# Signal denoising through topographic modularity of neural circuits

Barna Zajzon[1,2*], David Dahmen[1], Abigail Morrison[1,3], Renato Duarte[1,4]

**1** Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-BRAIN Institute I, Jülich Research Centre, 52425 Jülich, Germany
**2** Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen University, Aachen, Germany
**3** Department of Computer Science 3 - Software Engineering, RWTH Aachen University, Aachen, Germany
**4** Donders Institute for Brain, Cognition and Behavior, Radboud University, Nijmegen, the Netherlands

* b.zajzon@fz-juelich.de

**Abstract**

Information from the sensory periphery is conveyed to the cortex via structured projection pathways that spatially segregate stimulus features, providing a robust and efficient encoding strategy. Beyond sensory encoding, this prominent anatomical feature extends throughout the neocortex. However, the extent to which it influences cortical processing is unclear. In this study, we combine cortical circuit modeling with network theory to demonstrate that the sharpness of topographic projections acts as a bifurcation parameter, controlling the macroscopic dynamics and representational precision across a modular network. By shifting the balance of excitation and inhibition, topographic modularity gradually increases task performance and improves the signal-to-noise ratio across the system. We show that this is a robust and generic structural feature that enables a broad range of behaviorally-relevant operating regimes, and provide an in-depth theoretical analysis unravelling the dynamical principles underlying the mechanism.

# Introduction

Sensory inputs are often ambiguous, noisy and imprecise. Due to volatility in the environment and inaccurate peripheral representations, the sensory signals that arrive at the neocortical circuitry are often incomplete or corrupt [1, 2]. However, from these noisy input streams, the system is able to acquire reliable internal representations and extract relevant computable features at various degrees of abstraction [3, 4, 5]. Sensory perception in the mammalian neocortex thus relies on efficiently detecting the relevant input signals while minimizing the impact of noise.

Making sense of the environment also requires the estimation of features not explicitly represented by low-level sensory inputs. These inferential processes [6, 7] rely on the propagation of internal signals such as expectations and predictions, the accuracy of which must be evaluated against the ground truth, i.e. the sensory input stream. In a highly dynamic environment, this translates to a continuous process whose precision hinges on the fidelity with which external stimuli are encoded in the neural substrate. Additionally, as the system is modular and hierarchical (strikingly so in the sensory and motor components [8, 9]), it is critical that the external signal permeates the different processing modules despite the increasing distance from the sensory periphery (the input source) and the various transformations it is exposed to along the way, which degrade the signal via the interference of task-irrelevant and intrinsic, ongoing activity.

Accurate signal propagation can be achieved in a number of ways. One obvious solution is the direct routing and distribution of the signal, such that direct sensory input can be fed to different processing modules, which may be partially achieved through thalamocortical projections [10, 11]. Another possibility, which we explore in this study, is to propagate the input signal through tailored pathways that route the information throughout the system, allowing different processing stages to retrieve it without incurring much representational loss. Throughout the mammalian neocortex, the existence and characteristics of structured projections (topographic maps) presents a possible substrate for such signal routing. By preserving the relative organization of tuned neuronal populations, such maps imprint spatiotemporal features of (noisy) sensory inputs onto the cortex [12, 13, 14]. In a previous study [15], we discovered that structured projections can create feature-specific pathways that allow the external inputs to be faithfully represented and propagated throughout the system, but it remains unclear which connectivity properties are critical and what the underlying mechanism is. Moreover, beyond mere sensory representation, there is evidence that such structure-preserving mappings are also involved in more complex cognitive processes in associative and frontal areas [16, 17, 18], suggesting that topographic maps are a prominent structural feature of cortical organization.

In this study, we hypothesize that structured projection pathways allow sensory stimuli to be accurately reconstructed as they permeate multiple processing modules. We demonstrate that, by modulating effective connectivity and regional E/I balance, topographic projections additionally serve a *denoising* function, not merely allowing the faithful propagation of input signals, but systematically improving the system's internal representations and increasing signal-to-noise ratio. We identify a critical threshold in the degree of modularity in topographic projections, beyond which the system behaves effectively as a denoising autoencoder[1]. Additionally, we demonstrate that this phenomenon is robust, persisting across very different models, and hinges solely on the modularity of topographic projections. Our results suggest that modular structure in feedforward projection pathways can have a significant effect on the system's qualitative behavior, enabling a wide range of behaviorally-relevant and empirically supported dynamic regimes. This allows the system to: (i) maintain stable representations of multiple stimulus features [19]; (ii) amplify features of interest while suppressing others through winner-takes-all mechanisms [20, 21]; and (iii) dynamically represent different stimulus features as stable and metastable states and stochastically switch among active representations through a winnerless competition effect [22, 23, 24].

Our key finding, that the modulation of information processing dynamics and the fidelity of stimulus/feature representations results from the structure of topographic feedforward projections, provides new meaning and functional relevance to the pervasiveness of these projection maps throughout the mammalian neocortex. Beyond routing feature-specific information from sensory transducers through brainstem, thalamus and into primary sensory cortices (notably tonotopic, retinotopic and somatotopic maps), their maintenance within the neocortex [18] ensures that even cortical regions that are not directly engaged with the sensory input (higher-order cortex), can receive faithful representations of it, and that these internal signals, emanating from lower-order cortical areas, can dramatically skew and modulate the circuit's E/I balance and local functional connectivity, resulting in fundamental differences in the systems' responsiveness.

[1]Note that the parallel is established here on conceptual, not formal, grounds as the system is capable of retrieving the original, uncorrupted input from a noisy source, but bears no formal similarity to denoising autoencoder algorithms.

# Results

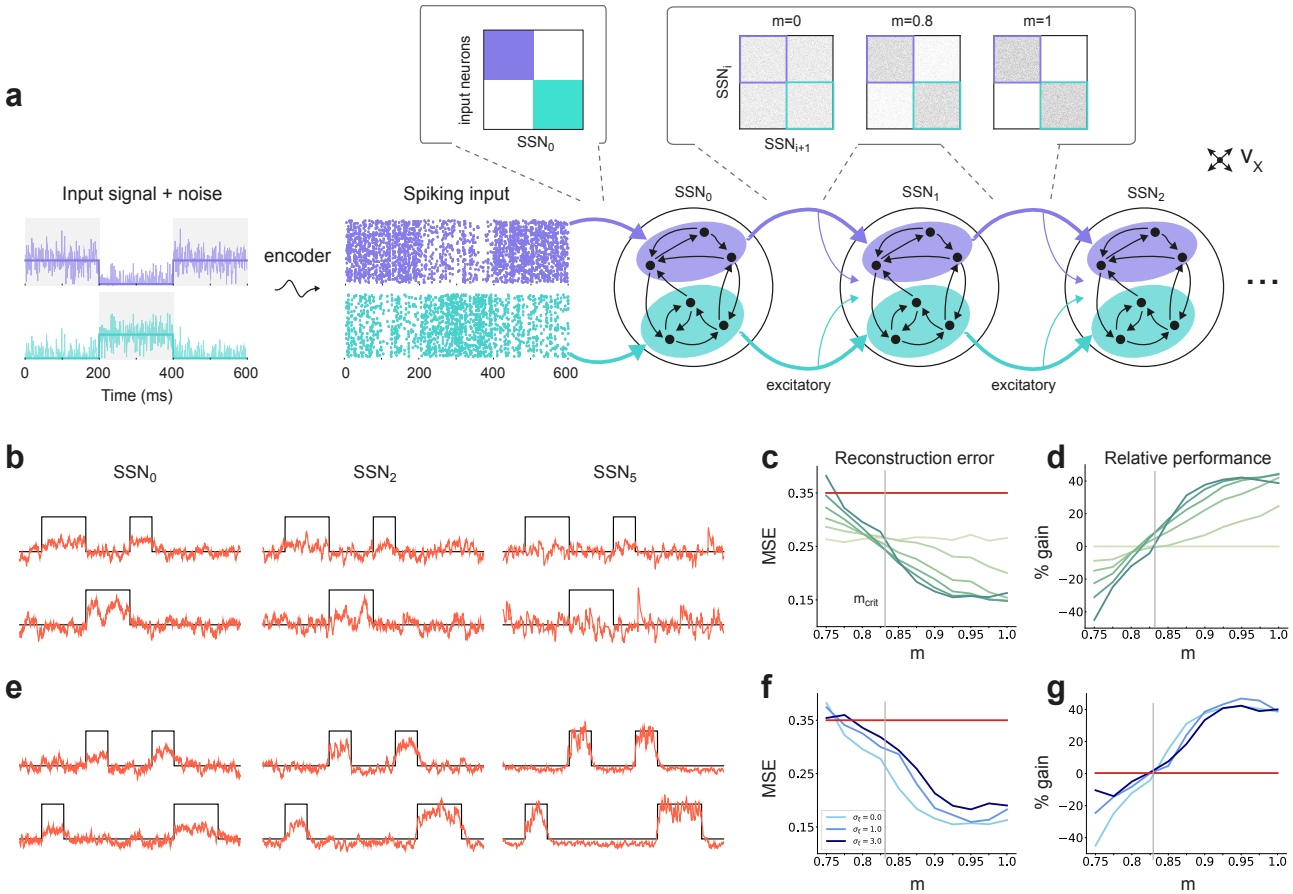


**Figure 1: Sequential denoising spiking architecture.** **(a)** A continuous step signal is used to drive the network. The input is spatially encoded in the first sub-network ($\mathrm{SSN}_0$), whereby each input channel is mapped exclusively onto a sub-population of stimulus-specific excitatory and inhibitory neurons (schematically illustrated by the colors; see also inset, top left). This exclusive encoding is retained to variable degrees across the network, through topographically structured feedforward projections (inset, top right) controlled by the modularity parameter $m$ (see Methods). This is illustrated explicitly for both topographic maps (purple and cyan arrows). Projections between SSNs are purely excitatory and target both excitatory and inhibitory neurons. **(b)** Signal reconstruction across the network. Single-trial illustration of target signal (black step function) and readout output (red curves) in 3 different SSNs, for $m = 0.75$ and no added noise ($\sigma_\xi = 0$). For simplicity, only two out of ten input channels are shown. **(c)** Signal reconstruction error in the different SSNs for the no-noise scenario shown in **(b)**. Color shade denotes network depth, from $\mathrm{SSN}_0$ (lightest) to $\mathrm{SSN}_5$ (darkest). The horizontal red line represents chance level, while the grey vertical line marks the critical transition point $m_{\mathrm{crit}} \approx 0.83$ (see main text). **(d)** Performance gain across the network, relative to $\mathrm{SSN}_0$, for the setup illustrated in **(b)**. **(e)** as in **(b)** but for $m = 0.9$. **(f)** Reconstruction error in $\mathrm{SSN}_5$ for the different noise intensities. Horizontal and vertical dashed lines as in **(c)**. **(g)** Performance gain in $\mathrm{SSN}_5$, relative to $\mathrm{SSN}_0$.

To investigate the role of structured pathways between processing modules in modulating the fidelity of stimulus representations, we study a network comprising up to six sequentially connected sub-networks (SSNs, see Methods and Fig. 1a). Each SSN is a *balanced random network* (see, e.g. [25]) of 10000, sparsely and randomly coupled leaky integrate-and-fire neurons (80% excitatory and 20% inhibitory). In each SSN, neurons are assigned to sub-populations associated with a particular stimulus. Excitatory neurons belonging to such stimulus-specific sub-populations then project to the subsequent SSN with a varying degree of specificity. We refer to a set of stimulus-specific sub-populations across the network and the structured feedforward projections among them as a *topographic map*. The specificity of the map is determined by the degree of *modularity* of the corresponding projections matrices (see, e.g. Fig. 1a). Modularity is thus defined as the relative density of connections within a stimulus-specific pathway (i.e. connecting sub-populations associated to the *same* stimulus; see Methods and Fig. 1a). In the following, we study the role of topographic specificity in modulating the system's functional and representational dynamics and its ability to cope with noise-corrupted input signals.

## Sequential denoising through structured projections

By systematically varying the degree of modular specialization in the feedforward projections (modularity parameter, $m$, see Methods and Fig. 1), we can control the segregation of stimulus-specific pathways across the network and investigate how it influences the characteristics of neural representations as the signal propagates. If the feedforward projections are unstructured or moderately structured ($m \lesssim 0.8$), information about the input fails to permeate the network, resulting in a chance-level reconstruction accuracy in the last sub-network, $\mathrm{SSN}_5$, even in the absence of noise (see Fig. 1b-c). However, as $m$ approaches a critical value $m_{\mathrm{crit}} \approx 0.83$, there is a qualitative transition in the system's behavior, leading to a consistently higher reconstruction accuracy across the sub-networks (Fig. 1b-e), regardless of the amount of noise added to the signal (Fig. 1f-g).

Beyond this critical transition point, reconstruction accuracy improves with depth, i.e. the signal is more accurately represented in $\mathrm{SSN}_5$ than in the initial sub-network, $\mathrm{SSN}_0$, with an effective accuracy gain of over 40% (Fig. 1d, g). While the addition of noise does impair the absolute reconstruction accuracy in all cases (see Supplementary Figure S1), the denoising effect persists even if the input is severely corrupted ($\sigma_\xi = 3\nu_{\mathrm{in}}$, see Fig. 1f-g). This is a counter-intuitive result, suggesting that topographic modularity is not only necessary for reliable communication across multiple populations (see [15]), but also supports an effective denoising effect, whereby representational precision increases with depth, even if the signal is profoundly distorted by noise.

