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 demonstrate that in biologically constrained networks, such a denoising behavior is contingent on recurrent inhibition. 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, with the qualitative behavior persisting across very different models. Theoretical considerations and network simulations show that it hinges solely on the modularity of topographic projections and the presence of recurrent inhibition, with the external input and single-neuron properties influencing where/when, but not if, denoising occurs. 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

[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.

areas, can dramatically skew and modulate the circuit's E/I balance and local functional connectivity, resulting in fundamental differences in the systems' responsiveness.

Results

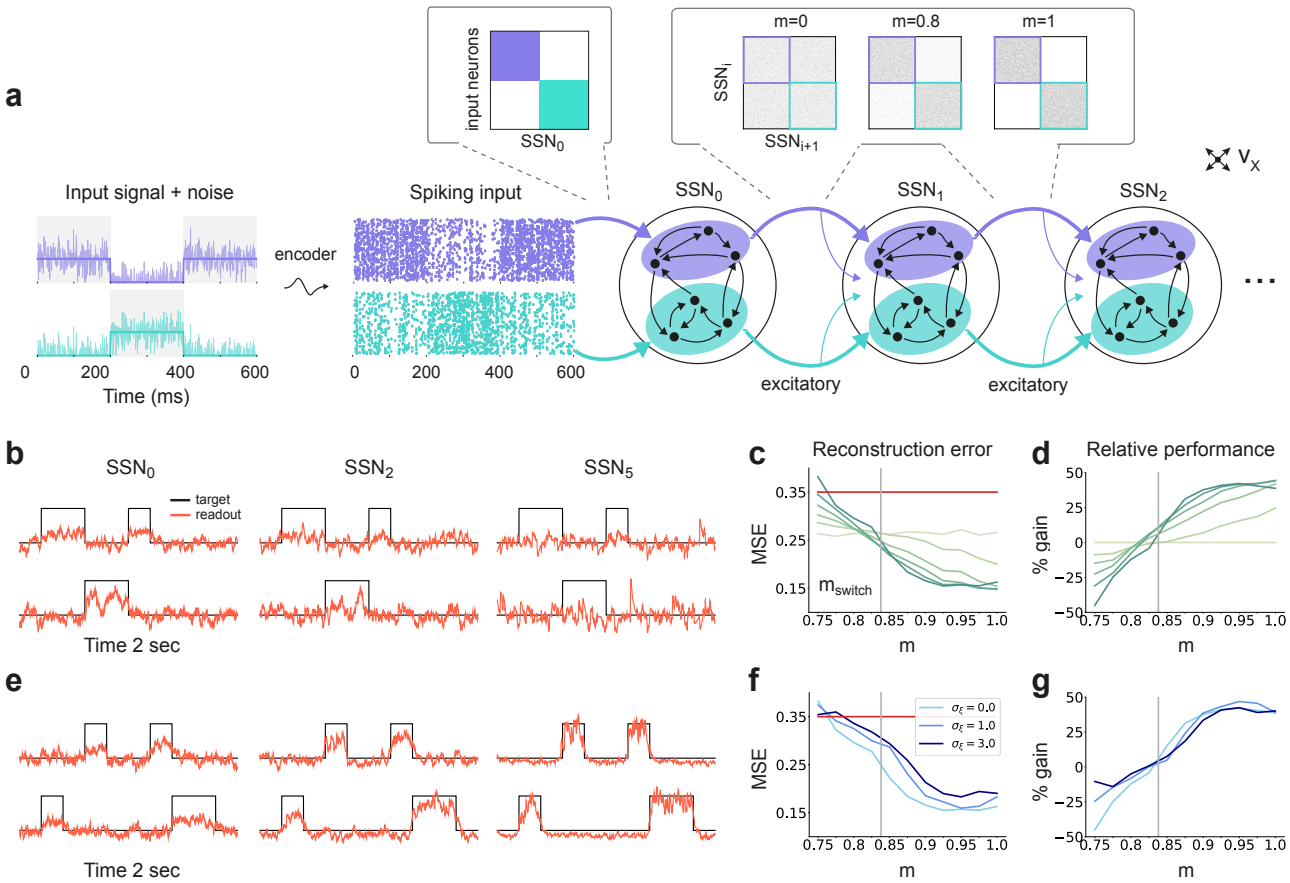

Figure 1: Sequential denoising spiking architecture. **(a)** A continuous step signal is used to drive the network. The input is spatially encoded in the first sub-network (SSN_0), whereby each input channel is mapped exclusively onto a sub-population of stimulus-specific excitatory and inhibitory neurons (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 SSN_0 (lightest) to SSN_5 (darkest). The horizontal red line represents chance level, while the grey vertical line marks the transition (switching) point $m_{\text{switch}} \approx 0.83$ (see main text). Figure 1 - figure supplement 1 shows the task performance for a broader range of parameters. **(d)** Performance gain across the network, relative to SSN_0, for the setup illustrated in **(b)**. **(e)** as in **(b)** but for $m = 0.9$. **(f)** Reconstruction error in SSN_5 for the different noise intensities. Horizontal and vertical dashed lines as in **(c)**. **(g)** Performance gain in SSN_5, relative to 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, SSN_5, even in the absence of noise (see Fig. 1b-c). However, as m approaches a switching value $m_{\mathrm{switch}} \approx 0.83$, 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 transition point, reconstruction accuracy improves with depth, i.e. the signal is more accurately represented in SSN_5 than in the initial sub-network, 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 Figure 1 - figure supplement 1), the denoising effect persists even if the input is severely corrupted ($\sigma_\xi = 3$, 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

The sequential denoising effect observed beyond the transition point $m_{\mathrm{switch}} \approx 0.83$ 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 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 $m < m_{\mathrm{switch}}$, 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 transition point $m_{\mathrm{switch}} \approx 0.83$, 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 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 (≈ 15), without being driven into implausible activity states.

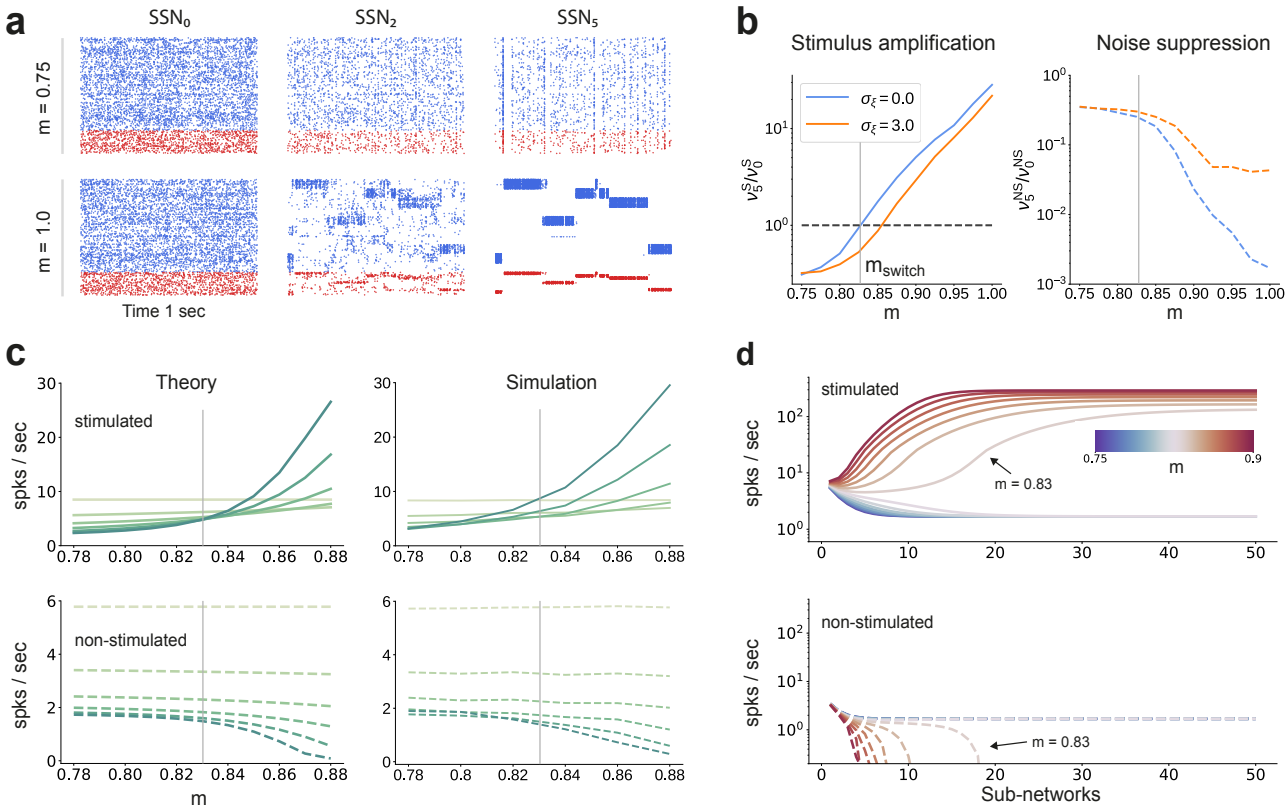

Figure 2: Activity modulation and representational precision. (a) 1 second of spiking activity observed across 1000 randomly chosen excitatory (blue) and inhibitory (red) neurons in SSN_0, SSN_2 and SSN_5, for $\sigma_\xi = 3$ and $m = 0.75$ (top) and $m = 1$ (bottom). **(b)** Mean quotient of firing rates in SSN_5 and SSN_0 (ν_5/ν_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. Figure 2 - figure supplement 1 shows the firing rates as a function of the input intensity λ.

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. 6). 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

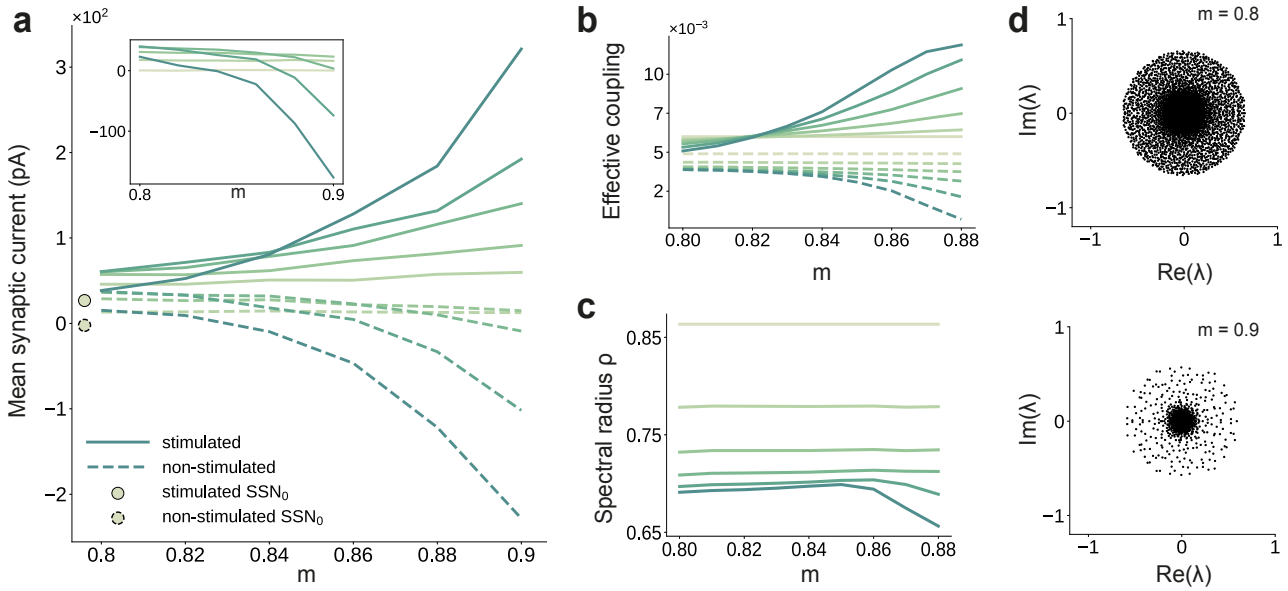

Figure 3: Asymmetric effective couplings modulate the E/I balance and support sequential denoising. **(a)** Mean synaptic input currents for neurons in the stimulated (solid curves) and non-stimulated (dashed curves) excitatory sub-populations in the different SSNs. To avoid clutter, data for SSN_0 is 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.

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 $m \lesssim 0.83$, 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 ($0.83 \lesssim m \lesssim 0.85$) the spectral radius ρ in the deep SSNs expands due to 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 ρ 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

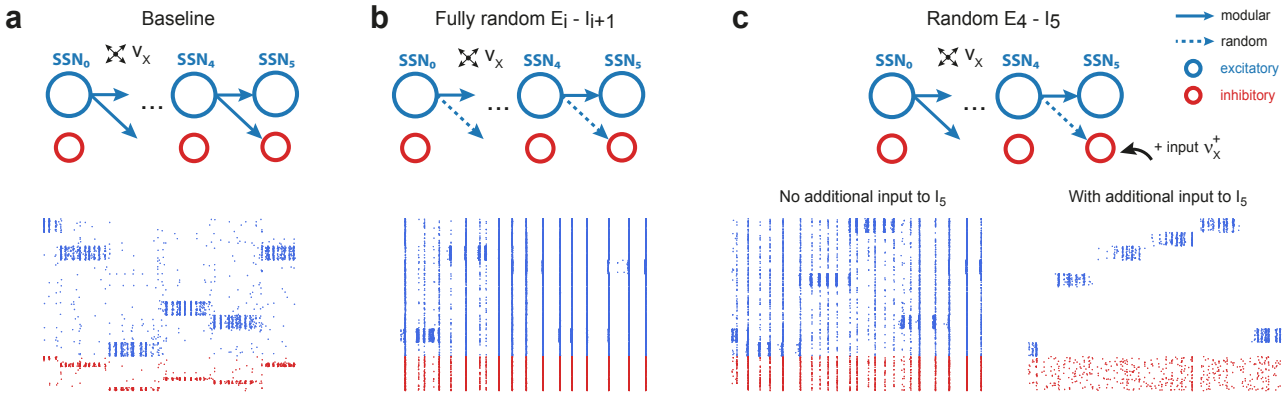

Figure 4: Modular projections to inhibitory populations stabilize network dynamics. Raster plots show 1 second of spiking activity of 1000 randomly chosen neurons in SSN_5, for different network configurations. **(a)** Baseline network with $m = 0.88$. **(b)** Unstructured feedforward projections to 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 $\mathrm{E}_4 \to \mathrm{I}_5$ and additional but unspecific (Poissonian) excitatory input to I_5 controlled via ν_{X}^{+}. Without such input ($\nu_{\mathrm{X}}^{+} = 0$, left), the activity is strongly synchronous, but this is compensated for by the additional excitation, reducing synchrony and restoring the denoising property ($\nu_{\mathrm{X}}^{+} = 10$ spikes/sec, right). Figure 4 - figure supplement 1 depicts the activity statistics in the last two modules, for the different scenarios.

sub-network $\mathrm{SSN}_{\mathrm{i}}$, 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].

