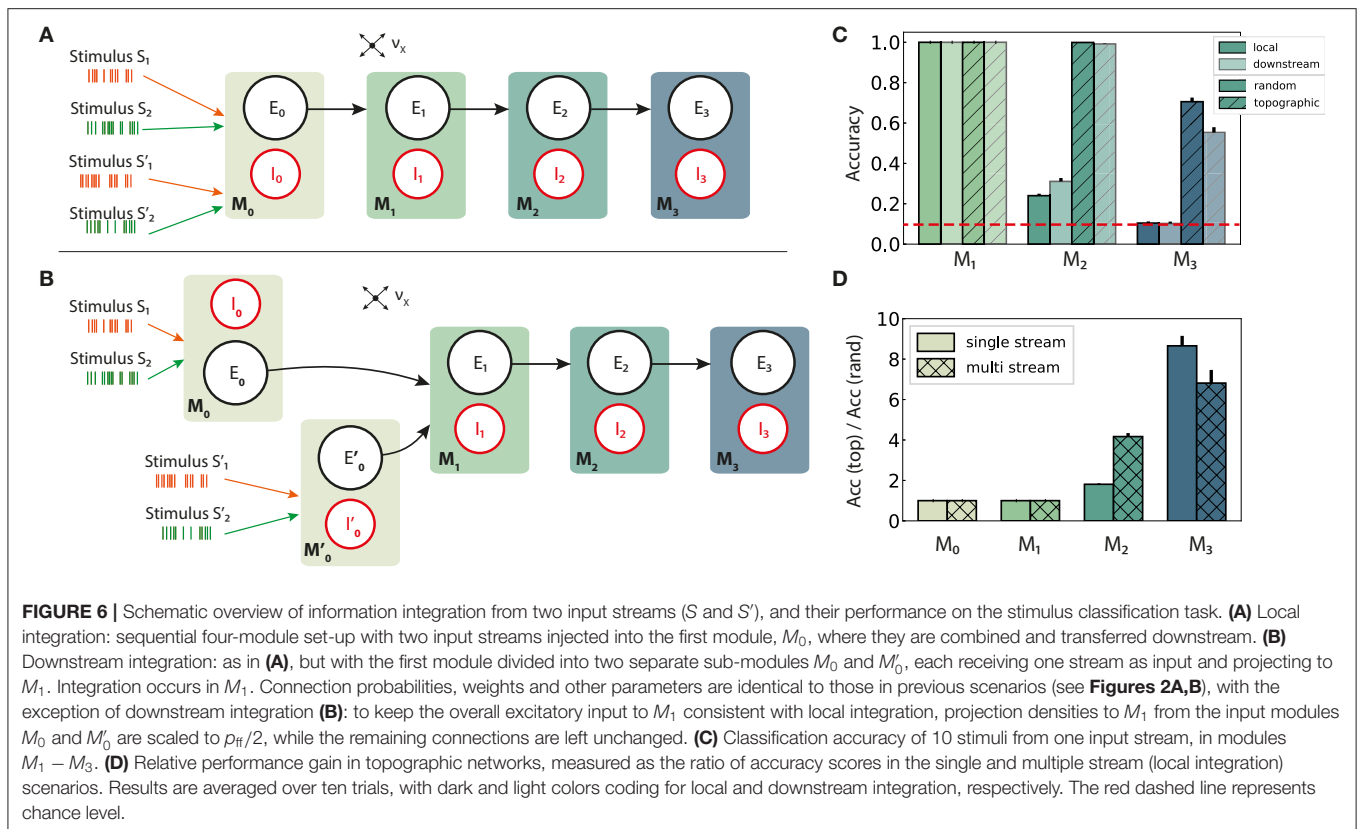
To determine the *stimulus sensitivity* of a population, we consider the extent of time where useful non-interfering representations are retained in each sub-network. This can be calculated as the area below the intersection of its memory and build-up curves. Following a similar trend to performance and memory, sensitivity to stimulus decreases with network depth and the existence of structured propagation pathways leads to clear benefits, particularly pronounced in the deeper modules (**Figure 5C**).

Overall, modules located deeper in the network forget faster and take longer (than the inter-module delays) to build up stimulus representations. No population is able to represent two sequential stimuli accurately for a significant amount of time (longer than 100 ms), although topographic maps improve memory capacity and stimulus sensitivity.

3.3. Integrating Multiple Input Streams

The previous section focuses on a single input stream, injected into a network with sequentially connected modules. Here, we examine the microcircuit's capability to integrate information from two different input streams, in two different scenarios with respect to the location of the integration. The set-up and results are illustrated in **Figure 6**.

In a first step, the set-up from **Figure 1A** is extended with an additional input stream S', without further alterations at population or connectivity level. The two stimulus sets, S and S', are in principle identical, each containing the same number of unique stimuli and connected to specific sub-populations in the networks. Since the inputs are combined locally in the first module and the mixed information transferred downstream, we refer to this setup, visualized in **Figure 6A**, as *local integration*. In a second scenario (**Figure 6B**), each input stream is injected into a separate sub-module (M_0 and M'_0), jointly forming the input module of the system. Here, computation on the combined input happens *downstream* from the first module, with the aim of simulating the integration of information that originated from