## Noise suppression and response amplification

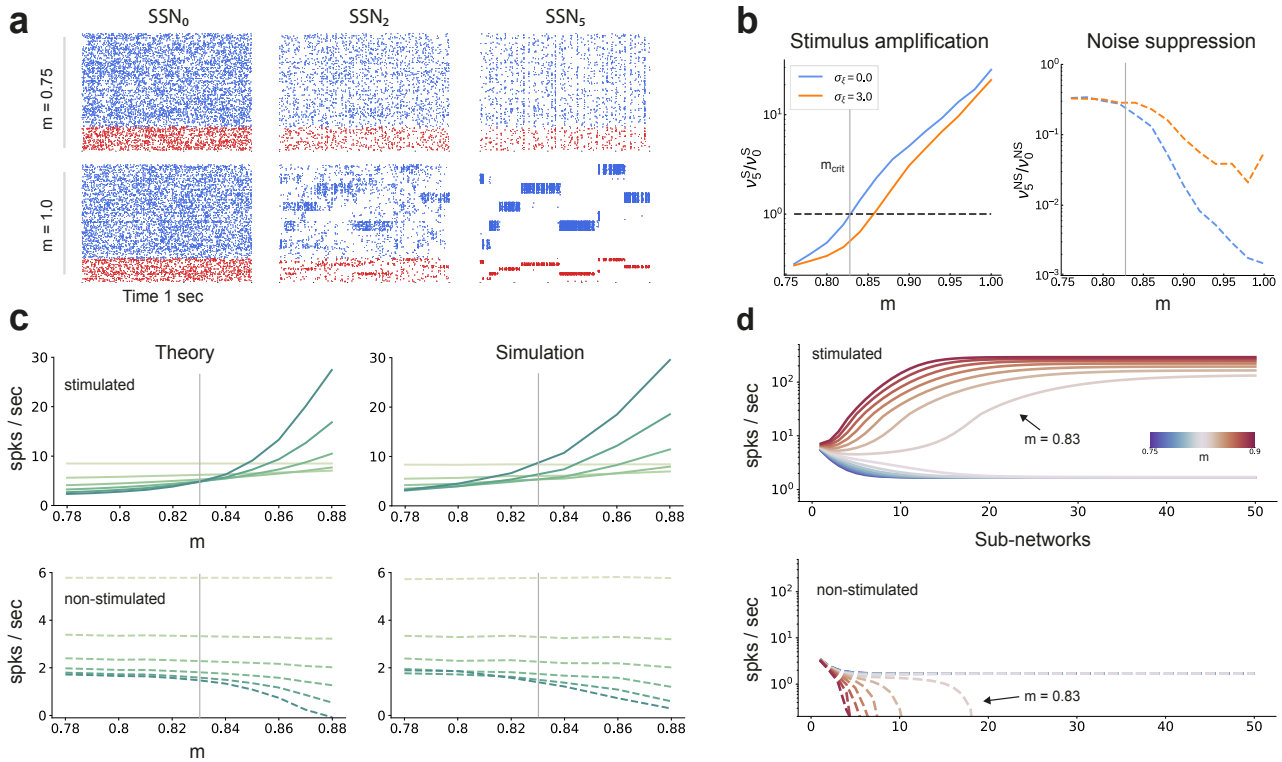


**Figure 2: Activity modulation and representational precision.** **(a)** 1 second of spiking activity observed across 1000 randomly chosen excitatory (blue) and inhibitory (red) neurons in $\mathrm{SSN}_0$, $\mathrm{SSN}_2$ and $\mathrm{SSN}_5$, for $\sigma_\xi = 3\nu_{\mathrm{in}}$ and $m = 0.75$ (top) and $m = 1$ (bottom). **(b)** Mean quotient of firing rates in $\mathrm{SSN}_5$ and $\mathrm{SSN}_0$ ($\nu_5/\nu_0$) for stimulated (S, left) and non-stimulated (NS, right) sub-populations for different input noise levels, describing response amplification and noise suppression, respectively. **(c)** Mean firing rates of the stimulated (top) and non-stimulated (bottom) excitatory sub-populations in the different SSNs (color shade as in Fig. 1), for $\sigma_\xi = 0$. For modularity values facilitating an asynchronous irregular regime across the network, the firing rates predicted by mean-field theory (left) closely match the simulation data (right). **(d)** Mean-field predictions for the stationary firing rates of the stimulated (top) and non-stimulated (bottom) sub-populations, in a system with 50 sub-networks and $\sigma_\xi = 0$. Note that all reported simulation data corresponds to the mean firing rates acquired over a period of 10 seconds and averaged across 5 trials per condition.

The sequential denoising effect observed beyond the critical transition point $m_{\mathrm{crit}} \approx 0.83$ results in an increasingly accurate input encoding through progressively more precise internal representations. In general, such a

phenomenon could be achieved either through noise suppression, stimulus-specific response amplification or both. In this section, we examine these possibilities by analysing and comparing the input-driven dynamics of the different sub-networks. The strict segregation of stimulus-specific sub-populations in $\mathrm{SSN}_0$ is only fully preserved across the system if $m = 1$, in which case signal encoding and transmission primarily rely on this spatial segregation. Spiking activity across the different SSNs (Fig. 2a) demonstrates that the system gradually sharpens the segregation of stimulus-specific sub-populations; indeed, in systems with fully modular feedforward projections, activity in the last sub-network is concentrated predominantly in the stimulated sub-populations. This effect can be observed in both excitatory (E) and inhibitory (I) populations, as both are equally targeted by the feedforward excitatory projections. The sharpening effect consists of both *noise suppression* and *response amplification* (Fig. 2b), measured as the relative firing rates of the non-stimulated $\nu_5^{\mathrm{NS}}/\nu_0^{\mathrm{NS}}$ and stimulated sub-populations $\nu_5^{\mathrm{S}}/\nu_0^{\mathrm{S}}$, respectively. For $m < m_{\mathrm{crit}}$, noise suppression is only marginal and responses within the stimulated pathways are not amplified ($\nu_5^{\mathrm{S}}/\nu_0^{\mathrm{S}} < 1$).

Mean-field analysis of the stationary network activity (see Methods and Appendix B) predicts that the firing rates of the stimulus-specific sub-populations increase systematically with modularity, whereas the untuned neurons are gradually silenced (Fig. 2c, left). At the critical transition point $m_{\mathrm{crit}} \approx 0.83$, mean firing rates across the different sub-networks converge, which translates into a globally uniform signal encoding capacity, corresponding to the zero-gain convergence point in Fig. 1d, g. As the degree of modularity increases beyond this point, the self-consistent state is lost again as the functional dynamics across the network shifts towards a gradual response sharpening, whereby the activity of stimulus-tuned neurons become increasingly dominant (Fig. 2a-c). The effect is more pronounced for the deeper sub-networks. Note that the analytical results match well with those obtained by numerical simulation (Fig. 2c, right).

In the limit of very deep networks (up to 50 SSNs, Fig. 2d) the system becomes bistable, with rates converging to either a high-activity state associated with signal amplification or a low-activity state driven by the background input. The critical transition point is observed at a modularity value of $m = 0.83$, matching the results reported so far. Below this value, elevated activity in the stimulated sub-populations can be maintained across the initial sub-networks ($< 10$), but eventually dies out; the rate of all neurons decays and information about the input cannot reach the deeper populations. Importantly, for $m = 0.83$, the transition towards the high-activity state is slower. This allows the input signal to faithfully propagate across a large number of sub-networks ($\approx 15$), without being driven into implausible activity states.

## E/I balance and asymmetric effective couplings

The departure from the balanced activity in the initial sub-networks can be better understood by zooming in at the synaptic level and analysing how topography influences the synaptic input currents. The segregation of feedforward projections into stimulus-specific pathways breaks the symmetry between excitation and inhibition (see Fig. 3a) that characterizes the balanced state [26, 27], for which the first two sub-networks were tuned (see Methods). E/I balance is thus systematically shifted towards excitation in the stimulated populations and inhibition in the non-stimulated ones. Neurons belonging to sub-populations associated with the active stimulus receive significantly more net overall excitation, whereas the other neurons become gradually more inhibited. This disparity grows not only with modularity but also with network depth. Overall, across the whole system, increasing modularity results in an increasingly inhibition-dominated dynamical regime (inset in Fig. 3a), whereby stronger effective inhibition silences non-stimulated populations, thus sharpening stimulus / feature representations by concentrating activity in the stimulus-driven sub-populations.

To gain an intuitive understanding of these effects from a dynamical systems perspective, we linearize the network dynamics around the stationary working points of the individual populations [28] in order to obtain the effective connectivity $W$ of the system (see Methods and Appendix B). The effective impact of a single spike from a presynaptic neuron $j$ on the firing rate of a postsynaptic neuron $i$ (the effective weight $w_{ij} \in W$) is determined not only by the synaptic efficacies $J_{ij}$, but also by the statistics of the synaptic input fluctuations to the target cell $i$ that determine its excitability (see Methods, Eq. 3). This analysis reveals that there is an increase in the effective synaptic input onto neurons in the stimulated sub-populations as a function of modularity (Fig. 3b). Conversely, non-stimulated neurons effectively receive weaker excitatory (and stronger inhibitory) drive and become increasingly less responsive (see Fig. 3a, b). The role of topographic modularity in denoising can thus be understood as a transient, stimulus-specific change in effective connectivity.

For low and moderate topographic precision ($m \lesssim 0.83$), denoising does not occur as the effective weights are sufficiently similar to maintain a stable E/I balance across all populations and sub-networks (Fig. 3a, b), resulting in a relatively uniform global dynamical state (indicated in Fig. 3c by a constant spectral radius for

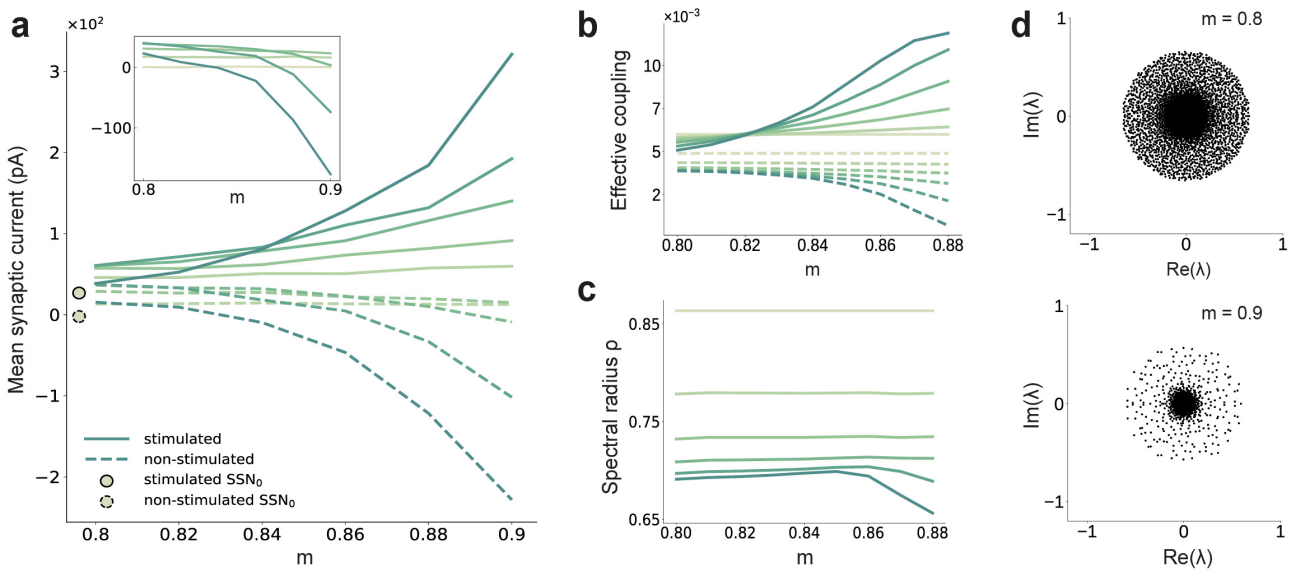


**Figure 3: Asymmetric effective couplings modulate the E/I balance and support sequential denoising.** **(a)** Mean synaptic input currents for neurons in the stimulated (solid curves) and non-stimulated (dashed curves) excitatory sub-populations in the different SSNs. To avoid clutter, data for $SSN_0$ is only shown by markers (independent of $m$). Inset shows the currents (in pA) averaged over all excitatory neurons in the different sub-networks; increasing modularity leads to a dominance of inhibition in the deeper sub-networks. Color shade represents depth, from $SSN_1$ (light) to $SSN_5$ (dark). **(b)** Mean-field approximation of the effective recurrent weights in $SSN_5$. Curve shade and style as in **(a)**. **(c)** Spectral radius of the effective connectivity matrices $\rho(W)$ as a function of modularity. **(d)** Eigenvalue spectra for the effective coupling matrices in $SSN_5$, for $m = 0.8$ (top) and $m = 0.9$ (bottom). The largest negative eigenvalue (outlier, see Methods), characteristic of inhibition-dominated networks, is omitted for clarity.

$m \lesssim 0.83$, see also Methods) and stable linearized dynamics ($\rho(W) < 1$).