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 $\mathrm{E}_{\mathrm{i}} \to \mathrm{I}_{\mathrm{i}+1}$, 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 Figure 4 - figure supplement 1a, 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 Figure 4 - figure supplement 1). 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 $\mathrm{E}_4 \to \mathrm{I}_5$ 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 ($\nu_{\mathrm{X}}^{+} \approx 10$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 Figure 4 - figure supplement 1c).

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

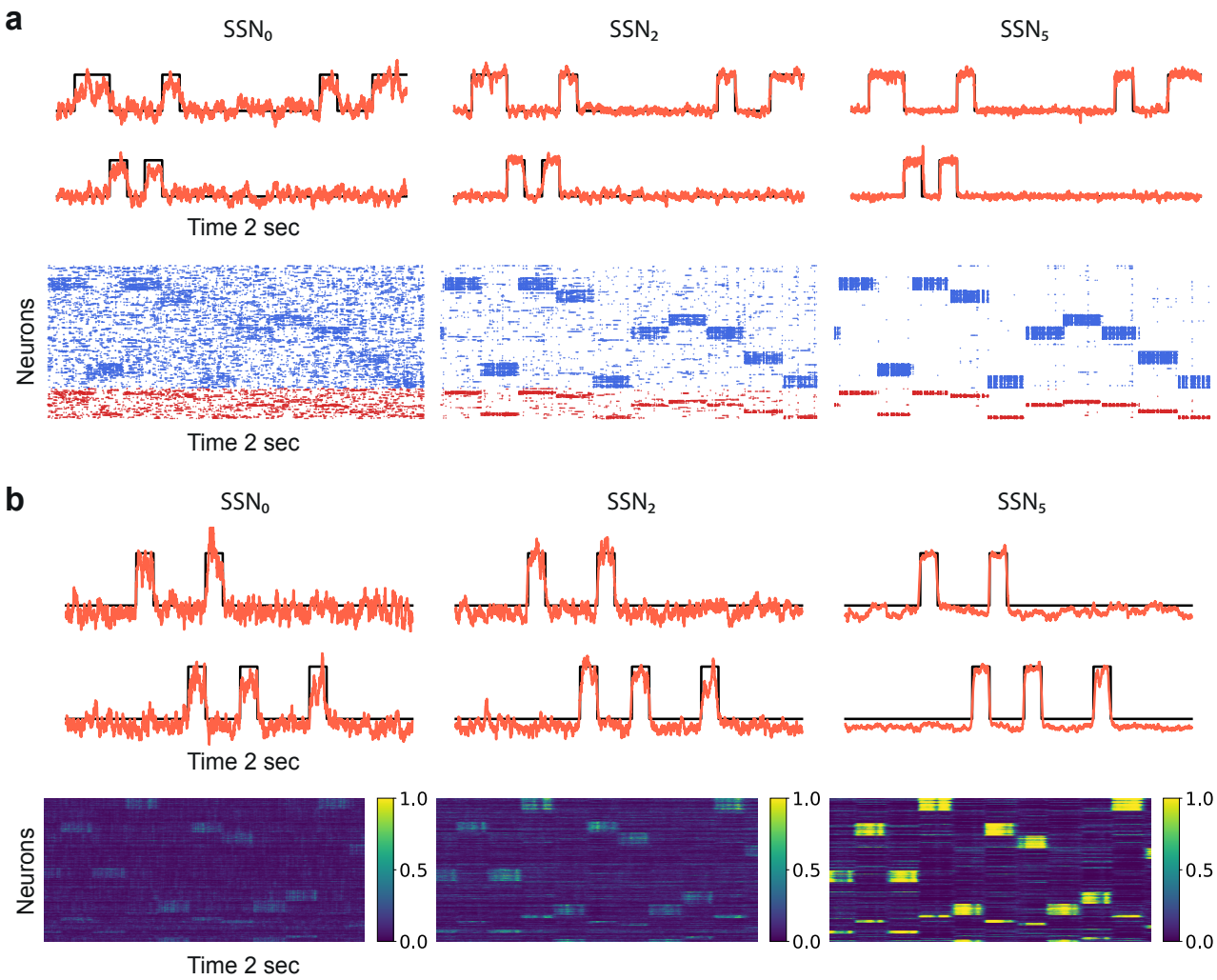

Figure 5: Denoising through modular topography is a robust structural effect. (a) Signal reconstruction (top) and corresponding network activity (bottom) for a network with LIF neurons and conductance-based synapses (see Methods). Single-trial illustration of target signal (black step function) and readout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

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. 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 modeled 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 SSN_2 and can accurately reconstruct the signal. Importantly, the relative improvement with increasing modularity and network depth is retained. In comparison to the baseline model, the transition occurs for a slightly different topographic configuration, $m_{\text{switch}} \approx 0.85$, 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 (Figure 5 - figure supplement 1).

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

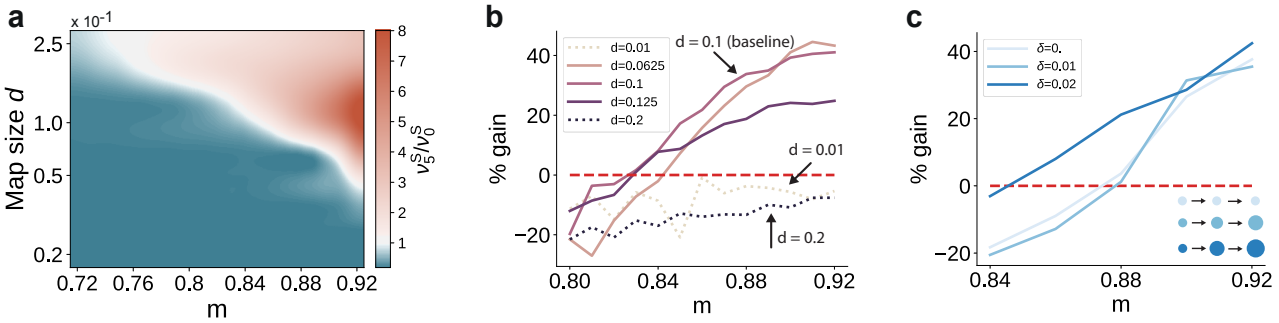

Figure 6: Variation in the map sizes. (a) Difference in the firing rates of the stimulated sub-populations in the first and last sub-networks, $\nu_5^{\mathrm{S}} - \nu_0^{\mathrm{S}}$, as a function of modularity and map size (parameterized by d and constant throughout the network, i.e. $\delta = 0$, 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 SSN_5 relative to 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 ($\sigma_\xi = 0$). Figure 6 - figure supplement 1 further illustrates how the activity varies across the modules as a function of the map size.

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.

First, we varied the size of stimulus-tuned sub-populations (parametrized by d_{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 Figure 6 - figure supplement 1). 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

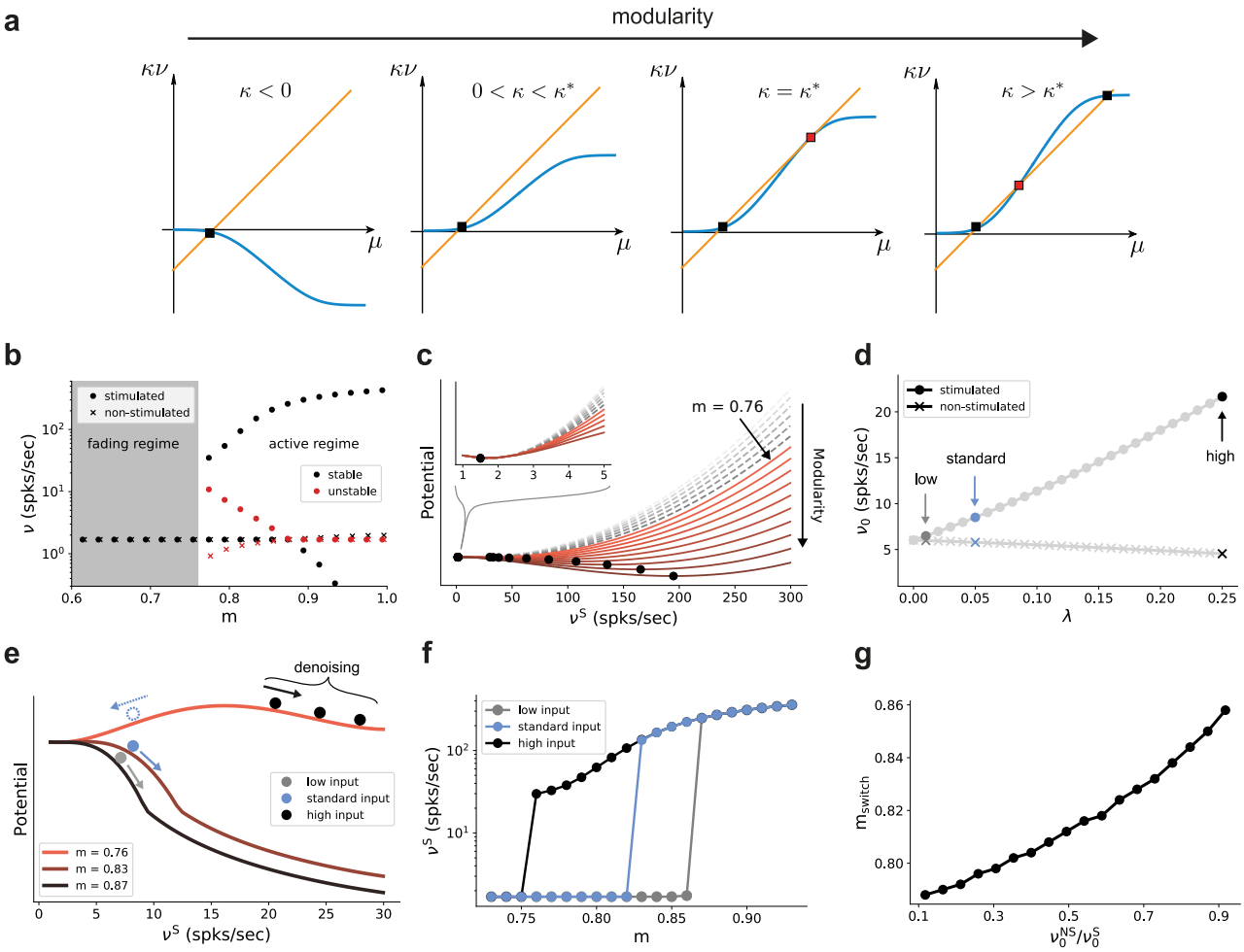

Figure 7: Modularity changes the fixed point structure of the system. (a) Sketch for self-consistent solution (for the full derivation, see Appendix B) for the firing rate of the stimulated sub-population (blue curves) and the linear relation $\kappa\nu = \mu - I$ (orange lines), in the limit of infinitely deep networks. Squares denote stable (black) and unstable (red) fixed points where input and output rates are the same. **(b)** Bifurcation diagram obtained from numerical evaluation of the mean-field self-consistency equations Eq. 9 and Eq. 10 showing a single stable fixed point in the fading regime, and multiple stable (black) and unstable (red) fixed points in the active regime where denoising occurs. **(c)** Potential energy of the mean activity (see Methods and Eq. 22 in Appendix B) 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)** Theoretical predictions for the stationary firing rates of the stimulated and non-stimulated sub-populations in SSN_0, as a function of stimulus intensity (λ, see Methods). Low, standard and high denote λ 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 λ 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. Figure 7 - figure supplement 1 show the dependency of the effective feedforward couplings on different parameters. Note that all panels (except **(a)**) show theoretical predictions obtained from numerical evaluation of the mean-field self-consistency equations.

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 μ to a group of neurons as a linear function of its firing rate ν with a slope κ 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.

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 κ 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), κ changes sign and gradually leads to two additional fixed points (as conceptually illustrated in Fig. 7a and quantified in Fig. 7b by numerically solving the self-consistent mean-field equations, 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 $m \geq m_{\mathrm{crit}} = 0.76$), 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. Note that this critical modularity represents the minimum modularity value for which bistability emerges. It typically differs from the actual switching point m_{switch}, which additionally depends on the input intensity.

In the potential energy landscape U (see Methods), where stable fixed points correspond to minima, the bistability that emerges for more structured topography $m \geq m_{\mathrm{crit}} = 0.76$ 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 (Fig. 7c). Even though the full dynamics of the spiking network away from the fixed point cannot be entirely understood in this simplified potential picture (see Appendix B), qualitatively, more strongly modular networks cause deeper potential wells, corresponding to more attractive dynamical states and higher firing rates (see Figure 9 - figure supplement 2).

Because the intensity of the input signal dictates the rate of different populations in the initial sub-network 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 towards the self-consistent firing rates.

As reported above, for the baseline network and (standard) input ($\lambda = 0.05$) used in Fig. 1 and Fig. 2, the 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 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.

Critical modularity for denoising

In addition to properties of the input, the critical modularity marking the onset of the active regime is also influenced by neuronal and connectivity features. To build some intuition, it is helpful to consider the sigmoidal activation function of spiking neurons (Fig. 8a). The nonlinearity of this function prohibits us from obtaining quantitative, closed-form analytical expressions for the critical modularity and requires a numerical solution of the self-consistency equations (Fig. 7b). However, since the continuous rate model shows a qualitatively similar behavior to the spiking baseline model (see Section "A generalizable structural effect"), we can study a fully analytically tractable model with piecewise linear activation function (Fig. 8a, b) to expose the dependence of

the critical modularity on both neuron and network properties (see detailed derivations in Appendix B).

In this simple model, the output is zero for inputs below $\mu_{\min} = 15$ and at maximum rate $\nu_{\max} = 150$ for inputs above $\mu_{\max} = 400$. In between these two bounds, the output is linearly interpolated $\nu(\mu) = \nu_{\max}(\mu - \mu_{\min})/(\mu_{\max} - \mu_{\min})$. As discussed before, successful denoising is achieved if the non-stimulated sub-populations are silent, $\nu^{\mathrm{NS}} = 0$, and the stimulated sub-populations are active, $\nu^{\mathrm{S}} > 0$. Note that in the following we focus on this ideal scenario representing perfect denoising, but, in principle, intermediate solutions with $\nu^{\mathrm{S}} \gg \nu^{\mathrm{NS}} > 0$ may also occur and could still be considered as successful denoising. Analysing for which neuron, network and input properties this scenario is achieved, we obtain multiple conditions for the modularity that need to be fulfilled.