FIGURE 6 | Schematic overview of information integration from two input streams (S and S'), and their performance on the stimulus classification task. **(A)** Local integration: sequential four-module set-up with two input streams injected into the first module, M_0, where they are combined and transferred downstream. **(B)** Downstream integration: as in **(A)**, but with the first module divided into two separate sub-modules M_0 and M'_0, each receiving one stream as input and projecting to M_1. Integration occurs in M_1. Connection probabilities, weights and other parameters are identical to those in previous scenarios (see **Figures 2A,B**), with the exception of downstream integration **(B)**: to keep the overall excitatory input to M_1 consistent with local integration, projection densities to M_1 from the input modules M_0 and M'_0 are scaled to $p_{ff}/2$, while the remaining connections are left unchanged. **(C)** Classification accuracy of 10 stimuli from one input stream, in modules $M_1 - M_3$. **(D)** Relative performance gain in topographic networks, measured as the ratio of accuracy scores in the single and multiple stream (local integration) scenarios. Results are averaged over ten trials, with dark and light colors coding for local and downstream integration, respectively. The red dashed line represents chance level.

more distant areas and had already been processed by two independent microcircuits.

Adding a second input stream significantly affects the network activity and the stimulus representations therein, which now must produce distinguishable responses for two stimuli concurrently. Compared to the same setup with a single input source (**Figure 2A**), the performance degrades in both random and topographic networks starting with M_2 (**Figures 6C,D**). This suggests that the mixture of two stimuli results in less separable responses as the two representations interfere with each other, with structured connectivity again proving to be markedly beneficial. These benefits become clearer in the deeper modules, as demonstrated in **Figure 6D** where the effects of topography can lead to an eight-fold gain in task accuracy in M_3.

As the spatio-temporal structure of the stimuli from both sources are essentially identical, it is to be expected that the mixed responses contain the same amount of information about both inputs. This is indeed the case, as reflected by comparable performance results when decoding from the second input stream (**Figure S1**).
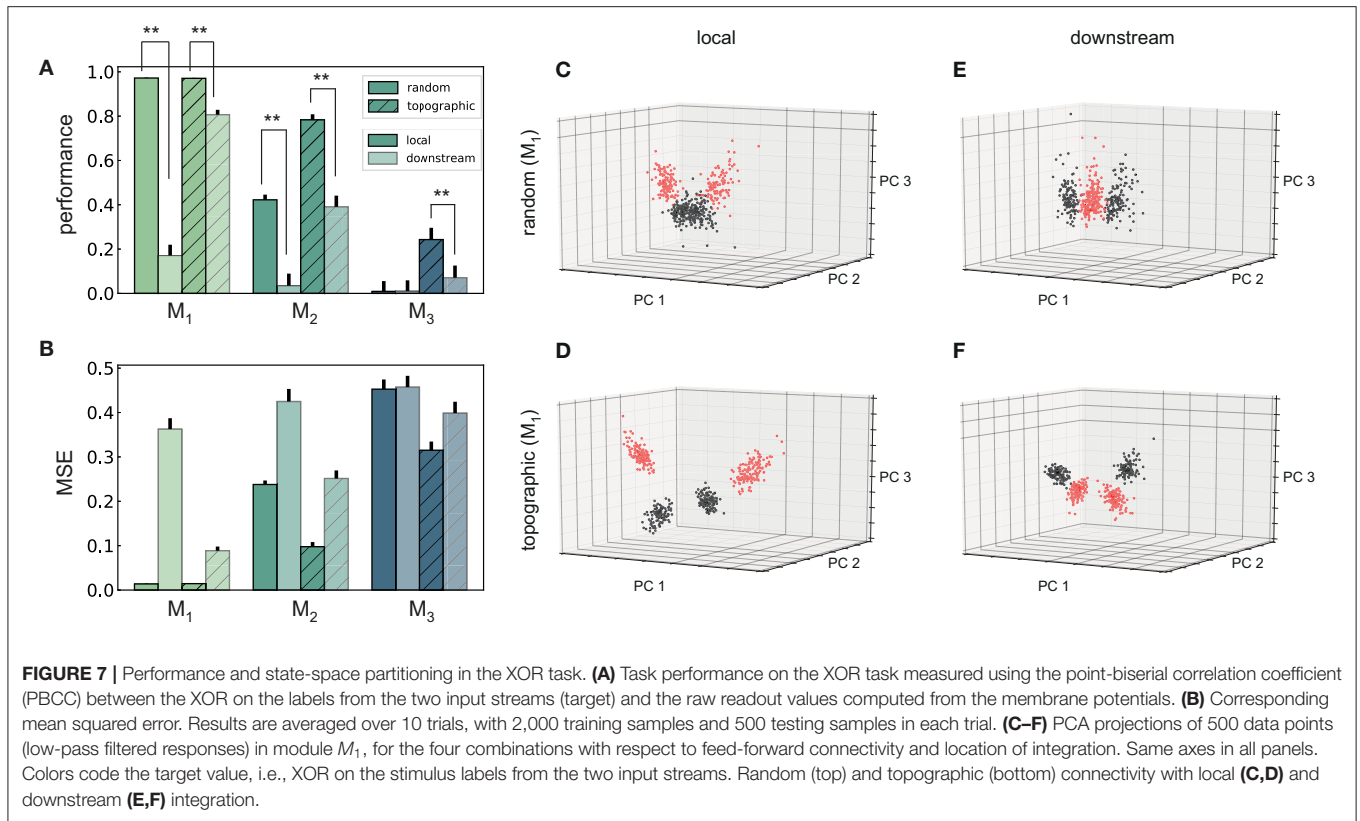
Interestingly, the location of the integration appears to play no major role for random networks. In networks with topographic maps, however, local integration improves the classification accuracy by around 25% in the last module compared to the downstream case. In the next section we investigate whether this phenomenon is set-up and task specific, or reflects a more generic computational principle.