However, as the feedforward projections become more structured, the system undergoes qualitative changes: after a weak transient ($0.83 \lesssim m \lesssim 0.85$) expansion of the spectral radius $\rho$ in the deep SSNs caused by the increased effective coupling to the stimulated sub-population (Fig. 3b), the spectral radius eventually ($m \gtrsim 0.85$) contracts with increasing modularity (Fig. 3c, d). Given that $\rho$ is determined by the variance of $W$, i.e. heterogeneity across connections [29], this behavior is expected: most weights are in the non-stimulated pathways, which decrease with larger $m$ and network depth (Fig. 3b). Strong inhibitory currents (Fig. 3a) suppress the majority of neurons, thereby reducing noise, as demonstrated by the collapse of the bulk of the eigenvalues towards the center for larger $m$ (Fig. 3d). Indicative of a more constrained state-space, this contractive effect suggests that population activity becomes gradually entrained by the spatially encoded input along the stimulated pathway, whereas the responses of the non-stimulated neurons have a diminishing influence on the overall behavior.

By biasing the effective connectivity of the system, precise topography can thus modulate the balance of excitation and inhibition in the different sub-networks, concentrating the activity along specific pathways. This results in both a systematic amplification of stimulus-specific responses and a systematic suppression of noise (Fig. 2b). The sharpness / precision of topographic specificity along these pathways thus acts as a critical control parameter that largely determines the qualitative behavior of the system and can dramatically alter its responsiveness to external inputs.

## Modulating inhibition

How can the system generate and maintain the elevated inhibition underlying such a noise-suppressing regime? On the one hand, feedforward excitatory input may increase the activity of certain excitatory neurons in $E_i$ of sub-network $SSN_i$, which, in turn, can lead to increased mean inhibition through local recurrent connections. On the other hand, denoising could depend strongly on the concerted topographic projections onto $I_i$. Such structured feedforward inhibition is known to play important functional roles in, e.g., sharpening the spatial contrast of somatosensory stimuli [30] or enhancing coding precision throughout the ascending auditory pathways [31].

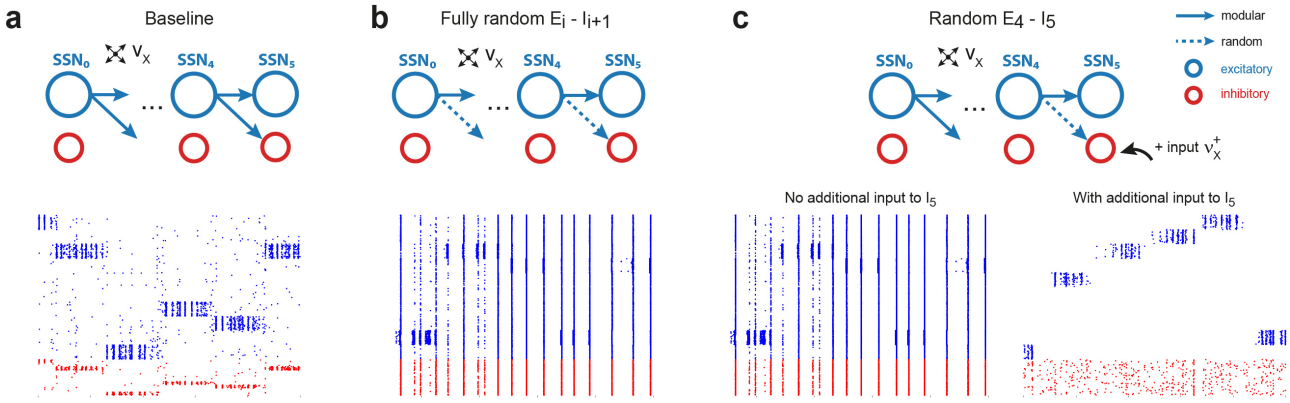


**Figure 4: Modular projections to inhibitory populations stabilize network dynamics.** Raster plots show 1 second of spiking activity of 1000 randomly chosen neurons in $SSN_5$, for different network configurations. **(a)** Baseline network with $m = 0.88$. **(b)** Unstructured feedforward projections to the inhibitory sub-populations lead to highly synchronized network activity, hindering signal representation. **(c)** Same as the baseline network in **(a)**, but with random projections for $E_4 \to I_5$ and additional but unspecific (Poissonian) excitatory input to $I_5$ controlled via $\nu_X^+$. Without such input ($\nu_X^+ = 0$, left), the activity is strongly synchronous, but this is compensated for by the additional excitation, reducing synchrony and restoring the denoising property ($\nu_X^+ = 10$ spikes/sec, right).

To investigate whether recurrent activity alone can generate sufficiently strong inhibition for signal transmission and denoising, we maintained the modular structure between the excitatory populations and randomized the feedforward projections onto the inhibitory ones ($m = 0$ for $E_i \to I_{i+1}$, compare top panels of Fig. 4a and b). This leads to unstable firing patterns in the downstream sub-networks, characterized by significant accumulation of synchrony and increased firing rates (see bottom panels of Fig. 4a and b and Supplementary Figure S3a, b). These effects, known to result from shared pre-synaptic excitatory inputs (see, e.g. [32, 33, 34]), are more pronounced for larger $m$ and network depth (see Supplementary Figure S3). Compared with the baseline network, whose activity shows clear spatially encoded stimuli (sequential activation of stimulus-specific sub-populations (Fig. 4a, bottom), removing structure from the projections onto inhibitory neurons abolishes the effect and prevents accurate signal transmission.

These effects of unstructured inhibitory projections are so marked that they can be observed even if a single set of projections is modified: this can be seen in Fig. 4c, where only the $E_4 \to I_5$ connections are randomized. It is worth noting, however, that the excessive synchronization that results from unstructured inhibitory projections (Fig. 4c bottom left, no additional input condition) can be easily counteracted by driving $I_5$ (the inhibitory population that receives only unstructured projections) with additional uncorrelated external input. If strong enough ($\nu_X^+ \approx 10$spk/sec), this additional external drive pushes the inhibitory population into an asynchronous regime that restores the sharp, stimulus-specific responses in the excitatory population of the corresponding sub-network (see Fig. 4c bottom right, and Supplementary Figure S3c).

These results emphasize the control of inhibitory neurons' responsiveness as the main causal mechanism behind the effects reported. Elevated local inhibition is strictly required, but whether this is achieved by tailored, stimulus-specific activation of inhibitory sub-populations, or by uncorrelated excitatory drive onto all inhibitory neurons appears to be irrelevant and both conditions result in sharp, stimulus-tuned responses in the excitatory populations.

## A generalizable structural effect

We have demonstrated that, by controlling the different sub-networks' operating point, the sharpness of feedforward projections allows the architecture to systematically improve the quality of internal representations and retrieve the input structure, even if profoundly corrupted by noise. In this section, we investigate the robustness of the phenomenon in order to determine whether it can be entirely ascribed to the topographic projections (a structural/architectural feature) or if the particular choices of models and model parameters for neuronal and synaptic dynamics contribute to the effect.

To do so, we study two alternative model systems on the signal denoising task. These are structured similarly to the baseline system explored so far, comprising separate sequential sub-networks with modular feedforward projections among them (see Fig. 1 and Methods), but vary in total size, neuronal and synaptic dynamics.

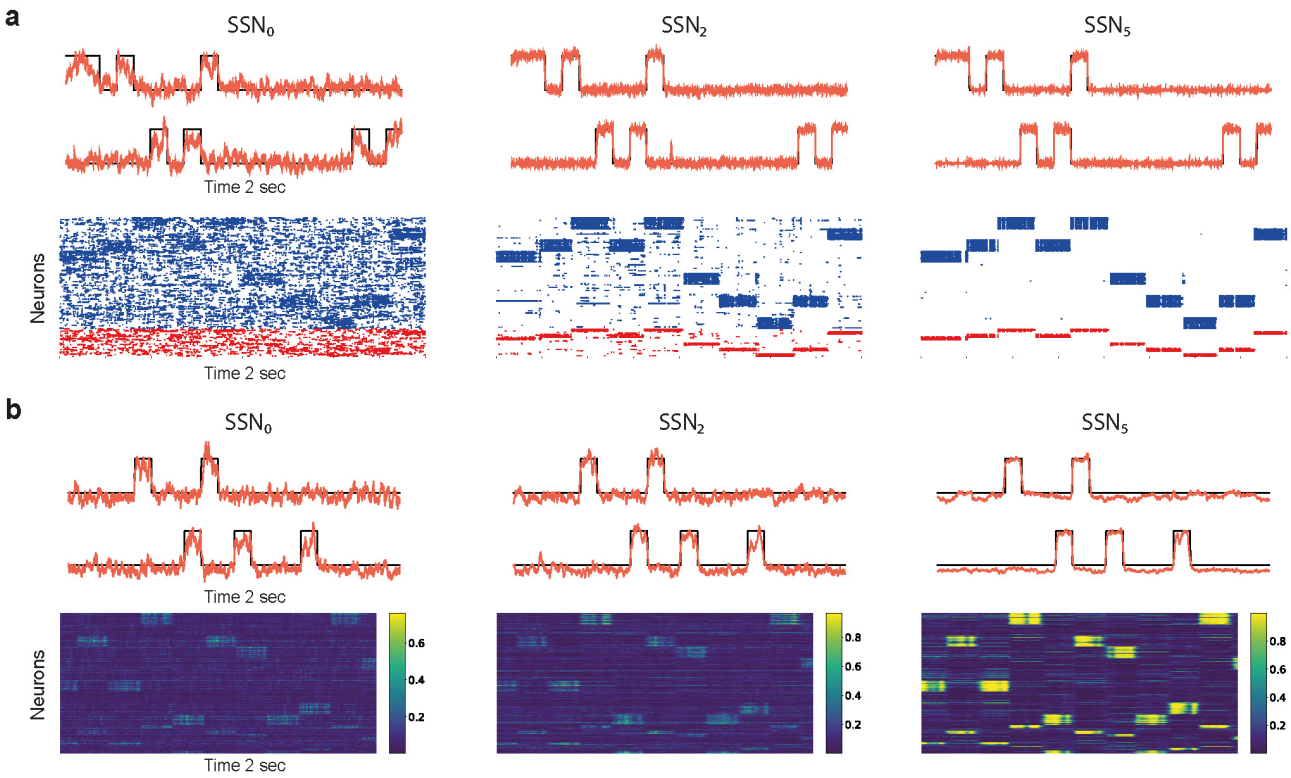


**Figure 5: Denoising through modular topography is a robust structural effect. (a)** Signal reconstruction (top) and corresponding network activity (bottom) for a network with LIF neurons and conductance-based synapses (see Methods). Single-trial illustration of target signal (black step function) and readout output (red curves) in three different SSNs, for $m = 0.9$ and strong noise corruption ($\sigma_\xi = 3\nu_{\text{in}}$). For simplicity, only two out of ten input channels are shown. **(b)** As in **(a)** for a rate-based model with $m = 1$ and $\sigma_\xi = 1\nu_{\text{in}}$ (see Methods for details).

In the first test case, only the models of synaptic transmission and corresponding parameters are altered. To increase biological verisimilitude and following [15], synaptic transmission is modelled as a conductance-based process, with different kinetics for excitatory and inhibitory transmission, corresponding to the responses of $AMPA$ and $GABA_a$ receptors, respectively, see Methods and Supplementary Tables S3 for details. The results, illustrated in Fig. 5a, demonstrate that task performance and population activity across the network follow a similar trend to the baseline model (Fig. 1 and Fig. 2a, b). Despite severe noise corruption, the system is able to generate a clear, discernible representation of the input as early as $\text{SSN}_2$ and can accurately reconstruct the signal. Importantly, the relative improvement with increasing modularity and network depth is retained. The transition occurs roughly for the same topographic configuration, $m \approx 0.85$, at which point the network dynamics converges towards a low-rate, stable asynchronous irregular regime across all populations, facilitating a linear firing rate propagation along the topographic maps (Supplementary Figure S4).

The second test case is a smaller and simpler network of nonlinear rate neuron models (see Fig. 5b and Methods) which interact via continuous signals (rates) rather than discontinuities (spikes). Despite these profound differences in the neuronal and synaptic dynamics, the same behavior is observed, demonstrating that sequential denoising is a structural effect, dependent on the population firing rates and thus less sensitive to fluctuations in the precise spike times. Moreover, the robustness with respect to the network size suggests that denoising could also be performed in smaller, localized circuits, possibly operating in parallel on different features of the input stimuli.

## Variable map sizes

Despite their ubiquity throughout the neocortex, the characteristics of structured projection pathways is far from uniform [13], exhibiting marked differences in spatial precision and specificity, aligned with macroscopic gradients of cortical organization. This non-uniformity may play an important functional role supporting feature aggregation [16] and the development of mixed representations [18] in higher (more anterior) cortical areas. Here, we consider two scenarios in the baseline (current-based) model to examine the robustness of our findings to more complex topographic configurations.