The first condition illustrates the dependence of the critical modularity on the neuron model (Fig. 8c, purple horizontal line)

$$m \geq \frac{(\mu_{\max} - \mu_{\min})N_{\mathrm{C}}}{(1-\alpha)\mathcal{J}\nu_{\max} + (\mu_{\max} - \mu_{\min})(N_{\mathrm{C}} - 1)}. \tag{1}$$

In particular, it depends on the dynamic range of input $\mu_{\max} - \mu_{\min}$ and output $\nu_{\max}$. The condition represents a lower bound on the modularity required for denoising. Importantly, while it depends on the effective coupling strength $\mathcal{J}$, the noise ratio in the deeper sub-networks α and the number of stimulus-specific sub-populations N_{C} (see Methods), it does not depend on the nature of the recurrent interactions (E/I ratio) and the strength of the external background input. In addition, we find two additional critical values of the modularity (cyan and green curves in Fig. 8c-e), both of which do depend on the strength of the external background input ν_X and the recurrent connectivity (E/I ratio γg):

$$m = \frac{N_{\mathrm{C}}}{N_{\mathrm{C}} - 1} - \frac{1}{N_{\mathrm{C}} - 1} \frac{(1-\alpha)\mathcal{J}\nu_{\max}}{\mu_{\max} - \alpha\mathcal{J}\nu_x - \frac{\mathcal{J}}{N_{\mathrm{C}}}(1+\gamma g)\,\nu_{\max}} \tag{2}$$

$$m = 1 - \frac{\left(\mu_{\min} - \alpha\mathcal{J}\nu_x - \frac{\mathcal{J}}{N_{\mathrm{C}}}(1+\gamma g)\,\nu_{\max}\right)}{\mathcal{J}(1-\alpha)\nu_{\max} - (N_{\mathrm{C}} - 1)\left(\mu_{\min} - \alpha\mathcal{J}\nu_x - \frac{\mathcal{J}}{N_{\mathrm{C}}}(1+\gamma g)\,\nu_{\max}\right)} \tag{3}$$

Depending on the external input strength νX, these are either upper or lower bounds. In the denominator of these expressions, the total input (recurrent and external) is compared to the limits of the dynamic range of the neuron model. The cancellation between recurrent and external inputs in the inhibition-dominated baseline model typically yields a total input within the dynamic range of the neuron, such that modularity in feedforward connections can decrease the input of the non-stimulated sub-populations to silence them, and increase the input of the stimulated sub-populations to support their activity. The competition between the excitatory and inhibitory contributions ensures that the total input does not lead to a saturating output activity. Thus, for inhibitory recurrence, denoising can be achieved at a moderate level of modularity over a large range of external background inputs (shaded black and hatched regions in Fig. 8c), which demonstrates a robust denoising mechanism even in the presence of changes in the input environment.

In contrast, if recurrent connections are absent, strong inhibitory external background input is required to counteract the excitatory feedforward input and achieve a denoising scenario (Fig. 8d). Fixed points at non-saturated activity $\nu^{\mathrm{S}} > 0$ are also present for low excitatory external input, but unstable due to the positive recurrent feedback. This is because in networks without recurrence, there is no competition between the recurrent input and the external and feedforward inputs. As a result, the input to both the stimulated and non-stimulated sub-populations is typically high, such that modulation of the feedforward input via topography cannot lead to a strong distinction between the pathways as required for denoising. In these networks, one typically observes high activity in all populations. A similar behavior can be observed in excitation-dominated networks (Fig. 8e), where the inhibitory external background input must be even stronger to compensate the excitatory feedforward and recurrent connectivity and reach a stable denoising regime.

Note that inhibitory external input is not in line with the excitatory nature of external inputs to local circuits in the brain and is therefore biologically implausible. One way to achieve denoising in excitation-dominated networks for excitatory background inputs would be to shift the dynamic range of the activation function (see Figure 8 - figure supplement 1), which is, however, not consistent with the biophysical properties of real neurons (distance between threshold and rest as compared to typical strengths of postsynaptic potentials). In summary, we find that recurrent inhibition is crucial to achieve denoising in biologically plausible settings.

These results on the role of recurrence and external input can be transferred to the behavior of the spiking model. While details of the fixed point behavior depend on the specific choice of the activation function, Fig. 8f and Fig. 8h show that there is also no denoising regime for the spiking model in case of no or excitation-dominated recurrence and a biologically plausible level of external input. Instead, one finds high activity in both stimulated and non-stimulated sub-populations, as confirmed by network simulations (Fig. 8g and Fig. 8i). Figure 8 - figure supplement 2 further confirms that even reducing the external input to zero does not avoid this high-activity state in both stimulated and non-stimulated sub-populations for $m < 1$.

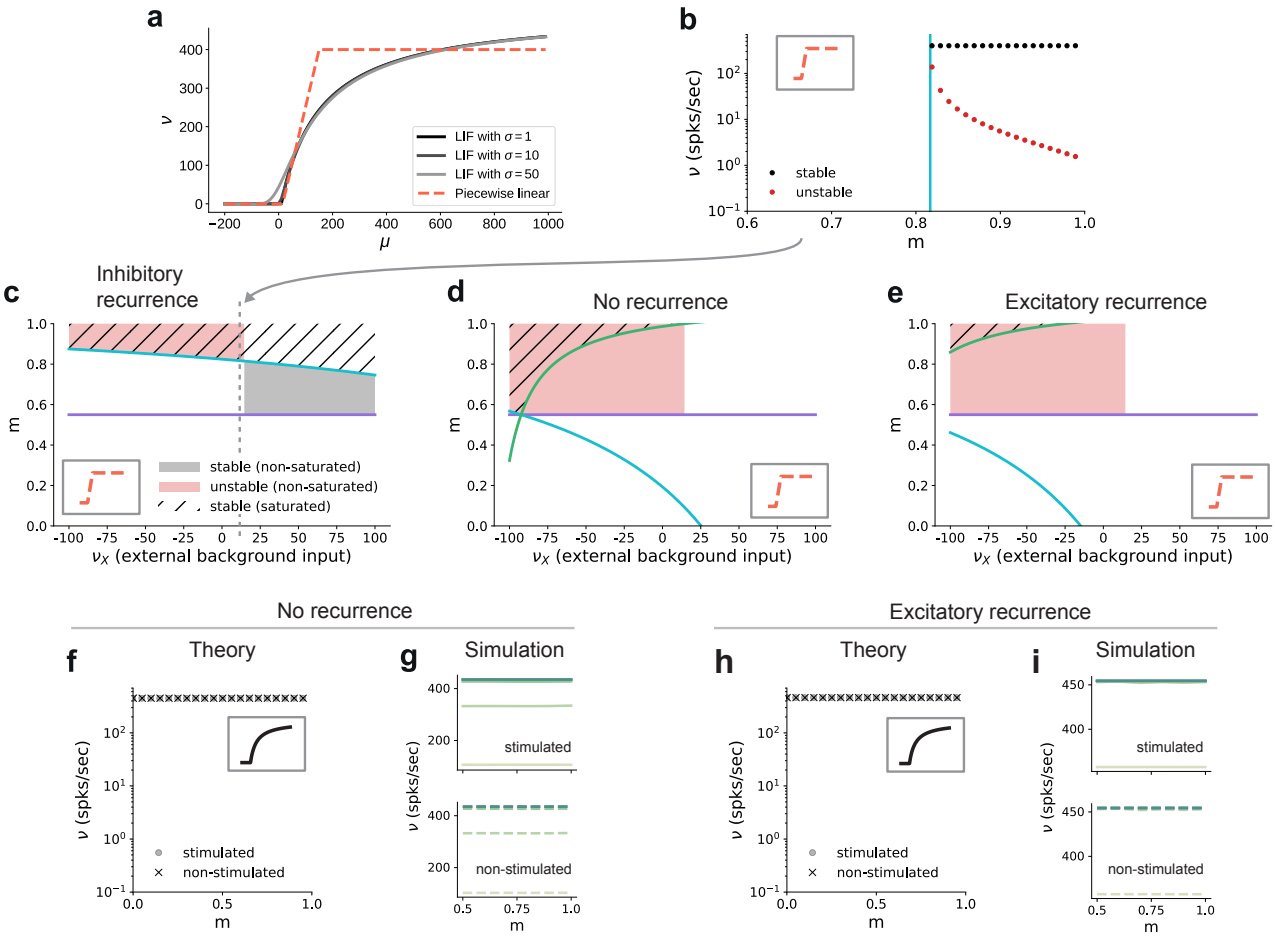

Figure 8: Dependence of critical modularity on neuron and connectivity features. **(a)** Activation function $\nu(\mu, \sigma)$ for leaky integrate-and-fire model as a function of the mean input μ for $\sigma = 1, 10, 50$ (black to gray) and piecewise linear approximation with qualitatively similar shape (red). **(b)** Bifurcation diagram as in Fig. 7b, but for piecewise linear activation function shown in inset. Low-activity fixed points at zero rate are not shown, which is the case throughout for the non-stimulated sub-populations. This panel corresponds to the cross-section marked by the gray dashed lines in **(c)**, at $\nu_{\mathrm{X}} = 12$. Likewise, the vertical cyan bar corresponds to the lower bound on modularity depicted by the cyan curve in **(c)** for the same value $\nu_{\mathrm{X}} = 12$. **(c)** Analytically derived bounds on modularity (purple line corresponds to Eq. 1, cyan curve to Eq. 2) as a function of external input for the baseline model with inhibition-dominated recurrent connectivity ($g = -12$). Shaded regions denote positions of stable (black) and unstable (red) fixed points with $0 < \nu^{\mathrm{S}} < \nu_{\max}$ and $\nu^{\mathrm{NS}} = 0$. Hatched area represents region with stable fixed points at saturated rates. Denoising occurs in all areas with stable fixed points (hatched and black shaded regions). Negative values on the x-axis correspond to inhibitory external background input with rate $|\nu x|$. **(d)** Same as panel **(c)** for networks with no recurrent connectivity within the SSNs (green curve defined by Eq. 3). **(e)** Same as panel **(c)**, for networks with excitation-dominated connectivity within SSNs ($g = -3$). **(f)** Same as Fig. 7b, obtained through numerical evaluation of the mean-field self-consistent equations for the spiking model. All non-zero fixed points are stable, with points representing stimulated (circle) and non-stimulated (cross) populations overlapping. **(g)** Mean firing rates across the SSNs in the current-based (baseline) model with no recurrent connections, obtained from 5 seconds of network simulations and averaged over 5 trials. **(h, i)** Same as **(f, g)**, for networks with excitation-dominated connectivity.

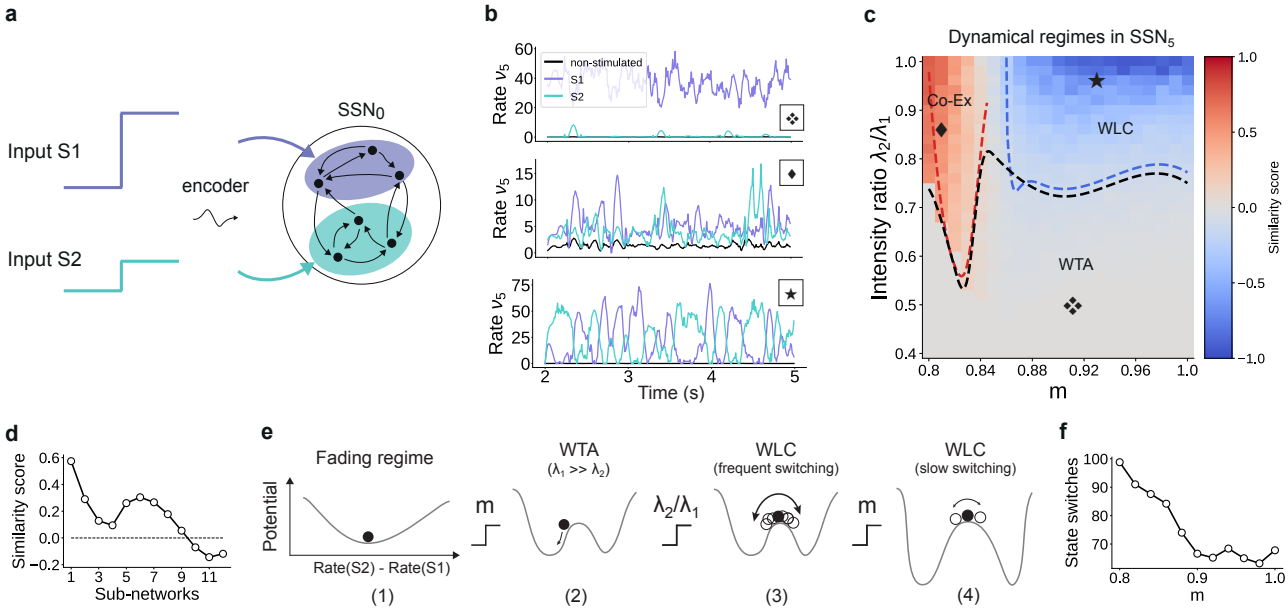

Figure 9: For multiple input streams, topography may elicit a wide range of dynamical regimes. **(a)** Two active input channels with corresponding stimulus intensities λ_1 and λ_2, mapped onto non-overlapping sub-populations, drive the network simultaneously. Throughout this section, $\lambda_1 = 0.05$ 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 λ_2/λ_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 $\lambda_1 \gg \lambda_2$; (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 $\lambda_2/\lambda_1 = 0.9$. Figure 9 - figure supplement 1 and Figure 9 - figure supplement 2 show the evolution of the Co-Ex region over 12 modules and the potential landscape, respectively.

Input integration and multi-stability

The analysis considered in the sections 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. 9a), 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. 9b, 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. 9b, 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 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 λ_2/λ_1 (see Methods and Fig. 9c). 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. 9c). 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. 9d).

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. 9e (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. 9c, grey area), one stimulus dominates (WTA) and the responses in the two populations are uncorrelated (Fig. 9b, 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. 9e, (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. 9b, bottom panel), i.e. the activation of the two stimulus-specific sub-populations switches, as they engage in a dynamic behavior reminiscent of winnerless competition (WLC) between multiple neuronal groups [42, 24]. The switching between the two states is driven by stochastic fluctuations (Fig. 9e, (3)). The depth of the wells and width of barrier (distance between fixed points) increase with modularity (see Fig. 9e, (4) and Figure 9 - figure supplement 2), suggesting a greater difficulty in moving between the two attractors and consequently fewer state changes. Numerical simulations confirm this slowdown in switching (Fig. 9f).

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.

Reconstruction and denoising of dynamical inputs

Until now, we have considered continuous but piecewise constant, step signals, with each step lasting for a relatively long and fixed period of 200 ms. This may give the impression that the denoising effects we report only works for static or slowly changing inputs, whereas naturalistic stimuli are continuously varying. Nevertheless, sensory perception across modalities relies on varying degrees of temporal and spatial discretization [43], with individual (sub-)features of the input encoded by specific (sub-)populations of neurons in the early stages of the sensory hierarchy. In this section, we will demonstrate that denoising is robust to the temporal properties of the input and its encoding, as we relax many of the assumptions made in previous sections.

We consider a sinusoidal input signal, which we discretize and map onto the network according to the depiction in Fig. 10a. This approach is similar to previous works, for instance it can mimic the movement of a light spot across the retina [44]. By varying the sampling interval dt and number of channels k, we can change the coarseness of the discretization from step-like signals to more continuous approximations of the input. If we choose a high sampling rate ($dt = 1$ ms) and sufficient channels ($k = 40$), we can accurately encode even fast changing signals (Fig. 10b). Given that each input-driven SSN is inhibition-dominated and therefore close to the balanced state, the network exhibits a fast tracking property [45] and can accurately represent and denoise the underlying continuous signal in the spiking activity (Fig. 10c, top). This is also captured by the readout, with the tracking precision increasing with network depth (Fig. 10c, bottom). In this condition, there is a performance gain of up to 50% in the noiseless case (Fig. 10d, top) and similar values for varying levels of noise (Fig. 10d, bottom).