3.3.1. Local Integration Improves Non-linear Computation

In addition to the linear classification task discussed above, we analyze the ability of the circuit to extract and combine information from the two concurrent streams in a more complex, nonlinear fashion. For this, we trained the readouts on the commonly used non-linear XOR task described in section 2.

We observe that the networks' computational capacity is considerably reduced compared to the simpler classification task, most noticeably in the deeper modules (**Figure 7A**). Although information about multiple stimuli from two input streams could be reasonably represented and transferred across the network, as shown in **Figure 6C**, it is substantially more difficult to perform complex transformations on even a small number of stimuli. This is best illustrated in the last module of topographic networks, where the stimulus identity can still be decoded with an accuracy of 70% (**Figure 6C**), but the XOR operation yields performance values close to chance level (PBCC of 0).

In contrast to the identity recognition, for XOR it is clearly more advantageous to fuse the two input streams in the first module (locally), rather than integrating only in M_1 (**Figures 7A,B**). The differences in performance are statistically significant (two-sided Kolmogorov–Smirnov (KS) test 1.0, $p < 0.01$ for M_1 and M_2, and KS-test 0.9 with $p < 0.01$ for M_3 in topographic networks) and consistent in every scenario and all modules from M_1 onwards, with the exception of M_3 in random networks.

FIGURE 7 | Performance and state-space partitioning in the XOR task. **(A)** Task performance on the XOR task measured using the point-biserial correlation coefficient (PBCC) between the XOR on the labels from the two input streams (target) and the raw readout values computed from the membrane potentials. **(B)** Corresponding mean squared error. Results are averaged over 10 trials, with 2,000 training samples and 500 testing samples in each trial. **(C–F)** PCA projections of 500 data points (low-pass filtered responses) in module M_1, for the four combinations with respect to feed-forward connectivity and location of integration. Same axes in all panels. Colors code the target value, i.e., XOR on the stimulus labels from the two input streams. Random (top) and topographic (bottom) connectivity with local **(C,D)** and downstream **(E,F)** integration.

One can gain a more intuitive understanding of the networks' internal dynamics by looking at the state-space partitioning (**Figures 7C–F**), which reveals four discernible clusters corresponding to the four possible label combinations. These low-dimensional projections illustrate two key computational aspects: the narrower spread of the clusters in topographic networks (**Figures 7D,F**) is an indication of their greater representational precision, while the significance of the integration location is reflected in the collapse along the third PC in the downstream scenario (**Figure 7F**). To a lesser extent, these differences are also visible for random networks (**Figures 7C,E**). A more compact representation of the clustering quality using silhouette scores, consistent with these observations, is depicted in **Figure S2**.

Altogether, these results suggest that it is computationally beneficial to perform non-linear transformations locally, as close to the input source as possible, and then propagate the result of the computation downstream instead of the other way around. The results were qualitatively similar for both the low-pass filtered spike trains and the membrane potential (see **Figure S3**). To rule out any possible bias arising from re-scaling the feed-forward projections to M_1 in the downstream scenario, we also ensured that these results still hold when each of the input sub-modules M_0 and M_0' projected to M_1 with the same unscaled probability p_{ff} as in **Figure 1B** (see **Figure S4**).

3.3.2. Effective Dimensionality Depends on the Architecture of Stimulus Integration

Previous studies have suggested that non-linear integration of multiple input streams is associated with high response dimensionality compared to areas in which little or only linear interactions occur (Barak et al., 2013; Rigotti and Fusi, 2016). To assess whether these predictions hold in our model, we consider different stimulus integration schemes and investigate whether the effective response dimensionality correlates with XOR accuracy, which is used to quantify the non-linear transformations performed by the system.

For simplicity, we focus only on random networks. To allow a better comparison between the integration schemes introduced in **Figure 6**, we explore two approaches to gradually interpolate the downstream scenario toward the local one in an attempt to approximate its properties. First, we distribute each input stream across the two segregated input sub-modules M_0 and M_0', referred to as *mixed input* (**Figure 8A**). Second, we maintain the input stream separation but progressively merge the two sub-modules into a single larger one by redistributing the recurrent connections (**Figure 8B**). We call this scenario *mixed connectivity*.

Relating these two scenarios is the *mixing factor* (m), which controls the input mapping or the connectivity between the sub-modules, respectively. A factor of 0 represents separated input sources and disconnected sub-modules as in **Figure 6B**; a value

FIGURE 8 | Mixed multi-stream integration and neural dimensionality in random networks. **(A)** Downstream integration (only M_0 shown) with progressive redistribution of the input across sub-modules M_0 and M'_0. The mixing ratio is parameterized by $m \in [0, 1]$, with $m = 0$ representing two information sources mapped exclusively onto the corresponding sub-module, and $m = 1$ indicating equally distributed inputs across the sub-modules. The overall input to the network is kept constant. Arrow thickness indicates connection density. **(B)** As in **(A)**, but gradually connecting M_0 and M'_0 while keeping the input streams separated. Here, m controls the ratio of inter- and intra-module connection probabilities, with the total number of connections kept constant. **(C)** XOR performance as a function of m for mixed input; **(D)** Corresponding effective dimensionality in the input modules M_0 and M'_0; **(E)** Corresponding effective dimensionality in the deeper modules M_1 and M_2. **(F–H)** As in **(C–E)**, but for mixed connectivity. Effective dimensionality is calculated for the membrane potential (mean and standard deviation over 10 trials, calculated using the first 500 PCs).

of $m = 1$ indicates that the input modules mix contributions from both sources equally (for mixed input), or that intra and inter-module connectivity for M_0 and M'_0 are identical (for mixed connectivity). In both cases, care was taken to keep the overall

input to the network unchanged, as well as the average in- and out-degree of the neurons.