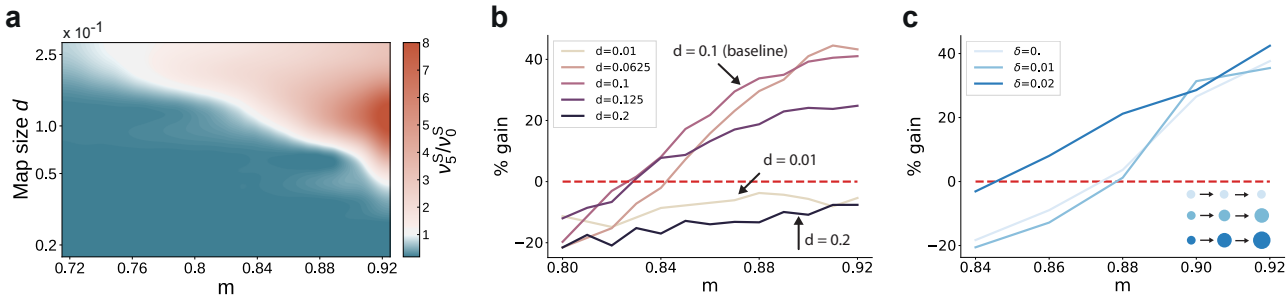


**Figure 6: Variation in the map sizes. (a)** Difference in the firing rates of the stimulated sub-populations in the first and last sub-networks, $\nu_5^{\mathrm{S}} - \nu_0^{\mathrm{S}}$, as a function of modularity and map size (parameterized by $d$ and constant throughout the network, i.e. $\delta = 0$, see Methods). Depicted values correspond to stationary firing rates predicted by mean-field theory, smoothed using a Lanczos filter. Note that, in order to ensure that every neuron was uniquely tuned, i.e. there is no overlap between stimulus-specific sub-populations, the number of sub-populations was chosen to be proportional to the map size ($N_{\mathrm{C}} = 1/d$). **(b-c)** Performance gain in $\mathrm{SSN}_5$ relative to $\mathrm{SSN}_0$ (ten stimuli, as in Fig. 1d, g), for varying properties of structural mappings: **(b)** fixed map size ($\delta = 0$) with color shade denoting map size, and **(c)** linearly increasing map size ($\delta > 0$) and a smaller initial map size $d_0 = 0.04$. The results depict the average performance gains measured across five trials, using the current-based model illustrated in Fig. 1 (ten stimuli) and no input noise ($\sigma_\xi = 0$).

First, we varied the size of stimulus-tuned sub-populations (parametrized by $d_{\mathrm{i}}$, see Methods) but kept them fixed across the network. For small sub-populations and intermediate degrees of topographic modularity, the activity along the stimulated pathway decays with network depth, suggesting that input information does not reach the deeper SSNs (see Fig. 6a and Supplementary Materials). These results place a lower bound on the size of stimulus-tuned sub-populations below which no signal propagation can occur, as reflected by the negative gain in performance for $d = 0.01$ (Fig. 6b). Whereas denoising is robust to variation around the baseline value of $d = 0.1$ that yielded perfect partitioning of the feedforward projections (see Supplementary Materials), an upper bound may emerge due to increasing overlap between the maps ($d = 0.2$ in Fig. 6b). In this case, the activity may "spill over" to other pathways than the stimulated one, corrupting the input representations and hindering accurate transmission and decoding. This can be alleviated by reduced or no overlap (as in Fig. 6a), in which case signal propagation and denoising is successful for larger map sizes ($\nu_5^{\mathrm{S}}/\nu_0^{\mathrm{S}} > 1$ also for $d > 0.1$). We thus observe a trade-off between map size, overlap and the degree of topographic precision that is required to accurately propagate stimulus representations (see Discussion).

Second, we took into account the fact that these structural features are known to vary with hierarchical depth resulting in increasingly larger sub-populations and, consequently, increasingly overlapping stimulus selectivity [35, 18, 13]. To capture this effect, we introduce a linear scaling of map size with depth ($d_{i+1} = \delta + d_i$ for $i \geq 1$, see Methods). The ability of the circuit to gradually clean the signal's representation is fully preserved, as illustrated in Fig. 6c. In fact, for intermediate modularity ($m < 0.9$) broadening the projections can further sharpen the reconstruction precision (compare curves for $\delta = 0.02$ and $\delta = 0$).

Taken together, these observations demonstrate that a gradual denoising of stimulus inputs can occur entirely as a consequence of the modular wiring between the subsequent processing circuits. Importantly, this effect generalizes well across diverse neuron and synapse models, as well as key system properties, making modular topography a potentially universal circuit feature for handling noisy data streams.

## Modularity as a bifurcation parameter

The results so far indicate that the modular topographic projections, more so than the individual characteristics of neurons and synapses, lead to a sequential denoising effect through a joint process of signal amplification and noise suppression. To better understand how the system transitions to such an operating regime, it is helpful to examine its macroscopic dynamics in the limit of many sub-networks [36, 37, 38]. We apply standard mean-field techniques [39, 40, 41] to find the asymptotic firing rates (fixed points across sub-networks) of the stimulated and non-stimulated sub-populations as a function of topography (Fig. 2d). For this, we can approximate the input $\mu$ to a group of neurons as a linear function of its firing rate $\nu$ with a slope $\kappa$ that is determined by the coupling within the group and an offset given by inputs from other groups of neurons (orange line in Fig. 7a). With an approximately sigmoidal rate transfer function, the self-consistent solutions are at the intersections marked in Fig. 7a.

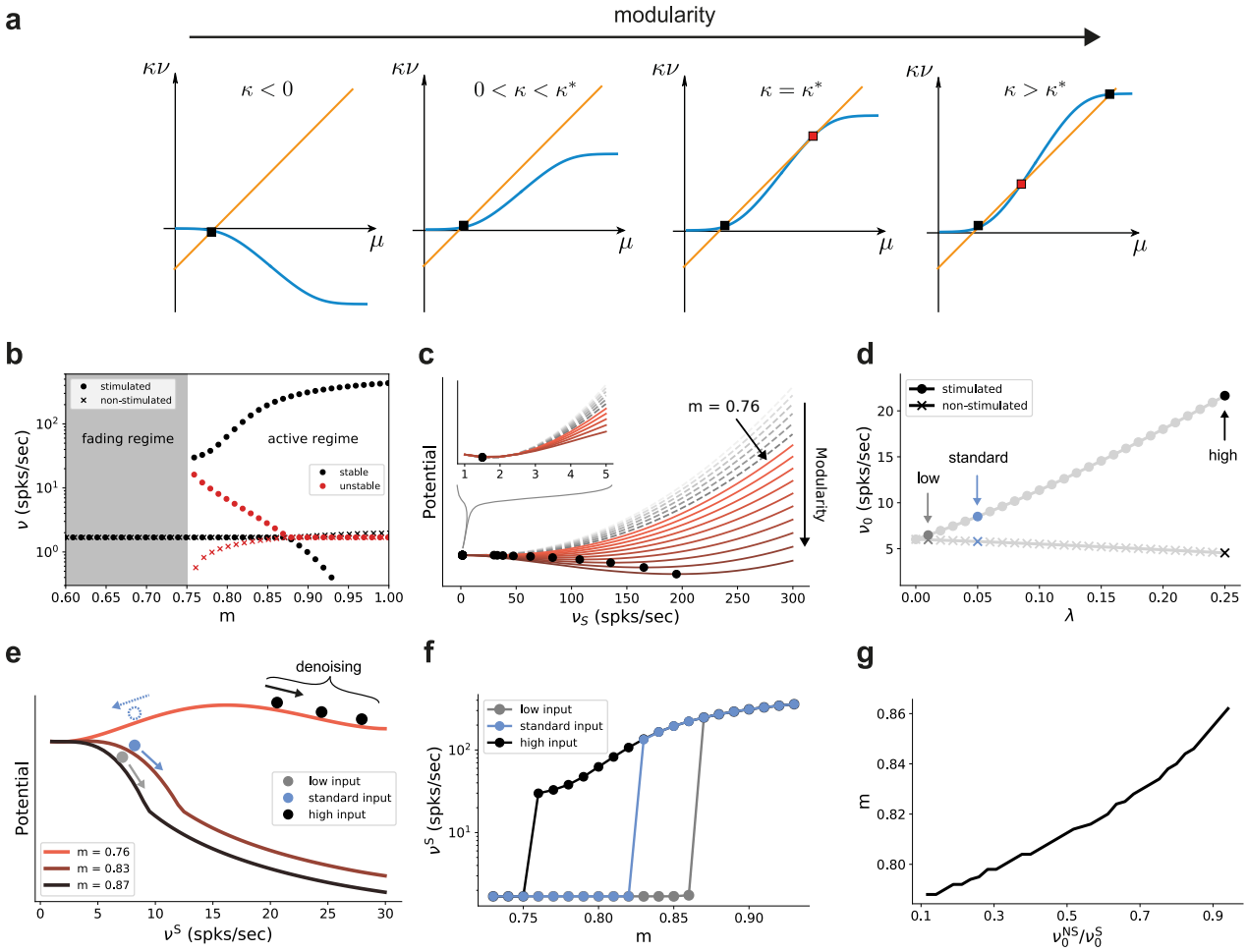


**Figure 7: Modularity changes the fixed point structure of the system. (a)** Self-consistent solution (for the full derivation, see Appendix B) for the firing rate of the stimulated sub-population (blue curves) and the linear relation $\kappa\nu = \mu - I$ (orange lines), in the limit of infinitely deep networks. Squares denote stable (black) and unstable (red) fixed points where input and output rates are the same. **(b)** Bifurcation diagram from numerical simulations showing a single stable fixed point in the fading regime, and both a stable (black) and an unstable (red) fixed point in the active regime where denoising occurs. **(c)** Potential energy of the mean activity (see Methods) for increasing topographic modularity. A stable state, corresponding to local minimum in the potential, exists at a low non-zero rate in every case, including for $m \leq 0.75$ (grey dashed curves, inset). For $m \geq 0.76$ (colored solid curves), a second fixed point appears at progressively larger firing rates. **(d)** Stationary firing rates of the stimulated and non-stimulated sub-populations in $SSN_0$, as a function of stimulus intensity ($\lambda$, see Methods). Low, standard and high denote $\lambda$ values of 0.01, 0.05 (baseline value used in Fig. 1) and 0.25, respectively. **(e)** Sketch of attractor basins in the potential for different values of $m$. Markers correspond to the highlighted initial states in **(d)**, with solid and dashed arrows indicating attraction towards the high- and low-activity state, respectively. **(f)** Firing rates of the stimulated sub-population as a function of modularity in the limit of infinite sub-networks, for the three different $\lambda$ marked in **(d)**. **(g)** Modularity threshold for the active regime shifts with increasing noise in the input, modeled as additional input to the non-stimulated sub-populations in $SSN_0$.

Formally, all neurons in the deep sub-networks of one topographic map form such a group as they share the same firing rate (asymptotic value). The coupling $\kappa$ within this group comprises not only recurrent connections of one sub-network but also modular feedforward projections across sub-networks. For small modularity, the group is in an inhibition-dominated regime ($\kappa < 0$) and we obtain only one fixed point at low activity (Fig. 7a, left). Importantly, the firing rate of this fixed point is the same for stimulated and non-stimulated topographic maps. Any influence of input signals applied to $SSN_0$ therefore vanishes in the deeper sub-networks and the signal cannot be reconstructed (*fading* regime). As topographic projections become more concentrated (larger $m$), $\kappa$ switches sign and gradually leads to two additional fixed points (as conceptually illustrated in Fig. 7a and numerically demonstrated in Fig. 7b, see also Appendix B): an unstable one (red) that eventually vanishes with increasing $m$ and a stable high-activity fixed point (black). The bistability opens the possibility to distinguish between stimulated and non-stimulated topographic maps and thereby reconstruct the signal in deep sub-networks: in the *active* regime beyond the critical modularity threshold (here $m > 0.75$), a sufficiently strong input signal can drive the activity along the stimulated map to the high-activity fixed point, such that it can permeate the

system, while the non-stimulated sub-populations still converge to the low-activity fixed point.

In the potential energy landscape $U$ (see Methods), where stable fixed points correspond to minima, the bistability that emerges for more structured topography can be understood as a transition from a single minimum at low rates (Fig. 7c, inset) to a second minimum associated with the high-activity state for $m > 0.75$ (Fig. 7c). More strongly modular networks cause deeper potential wells, corresponding to more attractive dynamical states and higher firing rates (see Supplementary Materials).

Because the intensity of the input signal dictates the rate of different populations in the initial sub-network $\mathrm{SSN}_0$ (Fig. 7d), it also determines, for any given modularity, whether the rate of the stimulated sub-population is in the basin of attraction of the high-activity (see Fig. 7e, solid markers and arrows) or low-activity (dashed, blue marker and arrow) fixed point. Denoising, and therefore increasing signal reconstruction, is thus achieved by successively (across sub-networks) pushing the population states down along the different potential gradients.

As reported above, for the baseline network and (standard) input ($\lambda = 0.05$) used in Fig. 1 and Fig. 2, the critical switching point between low and high activity is at $m = 0.83$ (blue markers in Fig. 7d, f). Stronger input signals move the switching point towards the minimal modularity $m = 0.76$ of the active regime (black markers in Fig. 7d, f), while weaker inputs only induce a switch at larger modularities (grey markers in Fig. 7d, f).

Noise in the input simply shifts the transition point to the high-activity state in a similar manner, with more modular connectivity required to compensate for stronger jitter (Fig. 7g). However, as long as the mean firing rate of the stimulated sub-population in $\mathrm{SSN}_0$ is slightly higher than that of the non-stimulated ones (up to 0.5 spks/sec), it is sufficient to position the system in the attracting basin of the high rate fixed point and the system is able to clean the signal representation. This indicates a remarkably robust denoising mechanism.