Note that due to the increased number of input channels (40 compared to 10) projecting to the same number of neurons in SSN_0 as before (800), for the same σ_ξ the effective amount of noise each neuron receives is, on average, four times larger than in the baseline network. Moreover, the task was made more difficult by the significant overlap between the maps ($N_\mathrm{C} = 20$) as well as the resulting decrease in neuronal input selectivity. Nevertheless, similar results were obtained for slower and more coarsely sampled signals (Fig. 10e-g).

We found comparable denoising dynamics for a large range of parameter combinations involving the map size, number of maps, number of channels and signal complexity. Although there are limits with respect to the frequencies (and noise intensity) the network can track (see Figure 10 - figure supplement 1), these findings indicate a very robust and flexible phenomenon for denoising spatially encoded sensory stimuli.

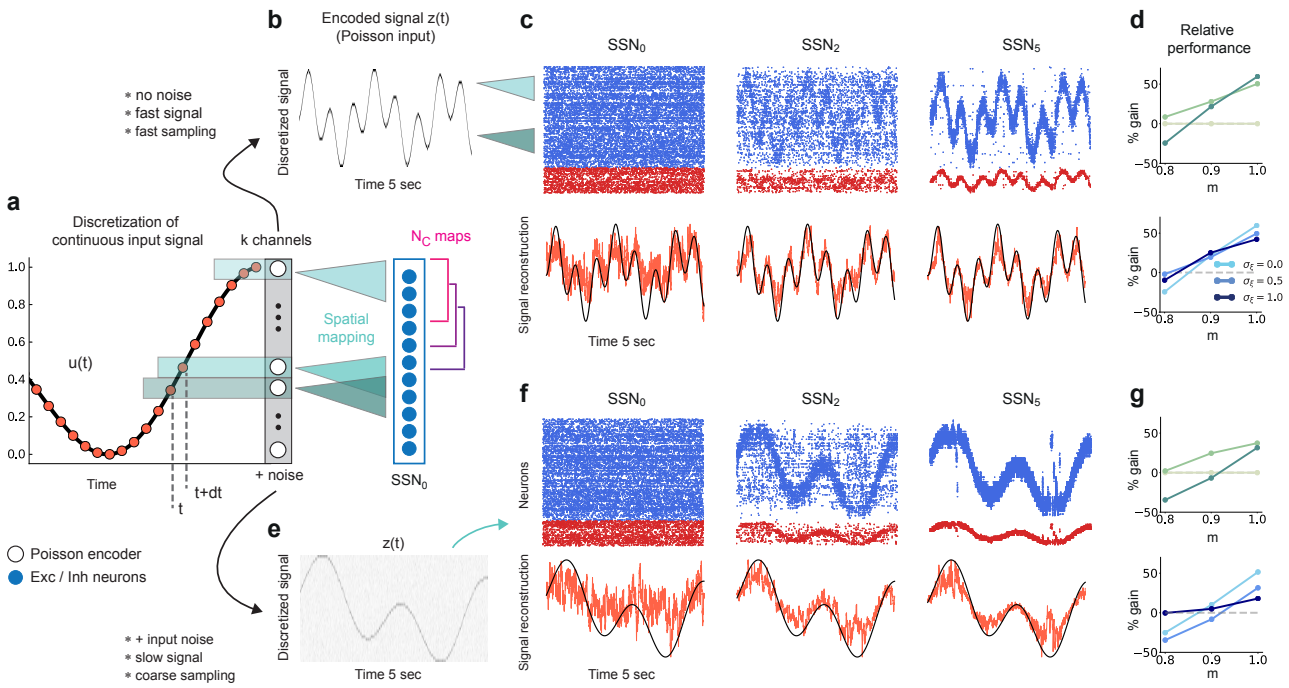

Figure 10: Reconstruction of a dynamic, continuous input signal. **(a)** Sketch of the encoding and mapping of a sinusoidal input $x(t)$ onto the current-based network model. The signal is sampled at regular time intervals dt, with each sample binned into one of k channels (which is then active for a duration of dt). This yields a temporally and spatially discretized k-dimensional binary signal $u(t)$, from which we obtain the final noisy input $z(t)$ similarly to the baseline network (see Fig. 1 and Methods). Unlike the one-to-one mapping in Fig. 1, here we decouple the number of channels $k = 40$ from that of topographic maps, $N_\mathrm{C} = 20$ (map size is unchanged, $C_\mathrm{i} = 800$). Because $N_\mathrm{C} < k$, the channels project to evenly spaced but overlapping sub-populations in SSN_0, while the maps themselves overlap significantly. **(b)** Discretized signal $z(t)$ and rate encoding for input $x(t) = \sin(10t) + \cos(3t)$, with $dt = 1$ ms and no noise ($\sigma_\xi = 0$). **(c)** Top panel shows the spiking activity of 500 randomly chosen excitatory (blue) and inhibitory (red) neurons in SSN_0, SSN_2 and SSN_5, for $m = 0.9$. Corresponding target signal $x(t)$ (black) and readout output (red) are shown in bottom panel. **(d)** Relative gain in performance in SSN_2 and SSN_5 for $\sigma_\xi = 0$ (top). Color shade denotes network depth. Bottom panel shows relative gain in SSN_5 for different levels of noise $\sigma_\xi \in \{0, 0.5, 1\}$. **(e-g)** Same as **(b-d)**, but for a slowly varying signal (sampled at $dt = 20$ ms), $\sigma_\xi = 0.5$ and $m = 1$. Performance results are averaged across five trials. We used 20 seconds of data for training and 10 seconds for testing (activity sampled every 1 ms, irrespective of input discretization dt).

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 [46, 47, 48] 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. We demonstrate that this holds for a variety of signals, from approximately static (stepwise) to smoothly and rapidly changing dynamic inputs (Fig. 10). Thanks to the balance of excitation and inhibition, the network is able to track spatially encoded signals on very short timescales, and is flexible with respect to the level of spatial and temporal discretization. Accurate tracking and denoising requires that the encoding is locally static/semi-stationary for only a few tens of milliseconds, which is roughly in line with psychophysics studies on the limits of sensory perception [49].

More generally, topographic modularity, in conjunction with other top-down processes [50], 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 [51, 52]. 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. 9 and [53]).

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 [54, 55]. 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 [56, 57]) and intermediate modularity, the firing rates remain within experimentally observed limits and do not saturate. The basic principle is similar to other approaches that alter the gain on specific pathways to facilitate stimulus propagation, for example through stronger synaptic weights [58], stronger nonlinearity [36], tuning of connectivity strength and neuronal thresholds [37], via detailed balance of local excitation and inhibition (amplitude gating [59]) or with additional sub-cortical structures [60]. Additionally, our model also displays some activity characteristics reported previously, such as the response sharpening observed for synfire chains [61] or (almost) linear firing rate propagation [62] (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 [63, 64, 59]. 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 [65, 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 [66, 67], also in the context of sequential denoising autoencoders [38] and stimulus classification [68], which used simplified neuron models or shallow networks, made no distinction between excitatory and inhibitory connections, or relied on specific, trained connection patterns (e.g., chosen by the pseudo-inverse model). Although the bistability underlying denoising can, in principle, also be achieved in such feedforward or networks without inhibition, our theoretical predictions and network simulations indicate that for biologically constrained circuits (i.e., where the background and long-range feedforward input is excitatory), inhibitory recurrence is indispensable for the spatial denoising studied here (see Section "Critical modularity for denoising"). Recurrent inhibition compensates for the feedforward and external excitation, generating competition between the topographic pathways and allowing the populations to rapidly track their input.

Moreover, our findings provide 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, 69, 70], 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. 9). Computation by switching is a functionally relevant principle [22, 71], 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 [72], and is conceptually similar to an attractor-based model of perceptual bistability [73]. 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 discernible in the initial modules. Previous studies reporting on similar dynamical states have usually considered either the synaptic weights as the main control parameter [42, 74, 58] or studied specific architectures with clustered connectivity [75, 76, 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., [77]). 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_{FF}.

In addition, all neurons receive K_{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, 78, 79]). To ensure that the total excitatory input to each neuron is consistent across the network, we scale K_{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_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_i by introducing a step-wise linear increment δ, 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 [80, 81] 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_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:

$$\tau_\mathrm{m} \frac{dV(t)}{dt} = (V_\mathrm{rest} - V(t)) + R\left(I^\mathrm{E}(t) + I^\mathrm{I}(t) + I^\mathrm{X}(t)\right) \tag{4}$$

where τ_m is the membrane time constant, and RI^β is the total synaptic input from population $\beta \in [E, I]$. The background input I^X is assumed to be excitatory and stochastic, modeled as a homogeneous Poisson process with constant rate ν_X. Synaptic weights J_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, 79], we choose the intensity of the stochastic input ν_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 ~ 3 spikes/s, coefficient of variation (CV_ISI) in the interval $[1.0, 1.5]$ and Pearson cross-correlation (CC) ≤ 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:

$$\begin{aligned} \tau_x \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{dt}} &= -\boldsymbol{x} + J\boldsymbol{r} + J^\mathrm{in}\boldsymbol{u} - \boldsymbol{b}^\mathrm{rec} + \sqrt{2\tau_{\boldsymbol{x}}}\sigma_\mathrm{X}\boldsymbol{\xi} \\ \boldsymbol{r} &= 0.5(1 + \tanh(\boldsymbol{x})) \end{aligned} \tag{5}$$

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_{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 J and J^{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 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 σ_{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 [82]. This is equivalent to probing the network's ability to function as a denoising autoencoder [83].

To generate the N_{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 adding a Gaussian white noise process with zero mean and variance σ_ξ^2 to $u(t)$, and scaling the sum with the input rate ν_{in}. 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 σ_ξ.

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 SSN_0 (see Fig. 2c).

Each input channel k is mapped onto one of the N_{C} stimulus-specific sub-populations of excitatory and inhibitory neurons in the first (input) sub-network SSN_0, chosen according to the procedure described above (see also Fig. 1a). This way, each stimulus S_{k} is mapped onto a specific set of sub-populations in the different sub-networks, i.e., the topographic map associated with S_{k}.

For each stimulus in the sequence, we sample the responses of the excitatory population in each $\mathrm{SSN}_{\mathrm{i}}$ 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 [84, 85]. 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 [86]. 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 $\mathrm{SSN}_{\mathrm{i}}$, we train the state matrix $X_{\mathrm{SSN}_{\mathrm{i}}}$ on a larger interval of delays and choose the one that minimizes the error, averaged across multiple trials.

In Section "Reconstruction and denoising of dynamical inputs", we generalize the input to a sinusoidal signal $x(t) = \sin(a \cdot t) + \cos(b \cdot t)$, with parameters a and b. From this, we obtain $u(t)$ through the sampling and discretization process described in the respective section, and compute the final input $z(t) = [\nu_{\mathrm{in}}(u(t) + \xi(t))]_+$ as above.

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 ν_{i}, to a small perturbation in the input rate ν_{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]:

$$
\begin{aligned}
w_{ij} &= \frac{\partial \nu_i}{\partial \nu_j} = \widetilde{\alpha} J_{ij} + \widetilde{\beta} J_{ij}^2 \\
\text{with} \quad \widetilde{\alpha} &= \sqrt{\pi}\,(\tau_{\mathrm{m}}\nu_i)^2\,\frac{1}{\sigma_i}\,(f(y_\theta) - f(y_{\mathrm{r}})) \\
\text{and } \tilde{\beta} &= \sqrt{\pi}\,(\tau_{\mathrm{m}}\nu_i)^2\,\frac{1}{2\sigma_i^2}\,(f(y_\theta)y_\theta - f(y_{\mathrm{r}})y_{\mathrm{r}})\,.
\end{aligned}
\tag{6}
$$

Note that in Fig. 3 we ignore the contribution $\tilde{\beta}$ resulting from the modulation in the input variance σ_{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 λ_{k}^* associated with the unit vector. The stability of the system is thus uniquely determined by the spectral radius ρ: 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_{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 ν^{NS} that is in general different from the rate of the stimulated SP ν^{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

$$
\begin{aligned}
\mu^{\mathrm{S}} &= (1+\lambda)\mathcal{J}\nu_{\mathrm{x}} + \frac{1}{N_{\mathrm{C}}}\mathcal{J}\,(1+\gamma g)\,\nu^{\mathrm{S}} + \frac{N_{\mathrm{C}}-1}{N_{\mathrm{C}}}\mathcal{J}\,(1+\gamma g)\,\nu^{\mathrm{NS}}\,, \\
\mu^{\mathrm{NS}} &= \qquad\quad \mathcal{J}\nu_{\mathrm{x}} + \frac{1}{N_{\mathrm{C}}}\mathcal{J}\,(1+\gamma g)\,\nu^{\mathrm{S}} + \frac{N_{\mathrm{C}}-1}{N_{\mathrm{C}}}\mathcal{J}\,(1+\gamma g)\,\nu^{\mathrm{NS}}
\end{aligned}
\tag{7}
$$

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

$$
\kappa\nu = \mu - I \tag{8}
$$

where κ is the effective self-coupling of a group of neurons with rate ν and input μ, and I denotes the external inputs from other groups. Equation (8) describes a linear relationship between the rate ν and the input μ. To find a self-consistent solution for the rates ν^{S} and ν^{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 σ^2 of inputs. The latter can be obtained analogous to the mean input μ (see Appendix B). Note that for general nonlinearity $\nu(\mu)$ there is no analytical closed-form solution for the fixed points.

Starting from SSN_1, networks are connected in a fixed pattern such that the rate ν_i in SSN_i also depends on the excitatory input from the previous sub-network SSN_{i-1} with rate ν_{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)

$$
\mu^{\mathrm{S}} = \alpha\mathcal{J}\nu_{\mathrm{x}} + \kappa_{\mathrm{S,S}}\;\nu^{\mathrm{S}} + \kappa_{\mathrm{S,NS}}\;\nu^{\mathrm{NS}} \tag{9}
$$

$$
\mu^{\mathrm{NS}} = \alpha\mathcal{J}\nu_{\mathrm{x}} + \kappa_{\mathrm{NS,S}}\;\nu^{\mathrm{S}} + \kappa_{\mathrm{NS,NS}}\;\nu^{\mathrm{NS}} \tag{10}
$$

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. 9 and Eq. 10 have the same linear form as Eq. 8 and can be solved numerically as described above. Again, for general nonlinear $\nu(\mu)$ there is no closed-form analytical solution, but see below for a piecewise linear activation function $\nu(\mu)$. The numerical solutions for fixed points are obtained using the root finding algorithm `root` of the `scipy.optimize` package [87]. The stability of the fixed points is obtained by inserting the corresponding firing rates into the effective connectivity Eq. 6. On the level of stimulated and non-stimulated sub-populations, the effective connectivity matrix reads

$$\frac{1}{\tau_{\mathrm{m}}}\begin{pmatrix} \kappa_{\mathrm{S,S}}(m)\tilde{\alpha}(\nu^{\mathrm{S}}) & \kappa_{\mathrm{S,NS}}(m)\tilde{\alpha}(\nu^{\mathrm{NS}}) \\ \kappa_{\mathrm{NS,S}}(m)\tilde{\alpha}(\nu^{\mathrm{S}}) & \kappa_{\mathrm{NS,NS}}(m)\tilde{\alpha}(\nu^{\mathrm{NS}}) \end{pmatrix} \tag{11}$$

from which we obtain the maximum eigenvalue ρ, which for stable fixed points must be smaller than 1.