Combining information from both input streams already in the first sub-modules ($m > 0$), either via mixed input or

mixed connectivity, significantly increases the task performance after convergence in the deeper modules. This is illustrated in **Figures 8C,F**, with $m > 0.5$ yielding similar values. Despite comparable gains in the nonlinear computational performance, the underlying mechanisms appear to differ in the two mixing approaches, as detailed in the following.

In M_0 and M_0', the effective dimensionality of the neural responses increases monotonically with the amount of information shared between the two modules (**Figures 8D,G**). This is expected, since the sub-modules are completely independent initially ($m = 0$) and can therefore use more compact state representations for single stimuli. However, diverging patterns emerge after convergence in M_1. While the dimensionality does increase with the coefficient m in the mixed connectivity scenario (**Figure 8H**), it remains fairly constant in the mixed input case (**Figure 8E**), despite comparable task performance. Thus, complex non-linear transformations do not necessarily involve the exploration of larger regions of state-space, but can also be achieved through more efficient representations.

These results also demonstrate the difficulty in defining a clear relation between the ability of the system to perform nonlinear transformations on the input and its response dimensionality. Particularly in the case of larger networks involving transmission across multiple modules, the effective dimensionality can depend on the system's architecture, such as the input mapping and connectivity structure in the initial stages.

4. DISCUSSION

Real-time interactions between a dynamic environment and a modular, hierarchical system like the mammalian neocortex strictly requires efficient and reliable mechanisms supporting the acquisition and propagation of adequate internal representations. Stable and reliable representations of relevant stimulus features must permeate the system in order to allow it to perform both local and distributed computations online. Throughout this study, we have analyzed the characteristics of state representations in modular spiking networks and the architectural and dynamical constraints that influence the system's ability to retain, transfer, and integrate stimulus information.

We have considered models of local microcircuits as state-dependent processing reservoirs whose computations are performed by the systems' high-dimensional transient dynamics (Mante et al., 2013; Sussillo, 2014), acting as a temporal expansion operator, and investigated how the features of long-range connectivity in a modular architecture influence the system's overall computational properties. By considering the network as a large modular reservoir, composed of multiple sub-systems, we have explored the role played by biologically-inspired connectivity features (conserved topographic projections) in the reliable information propagation across the modules, as well as the underlying dynamics that support the development and maintenance of such internal representations.

In addition to examining the temporal dynamics of the information transferred between sequentially connected modules, we have explored how different network characteristics enable information integration from two independent sources in a computationally useful manner. In these experiments, structural differences in the network were proven to greatly influence the dynamics and the downstream computation when combining inputs from two independent sources. In addition to the inter-module connectivity, the ability of the downstream modules to non-linearly combine the inputs was shown to depend on the location where the input converges, as well as on the extent to which the different input streams are mixed in the initial modules. We therefore anticipate that degree of mixed selectivity in early sensory stages is predictive of the computational outcome in deeper levels, particularly for non-linear processing tasks, as we describe in greater detail below.

4.1. Representation Transfer in Sequential Hierarchies

The proficiency of randomly coupled spiking networks (see e.g., Maass et al., 2002; Duarte and Morrison, 2014; Sussillo, 2014) demonstrates that random connectivity can be sufficient for local information processing. Successful signal propagation over multiple modules, however, appears to require some form of structured pathways for accurate and reliable transmission. Our results suggest that these requirements can be achieved by embedding simple topographic projections in the connectivity between the modules. Such mechanisms might be employed across the brain for fast and robust communication, particularly (but not exclusively) in the early sensory systems, where real-time computation is crucial and where the existence of topographic maps is well supported by anatomical studies (Kaas, 1997; Bednar and Wilson, 2016).

Purely random feed-forward connectivity allowed stimulus information to be decoded only up to the third module, whereas incorporating topographic projections ensured almost perfect accuracy in all modules (**Figures 2A,B**). These differences could be attributed to a decrease in the specificity of stimulus tuning with network depth, which is much more prominent for random networks (**Figure 4**). This result suggests that accurate information transmission over longer distances is not possible without topographic precision, thus uncovering an important functional role of this common anatomical feature.

Moreover, topography was shown to counteract the shared-input effect which leads to the development of synchronous regimes in the deeper modules. By doing so, stimulus information is allowed to propagate not only more robustly, but also more efficiently with respect to resources, in that the average spike emission is much lower (**Figures 3D,E**). Nevertheless, as the stimulus intensity invariably fades with network depth, the deeper modules capture fewer spatio-temporal features of the input and their response dimensionality increases. This process is clearer in random networks (**Figures 4C,D**), a further indication that topography enforces more stereotypical, lower-dimensional and stimulus-specific response trajectories. The

input-state mappings are also retained longer and built up more rapidly in topographic networks (**Figure 5**).