## Input integration and multi-stability

The analysis considered in the section above is restricted to a system driven with a single external stimulus. However, to adequately understand the system's dynamics, we need to account for the fact that it can be concurrently driven by multiple input streams. If two simultaneously active stimuli drive the system (see illustration in Fig. 8a), the qualitative behavior where the responses along the stimulated (non-stimulated) maps are enhanced (silenced) is retained if the strength of the two input channels is sufficiently different (Fig. 8b, top panel). In this case, the weaker stimulus is not strong enough to drive the sub-population it stimulates towards the basin of attraction of the high-activity fixed point. Consequently, the sub-population driven by this second stimulus behaves as a non-stimulated sub-population and the system remains responsive to only one of the two inputs, acting as a winner-takes-all (WTA) circuit. If, however, the ratio of stimulus intensities varies, two active sub-populations may co-exist (Fig. 8b, center) and/or compete (bottom panel), depending also on the degree of topographic modularity.

To quantify these variations in macroscopic behavior, we focus on the dynamics of $\mathrm{SSN}_5$ and measure the similarity (correlation coefficient) between the firing rates of the two stimulus-specific sub-populations as a function of modularity and ratio of input intensities $\lambda_2/\lambda_1$ (see Methods and Fig. 8c). In the case that both inputs have similar intensities but the feedforward projections are not sufficiently modular, both sub-populations are activated simultaneously (Co-Ex, red area in Fig. 8c). This is the dynamical regime that dominates the earlier sub-networks. However, this is a transient state, and the Co-Ex region gradually shrinks with network depth until it vanishes completely after approximately 9-10 SSNs (see Fig. 8d).

For low modularity, the system settles in the single stable state associated with near-zero firing rates, as illustrated schematically in the energy landscape in Fig. 8e (1) (see Methods, Appendix B and Supplementary Materials for derivations and numerical simulations). Above the critical modularity value, the system enters one of two different regimes. For $m > 0.84$ and an input ratio below 0.7 (Fig. 8c, grey area), one stimulus dominates (WTA) and the responses in the two populations are uncorrelated (Fig. 8b, top panel). Although the potential landscape contains two minima corresponding to either population being active, the system always settles in the high-activity attractor state corresponding to the dominating input (Fig. 8e, (2)).

If, however, the two inputs have comparable intensities and the topographic projections are sharp enough ($m > 0.84$), the system transitions into a different dynamical state where neither stimulus-specific sub-population can maintain an elevated firing rate for extended periods of time. In the extreme case of nearly identical intensities ($\lambda_2/\lambda_1 \geq 0.9$) and high modularity, the responses become anti-correlated (Fig. 8b, bottom panel), i.e. the activation of the two stimulus-specific sub-populations switches, as they engage in a dynamic behavior reminiscent

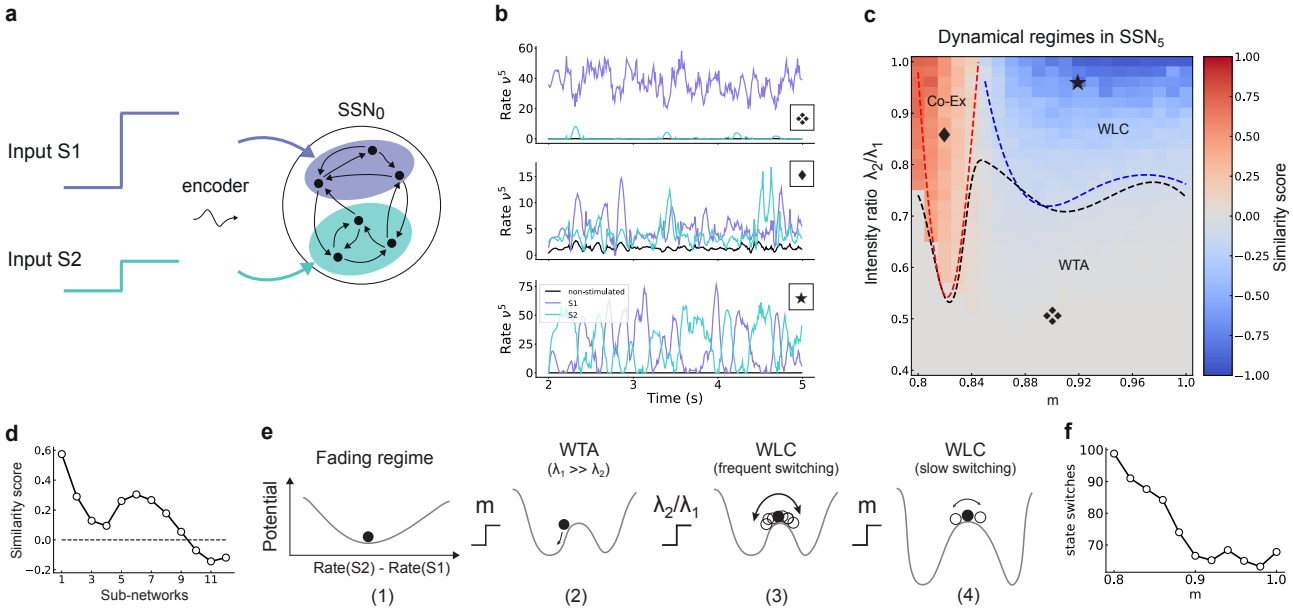


**Figure 8: For multiple input streams, topography may elicit a wide range of dynamical regimes.** **(a)** Two active input channels with corresponding stimulus intensities $\lambda_1$ and $\lambda_2$, mapped onto non-overlapping sub-populations, drive the network simultaneously. Throughout this section, $\lambda_1 = 0.05$ is fixed to the previous baseline value. **(b)** Mean firing rates of the two stimulated sub-populations (purple and cyan), as well as the non-stimulated sub-populations (black) for three different combinations of $m$ and ratios $\lambda_2/\lambda_1$ (as marked in **(c)**).**(c)** Correlation based similarity score shows three distinct dynamical regimes in $SSN_5$ when considering the firing rates of two, simultaneously stimulated sub-populations associated with $S_1$ and $S_2$, respectively: coexisting (Co-Ex, red area), winner-takes-all (WTA, grey) and winnerless competition (WLC, blue). Curves mark the boundaries between the different regimes (see Methods). Activity for marked parameter combinations shown in **(b)**. **(d)** Evolution of the similarity score with increasing network depth, for $m = 0.83$ and input ratio of 0.86. For deep networks, the Co-Ex region vanishes and the system converges to either WLC or WTA dynamics. **(e)** Schematic showing the influence of modularity and input intensity on the system's potential energy landscape (see Methods): (1) in the fading regime there is a single low-activity fixed point (minimum in the potential); (2) increasing modularity creates two high-activity fixed points associated with S1 and S2, with the dynamics always converging to the same minimum due to $\lambda_1 >> \lambda_2$; (3) strengthening S2 balances the initial conditions, resulting in frequent, fluctuation-driven switching between the two states. (4) for larger $m$ values, switching speed decreases as the wells become deeper and the barrier between the wells wider. **(f)** Switching frequency between the dominating sub-populations in $SSN_5$ decays with increasing modularity. Data computed over 10 sec, for $\lambda_2/\lambda_1 = 0.9$.

of winnerless competition (WLC) between multiple neuronal groups [42, 24]. The switching between the two states is driven by stochastic fluctuations (Fig. 8e, (3)). The depth of the wells and width of barrier (distance between fixed points) increase with modularity (see Fig. 8e, (4) and Supplementary Materials), suggesting a greater difficulty in moving between the two attractors and consequently fewer state changes. Numerical simulations confirm this slowdown in switching (Fig. 8f).

We wish to emphasize that the different dynamical states arise primarily from the feedforward connectivity profile. Nevertheless, even though the synaptic weights are not directly modified, varying the topographic modularity does translate to a modification of the effective connectivity weights (Fig. 3b). The ratio of stimulus intensities also plays a role in determining the dynamics, but there is a (narrow) range (approximately between 0.75 and 0.8) for which all 3 regions can be reached through sole modification of the modularity. Together, these results demonstrate that topography can not only lead to spatial denoising but also enable various, functionally important network operating points.

# Discussion

The presence of stimulus- or feature-tuned sub-populations of neurons in primary sensory cortices (as well as in downstream areas), provides an efficient spatial encoding strategy [43, 44, 45] that ensures the relevant computable features are accurately represented. Here, we propose that beyond primary sensory areas, modular topographic projections play a key role in preserving accurate representations of sensory inputs across many processing modules. Acting as a structural scaffold for a sequential denoising mechanism, we show how they simultaneously

enhance relevant stimulus features and remove noisy interference. We demonstrate this phenomenon in a variety of network models and provide a theoretical analysis that indicates its robustness and generality.

When reconstructing a spatially encoded input signal corrupted by noise in a network of sequentially connected populations, we find that a convergent structure in the feedforward projections is not only critical for successfully solving the task, but that the performance increases significantly with network depth beyond a certain modularity (Fig. 1). Through this mechanism, the response selectivity of the stimulated sub-populations is sharpened within each subsequent sub-network, while others are silenced (Fig. 2). Such wiring may support efficient and robust information transmission from the thalamus to deeper cortical centers, retaining faithful representations even in the presence of strong noise. More generally, topographic modularity, in conjunction with other top-down processes [46], could provide the anatomical substrate for the implementation of a number of behaviorally relevant processes. For example, feedforward topographic projections on the visual pathway could contribute, together with various attentional control processes, to the widely observed *pop-out effect* in the later stages of the visual hierarchy [47, 48]. The pop-out effect, at its core, assumes that in a given context some neurons exhibit sharper selectivity to their preferred stimulus feature than the neighboring regions, which can be achieved through a winner-take-all (WTA) mechanism (see Fig. 8 and [49]).

The WTA behavior underlying the denoising is caused by a re-shaping of the E/I balance across the network (see Fig. 3). As the excitatory feedforward projections become more focused, they modulate the system's effective connectivity and thereby the gain on the stimulus-specific pathways, gating or allowing (and even enhancing) signal propagation. This change renders the stimulated pathway excitatory in the active regime (see Fig. 7), leading to multiple fixed points such as those observed in networks with local recurrent excitation [50, 51]. While the high-activity fixed point of such clustered networks is reached over time, in our model it unfolds progressively in space, across multiple populations. Importantly, in the range of biologically plausible numbers of cortical areas relevant for signal transmission (up to ten for some visual stimuli, see [52, 53]) and intermediate modularity, the firing rates remain within experimentally observed limits and do not saturate. The basic principle and net result is similar to other approaches that alter the gain on specific pathways to facilitate stimulus propagation, for example through stronger synaptic weights [54], stronger nonlinearity [36], tuning of connectivity strength and neuronal thresholds [37], via detailed balance of local excitation and inhibition (amplitude gating [55]) or with additional sub-cortical structures [56]. Additionally, our model also displays some activity characteristics reported previously, such as the response sharpening observed for synfire chains [57] or (almost) linear firing rate propagation [58] (for intermediate modularity).

However, due to the reliance on increasing inhibitory activity at every stage, we speculate that denoising, as studied here, would not occur in such a system containing a single, shared inhibitory pool with homogeneous connectivity. In this case inhibition would affect all excitatory populations uniformly, with stronger activity potentially preventing accurate stimulus transmission from the initial sub-networks. Nevertheless, this problem could be alleviated using a more realistic, localized spatial connectivity profile as in [34], or by adding shadow pools (groups of inhibitory neurons) for each layer of the network, carefully wired in a recurrent or feedforward manner [59, 60, 55]. In such networks with non-random or spatially-dependent connectivity, structured (modular) topographic projections onto the inhibitory populations will likely be necessary to maintain stable dynamics and attain the appropriate inhibition-dominated regimes (Fig. 3). Alternatively, these could be achieved through additional, targeted inputs from other areas (Fig. 4), with feedforward inhibition known to provide a possible mechanism for context-dependent gating or selective enhancement of certain stimulus features [61, 31].

While our findings build on the above results, we here show that the experimentally observed topographic maps may serve as a structural denoising mechanism for sensory stimuli. In contrast to most works on signal propagation where noise mainly serves to stabilize the dynamics and is typically avoided in the input, here the system is driven by a continuous signal severely corrupted by noise. Taking a more functional approach, this input is reconstructed using linear combinations of the full network responses, rather than evaluating the correlation structure of the activity or relying on precise firing rates. Focusing on the modularity of such maps in recurrent spiking networks, our model also differs from previous studies exploring optimal connectivity profiles for minimizing information loss in purely feedforward networks [62, 63], also in the context of sequential denoising autoencoders [38], which used simplified neuron models, shallow networks or relied on specific, trained connection patterns (e.g., chosen by the pseudo-inverse model). It provides an explanation for how low-intensity stimuli (1-2 spks/sec above background activity, see Fig. 2 and Supplementary Materials) could be amplified across the cortex despite significant noise corruption, and relies on a generic principle that persists across different network models (Fig. 5) while also being robust to variations in the map size (Fig. 6). We demonstrated both the existence of a lower and upper (due to increased overlap) bound on their spatial extent for signal transmission, as well as an optimal region for which denoising was most pronounced.