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 $\frac{d\nu^{\mathrm{S}}}{dt} = F$ [88, 55, 89]. Note that μ and σ^2 are again functions of ν^{S} and ν^{NS}, where the latter is the self-consistent rate of the non-stimulated sub-populations for given rate ν^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. 9 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 λ_2/λ_1 (with λ_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. 9c 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. 9d), 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 [90], 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 [91], 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 [92], we provide a complete code package that implements project-specific functionality within NMSAT (see Supplementary Materials) using NEST 2.18.0 [93].

Data availability

All the relevant data and code is available in a public zenodo repository at https://doi.org/10.5281/zenodo.6326496 (see also Supplementary Materials). Source data and code files are also attached as zip folders to the individual main figures of this submission.

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.

Supplementary Figures

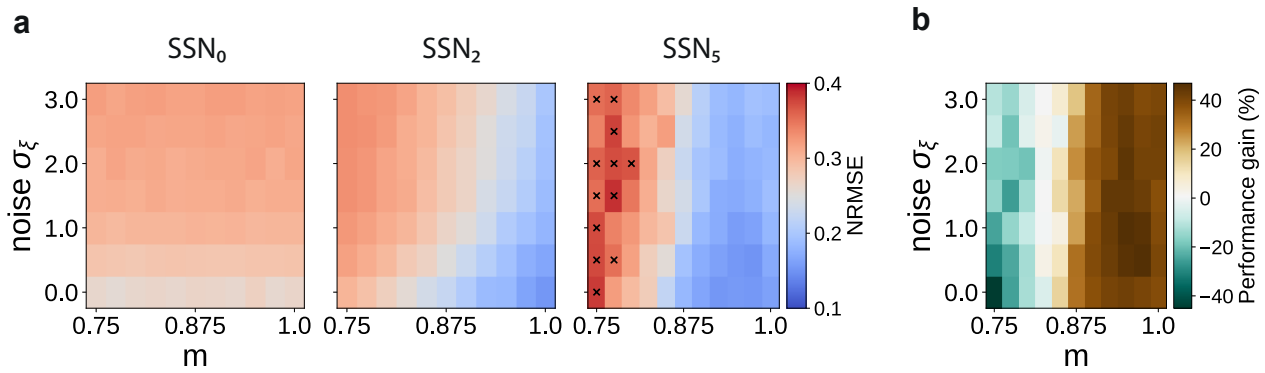

Figure 1 - supplement 1: Sequential denoising effect. **(a)** Reconstruction error (NRMSE) in 3 different sub-networks as a function of modularity (m) and noise amplitude (σ_ϵ). The points marked in the rightmost panel correspond to chance-level reconstruction accuracy. **(b)** Relative reconstruction performance gain in SSN_5 compared to SSN_0, expressed as percentage of error decrease.

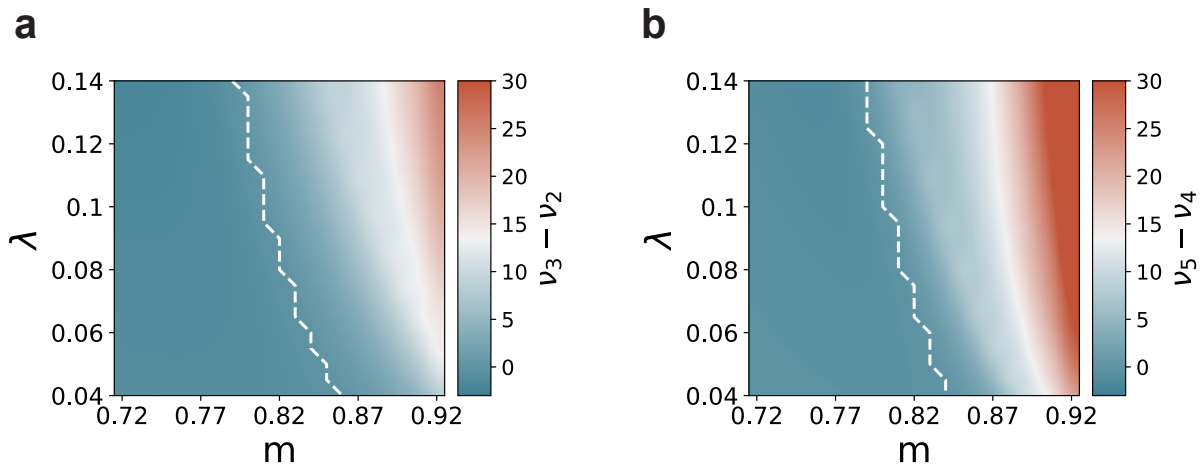

Figure 2 - supplement 1: Mean-field predictions for the gain in the firing rates of stimulated sub-populations (**(a)** $\nu_3^S - \nu_2^S$ and **(b)** $\nu_5^S - \nu_4^S$), as a function of modularity m and input intensity, scaled by λ (see Methods). Dashed lines demarcate the transition to positive gain.

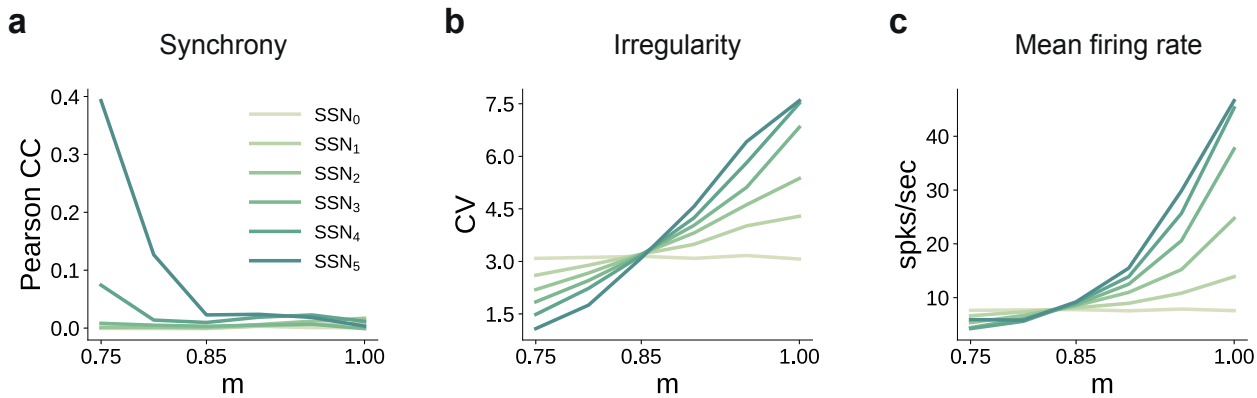

Figure 5 - supplement 1: Spiking statistics for the conductance-based model: **(a)** synchrony (Pearson's correlation coefficient, computed pairwise over spikes binned into 2 ms bins and averaged across 500 pairs); **(b)** irregularity measured as the coefficient of variation (CV); **(c)** mean firing rate across the excitatory populations. All depicted statistics were averaged over five simulations, each lasting 5 s, with 10 input stimuli.

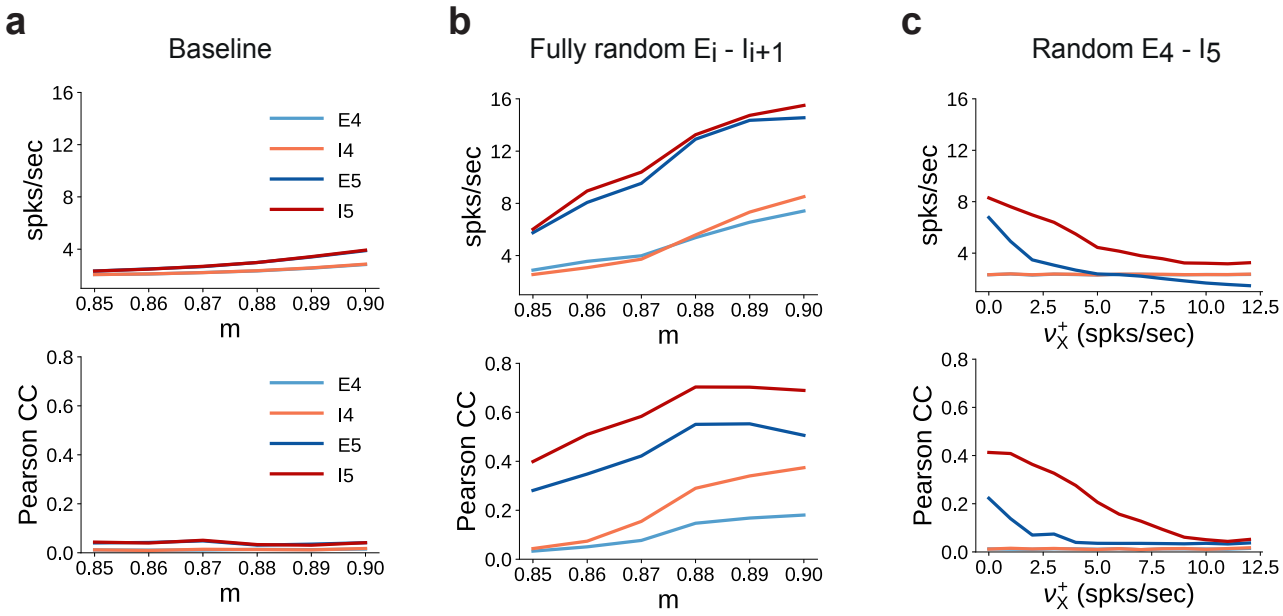

Figure 4 - supplement 1: Spiking statistics for different feedforward wiring to inhibitory neurons. (a) Mean firing rates (top panel) and synchrony (Pearson's correlation coefficient, computed pairwise over spikes binned into 2 ms bins and averaged across 500 pairs, lower panel) in SSN_4 and SSN_5, as a function of modularity. **(b)** Same as **(a)**, except with random feedforward projections to the inhibitory pools, i.e., $m = 0$ for all $E_i \rightarrow I_{i+1}$ connections, $i = \{0..4\}$. **(c)** Same as the baseline network in **(a)**, with $m = 0$ only for $E_4 \rightarrow I_5$. In addition, each neuron in I_5 receives further excitatory background input with intensity $\nu'_X = \nu_X + \nu_X^+$. Statistics are computed as a function of the additional rate ν_X^+.

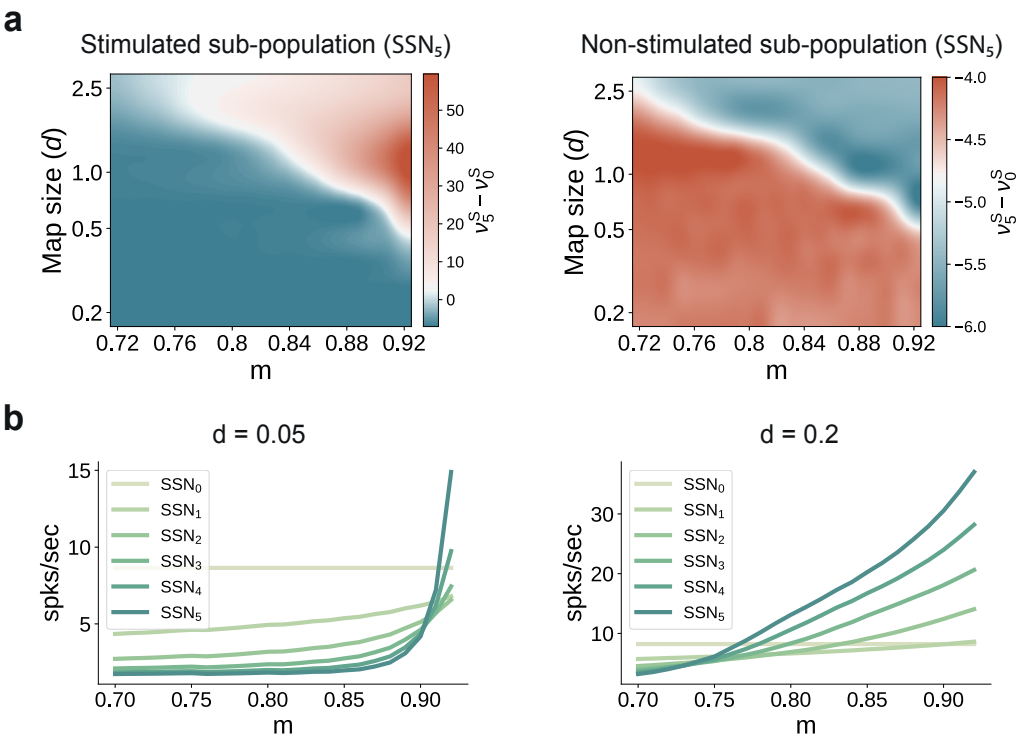

Figure 6 - supplement 1: Transition point in modularity decreases with larger map sizes. (a) Mean-field predictions for the stationary firing rates of the stimulated (left) and non-stimulated sub-populations (right) in SSN_5, as a function of modularity (m) and fixed map size (parametrized by d, see Methods) across the modules ($\delta = 0$). To limit the impact of additional parameters when varying the map sizes (e.g., overlap), the number of stimulus-specific sub-populations and d where chosen such that every neuron in each population belonged to exactly one stimulus-specific sub-population (see main text for more details). **(b)** Predicted firing rates in the stimulated sub-populations of the different sub-networks, for $d = 0.05$ (left) and $d = 0.2$ (right), with $\delta = 0$.