4.2. Network Architecture and Input Integration

In biological microcircuits, local connections are complemented by long range projections which either stem from other cortical regions (cortico-cortical), or from different sub-cortical nuclei (e.g., thalamo-cortical). These different projections carry different information content and thus require the processing circuits to integrate multiple input streams during online processing. The ability of local modules to process information from multiple sources simultaneously and effectively is thus a fundamental building block of cortical processing.

Including a second input source into our sequential networks leads to less discriminable responses, as reflected in a decreased classification performance (**Figure 6C**). Integrating information from the two sources as early as possible in the system (i.e., in the modules closest to the input) was found to be clearly more advantageous for non-linear computations (**Figure 6C**) and, to a lesser extent, also to linear computations. For both tasks, however, topographic networks achieved better overall performance.

One of our main results thus suggests that computing locally, within a module, and transmitting the outcome of such computation (local integration scenario) is more effective than transmitting partial information and computing downstream. Accordingly, even a single step of non-linear transformation on individual inputs (downstream integration scenario) hinders the ability of subsequent modules to exploit non-trivial dependencies and features in the data. Therefore, it might be more efficient to integrate information and extract relevant features within local microcircuits that can act as individual computational units (e.g., cortical columns Mountcastle, 1997). By combining the inputs locally, the population can create more stable representations which can then be robustly transferred across multiple modules. We speculate that in hierarchical cortical microcircuits, contextual information (simply modeled as a second input stream here) must be present in the early processing stages to enable more accurate computations in the deeper modules. This could, in part, explain the role of feedback connections from higher to lower processing centers.

4.3. Degree of Mixed Selectivity Predicts Computational Performance

We have further shown that the effective dimensionality of the neural responses does not correlate with the non-linear computational capabilities, except in the very first modules (**Figure 8**). These insights are in agreement with previous studies (Barak et al., 2013; Rigotti and Fusi, 2016), based on fMRI data that predict a high response dimensionality in areas involved in nonlinear multi-stream integration, and lower in areas where inputs from independent sources do not interact at all or solely overlap linearly. These studies considered single circuits driven by input from two independent sources, focusing on the role of mixed selectivity neurons in the convergent population. Mixed selectivity refers to neurons being tuned to mixtures of multiple task-related aspects (Warden and Miller, 2010; Rigotti et al., 2013), which we approximated as a differential driving of the neurons with a variable degree of input from both sources.

Although we did not specifically examine mixed selectivity at a single neuron level, one can consider both the mixed input and mixed connectivity scenarios (**Figures 8A,B**, respectively) to approximate this behavior at a population level. This is particularly the case for the input sub-modules M_0 and M'_0, where the network's response dimensionality, as expected, increases with the mixing ratio (**Figures 8D,G**). However, the different results we obtained for the deeper modules (**Figures 8E,H**), suggest that the effective dimensionality measured at the neuronal level is not a reliable evidence for non-linear processing in downstream convergence areas (despite similar performance), but instead depends on how information is mixed in the early stages of the system. Further research in this direction, possibly resorting to multimodal imaging data, is needed to determine a clear relation between functional performance, integration schemes, and response dimensionality.

In our models, the task performance improved (and plateaued) with increased mixing factors, suggesting no obvious computational disadvantages for large factor values. While this holds for the discrimination capability of the networks, we did not address their ability to generalize. Since the sparsity of mixed selectivity neurons has been previously shown to control the discrimination-generalization trade-off, along with the existence of an optimal sparsity for neural representations (Barak et al., 2013), it would be interesting to analyze the effect of this parameter more thoroughly in the context of hierarchical processing.

Based on the presented findings, we expect that the degree of mixed selectivity in early sensory stages can predict the computational performance in the deeper levels, particularly for non-linear processing tasks. This might be the case for some components in the initial stages of visual processing, for instance when multiple features are combined. Whereas, the retinotopic maps are mostly conserved in the primary visual cortex (Girman et al., 1999; Adams and Horton, 2003), these gradually overlap (approximated in the mixed input scenario) in the subsequent areas, giving rise to more complex receptive fields and tuning properties (Hubel, 1988). Our results suggest that topographic maps may play a vital role in balancing between accurate transmission of state representations as well as controlling where and how information is integrated.

Despite the limitations of our models, we have highlighted the importance of biologically plausible structural patterning for information processing in modular spiking networks. Even simple forms of topography were shown to significantly enhance computational performance in the deeper modules. Additionally, architectural constraints have a considerable impact on the effectiveness with which different inputs are integrated, with early mixing being clearly advantageous and highlighting a possibly relevant feature of hierarchical processing. Taken together, these results provide useful constraints for building modular systems composed of spiking balanced networks that enable accurate information transmission.

4.4. Limitations and Future Work

Our analysis consisted of a relatively simple implementation both in terms of the microcircuit composition and the characteristics of topographic maps. Even though abstractions are required in any modeling study, it is important to highlight the inherent limitations and drawbacks.

The network we referred to as *random* in this study (**Figure 1A**) was considered to be the most appropriate to serve as a baseline for the unstructured architecture, due to its simplicity. However, there are many other classes of non-modular networks, such as small-world or scale-free networks, which are likely to display similar or even superior computational characteristics than our baseline. Investigating the behavior and impact of such alternative network structures could be an interesting topic for future research, as they constitute intermediate steps between fully random and modular architectures.