These results indicate a trade-off between modularity and map size, with larger maps sustaining stimulus propagation at lower modularity values, whereas smaller maps must compensate through increased topographic density (see Fig. 6a and Supplementary Materials). In the case of smaller maps, progressively enlarging the receptive fields enhanced the denoising effect and improved task performance (Fig. 6c), suggesting a functional benefit for the anatomically observed decrease in topographic specificity with hierarchical depth [13, 35]. One advantage of such a wiring could be spatial efficiency in the initial stages of the sensory hierarchy due to anatomical constraints, for instance the retina or the lateral geniculate nucleus. While we get a good qualitative description of how the spatial variation of topographic maps influences the system's computational properties, the numerical values in general are not necessarily representative. Cortical maps are highly dynamic and exhibit more complex patterning, making (currently scarce) precise anatomical data a prerequisite for more detailed investigations. For instance, despite abundant information on the size of receptive fields [35, 64, 65], there is relatively little data on the connectivity between neurons tuned to related or different stimulus features across distinct cortical circuits. Should such experiments become feasible in the future, our model provides a testable prediction: the projections must be denser (or stronger) between smaller maps to allow robust communication whereas for larger maps fewer connections may be sufficient.

Finally, our model relates topographic connectivity to competition-based network dynamics. For two input signals of comparable intensities, moderately structured projections allow both representations to coexist in a decodable manner up to a certain network depth, whereas strongly modular connections elicit winnerless competition (WLC) like behavior characterized by stochastic switching between the two stimuli (see Fig. 8). Computation by switching is a functionally relevant principle [22, 66], which relies on fluctuation- or input-driven competition between different metastable (unstable) or stable attractor states. In the model studied here, modular topography induced multi-stability (uncertainty) in representations, alternating between two stable fixed points corresponding to the two input signals. Structured projections may thus partially explain the experimentally observed competition between multiple stimulus representations across the visual pathway [67], and is conceptually similar to an attractor-based model of perceptual bistability [68]. Moreover, this multi-stability across sub-networks can be "exploited" at any stage by control signals, i.e. additional modulation (inihibitory) could suppress one and amplify (bias) another.

Importantly, all these different dynamical regimes emerge progressively through the hierarchy and are not discernable in the initial modules. Previous studies reporting on similar dynamical states have usually considered either the synaptic weights as the main control parameter [42, 69, 54] or studied specific architectures with clustered connectivity [70, 71, 24]. Our findings suggest that in a hierarchical circuit a similar palette of behaviors can be also obtained given appropriate effective connectivity patterns modulated exclusively through modular topography. Although we used fixed projections throughout this study, these could also be learned and shaped continuously through various forms of synaptic plasticity (see e.g., [72]). To achieve such a variety of dynamics, cortical circuits most likely rely on a combination of all these mechanisms, i.e., pre-wired modular connections (within and between distant modules) and heterogeneous gain adaptation through plasticity, along with more complex processes such as targeted inhibitory gating.

Overall, our results highlight a novel functional role for topographically structured projection pathways in constructing reliable representations from noisy sensory signals, and accurately routing them across the cortical circuitry despite the plethora of noise sources along each processing stage.

# Methods

## Network architecture

We consider a feedforward network architecture where each sub-network (SSN) is a balanced random network [25] composed of $N = 10000$ homogeneous leaky integrate-and-fire neurons, grouped into a population of $N^{\mathrm{E}} = 0.8N$ excitatory and $N^{\mathrm{I}} = 0.2N$ inhibitory units. Within each sub-network, neurons are connected randomly and sparsely, with a fixed number of $K_{\mathrm{E}} = \epsilon N^{\mathrm{E}}$ local excitatory and $K_{\mathrm{I}} = \epsilon N^{\mathrm{I}}$ local inhibitory inputs per neuron. The sub-networks are arranged sequentially, i.e. the excitatory neurons $\mathrm{E_i}$ in $\mathrm{SSN_i}$ project to both $\mathrm{E_{i+1}}$ and $\mathrm{I_{i+1}}$ populations in the subsequent sub-network $\mathrm{SSN_{i+1}}$ (for an illustrative example, see Fig. 1a). There are no inhibitory feedforward projections. Although projections between sub-networks have a specific, non-uniform structure (see next section), each neuron in $\mathrm{SSN_{i+1}}$ receives the same total number of synapses from the previous SSN, $K_{\mathrm{FF}}$.

In addition, all neurons receive $K_{\mathrm{X}}$ inputs from an external source representing stochastic background noise. For the first sub-network, we set $K_{\mathrm{X}} = K_{\mathrm{E}}$, 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 (see e.g. [25, 73, 74]).

To ensure that the total excitatory input to each neuron is consistent across the network, we scale $K_{\mathrm{X}}$ by a factor of $\alpha = 0.25$ for the deeper SSNs and set $K_{\mathrm{FF}} = (1-\alpha)K_{\mathrm{E}}$, resulting in a ratio of 3:1 between the number of feedforward and background synapses.

## Modular feedforward projections

Within each SSN, each neuron is assigned to one or more of $N_{\mathrm{C}}$ sub-populations $SP$ associated with a specific stimulus ($N_{\mathrm{C}} = 10$ unless otherwise stated). This is illustrated in Fig. 1a for $N_{\mathrm{C}} = 2$. We choose these sub-populations so as to minimize their overlap within each $\mathrm{SSN_i}$, and control their effective size $C_{\mathrm{i}}^{\beta} = d_{\mathrm{i}} N^{\beta}, \beta \in [\mathrm{E}, \mathrm{I}]$, through the scaling parameter $d_{\mathrm{i}} \in [0, 1]$. Depending on the size and number of sub-populations, it is possible that some neurons are not part of any or that some neurons belong to multiple such sub-populations (overlap).

**Map size.** In what follows, a topographic map refers to the sequence of sub-populations in the different sub-networks associated with the same stimulus. To enable a flexible manipulation of the map sizes, we constrain the scaling factor $d_{\mathrm{i}}$ by introducing a step-wise linear increment $\delta$, such that $d_{\mathrm{i}} = d_0 + i\delta, i \geq 1$. Unless otherwise stated, we set $d_0 = 0.1$ and $\delta = 0$. Note that all SPs within a given SSN have the same size. In this study, we will only explore values in the range $0 \leq \delta \leq 0.02$ to ensure consistent map sizes across the system, i.e., $0 \leq d_{\mathrm{i}} \leq 1$ for all $\mathrm{SSN_i}$ (see constraints in Appendix A).

**Modularity.** To systematically modify the degree of modular segregation in the topographic projections, we define a modularity parameter that determines the relative probability for feedforward connections from a given SP in $\mathrm{SSN_i}$ to target the corresponding SP in $\mathrm{SSN_{i+1}}$. Specifically, we follow [75, 76] and define $m = 1 - \frac{p_0}{p_{\mathrm{c}}} \in [0, 1]$ as the ratio of the feedforward projection probabilities between neurons belonging to different SPs ($p_0$) and between neurons on the same topographic map ($p_{\mathrm{c}}$). According to the above definition, the feedforward connectivity matrix is random and homogeneous (Erdős-Rényi graph) if $m = 0$ or $d_{\mathrm{i}} = 1$ (see Fig. 1a). For $m = 1$ it is a block-diagonal matrix, where the individual SPs overlap only when $d_{\mathrm{i}} > 1/N_{\mathrm{C}}$. In order to isolate the effects on the network dynamics and computational performance attributable exclusively to the topographic structure, the overall density of the feedforward connectivity matrix is kept constant at $(1-\alpha) * \epsilon = 0.075$ (see also previous section). We note that, while providing the flexibility to implement the variations studied in this manuscript, this formalism has limitations (see Appendix A).

## Neuron and synapse model

We study networks composed of leaky integrate-and-fire (LIF) neurons with fixed voltage threshold and static synapses with exponentially decaying postsynaptic currents or conductances. The sub-threshold membrane potential dynamics of such a neuron evolves according to:

$$\tau_{\mathrm{m}} \frac{dV(t)}{dt} = (V_{\mathrm{rest}} - V(t)) + R\left(I^{\mathrm{E}}(t) + I^{\mathrm{I}}(t) + I^{\mathrm{X}}(t)\right) \tag{1}$$

where $\tau_{\mathrm{m}}$ is the membrane time constant, and $RI^{\beta}$ is the total synaptic input from population $\beta \in [E, I]$. The background input $I^{\mathrm{X}}$ is assumed to be excitatory and stochastic, modeled as a homogeneous Poisson process with constant rate $\nu_{\mathrm{X}}$. Synaptic weights $J_{\mathrm{ij}}$, representing the efficacy of interaction from presynaptic neuron $j$ to postsynaptic neuron $i$, are equal for all realized connections of a given type, i.e., $J_{\mathrm{EE}} = J_{\mathrm{IE}} = J$ for excitatory and $J_{\mathrm{EI}} = J_{\mathrm{II}} = -gJ$ for inhibitory synapses. All synaptic delays and time constants are equal in this setup. For a complete, tabular description of the models and model parameters used throughout this study, see Supplementary Tables S1-S5.

Following previous works [15, 74], we choose the intensity of the stochastic input $\nu_{\mathrm{X}}$ and the E-I ratio $g$ such that the first two sub-networks operate in a balanced, asynchronous irregular regime when driven solely by background input. This is achieved with $\nu_{\mathrm{X}} = 12$ spikes/s and $g = 12$, resulting in average firing rates of $\sim 3$ spikes/s, coefficient of variation ($CV_{\mathrm{ISI}}$) in the interval $[1.0, 1.5]$ and Pearson cross-correlation (CC) $\leq 0.01$ in $\mathrm{SSN_0}$ and $\mathrm{SSN_1}$.

In Section "A generalizable structural effect" we consider two additional systems, a network of LIF neurons with conductance-based synapses and a continuous firing rate model. The LIF network is described in detail in [15]. Spike-triggered synaptic conductances are modeled as exponential functions, with fixed and equal conduction delays for all synapses. Key differences to the current-based model include, in addition to the biologically more plausible synapse model, longer synaptic time constants and stronger input (see also [15] and Table S3 for the numerical values of all parameters).

The continuous rate model contains $N = 3000$ nonlinear units, the dynamics of which are governed by:

$$\begin{aligned} \tau_x \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{dt}} &= -\boldsymbol{x} + W\boldsymbol{r} + W^{\mathrm{in}}\boldsymbol{u} - \boldsymbol{b}^{\mathrm{rec}} + \sqrt{2\tau_{\boldsymbol{x}}}\sigma_{\mathrm{X}}\boldsymbol{\xi} \\ \boldsymbol{r} &= 0.5(1 + \tanh(\boldsymbol{x})) \end{aligned} \tag{2}$$

where $\boldsymbol{x}$ represents the activation and $\boldsymbol{r}$ the output of all units, commonly interpreted as the synaptic current variable and the firing rate estimate, respectively. The rates $r_{\mathrm{i}}$ are obtained by applying the nonlinear transfer function $\tanh(x_{\mathrm{i}})$, modified here to constrain the rates to the interval $[0, 1]$. $\tau_x = 10$ms is the neuronal time constant, $\boldsymbol{b}^{\mathrm{rec}}$ is a vector of individual neuronal bias terms (i.e. a baseline activation), and $W$ and $W^{\mathrm{in}}$ are the recurrent (including feedforward) and input weight matrices, respectively. These are constructed in the same manner as for the spiking networks, such that the overall connectivity, including the input mapping onto $\mathrm{SSN}_0$, is identical for all three models. Input weights are drawn from a uniform distribution, while the rest follow a normal distribution. Finally, $\boldsymbol{\xi}$ is a vector of $N$ independent realizations of Gaussian white noise with zero mean and variance scaled by $\sigma_{\mathrm{X}}$. The differential equations are integrated numerically, using the Euler–Maruyama method with step $\delta t = 1$ ms, with specific parameter values given in Table S5.

## Signal reconstruction task

We evaluate the system's ability to recover a simple, continuous step signal from a noisy variation using linear combinations of the population responses in the different SSNs [77]. This is equivalent to probing the network's ability to function as a denoising autoencoder [78].

To generate the $N_{\mathrm{C}}$-dimensional input signal $u(t)$, we randomly draw stimuli from a predefined set $S = \{S_1, S_2, ..., S_{N_{\mathrm{C}}}\}$ and set the corresponding channel to active for a fixed duration of 200 ms (Fig. 1a, left). This binary step signal $u(t)$ is also the target signal to be reconstructed. The effective input is obtained by scaling $u(t)$ with the input rate $\nu_{\mathrm{in}}$, and adding a Gaussian white noise process with zero mean and variance $\sigma_\xi^2$. Rectifying the resulting signal leads to the final form of the continuous input signal $z(t) = (\nu_{\mathrm{in}}u(t) + \xi(t))_+$. This allows us to control the amount of noise in the input, and thus the task difficulty, through a single parameter $\sigma_\xi$.

To deliver the input to the circuit, the analog signal $z(t)$ is converted into spike trains, with its amplitude serving as the rate of an inhomogeneous Poisson process generating independent spike trains. We set the scaling amplitude to $\nu_{\mathrm{in}} = K_E\lambda\nu_{\mathrm{X}}$, modelling stochastic input with fixed rate $\lambda\nu_{\mathrm{X}}$ from $K_E = 800$ neurons. If not otherwise specified, $\lambda = 0.05$ holds, resulting in a mean firing rate below 8 spks/sec in $\mathrm{SSN}_0$ (see Fig. 2c).

Each input channel $k$ is mapped onto one of the $N_{\mathrm{C}}$ stimulus-specific sub-populations of excitatory and inhibitory neurons in the first (input) sub-network $\mathrm{SSN}_0$, chosen according to the procedure described above (see also Fig. 1a). This way, each stimulus $S_{\mathrm{k}}$ is mapped onto a specific set of sub-populations in the different sub-networks, i.e., the topographic map associated with $S_{\mathrm{k}}$.