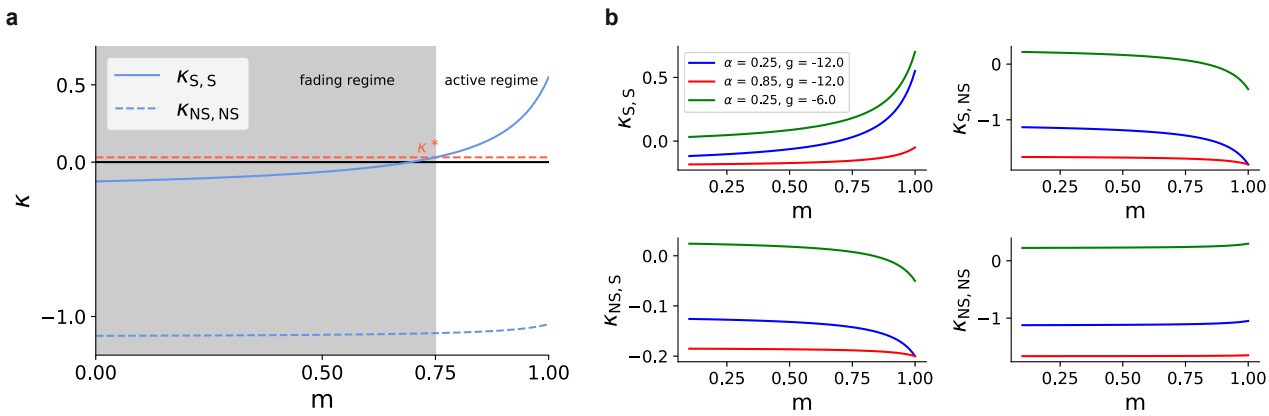

Figure 7 - supplement 1: Effective couplings. (a) The effective coupling between stimulated sub-populations, $\kappa_{\mathrm{S,S}}$ increases with modularity, with κ^* marking the critical transition point between the fading and active regime. Dashed red line represents this κ^* value. **(b)** $\kappa_{\mathrm{S,S}}$ (top left panel), $\kappa_{\mathrm{S,NS}}$ (top right), $\kappa_{\mathrm{NS,S}}$ (bottom left), $\kappa_{\mathrm{NS,NS}}$ (bottom right), as a function of modularity for different values of α and E/I weight ratio g. The parameters used in this study (blue) yield $\kappa_{\mathrm{S,S}} > 0$ only for large modularity, with the other couplings being negative for all m. Increasing the signal-to-noise ratio to $\mathrm{SSN}_{\geq 1}$ (red), i.e., increasing the background external input while reducing the feedforward connection density (directly coupled, see Methods), destroys bistability (all couplings are negative for all values of m) and leads to extinction of activity in all sub-populations in the limit of very deep networks. Decreasing inhibition (green) also creates possible bistability for non-stimulated sub-populations ($\kappa_{\mathrm{NS,NS}} > 0$) such that their activity might approach a high-activity fixed point, leading to destruction of task performance. Note that in all panels, the values κ are scaled by $1/\mathcal{J}$ to highlight the transitions around zero (see Eq. 18 in Appendix B).

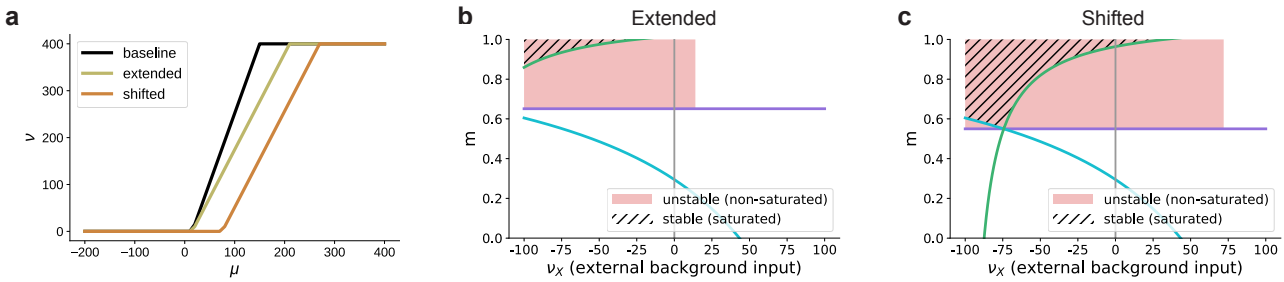

Figure 8 - supplement 1: Influence of the activation function's dynamic range on the bifurcation behavior in excitation-dominated networks ($g = -3$, see also Fig. 8e) (a) The baseline dynamic range $[15, 150]$ is extended to $[15, 210]$ or shifted to $[75, 210]$. **(b)** Given that $\mu_{\max}$ does not enter the lower bound on the modularity determined by Eq. 3 (green curve), extending the dynamic range (see panel **(a)**) does not affect the region of stable fixed points in the parameter space. For positive background input, there are no stable fixed points, only unstable ones at non-saturated rates for low values of ν_{X}, due to excitatory recurrent fluctuations in the activity. For stronger background input, no fixed points exists where $\nu_{\mathrm{S}} > \nu_{\mathrm{NS}}$. In this case, the activity of the non-stimulated populations (non-zero) dominates the recurrent dynamics and denoising can not be achieved. **(c)** Shifting the dynamic range altogether (see panel **(a)**) leads to the emergence of stable fixed points at saturated rates also for positive external input, but the region in which denoising occurs is still significantly smaller than for networks with recurrent inhibition (see Fig. 8c). For these values, $\nu_{\mathrm{NS}} = 0$ is ensured because the total input to the non-stimulated populations remains below the shifted dynamical range, in contrast to just extending the dynamic range where even low inputs can lead to non-zero activity. Moreover, the shifted activation function requires a biologically implausible strength of input for activation. The firing threshold of biological neurons is typically $15 - 20$mV above the resting membrane potential, which is much less than the shifted $\mu_{\min}$. Note that similarly to Fig. 8, here we plot only the fixed points for $\nu_{\mathrm{S}} > 0$ and $\nu_{\mathrm{NS}} = 0$.

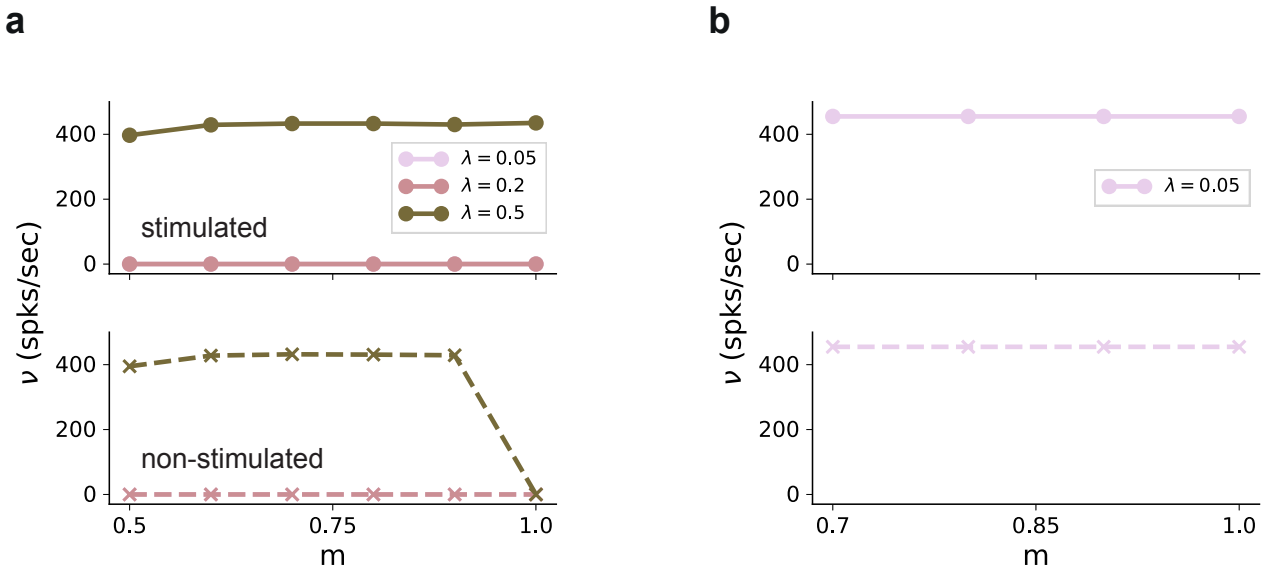

Figure 8 - supplement 2: Firing rates in SSN_5 in the absence of external background noise ($\nu_{\mathrm{X}} = 0$)**.** **(a)** Firing rates of stimulated (top) and non-stimulated (bottom) populations obtained from simulations of a network without any recurrent connections, for different input intensities λ. Successful denoising only occurs for the extreme case of $m = 1$, in which case the pathways are completely segregated. Note that the input rate was unchanged, $\nu_{\mathrm{in}} = 12\lambda K_{\mathrm{E}}$. **(b)** Same as **(a)**, but for excitatory recurrence ($g = -3$). Recurrent excitation spreads the input from the stimulated pathway to non-stimulated neurons. Results shown only for $\lambda = 0.05$, with larger values leading to similar results.

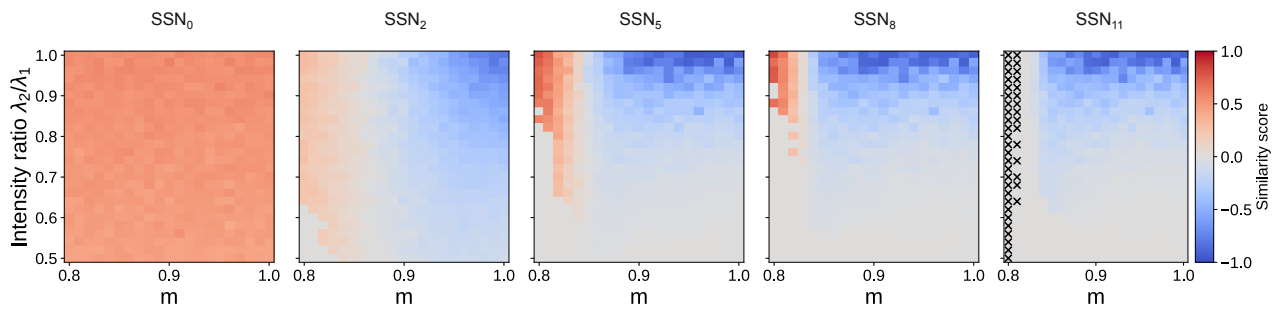

Figure 9 - supplement 1: Evolution of similarity score for 12 sub-networks. Correlation-based similarity score illustrates the three dynamical regimes observed across the different sub-networks, for two input streams: coexisting (CoEx, red area and positive score), winner-take-all behavior (grey, score near 0) and winnerless competition (WLC, blue and negative score). As predicted by the mean-field analysis, the CoEx region vanishes with increasing network depth. The calculation of the similarity score is detailed in Methods. If either stimuli could not be decoded, we set the score to 0. In SSN_{11}, 'X' indicates parameter combinations where none of the stimuli could be decoded.

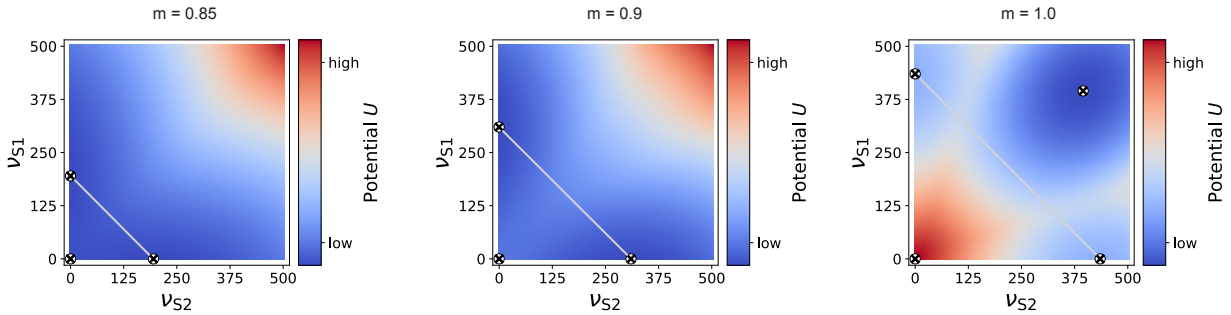

Figure 9 - supplement 2: Potential landscape for two input streams. For intermediate modularity ($m = 0.85$, left and $m = 0.9$, right), there are two high-activity fixed points (circled cross markers) in addition to the low-activity one near zero (marker added manually here, as it is not observable due to the larger integration step of 5 spks/sec used here). If the projections are almost fully modular ($m \approx 1$), an additional high-activity fixed point can be observed for identical ν^{S1} and ν^{S2}. In this case, the two stimulated sub-populations can be considered as one larger population, for which the common $\kappa_{\mathrm{S,S}}$ becomes positive, as in the case of a single input stream (see Figure 7 - figure supplement 1), just for larger m. Grey, anti-diagonal lines represent the one-dimensional sections illustrated in Fig. 9e.

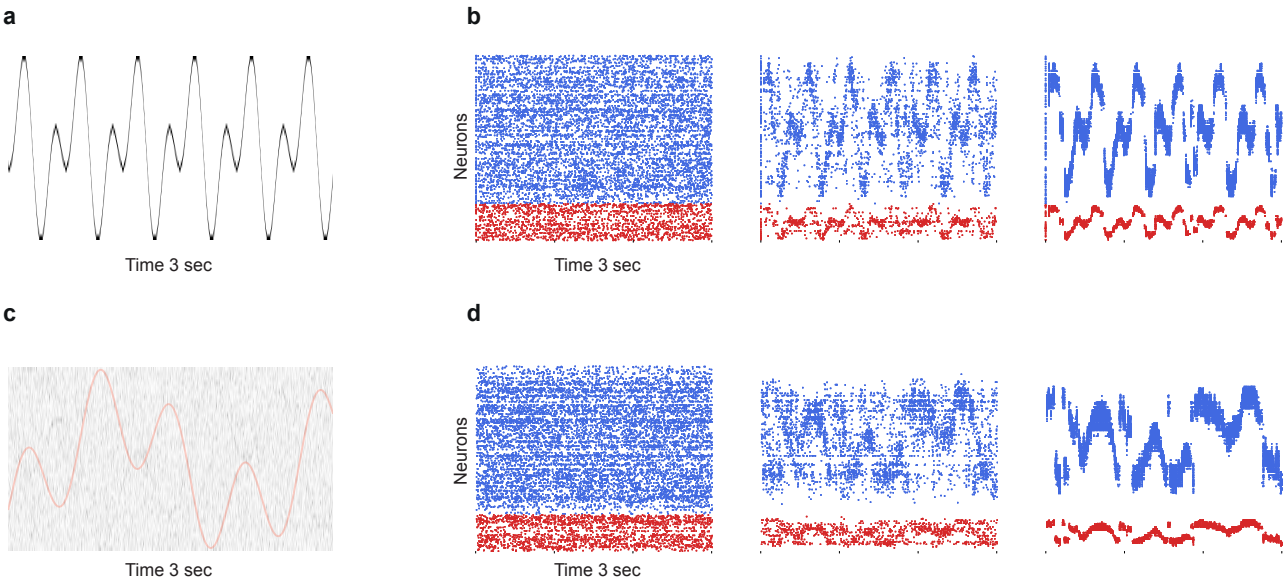

Figure 10 - supplement 1: Limits of denoising for rapidly changing and noisy dynamical inputs. (a) A fast changing input signal $x(t) = \sin(24t) + \cos(12t)$ sampled at $dt = 1$ms, with no additional noise ($\sigma_\xi = 0$). **(b)** While portions of the signal can be successfully transmitted and denoised in SSN_5, there are significant periods (steep slopes) where the signal representation is lost. **(c)** Slower signal $x(t) = \sin(10t) + \cos(3t)$ with significant noise corruption ($\sigma_\xi = 2\nu_{\mathrm{in}}$). Continuous red curve denotes the input signal $u(t)$. **(d)** Strong noise in the input leads to heavy fluctuations in the activity of the deeper populations, corrupting the signal representations.