We have employed a simple process to embed topographic maps in unstructured networks (see section 2), whereby the map size (i.e., size of a population involved in a specific pathway) was kept constant in all modules. Cortical maps, however, exhibit more structured and complex spatial organization (Bednar and Wilson, 2016), characterized by a decrease in topographic specificity with hierarchical depth. This, in turn, is likely a consequence of increasingly overlapping projections and increasing map sizes and is considered to have significant functional implications (see e.g., Rigotti et al., 2013), which we did not explore in more detail here. Nevertheless, our results (**Figure 2F**) suggest that, at least for the relatively simple and low-dimensional (considering the network size) tasks employed in this study, overlapping projections have a detrimental effect on the network's discrimination ability. More complex tasks involving high-dimensional mappings would therefore negatively impact the performance of our modular networks. Assuming a one-to-one mapping between input dimensions and stimulus-specific neuronal clusters, a larger task dimensionality would require either fewer neurons per cluster, or some compensation mechanism (e.g., stronger or denser projections between the clusters), possibly limiting the task complexity smaller local circuits can handle. Alternatively, cortical circuits might solve this dimensionality problem by combining multiple modules dynamically, in a task-dependent manner (Yang et al., 2019).

In addition, cortical systems also display an abundance of feedback loops that exhibit, similarly to the feed-forward cortico-cortical connections, a high degree of specificity and spatial segregation (Markov and Kennedy, 2013; Markov et al., 2014). Such feedback connections from more anterior cortical regions (typically associated with more abstract or "higher" cognitive functions) have been shown to play a central role in top-down control and modulation of sensory processing by providing contextual information and facilitating multisensory integration (see e.g., Clavagnier et al., 2004; Markopoulos et al., 2012; Revina et al., 2018). In addition, important theoretical frameworks of cortical processing, known as predictive coding theories (Friston and Kiebel, 2009; Bastos et al., 2012) place a fundamental importance in the role of such top-down feedback as a pathway through which internal predictions from higher cortical regions are propagated downstream and used as an explicit error signal, guiding, and structuring the nature of internal representation in the hierarchically lower cortical modules.

Although their functional role is not entirely unambiguous and depends on specific functional interpretations, a recent study (Joglekar et al., 2018) found that these feedback projections have a destabilizing effect on long-range signal propagation. Failure to account for feedback projections will therefore limit the scope and generalizability of our models. Nevertheless, these limitations do not invalidate the main conclusions of this study pertaining to the importance of structured projections in signal propagation and integration in the context of feed-forward network architectures.

An additional aspect of anatomical connectivity concerns the presence of long-range projections, directly linking distant cortical modules (commonly referred to as skip, or "jump" connections Knösche and Tittgemeyer, 2011). Such projections between non-adjacent areas were found to significantly improve the short-term memory capacity of a biologically realistic spiking network model (Schomers et al., 2017), suggesting that a similar effect could be expected in our model. Additionally, in the domain of artificial neural networks, there is an entire class of architectures that exploit this principle (residual networks, ResNet), that demonstrate their functional significance as a way to eliminate singularities during training and ameliorate the problem of vanishing gradients (Orhan and Pitkow, 2017), as well as improving performance in image standard recognition tasks (He et al., 2016). Even though these aspects were not explicitly explored in this study as they would greatly extend its scope, these studies support the crucial role of network architecture and the nature of inter-modular connections in determining the system's computational characteristics.

Ultimately, understanding the core principles of cortical computation requires bridging neuro-anatomy and physiology with cognitively relevant computations. The classification and XOR problems we have employed here provided a convenient method to investigate information transfer across multiple spiking modules, and allowed us to shed light on the functional implications of the wiring architecture. However, it is imperative that future works tackle more complex, behaviorally relevant tasks and possibly more detailed anatomical and physiological observations to help disentangle the nature of cognitive processing across cortical hierarchies.

DATA AVAILABILITY STATEMENT

All the relevant data is available from the Open Science Framework database (https://osf.io/nywc2/, see **Supplementary Material**).

AUTHOR CONTRIBUTIONS

BZ, SM, AM, and RD designed the study. BZ and RD performed the analysis. BZ performed the simulations and

implemented the models. BZ, AM, and RD contributed to writing of manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncom. 2019.00079/full#supplementary-material

REFERENCES

Abbott, L., Rajan, K., and Sompolinsky, H. (2009). Interactions between intrinsic and stimulus-evoked activity in recurrent neural networks. *arXiv [Preprint]. arXiv:0912.3832*. doi: 10.1093/acprof:oso/9780195393798.003.0004

Adams, D. L., and Horton, J. C. (2003). A precise retinotopic map of primate striate cortex generated from the representation of angioscotomas. *J. Neurosci.* 23, 3771–3789. doi: 10.1523/JNEUROSCI.23-09-03771.2003

Barak, O., Rigotti, M., and Fusi, S. (2013). The sparseness of mixed selectivity neurons controls the generalization–discrimination trade-off. *J. Neurosci.* 33, 3844–3856. doi: 10.1523/JNEUROSCI.2753-12.2013

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

Bednar, J. A., and Wilson, S. P. (2016). Cortical maps. *Neuroscientist* 22, 604–617. doi: 10.1177/1073858415597645

Brunel, N. (2000). Dynamics of networks of randomly connected excitatory and inhibitory spiking neurons. *J. Physiol. Paris* 94, 445–463. doi: 10.1016/S0928-4257(00)01084-6