For each stimulus in the sequence, we sample the responses of the excitatory population in each $\mathrm{SSN_i}$ at fixed time points (once every ms) relative to stimulus onset. We record from the membrane potentials $V_m$ as they represent a parameter-free and direct measure of the population state [79, 80]. The activity vectors are then gathered in a state matrix $X_{\mathrm{SSN}_i} \in \mathbb{R}^{N^{\mathrm{E}} \times T}$, which is then used to train a linear readout to approximate the target output of the task [81]. We divide the input data, containing a total of 100 stimulus presentations (yielding $T = 20000$ samples), into a training and a testing set (80/20 %), and perform the training using ridge regression (L2 regularization), with the regularization parameter chosen by leave-one-out cross-validation on the training dataset.

Reconstruction performance is measured using the normalized root mean squared error (NRMSE). For this particular task, the effective delay in the build-up of optimal stimulus representations varies greatly across the sub-networks. In order to close in on the optimal delay for each $\mathrm{SSN_i}$, we train the state matrix $X_{\mathrm{SSN_i}}$ on a larger interval of delays and choose the one that minimizes the error, averaged across multiple trials.

## Effective connectivity and stability analysis

To better understand the role of structural variations on the network's dynamics, we determine the network's effective connectivity matrix $W$ analytically by linear stability analysis around the system's stationary working points (see Appendix B for the complete derivations). The elements $w_{ij} \in W$ represent the integrated linear response of a target neuron $i$, with stationary rate $\nu_{\mathrm{i}}$, to a small perturbation in the input rate $\nu_{\mathrm{j}}$ caused by a spike from presynaptic neuron $j$. In other words, $w_{ij}$ measures the average number of additional spikes emitted by a target neuron $i$ in response to a spike from the presynaptic neuron $j$, and its relation to the synaptic weights is defined by [28, 40]:

$$
\begin{aligned}
w_{ij} &= \frac{\partial \nu_i}{\partial \nu_j} = \widetilde{\alpha} J_{ij} + \widetilde{\beta} J_{ij}^2 \\
\text{with} \quad \widetilde{\alpha} &= \sqrt{\pi}\,(\tau_{\mathrm{m}}\nu_i)^2\,\frac{1}{\sigma_i}\,(f(y_\theta) - f(y_{\mathrm{r}})) \\
\text{and } \tilde{\beta} &= \sqrt{\pi}\,(\tau_{\mathrm{m}}\nu_i)^2\,\frac{1}{2\sigma_i^2}\,(f(y_\theta)y_\theta - f(y_{\mathrm{r}})y_{\mathrm{r}})\,.
\end{aligned}
\tag{3}
$$

Note that in Fig. 3 we ignore the contribution $\tilde{\beta}$ resulting from the modulation in the input variance $\sigma_{\mathrm{j}}^2$ which is significantly smaller due to the additional factor $1/\sigma_i \sim \mathcal{O}(1/\sqrt{N})$. Importantly, the effective connectivity matrix $W$ allows us to gain insights into the stability of the system by eigenvalue decomposition. For large random coupling matrices, the effective weight matrix has a spectral radius $\rho = \max_k\,(\mathrm{Re}\{\lambda_{\mathrm{k}}\})$ which is determined by the variances of $W$ [29]. For inhibition-dominated systems, such as those we consider, there is a single negative outlier representing the mean effective weight, given the eigenvalue $\lambda_{\mathrm{k}}^*$ associated with the unit vector. The stability of the system is thus uniquely determined by the spectral radius $\rho$: values smaller than unity indicate stable dynamics, whereas $\rho > 1$ lead to unstable linearized dynamics.

## Fixed point analysis

For the mean-field analysis, the $N_{\mathrm{C}}$ sub-populations in each sub-network can be reduced to only two groups of neurons, the first one comprising all neurons of the stimulated SPs and the second one comprising all neurons in all non-stimulated SPs. This is possible because 1) the firing rates of the excitatory and inhibitory neurons within one SP are identical, owing to homogeneous neuron parameters and matching incoming connection statistics, and 2) all neurons in non-stimulated SPs have the same rate $\nu^{\mathrm{NS}}$ that is in general different from the rate of the stimulated SP $\nu^{\mathrm{S}}$. Here we only sketch the main steps, with a detailed derivation given in Appendix B.

The mean inputs to the first sub-network can be obtained via

$$
\begin{aligned}
\mu^{\mathrm{S}} &= (1+\lambda)\mathcal{J}\nu_{\mathrm{x}} + \frac{1}{N_{\mathrm{C}}}\mathcal{J}\,(1+\gamma g)\,\nu^{\mathrm{S}} + \frac{N_{\mathrm{C}}-1}{N_{\mathrm{C}}}\mathcal{J}\,(1+\gamma g)\,\nu^{\mathrm{NS}}\,, \\
\mu^{\mathrm{NS}} &= \qquad\quad \mathcal{J}\nu_{\mathrm{x}} + \frac{1}{N_{\mathrm{C}}}\mathcal{J}\,(1+\gamma g)\,\nu^{\mathrm{S}} + \frac{N_{\mathrm{C}}-1}{N_{\mathrm{C}}}\mathcal{J}\,(1+\gamma g)\,\nu^{\mathrm{NS}}
\end{aligned}
\tag{4}
$$

where $\gamma = K_{\mathrm{I}}/K_{\mathrm{E}}$ and $\mathcal{J} = \tau K_{\mathrm{E}} J$. Both equations are of the form

$$
\kappa\nu = \mu - I \tag{5}
$$

where $\kappa$ is the effective self-coupling of a group of neurons with rate $\nu$ and input $\mu$, and $I$ denotes the external inputs from other groups. Equation (5) describes a linear relationship between the rate $\nu$ and the input $\mu$. To find a self-consistent solution for the rates $\nu^{\mathrm{S}}$ and $\nu^{\mathrm{NS}}$, the above equations need to be solved numerically, taking into account in addition the f-I curve $\nu(\mu)$ of the neurons that in the case of leaky integrate-and-fire model neurons also depends on the variance $\sigma^2$ of inputs. The latter can be obtained analogous to the mean input $\mu$ (see Appendix B).

Starting from $\mathrm{SSN}_1$, networks are connected in a fixed pattern such that the rate $\nu_i$ in $\mathrm{SSN}_i$ also depends on the excitatory input from the previous sub-network $\mathrm{SSN}_{i-1}$ with rate $\nu_{i-1}$. For a fixed point, we have $\nu_i = \nu_{i-1}$ [36]. In this case, we can effectively group together stimulated/non-stimulated neurons in successive sub-networks and re-group equations for the mean input in the limit of many sub-networks, obtaining the simplified description (details see Appendix B)

$$
\mu^{\mathrm{S}} = \alpha\mathcal{J}\nu_{\mathrm{x}} + \kappa_{\mathrm{S,S}}\;\nu^{\mathrm{S}} + \kappa_{\mathrm{S,NS}}\;\nu^{\mathrm{NS}} \tag{6}
$$

$$
\mu^{\mathrm{NS}} = \alpha\mathcal{J}\nu_{\mathrm{x}} + \kappa_{\mathrm{NS,S}}\;\nu^{\mathrm{S}} + \kappa_{\mathrm{NS,NS}}\;\nu^{\mathrm{NS}} \tag{7}
$$

The scaling terms of the firing rates incorporate the recurrent and feedforward contributions from the stimulated and non-stimulated groups of neurons. They depend solely on some fixed parameters of the system, including modularity $m$ (see Appendix B). Importantly, Eq. 6 and Eq. 7 have the same linear form as Eq. 5 and can be solved numerically as described above. The numerical solutions for fixed points are obtained using the root finding algorithm `root` of the `scipy.optimize` package [82]. The structure of fixed points for the stimulated sub-population (see discussion in "Modularity as a bifurcation parameter") can furthermore be intuitively understood by studying the potential landscape of the system. The potential $U$ is thereby defined via the conservative force $F = -\frac{dU}{d\nu^{\mathrm{S}}} = -\nu^{\mathrm{S}} + \nu(\mu, \sigma^2)$ that drives the system towards its fixed points via the equation of motion $\frac{d\nu^{\mathrm{S}}}{dt} = F$ [83, 51, 84]. Note that $\mu$ and $\sigma^2$ are again functions of $\nu^{\mathrm{S}}$ and $\nu^{\mathrm{NS}}$, where the latter is the self-consistent rate of the non-stimulated sub-populations for given rate $\nu^S$ of the stimulated sub-population, $\nu^{\mathrm{NS}} = \nu^{\mathrm{NS}}(\nu^{\mathrm{S}})$ (details see Appendix B).

### Multiple inputs and correlation-based similarity score

In Fig. 8 we consider two stimuli $S_1$ and $S_2$ to be active simultaneously for 10 s. Let $SP_1$ and $SP_2$ be the two corresponding SPs in each sub-network. The firing rate of each SP is estimated from spike counts in time bins of 10 ms and smoothed with a Savitzky-Golay filter (length 21 and polynomial order 4). We compute a similarity score based on the correlation between these rates, scaled by the ratio of the input intensities $\lambda_2/\lambda_1$ (with $\lambda_1$ fixed). This scaling is meant to introduce a gradient in the similarity score based on the firing rate differences, ensuring that high (absolute) scores require comparable activity levels in addition to strong correlations. To ensure that both stimuli are decodable where appropriate, we set the score to 0 when the difference between the rate of $SP_2$ and the non-stimulated SPs was $<$ 1 spks/sec ($SP_1$ had significantly higher rates). The curves in Fig. 8c mark the regime boundaries: coexisting (Co-Ex) where score is $>$ 0.1 (red curve); winnerless competition (WLC) where score is $<$ -0.1 (blue); winner-takes-all (WTA, grey) and where the score is in the interval (-0.1, 0.1), and either $\lambda_2/\lambda_1 < 0.5$ holds or the score is 0. While the Co-Ex region is a dynamical regime that only occurs in the initial sub-networks (Fig. 8d), the WTA and WLC regimes persist and can be understood again with the help of a potential $U$, which is in this case a function of the rates of the two SPs (details see Appendix B).

### Numerical simulations and analysis

All numerical simulations were conducted using the Neural Microcircuit Simulation and Analysis Toolkit (NMSAT) v0.2 [85], 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 [86], 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 [87], we provide a complete code package that implements project-specific functionality within NMSAT (see Supplementary Materials) using NEST 2.18.0 [88].

## Data availability

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

## Author contributions

BZ, AM, and RD designed the study. BZ, DD and RD performed the analysis. BZ performed the simulations and implemented the models. BZ and DD derived and implemented the theory. BZ, DD, AM, and RD contributed to writing of manuscript.


## Funding

This work has received partial support from the the Initiative and Networking Fund of the Helmholtz Association, the Helmholtz Portfolio theme Supercomputing and Modeling for the Human Brain, and the Excellence Initiative of the German federal and state governments (G:(DE-82)EXS-SF-neuroIC002).


## Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.


## Acknowledgements

The authors gratefully acknowledge the computing time granted by the JARA-HPC Vergabegremium on the supercomputer JURECA at Forschungszentrum Jülich.

# References


[1] A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. Noise in the nervous system. *Nat. Rev. Neurosci.*, 9(4):292–303, 2008.

[2] Alfonso Renart and Christian K Machens. Variability in neural activity and behavior. *Current Opinion in Neurobiology*, 25:211–220, 2014. Theoretical and computational neuroscience.

[3] Karl Friston. A theory of cortical responses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(1456):815–836, 2005.

[4] Kayoko Okada, Feng Rong, Jon Venezia, William Matchin, I. Hui Hsieh, Kourosh Saberi, John T. Serences, and Gregory Hickok. Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, 20(10):2486–2495, 2010.

[5] James J. DiCarlo, Davide Zoccolan, and Nicole C. Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415–434, 2012.

[6] Wiktor F Młynarski and Ann M Hermundstad. Adaptive coding for dynamic sensory inference. *eLife*, 7:e32055, jul 2018.

[7] Thomas Parr, Andrew W Corcoran, Karl J Friston, and Jakob Hohwy. Perceptual awareness and active inference. *Neuroscience of Consciousness*, 2019(1), 09 2019. niz012.

[8] David Meunier, Renaud Lambiotte, and Edward Bullmore. Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience*, 4:200, 2010.

[9] Hae-Jeong Park and Karl Friston. Structural and functional brain networks: From connections to cognition. *Science*, 342(6158), 2013.

[10] S. Murray Sherman and R. W. Guillery. The role of the thalamus in the flow of information to the cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1428):1695–1708, Dec. 2002.

[11] Miho Nakajima and Michael M Halassa. Thalamic control of functional cortical connectivity. *Current Opinion in Neurobiology*, 44:127–131, 2017. Neurobiology of Sleep.

[12] Jon H Kaas. Topographic maps are fundamental to sensory processing. *Brain Research Bulletin*, 44(2):107 – 112, 1997.

[13] James A. Bednar and Stuart P. Wilson. Cortical maps. *The Neuroscientist*, 22(6):604–617, 2016. PMID: 26290447.

[14] Brian A. Wandell and Jonathan Winawer. Imaging retinotopic maps in the human brain. *Vision Research*, 51(7):718 – 737, 2011. Vision Research 50th Anniversary Issue: Part 1.

[15] Barna Zajzon, Sepehr Mahmoudian, Abigail Morrison, and Renato Duarte. Passing the message: Representation transfer in modular balanced networks. *Frontiers in Computational Neuroscience*, 13:79, 2019.