Supplementary Tables

A: Model Summary	
Populations	Multiple modules, each one composed of 1 excitatory and 1 inhibitory sub-population
Topology	None
Connectivity	Sparse, random recurrent connectivity with random or topographically structured feed-forward projections (fixed in-degrees)
Neuron Model	Leaky integrate-and-fire, fixed voltage threshold, fixed absolute refractory time, no adaptation
Synapse Model	Exponentially decaying postsynaptic currents, static synaptic weights, fixed delays
Plasticity	None
Input	Stochastic background spikes and inhomogeneous Poisson spikes onto $d_0 N^{\mathrm{E}}$ excitatory and $d_0 N^{\mathrm{E}}$ inhibitory neurons in SSN_0
Measurements	Spiking activity, membrane potentials

B: Populations		
Name	**Elements**	**Size**
E_i	`iaf_psc_exp`	8000
I_i	`iaf_psc_exp`	2000

C: Neuron Models	
Name	Leaky integrate-and-fire (LIF) neuron (`iaf_psc_exp`)
Subthreshold Dynamics	if $(t > t^* + \tau_{\mathrm{ref}})$ $\tau_{\mathrm{m}} \frac{dV_i(t)}{dt} = (V_{\mathrm{rest}} - V_i(t)) + R_{\mathrm{m}} \left(I_i^{\mathrm{E}}(t) + I_i^{\mathrm{I}}(t) + I_i^{\mathrm{X}}(t)\right)$ else $V(t) = V_{\mathrm{reset}}$
Spiking	If $V(t-) < V_{\mathrm{th}}$ OR $V(t+) \geq V_{\mathrm{th}}$ 1. set $t^* = t$ 2. emit spike with time stamp t^*

D: Synapse Models	
Synaptic Transmission	$\tau_\beta \frac{dI_i^\beta(t)}{dt} = -I_i(t) + \tau_\beta \hat{I}_\beta \sum_j \sum_k \delta(t - t_j^k)$ with postsynaptic potential $\mathrm{PSP}_{ij}(t) = \hat{I}_\beta \frac{R_{\mathrm{m}}\tau_\beta}{\tau_\beta - \tau_{\mathrm{m}}} \left(e^{-t/\tau_\beta} - e^{-t/\tau_{\mathrm{m}}}\right) \Theta(t)$ and Heaviside function $\Theta(t) = \begin{cases} 1 & t \geq 0 \\ 0 & \text{else} \end{cases}$. The synaptic efficacy (weight) corresponds to the PSP amplitude: $J_\beta = \hat{I}_\beta R_{\mathrm{m}} \frac{\tau_\beta}{\tau_\beta - \tau_{\mathrm{m}}} \left(\left[\frac{\tau_{\mathrm{m}}}{\tau_\beta}\right]^{\frac{-\tau_m}{\tau_m - \tau_\beta}} - \left[\frac{\tau_{\mathrm{m}}}{\tau_\beta}\right]^{\frac{-\tau_\beta}{\tau_{\mathrm{m}} - \tau_\beta}} \right)$

E: Input		
Type	**Target**	**Description**
poisson_generator	E_0, I_0	Total rate $\nu_{\mathrm{X}} \cdot \mathrm{K}_{\mathrm{X}}$
poisson_generator	E_i, I_i for $i > 0$	Total rate $0.25 \cdot \nu_{\mathrm{X}} \cdot \mathrm{K}_{\mathrm{X}}$
inhomogeneous_poisson_generator	$\mathrm{E}_0^{(k)}$, $\mathrm{I}_0^{(k)}$ for $S_k \in S$ $\mathrm{E}_0^{'(j)}$, $\mathrm{I}_0^{'(j)}$ for $S'_j \in S'$	Inhomogeneous Poisson process with rate ν_{stim}, changing every 200 ms

F: Measurements
Spiking activity, membrane potentials

Table 1: Tabular description of current-based (baseline) network model after [94].

A: Populations		
Name	**Value**	**Description**
N^{E}	8000	Excitatory population size in each module
N^{I}	2000	Inhibitory population size in each module
B: Connectivity		
Name	**Value**	**Description**
ϵ	0.1	Baseline connection probability
α	0.25	Connection scaling factor for $\mathrm{SSN}_{i>0}$
p_{x}	ϵ	Connection probability for background noise input in SSN_0
	$\alpha\epsilon$	Scaled connection probability for background input in $\mathrm{SSN}_i, i > 0$
σ_{i}	$(1-\alpha) * \epsilon$	Fixed density of feed-forward projection matrices
p_{c}	$(1-m) * p_0$	Feed-forward connection probability within topographic maps
p_0	$(1-m) * p_{\mathrm{c}}$	Feed-forward connection probability between SPs on different topographic maps
B: Neuron Model		
Name	**Value**	**Description**
C_{m}	250 pF	Membrane capacitance
E_L	-70 mV	Resting membrane potential
τ_{m}	20 ms	Membrane time constant
V_{th}	-55 mV	Membrane potential threshold for action-potential firing
V_{reset}	-60 mV	Reset potential
τ_{ref}	2 ms	Absolute refractory period
C: Synapse Model		
τ_{E}	2 ms	Synaptic decay time constant for excitatory synapses
τ_{I}	2 ms	Synaptic decay time constant for inhibitory synapses
d	1.5 ms	Synaptic transmission delay
$\hat{I}_{\mathrm{E}}$	32.78 pA	Peak excitatory current
$\hat{I}_{\mathrm{I}}$	g32.78 pA	Peak inhibitory current
J_{E}	0.2 mV	EPSP amplitude
J_{I}	g0.2 mV	IPSP amplitude
g	-12	Scaling factor for the inhibitory synapses

Table 2: Summary of all the model parameters for the current-based network. For more details, see [15].

A: Populations		
Name	**Value**	**Description**
N^{E}	8000	Excitatory population size in each module
N^{I}	2000	Excitatory population size in each module
B: Connectivity		
Name	**Value**	**Description**
d	1.5 ms	Synaptic transmission delay
$\overline{g}_{\mathrm{E}}$	1nS	Excitatory synaptic conductance
$\overline{g}_{\mathrm{I}}$	$g\overline{g}_{\mathrm{E}}$nS	Inhibitory synaptic conductance
g	-16	Scaling factor for the inhibitory synapses
ϵ	0.1	Baseline connection probability
α	0.25	Connection scaling factor for $\mathrm{SSN}_{i>0}$
p_{x}	ϵ	Connection probability for background noise input in SSN_0
	$\alpha\epsilon$	Scaled connection probability for background input in $\mathrm{SSN}_i, i > 0$
σ_{i}	$(1-\alpha) * \epsilon$	Fixed density of feed-forward projection matrices
p_{c}	$(1-m) * p_0$	Feed-forward connection probability within topographic maps
p_0	$(1-m) * p_{\mathrm{c}}$	Feed-forward connection probability between SPs on different topographic maps
B: Neuron Model		
Name	**Value**	**Description**
C_{m}	250 pF	Membrane capacitance
E_L	-70 mV	Resting membrane potential
τ_{m}	15 ms	Membrane time constant
V_{th}	-50 mV	Membrane potential threshold for action-potential firing
V_{reset}	-60 mV	Reset potential
τ_{ref}	2 ms	Absolute refractory period
g_{L}	16.7nS	Leak conductance
C: Synapse Model		
τ_{E}	5 ms	Synaptic decay time constant for excitatory synapses
τ_{I}	10 ms	Synaptic decay time constant for inhibitory synapses
V_{E}	0 mV	Excitatory reversal potential
V_{I}	-80 mV	Inhibitory reversal potential

Table 3: Parameter values for the conductance-based model.

A: Model Summary	
Populations	Multiple modules, each one composed of 1 excitatory and 1 inhibitory sub-population
Topology	None
Connectivity	Sparse, random recurrent connectivity with modular topographic feed-forward projections (fixed in-degrees)
Neuron Model	Rate neuron with shifted tanh gain function
Synapse Model	Delayed rate connection
Plasticity	None
Input	Uniformly distributed input onto $d_0 N^{\mathrm{E}}$ excitatory and $d_0 N^{\mathrm{E}}$ inhibitory neurons in SSN_0
Measurements	Unit output (rate)

B: Populations		
Name	**Elements**	**Size**
E_i	rate neuron	2400
I_i	rate neuron	600

C: Neuron Models	
Name	Rate neuron
Differential equation	$\boldsymbol{\tau_x} \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{dt}} = -\boldsymbol{x} + W\boldsymbol{r} + W^{\mathrm{in}}\boldsymbol{u} - \boldsymbol{b}^{\mathrm{rec}} + \sqrt{2\boldsymbol{\tau_x}}\sigma_{\mathrm{X}}\xi$ $\boldsymbol{r} = 0.5(1 + \tanh(\boldsymbol{x}))$

D: Input		
Type	**Target**	**Description**
random uniform distribution	E_0, I_0	Step signal input to SSN_0, changing every 200 ms, with amplitude 0.8
Gaussian white noise	E_i, I_i for $i \in \{0..5\}$	Intrinsic unit noise

E: Measurements
Unit output (rate)

Table 4: Description of rate model.

A: Populations		
Name	**Value**	**Description**
N^{E}	2400	Number of excitatory units in each module
N^{I}	600	Number of inhibitory units in each module

B: Connectivity		
Name	**Value**	**Description**
d	1 ms	Synaptic transmission delay
ϵ	0.2	Baseline connection probability
w_{in}	$\sim U(0.9, 1.0)$	Input weights
w_{in}	$\sim N_{\mathrm{tr}}(0, 1/\sqrt{\epsilon N}) > 0$	Recurrent and feed-forward weights drawn from a normal distribution truncated to positive values
g	-6	Scaling factor for the inhibitory synapses

B: Neuron Model		
Name	**Value**	**Description**
τ	10 ms	Unit time constant
b	1	Bias term
σ_{X}	1.5	Scaling term for unit noise

Table 5: Rate model parameters.

Appendix A: Constraints on feedforward connectivity

This section expands on the limitations arising from the definitions of topographic modularity and map sizes used in this study. By imposing a fixed connection density on the feedforward connection matrices, the projection probabilities between neurons tuned to the same (p_c) and different (p_0) stimuli are uniquely determined by the modularity m and the parameter d_0 and δ, which control the size of stimulus-specific sub-populations (see Methods). For notational simplicity, here we consider the merged excitatory and inhibitory sub-populations tuned to a particular stimulus in a given sub-network SSN_i, with a total size $C_\mathrm{i} = C_\mathrm{i}^\mathrm{E} + C_\mathrm{i}^\mathrm{I}$.

Under the constraints applied in this work, the total density of a feedforward adjacency matrix between $\mathrm{SSN_i}$ and $\mathrm{SSN_{i+1}}$ can be computed as:

$$\sigma_\mathrm{i} = \frac{p_\mathrm{c} U_\mathrm{c}^\mathrm{i} + p_0 U_0^\mathrm{i}}{N^2} \tag{12}$$

where U_0^i and U_c^i are the number of *realizable* connections between similarly and differently tuned sub-populations, respectively. Since $U_\mathrm{c}^\mathrm{i} = N^2 - U_0^\mathrm{i}$, we can simplify the notation and focus only on U_0^i. We distinguish between the cases of non-overlapping and overlapping stimulus-specific sub-populations:

$$U_0^\mathrm{i} = \begin{cases} N^2 - N_\mathrm{C} C_\mathrm{i} C_\mathrm{i+1} & \text{if } d_\mathrm{i} < \frac{1}{N_\mathrm{C}} \\ \frac{N_\mathrm{C}}{N_\mathrm{C}-1}(N - C_\mathrm{i})(N - C_\mathrm{i+1}) & \text{if } d_\mathrm{i} \geq \frac{1}{N_\mathrm{C}} \end{cases},$$

where each potential synapse is counted only once, regardless of whether the involved neurons belong to any or multiple overlapping sub-populations. This ensures consistency with the definitions of the probabilities p_c and p_0. Alternatively, we can express U_0^i as:

$$U_0^\mathrm{i} = \frac{N^2 N_\mathrm{stim}}{N_\mathrm{stim} - 1}(1 - i\delta - d_0)(1 - (i-1)\delta - d_0)$$

For the case with no overlap, we can derive an additional constraint on the minimum sub-populations size C_i for the required density σ_i to be satisfied, which we define in relation to the total number of sub-populations N_C:

$$d_\mathrm{i} \geq \sqrt{\frac{\sigma_\mathrm{i}}{N_\mathrm{C}}} \tag{13}$$

The equality holds in the case of $m = 1$ and all-to-all feedforward connectivity between similarly tuned sub-populations, i.e., $p_\mathrm{c} = 1$.

Appendix B: Mean-field analysis of network dynamics

For an analytical investigation of the role of topographic modularity on the network dynamics, we used mean field theory [39, 40, 41]. Under the assumptions that each neuron receives a large number of small amplitude inputs at every time step, the synaptic time constants τ_s are small compared to the membrane time constant τ_m, and that the network activity is sufficiently asynchronous and irregular, we can make use of theoretical results obtained from the diffusion approximation of the LIF neuron model to determine the stationary population dynamics. The equations in this section were partially solved using a modified version of the LIF Meanfield Tools library [95].

Stationary firing rates and fixed points

In the circumstances described above, the total synaptic input to each neuron can be replaced by a Gaussian white noise process (independent across neurons) with mean $\mu(t)$ and variance $\sigma^2(t)$. In the stationary state, these quantities, along with the firing rates of each afferent, can be well approximated by their constant time average. The stationary firing rate of the LIF neuron in response to such input is:

$$\nu = \left(\tau_\mathrm{ref} + \sqrt{\pi}\tau_\mathrm{eff}\int_{y_\mathrm{r}}^{y_\theta} \exp(u^2)\left[1+\mathrm{erf}(u)\right]du\right)^{-1} \tag{14}$$

where erf is the error function and the integration limits are defined as $y_\mathrm{r} = (V_\mathrm{reset} - \mu)/\sigma + \frac{q}{2}\sqrt{\tau_\mathrm{s}/\tau_\mathrm{eff}}$ and $y_\theta = (\theta - \mu)/\sigma + \frac{q}{2}\sqrt{\tau_\mathrm{s}/\tau_\mathrm{eff}}$, with $q = \sqrt{2}|\zeta(1/2)|$ and Riemann zeta function ζ (see [96], eq. 4.33). As we will see below, the mean μ and variance σ^2 of the input also depend on the stationary firing rate ν, rendering Eq. 14 an implicit equation that needs to be solved self-consistently using fixed-point iteration.