Clavagnier, S., Falchier, A., and Kennedy, H. (2004). Long-distance feedback projections to area v1: implications for multisensory integration, spatial awareness, and visual consciousness. *Cogn. Affect. Behav. Neurosci.* 4, 117–126. doi: 10.3758/CABN.4.2.117

Diesmann, M., Gewaltig, M. O., and Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature* 402, 529–533. doi: 10.1038/990101

Duarte, R. (2015). Expansion and state-dependent variability along sensory processing streams. *J. Neurosci.* 35, 7315–7316. doi: 10.1523/JNEUROSCI.0874-15.2015

Duarte, R., Seeholzer, A., Zilles, K., and Morrison, A. (2017). Synaptic patterning and the timescales of cortical dynamics. *Curr. Opin. Neurobiol.* 43, 156–165. doi: 10.1016/j.conb.2017.02.007

Duarte, R., Uhlmann, M., den van Broek, D., Fitz, H., Petersson, K. M., and Morrison, A. (2018). "Encoding symbolic sequences with spiking neural reservoirs," in *2018 International Joint Conference on Neural Networks (IJCNN)* (Rio de Janeiro), 1–8.

Duarte, R. C., and Morrison, A. (2014). Dynamic stability of sequential stimulus representations in adapting neuronal networks. *Front. Comput. Neurosci.* 8:124. doi: 10.3389/fncom.2014.00124

Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K., and Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science* 327, 584–587. doi: 10.1126/science.1179867

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierachical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1

Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1211–1221. doi: 10.1098/rstb.2008.0300

Gewaltig, M.-O., and Diesmann, M. (2007). Nest (neural simulation tool). *Scholarpedia* 2:1430. doi: 10.4249/scholarpedia.1430

Girman, S. V., Sauvé, Y., and Lund, R. D. (1999). Receptive field properties of single neurons in rat primary visual cortex. *J. Neurophysiol.* 82, 301–311. doi: 10.1152/jn.1999.82.1.301

Haeusler, S., and Maass, W. (2007). A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models. *Cereb. Cortex* 17, 149–162. doi: 10.1093/cercor/bhj132

Hagler, D. J., and Sereno, M. I. (2006). Spatial maps in frontal and prefrontal cortex. *NeuroImage* 29, 567–577. doi: 10.1016/j.neuroimage.2005.08.058

Harris, K. D., and Shepherd, G. M. (2015). The neocortical circuit: themes and variations. *Nat. Neurosci.* 18, 170–181. doi: 10.1038/nn.3917

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 770–778.

Hubel, D. H. (1988). *Eye, Brain, and Vision / David H. Hubel*. New York, NY: Scientific American Library.

Joglekar, M. R., Mejias, J. F., Yang, G. R., and Wang, X. J. (2018). Inter-areal balanced amplification enhances signal propagation in a large-scale circuit model of the primate cortex. *Neuron* 98, 222–234. doi: 10.1016/j.neuron.2018.02.031

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Res. Bull.* 44, 107–112. doi: 10.1016/S0361-9230(97)00094-4

Keller, G., Bonhoeffer, T., and Hübener, M. (2012). Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* 74, 809–815. doi: 10.1016/j.neuron.2012.03.040

Klampfl, S., and Maass, W. (2013). Emergence of dynamic memory traces in cortical microcircuit models through STDP. *J. Neurosci.* 33, 11515–11529. doi: 10.1523/JNEUROSCI.5044-12.2013

Knösche, T. R., and Tittgemeyer, M. (2011). The role of long-range connectivity for the characterization of the functional–anatomical organization of the cortex. *Front. Syst. Neurosci.* 5:58. doi: 10.3389/fnsys.2011.00058

Kremkow, J., Aertsen, A., and Kumar, A. (2010). Gating of signal propagation in spiking neural networks by balanced and correlated excitation and inhibition. *J. Neurosci.* 30, 15760–15768. doi: 10.1523/JNEUROSCI.3874-10.2010

Kumar, A., Rotter, S., and Aertsen, A. (2008). Conditions for propagating synchronous spiking and asynchronous firing rates in a cortical network model. *J. Neurosci.* 28, 5268–5280. doi: 10.1523/JNEUROSCI.2542-07.2008

Kumar, A., Rotter, S., and Aertsen, A. (2010). Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nat. Rev. Neurosci.* 11, 615–627. doi: 10.1038/nrn2886

Kunkel, S., Morrison, A., Weidel, P., Eppler, J. M., Sinha, A., Schenck, W., et al. (2017). *NEST 2.12.0*. Jülich.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Lukoševičius, M., and Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* 3, 127–149. doi: 10.1016/j.cosrev.2009.03.005

Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* 14, 2531–2560. doi: 10.1162/089976602760407955

Maass, W., Natschläger, T., and Markram, H. (2004). Fading memory and kernel properties of generic cortical microcircuit models. *J. Physiol. Paris* 98, 315–330. doi: 10.1016/j.jphysparis.2005.09.020

Macaluso, E., Frith, C. D., and Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science* 289, 1206–1208. doi: 10.1126/science.289.5482.1206

Mante, V., Sussillo, D., Shenoy, K. V., and Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503, 78–84. doi: 10.1038/nature12742

Markopoulos, F., Rokni, D., Gire, D., and Murthy, V. (2012). Functional properties of cortical feedback projections to the olfactory bulb. *Neuron* 76, 1175–1188. doi: 10.1016/j.neuron.2012.10.028

Markov, N. T., and Kennedy, H. (2013). The importance of being hierarchical. *Curr. Opin. Neurobiol.* 23, 187–194. doi: 10.1016/j.conb.2012.12.008

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., et al. (2014). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J. Compar. Neurol.* 522, 225–259. doi: 10.1002/cne.23458

Mazzucato, L., Fontanini, A., and La Camera, G. (2016). Stimuli reduce the dimensionality of cortical activity. *Front. Syst. Neurosci.* 10:11. doi: 10.3389/fnsys.2016.00011

Miller, K. D. (2016). Canonical computations of cerebral cortex. *Curr. Opin. Neurobiol.* 37, 75–84. doi: 10.1016/j.conb.2016.01.008

Mountcastle, V. (1978). *An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System.* Cambridge, MA: MIT Press.

Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain* 120, 701–722. doi: 10.1093/brain/120.4.701

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., et al. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* 17, 1661–1663. doi: 10.1038/nn.3862

Nikolić, D., Hausler, S., Singer, W., and Maass, W. (2009). Distributed fading memory for stimulus properties in the primary visual cortex. *PLoS Biol.* 7:e1000260. doi: 10.1371/journal.pbio.1000260

Orhan, A. E., and Pitkow, X. (2017). Skip connections eliminate singularities. *arXiv [Preprint]. arXiv:1701.09175*.

Park, H. J., and Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science* 342, 1238411–1238411. doi: 10.1126/science.1238411

Pauli, R., Weidel, P., Kunkel, S., and Morrison, A. (2018). Reproducing polychronization: a guide to maximizing the reproducibility of spiking network models. *Front. Neuroinformatics* 12:46. doi: 10.3389/fninf.2018.00046

Revina, Y., Petro, L. S., and Muckli, L. (2018). Cortical feedback signals generalise across different spatial frequencies of feedforward inputs. *NeuroImage* 180, 280–290. doi: 10.1016/j.neuroimage.2017.09.047

Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K., et al. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–590. doi: 10.1038/nature12160

Rigotti, M., and Fusi, S. (2016). Estimating the dimensionality of neural responses with fMRI repetition suppression. *arXiv [Preprint]. arXiv:1605.03952*.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. doi: 10.1016/0377-0427(87)90125-7

Schomers, M. R., Garagnani, M., and Pulvermüller, F. (2017). Neurocomputational consequences of evolutionary connectivity changes in perisylvian language cortex. *J. Neurosci.* 37, 3045–3055. doi: 10.1523/JNEUROSCI.2693-16.2017

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

Shinomoto, S., Kim, H., Shimokawa, T., Matsuno, N., Funahashi, S., Shima, K., et al. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Comput. Biol.* 5:e1000433. doi: 10.1371/journal.pcbi.1000433

Shuler, M. G., and Bear, M. F. (2006). Reward timing in the primary visual cortex. *Science* 311, 1606–1609. doi: 10.1126/science.1123513

Silver, M. A., Ress, D., and Heeger, D. J. (2005). Topographic maps of visual spatial attention in human parietal cortex. *J. Neurophysiol.* 94, 1358–1371. doi: 10.1152/jn.01316.2004

Sussillo, D. (2014). Neural circuits as computational dynamical systems. *Curr. Opin. Neurobiol.* 25, 156–163. doi: 10.1016/j.conb.2014.01.008

Tetzlaff, T., Buschermöhle, M., Geisel, T., and Diesmann, M. (2003). The spread of rate and correlation in stationary cortical networks. *Neurocomputing* 52, 949–954. doi: 10.1016/S0925-2312(02)00854-8

Thivierge, J. P., and Marcus, G. F. (2007). The topographic brain: from neural connectivity to cognition. *Trends Neurosci.* 30, 251–259. doi: 10.1016/j.tins.2007.04.004

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., et al. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* 373, 515–518. doi: 10.1038/373515a0

van den Broek, D., Uhlmann, M., Fitz, H., Duarte, R., Hagoort, P., and Petersson, K. M. (2017). "The best spike filter kernel is a neuron," in *Conference on Cognitive Computational Neuroscience* (New York, NY).

van Rossum, M. C., Turrigiano, G. G., and Nelson, S. B. (2002). Fast propagation of firing rates through layered networks of noisy neurons. *J. Neurosci.* 22, 1956–1966. doi: 10.1523/JNEUROSCI.22-05-01956.2002

Vogels, T. P., and Abbott, L. F. (2005). Signal propagation and logic gating in networks of integrate-and-fire neurons. *J. Neurosci.* 25, 10786–10795. doi: 10.1523/JNEUROSCI.3508-05.2005

Vogels, T. P., and Abbott, L. F. (2009). Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nat. Neurosci.* 12, 483–491. doi: 10.1038/nn.2276

Warden, M. R., and Miller, E. K. (2010). Task-dependent changes in short-term memory in the prefrontal cortex. *J. Neurosci.* 30, 15801–15810. doi: 10.1523/JNEUROSCI.1569-10.2010

Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T., and Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nat. Neurosci.* 22:297. doi: 10.1038/s41593-018-0310-2

Zajzon, B., Duarte, R., and Morrison, A. (2018). "Transferring state representations in hierarchical spiking neural networks," in *2018 International Joint Conference on Neural Networks (IJCNN)* (Rio de Janeiro), 1–9.

Zajzon, B., Mahmoudian, S., Morrison, A., and Duarte, R. (2019). Passing the message: representation transfer in modular balanced networks. *bioRxiv*. 1–26. doi: 10.1101/716118