[16] Donald J. Hagler and Martin I. Sereno. Spatial maps in frontal and prefrontal cortex. *NeuroImage*, 29(2):567–577, jan 2006.

[17] Michael A Silver and Sabine Kastner. Topographic maps in human frontal and parietal cortex. *Trends in cognitive sciences*, 13(11):488–495, 2009.

[18] Gaurav H Patel, David M Kaplan, and Lawrence H Snyder. Topographic organization in the brain: searching for general principles. *Trends in cognitive sciences*, 18(7):351–363, 2014.

[19] Søren K. Andersen, Steven A. Hillyard, and Matthias M. Müller. Attention facilitates multiple stimulus features in parallel in human visual cortex. *Current Biology*, 18(13):1006–1009, Jul. 2008.

[20] R. J. Douglas and K. A. C. Martin. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, 27:419–451, 2004.

[21] M Carandini and Dj Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(November):1–12, 2012.

[22] David A McCormick. Neuronal networks: Flip-flops in the brain. *Current Biology*, 15(8):R294–R296, 2005.

[23] Mikhail I. Rabinovich, Ramón Huerta, Pablo Varona, and Valentin S. Afraimovich. Transient cognitive dynamics, metastability, and decision making. *PLoS Computational Biology*, 4(5):e1000072, may 2008.

[24] Thomas Rost, Moritz Deger, and Martin P Nawrot. Winnerless competition in clustered balanced networks: inhibitory assemblies do the trick. *Biological cybernetics*, 112(1):81–98, 2018.

[25] Nicolas Brunel. Dynamics of networks of randomly connected excitatory and inhibitory spiking neurons. *Journal of Physiology Paris*, 94:445–463, 2000.

[26] Bilal Haider, Alvaro Duque, Andrea R. Hasenstaub, and David A. McCormick. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *Journal of Neuroscience*, 26(17):4535–4545, 2006.

[27] Michael N. Shadlen and William T. Newsome. Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, 4(4):569–579, 1994.

[28] Tom Tetzlaff, Moritz Helias, Gaute T. Einevoll, and Markus Diesmann. Decorrelation of neural-network activity by inhibitory feedback. *PLOS Computational Biology*, 8(8):1–29, 08 2012.

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

[30] V. B. MOUNTCASTLE and T. P. POWELL. Neural mechanisms subserving cutaneous sensibility, with special reference to the role of afferent inhibition in sensory perception and discrimination. *Bull Johns Hopkins Hosp*, 105:201–232, Oct 1959.

[31] Michael T. Roberts, Stephanie C. Seeman, and Nace L. Golding. A mechanistic understanding of the role of feedforward inhibition in the mammalian sound localization circuitry. *Neuron*, 78(5):923–935, 2013.

[32] Michael N. Shadlen and William T. Newsome. The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10):3870–3896, May 1998.

[33] Tom Tetzlaff, Michael Buschermöhle, Theo Geisel, and Markus Diesmann. The spread of rate and correlation in stationary cortical networks. *Neurocomputing*, 52–54:949–954, 2003.

[34] A. Kumar, S. Rotter, and A. Aertsen. Conditions for propagating synchronous spiking and asynchronous firing rates in a cortical network model. *Journal of Neuroscience*, 28(20):5268–5280, 2008.

[35] A.T. Smith, K.D. Singh, A.L. Williams, and M.W. Greenlee. Estimating Receptive Field Size from fMRI Data in Human Striate and Extrastriate Visual Cortex. *Cerebral Cortex*, 11(12):1182–1190, 12 2001.

[36] Taro Toyoizumi. Nearly extensive sequential memory lifetime achieved by coupled nonlinear neurons. *Neural Computation*, 24(10):2678–2699, Oct. 2012.

[37] Natasha Cayco Gajic and Eric Shea-Brown. Neutral stability, rate propagation, and critical branching in feedforward networks, 2012.

[38] Jonathan Kadmon and Haim Sompolinsky. Optimal architectures in a solvable model of deep networks. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

[39] Nicolas Fourcaud and Nicolas Brunel. Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural computation*, 14(9):2057–2110, 2002.

[40] M Helias, T Tetzlaff, and M Diesmann. Echoes in correlated neural systems. *New Journal of Physics*, 15(2):023002, feb 2013.

[41] Jannis Schuecker, Markus Diesmann, and Moritz Helias. Modulated escape from a metastable state driven by colored noise. *Phys. Rev. E*, 92:052119, Nov 2015.

[42] Fereshteh Lagzi and Stefan Rotter. Dynamics of competition between subnetworks of spiking neuronal networks in the balanced state. *PLOS ONE*, 10(9):1–28, 09 2015.

[43] Alexandre Pouget, Sophie Deneve, Jean-Christophe Ducom, and Peter E. Latham. Narrow versus wide tuning curves: What's best for a population code? *Neural Computation*, 11(1):85–90, 1999.

[44] Peggy Seriès, Peter E Latham, and Alexandre Pouget. Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nature Neuroscience*, 7(10):1129–1135, Sep. 2004.

[45] Gašper Tkačik, Jason S. Prentice, Vijay Balasubramanian, and Elad Schneidman. Optimal population coding by noisy spiking neurons. *Proceedings of the National Academy of Sciences*, 107(32):14419–14424, 2010.

[46] Peter Kok, Janneke F.M. Jehee, and Floris P. de Lange. Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, 75(2):265 – 270, 2012.

[47] Julie A. Brefczynski-Lewis, Ritobrato Datta, James W. Lewis, and Edgar A. DeYoe. The topography of visuospatial attention as revealed by a novel visual field mapping technique. *Journal of Cognitive Neuroscience*, 21(7):1447–1460, 2009. PMID: 18752412.

[48] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.

[49] Kevin D. Himberger, Hsiang-Yun Chien, and Christopher J. Honey. Principles of temporal processing across the cortical hierarchy. *Neuroscience*, 389:161 – 174, 2018. Sensory Sequence Processing in the Brain.

[50] Alfonso Renart, Rubén Moreno-Bote, Xiao-Jing Wang, and Néstor Parga. Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural Computation*, 19(1):1–46, Jan. 2007.

[51] Ashok Litwin-Kumar and Brent Doiron. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature Neuroscience*, 15(11):1498–1505, Sep. 2012.

[52] D J Felleman and D C Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, N.Y. : 1991)*, 1(1):1–47, 1991.

[53] Jay Hegdé and Daniel J. Felleman. Reappraising the functional implications of the primate visual anatomical hierarchy. *The Neuroscientist*, 13(5):416–421, Oct. 2007.

[54] Tim P. Vogels and L. F. Abbott. Signal propagation and logic gating in networks of integrate-and-fire neurons. *Journal of Neuroscience*, 25(46):10786–10795, 2005.

[55] Tim P Vogels and L F Abbott. Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nat. Neurosci.*, 12(4):483–491, 2009.

[56] Nelson Cortes and Carl van Vreeswijk. Pulvinar thalamic nucleus allows for asynchronous spike propagation through the cortex. *Frontiers in Computational Neuroscience*, 9, May 2015.

[57] Markus Diesmann, Marc-Oliver Gewaltig, and Ad Aertsen. Stable propagation of synchronous spiking in cortical neural networks. *Nature*, 402(6761):529–533, 1999.

[58] Arvind Kumar, Stefan Rotter, and Ad Aertsen. Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nat Rev Neurosci*, 11:615–627, 2010.

[59] Yuval Aviel, Carsten Mehring, Moshe Abeles, and David Horn. On embedding synfire chains in a balanced network. *Neural Comput.*, 15(6):1321–1340, 2003.

[60] Yuval Aviel, David Horn, and Moshe Abeles. Memory capacity of balanced networks. *Neural Comput.*, 17(3):691–713, Mar 2005. Comparative Study.

[61] Michele Ferrante, Michele Migliore, and Giorgio A. Ascoli. Feed-forward inhibition as a buffer of the neuronal input-output relation. *Proceedings of the National Academy of Sciences*, 106(42):18004–18009, 2009.

[62] Alfonso Renart and Mark C. W. van Rossum. Transmission of Population-Coded Information. *Neural Computation*, 24(2):391–407, 02 2012.

[63] Joel Zylberberg, Alexandre Pouget, Peter E. Latham, and Eric Shea-Brown. Robust information propagation through noisy neural circuits. *PLOS Computational Biology*, 13(4):1–35, 04 2017.

[64] Lu Liu, Liang She, Ming Chen, Tianyi Liu, Haidong D. Lu, Yang Dan, and Mu-ming Poo. Spatial structure of neuronal receptive field in awake monkey secondary visual cortex (v2). *Proceedings of the National Academy of Sciences*, 113(7):1913–1918, 2016.

[65] Georgios A. Keliris, Qinglin Li, Amalia Papanikolaou, Nikos K. Logothetis, and Stelios M. Smirnakis. Estimating average single-neuron visual receptive field sizes by fmri. *Proceedings of the National Academy of Sciences*, 116(13):6425–6434, 2019.

[66] Fabio Schittler Neves and Marc Timme. Computation by switching in complex networks of states. *Phys. Rev. Lett.*, 109:018701, Jul 2012.

[67] Kang Li, Vladislav Kozyrev, Søren Kyllingsbæk, Stefan Treue, Susanne Ditlevsen, and Claus Bundesen. Neurons in primate visual cortex alternate between responses to multiple stimuli in their receptive field. *Frontiers in Computational Neuroscience*, 10:141, 2016.

[68] Rubén Moreno-Bote, John Rinzel, and Nava Rubin. Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of Neurophysiology*, 98(3):1125–1139, 2007. PMID: 17615138.

[69] Fereshteh Lagzi, Fatihcan M Atay, and Stefan Rotter. Bifurcation analysis of the dynamics of interacting subnetworks of a spiking network. *Scientific reports*, 9(1):1–17, 2019.

[70] Michael T. Schaub*, Yazan Billeh*, Costas A. Anastassiou, Christof Koch, and Mauricio Barahona. Emergence of slow-switching assemblies in structured neuronal networks. *PLOS Comput. Biol.*, 11(7):e1004196, 2015.

[71] Ashok Litwin-Kumar and Brent Doiron. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.*, 15(11):1498–1505, Sep. 2012.

[72] Rosario Tomasello, Max Garagnani, Thomas Wennekers, and Friedemann Pulvermüller. A neurobiologically constrained cortex model of semantic grounding with spiking neurons and brain-like connectivity. *Frontiers in Computational Neuroscience*, 12, Nov. 2018.

[73] A. Kumar, S. Schrader, A. Aertsen, and S. Rotter. The high-conductance state of cortical networks. *Neural Comput.*, 20(1):1–43, 2008.

[74] Renato C. F. Duarte and Abigail Morrison. Dynamic stability of sequential stimulus representations in adapting neuronal networks. *Frontiers in Computational Neuroscience*, 8(124), 2014.

[75] M. E. J. Newman. Random graphs with clustering. *Physical Review Letters*, 103(5), Jul. 2009.

[76] Neeraj Pradhan, Subinay Dasgupta, and Sitabhra Sinha. Modular organization enhances the robustness of attractor network dynamics. *EPL (Europhysics Letters)*, 94(3):38004, Apr. 2011.

[77] Wolfgang Maass, Thomas Natschläger, and Henry Markram. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, 2002.

[78] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

[79] R. Duarte, M. Uhlmann, D. den van Broek, H. Fitz, K. M. Petersson, and A. Morrison. Encoding symbolic sequences with spiking neural reservoirs. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, July 2018.

[80] Dick van den Broek, Marvin Uhlmann, Hartmut Fitz, Renato Duarte, Peter Hagoort, and Karl Magnus Petersson. The best spike filter kernel is a neuron. 2017.

[81] Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, aug 2009.

[82] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[83] K.-F. Wong. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4):1314–1328, Jan. 2006.

[84] Jannis Schuecker, Maximilian Schmidt, Sacha J. van Albada, Markus Diesmann, and Moritz Helias. Fundamental activity constraints lead to specific interpretations of the connectome. *PLOS Computational Biology*, 13(2):e1005179, Feb. 2017.

[85] Renato Duarte, Barna Zajzon, and Abigail Morrison. Neural Microcircuit Simulation And Analysis Toolkit. 2017.

[86] Marc-Oliver Gewaltig and Markus Diesmann. Nest (neural simulation tool). *Scholarpedia*, 2(4):1430, 2007.

[87] Robin Pauli, Philipp Weidel, Susanne Kunkel, and Abigail Morrison. Reproducing polychronization: A guide to maximizing the reproducibility of spiking network models. *Frontiers in Neuroinformatics*, 12:46, 2018.

[88] Jakob Jordan, Håkon Mørk, Stine Brekke Vennemo, Dennis Terhorst, Alexander Peyser, Tammo Ippen, Rajalekshmi Deepu, Jochen Martin Eppler, Alexander van Meegen, Susanne Kunkel, Ankur Sinha, Tanguy Fardet, Sandra Diaz, Abigail Morrison, Wolfram Schenck, David Dahmen, Jari Pronold, Jonas Stapmanns, Guido Trensch, Sebastian Spreizer, Jessica Mitchell, Steffen Graber, Johanna Senk, Charl Linssen, Jan Hahne, Alexey Serenko, Daniel Naoumenko, Eric Thomson, Itaru Kitayama, Sebastian Berns, and Hans Ekkehard Plesser. Nest 2.18.0, Jun. 2019.