For simplicity, throughout the mean-field analyses we consider perfectly partitioned networks where each neuron belongs to exactly one topographic map, that is, to one of the N_C stimulus-specific, identically sized sub-populations *SP* (no overlap condition). We denote the firing rate of a neuron in the currently stimulated SP (receiving stimulus input in SSN_0) in sub-network SSN_i by $\nu_\mathrm{i}^\mathrm{S}$, and by $\nu_\mathrm{i}^\mathrm{NS}$ that of neurons not associated with the stimulated pathway. Since the firing rates of excitatory and inhibitory neurons are equal (due to identical synaptic time constants and input statistics), we can write the constant mean synaptic input to neurons in the input sub-network as

$$\begin{aligned}
\mu_0^\mathrm{S} &= \left(\overbrace{K_\mathrm{X}J_\mathrm{X}\nu_\mathrm{X}}^{\text{noise}} + \overbrace{(\frac{1}{N_\mathrm{C}}K_\mathrm{E}J_\mathrm{E} + \frac{1}{N_\mathrm{C}}K_\mathrm{I}J_\mathrm{I})\nu_0^\mathrm{S}}^{\text{rec. stimulated}} + \overbrace{(N_\mathrm{C}-1)(\frac{1}{N_\mathrm{C}}K_\mathrm{E}J_\mathrm{E} + \frac{1}{N_\mathrm{C}}K_\mathrm{I}J_\mathrm{I})\nu_0^\mathrm{NS}}^{\text{rec. non-stimulated}} + \overbrace{J_\mathrm{X}\nu_\mathrm{in}}^{\text{stimulus}}\right)\tau_\mathrm{m} \\
\mu_0^\mathrm{NS} &= \left(\overbrace{K_\mathrm{X}J_\mathrm{X}\nu_\mathrm{X}}^{\text{noise}} + \overbrace{(\frac{1}{N_\mathrm{C}}K_\mathrm{E}J_\mathrm{E} + \frac{1}{N_\mathrm{C}}K_\mathrm{I}J_\mathrm{I})\nu_0^\mathrm{S}}^{\text{rec. stimulated}} + \overbrace{(N_\mathrm{C}-1)(\frac{1}{N_\mathrm{C}}K_\mathrm{E}J_\mathrm{E} + \frac{1}{N_\mathrm{C}}K_\mathrm{I}J_\mathrm{I})\nu_0^\mathrm{NS}}^{\text{rec. non-stimulated}}\right)\tau_\mathrm{m},
\end{aligned} \tag{15}$$

The variances $(\sigma_0^\mathrm{S})^2$ and $(\sigma_0^\mathrm{NS})^2$ can be obtained by squaring each weight J in the above equation. To derive these equations for the deeper sub-networks $\mathrm{SSN}_{\mathrm{i}>0}$, it is helpful to include auxiliary variables K_S and K_NS, representing the number of feed-forward inputs to a neuron in SSN_i from its own SP in $\mathrm{SSN}_{\mathrm{i}-1}$, and from one different SP (there are $N_\mathrm{C}-1$ such sub-populations), respectively. Both K_S and K_NS are uniquely defined by the modularity m and projection density d, and $K_\mathrm{NS} = (1-m)K_\mathrm{S} = (1-m)(1-\alpha)K_\mathrm{E}$ holds as well. The mean synaptic inputs to the neurons in the deeper sub-networks can thus be written as:

$$
\begin{aligned}
\mu_{\mathrm{i}}^{\mathrm{S}} = \Bigg(& \overbrace{\alpha K_{\mathrm{X}} J_{\mathrm{X}} \nu_{\mathrm{X}}}^{\text{noise}} + \overbrace{(\frac{1}{N_{\mathrm{C}}} K_{\mathrm{E}} J_{\mathrm{E}} + \frac{1}{N_{\mathrm{C}}} K_{\mathrm{I}} J_{\mathrm{I}}) \nu_{\mathrm{i}}^{\mathrm{S}}}^{\text{rec. stimulated}} \\
& + \overbrace{(N_{\mathrm{C}} - 1)(\frac{1}{N_{\mathrm{C}}} K_{\mathrm{E}} J_{\mathrm{E}} + \frac{1}{N_{\mathrm{C}}} K_{\mathrm{I}} J_{\mathrm{I}}) \nu_{\mathrm{i}}^{\mathrm{NS}}}^{\text{rec. non-stimulated}} \\
& + \overbrace{K_{\mathrm{S}} J_{\mathrm{E}} \nu_{\mathrm{i-1}}^{\mathrm{S}}}^{\text{stimulated FF}} + \overbrace{(N_{\mathrm{C}} - 1) K_{\mathrm{NS}} J_{\mathrm{E}} \nu_{\mathrm{i-1}}^{\mathrm{NS}}}^{\text{non-stimulated FF}} \Bigg) \tau_{\mathrm{m}} \\
\mu_{\mathrm{i}}^{\mathrm{NS}} = \Bigg(& \overbrace{\alpha K_{\mathrm{X}} J_{\mathrm{X}} \nu_{\mathrm{X}}}^{\text{noise}} + \overbrace{(\frac{1}{N_{\mathrm{C}}} K_{\mathrm{E}} J_{\mathrm{E}} + \frac{1}{N_{\mathrm{C}}} K_{\mathrm{I}} J_{\mathrm{I}}) \nu_{\mathrm{i}}^{\mathrm{S}}}^{\text{rec. stimulated}} \\
& + \overbrace{(N_{\mathrm{C}} - 1)(\frac{1}{N_{\mathrm{C}}} K_{\mathrm{E}} J_{\mathrm{E}} + \frac{1}{N_{\mathrm{C}}} K_{\mathrm{I}} J_{\mathrm{I}}) \nu_{\mathrm{i}}^{\mathrm{NS}}}^{\text{rec. non-stimulated}} \\
& + K_{\mathrm{NS}} J_{\mathrm{E}} \nu_{1}^{\mathrm{S}} + ((N_{\mathrm{C}} - 2) K_{\mathrm{NS}} + K_{\mathrm{S}}) J_{\mathrm{E}} \nu_{\mathrm{i-1}}^{\mathrm{NS}} \Big) \tau_{\mathrm{m}}
\end{aligned}
\tag{16}
$$

Again, one can obtain the variances by squaring each weight J. The stationary firing rates for the stimulated and non-stimulated sub-populations in all sub-networks are then found by first solving Eq. 15 and Eq. 14 for the first sub-network and then Eq. 16 and Eq. 14 successively for deeper sub-networks.

For very deep networks, one can ask the question, whether firing rates approach fixed points across sub-networks. If there are multiple fixed points, the initial condition, that is the externally stimulated activity of sub-populations in the first sub-network, decides in which of the fixed points the rates evolve, in a similar spirit as in recurrent networks after a start-up transient. For a fixed point, we have $\nu_{i-1} = \nu_i$. In effect, we can re-group terms in Eq. 16 that have the same rates such that formally we obtain an effective new group of neurons from the excitatory and inhibitory SPs of the current sub-network and the corresponding excitatory SPs of the previous sub-network, as indicated by the square brackets in the following formulas:

$$
\begin{aligned}
\mu^{\mathrm{S}} = \alpha\beta\mathcal{J}\nu_x + & \underbrace{\mathcal{J}\left[\frac{1}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{1}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{S,S}}} \nu^{\mathrm{S}} \\
& + \underbrace{\mathcal{J}\left[\frac{N_{\mathrm{C}}-1}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{(N_{\mathrm{C}}-1)(1-m)}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{S,NS}}} \nu^{\mathrm{NS}}
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
\mu^{\mathrm{NS}} = \alpha\beta\mathcal{J}\nu_x + & \underbrace{\mathcal{J}\left[\frac{1}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{(1-m)}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{NS,S}}} \nu^{\mathrm{S}} \\
& + \underbrace{\mathcal{J}\left[\frac{N_{\mathrm{C}}-1}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{1+(N_{\mathrm{C}}-2)(1-m)}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{NS,NS}}} \nu^{\mathrm{NS}}
\end{aligned}
\tag{18}
$$

with $\beta = K_{\mathrm{X}}/K_{\mathrm{E}}$, $\gamma = K_{\mathrm{I}}/K_{\mathrm{E}}$ and $\mathcal{J} = \tau K_{\mathrm{E}} J$.

For the parameters g and γ chosen here, $\kappa_{\mathrm{S,NS}}$, $\kappa_{\mathrm{NS,S}}$ and $\kappa_{\mathrm{NS,NS}}$ in Eq. 17 and Eq. 18 are always negative for any modularity m due to the large recurrent inhibition. Therefore, for the non-stimulated group, $\kappa < 0$ in Eq. 8 (see main text), such that one always finds a single fixed point, which, as desired, is at a low rate. Interestingly, the excitatory feed-forward connections can switch the sign of $\kappa_{\mathrm{S,S}}$ from negative to positive for large values of m, thereby rendering the active group effectively excitatory, leading to a saddle-node bifurcation and the emergence of a stable high-activity fixed point (see Fig. 7b in the main text).

The structure of fixed points can also be understood by studying the potential landscape of the system: Eq. 14 can be regarded as the fixed-point solution of the following evolution equations for the stimulated and non-stimulated sub-populations [88, 89]

$$\tau_{\mathrm{S}} \frac{d\nu^{\mathrm{S}}}{dt} = -\nu^{\mathrm{S}} + \Phi_{\mathrm{S}}(\nu^{\mathrm{S}}, \nu^{\mathrm{NS}}) \,, \tag{19}$$

$$\tau_{\mathrm{NS}} \frac{d\nu^{\mathrm{NS}}}{dt} = -\nu^{\mathrm{NS}} + \Phi_{\mathrm{NS}}(\nu^{\mathrm{S}}, \nu^{\mathrm{NS}}) \,, \tag{20}$$

where Φ_{S} and Φ_{NS} are defined via the right-hand side of Eq. 14 with μ^{S} and μ^{NS} inserted as defined in Eq. 17 and Eq. 18 (and likewise for σ^{S} and σ^{NS}). Due to the asymmetry in connections between stimulated and non-stimulated sub-populations, the right-hand side of Eq. 19 and Eq. 20 cannot be interpreted as a conservative force. Following the idea of effective response functions [97], a potential $U(\nu^{\mathrm{S}})$ for the stimulated sub-population alone can, however, be defined by inserting the solution $\nu^{\mathrm{NS}} = f(\nu^{\mathrm{S}})$ of Eq. 20 into Eq. 19

$$\tau_{\mathrm{S}} \frac{d\nu^{\mathrm{S}}}{dt} = -\nu^{\mathrm{S}} + \Phi_{\mathrm{S}}(\nu^{\mathrm{S}}, f(\nu^{\mathrm{S}})) \tag{21}$$

and interpreting the right-hand side as a conservative force $F = -\frac{dU}{d\nu^{\mathrm{S}}}$ [55]. The potential then follows from integration as

$$U(\nu^{\mathrm{S}}) - U(0) = \frac{1}{2}(\nu^{\mathrm{S}})^2 - \int_0^{\nu^{\mathrm{S}}} \Phi_{\mathrm{S}}(\nu, f(\nu)) d\nu \,, \tag{22}$$

where $U(0)$ is an inconsequential constant. We solved the latter integral numerically using the `scipy.integrate.trapz` function of `SciPy` [87]. The minima and maxima of the resulting potential correspond to locally stable and unstable fixed points, respectively. Note that while this single-population potential is useful to study the structure of fixed points, the full dynamics of all populations and global stability cannot be straight-forwardly infered from this reduced picture [97, 24], here for two reasons : 1. For spiking networks, Eq. 19 and Eq. 20 do not describe the real dynamics of the mean activity. Their right hand side only defines the stationary state solution. 2. The global stability of fixed points also depends on the time constants of all sub-populations' mean activities (here τ_{S} and τ_{NS}), but the temporal dynamics of the non-stimulated sub-populations is neglected here.

Mean-field analysis for two input streams

In the case of two simultaneously active stimuli (see Section "Input integration and multi-stability"), if the stimulated group 1 is in the high-activity state with rate ν^{S1}, the second stimulated group 2 will receive an additional non-vanishing input of the form

$$\left[\frac{1}{N_{\mathrm{C}}}(1 + \gamma g) + (1 - \alpha) \frac{(1 - m)}{(N_{\mathrm{C}} - 1)(1 - m) + 1} \right] \nu^{\mathrm{S1}} < 0, \tag{23}$$

which is negative for all values of m and can therefore lead to the silencing of group 2. If the stimuli are similarly strong, network fluctuations can dynamically switch the roles of the stimulated groups 1 and 2.

The dynamics and fixed-point structure in deep sub-networks can be studied using a two-dimensional potential landscape that is defined via the following evolution equations

$$\frac{d\nu^{\mathrm{S1}}}{dt} = -\nu^{\mathrm{S1}} + \Phi_{\mathrm{S1}}(\nu^{\mathrm{S1}}, \nu^{\mathrm{S2}}, f(\nu^{\mathrm{S1}}, \nu^{\mathrm{S2}})) \,, \tag{24}$$

$$\frac{d\nu^{\mathrm{S2}}}{dt} = -\nu^{\mathrm{S2}} + \Phi_{\mathrm{S2}}(\nu^{\mathrm{S1}}, \nu^{\mathrm{S2}}, f(\nu^{\mathrm{S1}}, \nu^{\mathrm{S2}})) \,, \tag{25}$$

where $f(\nu^{\mathrm{S1}}, \nu^{\mathrm{S2}}) = \nu^{\mathrm{NS}}$ is the fixed-point of the non-stimulated sub-populations for given rates $\nu^{\mathrm{S1}}, \nu^{\mathrm{S2}}$ of the two stimulated sub-populations, respectively. The functions Φ_{S1} and Φ_{S2} are again defined via the right-hand side of Eq. 14 with inserted μ^{S1}, μ^{S2} and μ^{NS} that are defined as follows (derivation analogous to the single-input case):

$$\mu^{\mathrm{S1}} = \alpha\mathcal{J}\nu_x + \underbrace{\mathcal{J}\left[\frac{1}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{1}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{S1,S1}}}\nu^{\mathrm{S1}} \tag{26}$$

$$+\underbrace{\mathcal{J}\left[\frac{1}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{1-m}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{S1,S2}}}\nu^{\mathrm{S2}}$$

$$+\underbrace{\mathcal{J}\left[\frac{N_{\mathrm{C}}-2}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{(N_{\mathrm{C}}-2)(1-m)}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{S1,NS}}}\nu^{\mathrm{NS}}$$

$$\mu^{\mathrm{S2}} = \alpha\mathcal{J}\nu_x + \underbrace{\mathcal{J}\left[\frac{1}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{1-m}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{S2,S1}}}\nu^{\mathrm{S1}} \tag{27}$$

$$+\underbrace{\mathcal{J}\left[\frac{1}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{1}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{S2,S2}}}\nu^{\mathrm{S2}}$$

$$+\underbrace{\mathcal{J}\left[\frac{N_{\mathrm{C}}-2}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{(N_{\mathrm{C}}-2)(1-m)}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{S1,NS}}}\nu^{\mathrm{NS}}$$

$$\mu^{\mathrm{NS}} = \alpha\mathcal{J}\nu_x + \underbrace{\mathcal{J}\left[\frac{1}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{(1-m)}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{NS,S1}}}\nu^{\mathrm{S1}} \tag{28}$$

$$+\underbrace{\mathcal{J}\left[\frac{1}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{(1-m)}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{NS,S2}}}\nu^{\mathrm{S2}} \tag{29}$$

$$+\underbrace{\mathcal{J}\left[\frac{N_{\mathrm{C}}-2}{N_{\mathrm{C}}}(1+\gamma g) + (1-\alpha)\frac{1+(N_{\mathrm{C}}-3)(1-m)}{(N_{\mathrm{C}}-1)(1-m)+1}\right]}_{\kappa_{\mathrm{NS,NS}}}\nu^{\mathrm{NS}}$$

Due to the symmetry between the two stimulated sub-populations, the right-hand side of Eq. 24 and Eq. 25 can be viewed as a conservative force $\boldsymbol{F}$ of the potential $U(\nu^{\mathrm{S1}}, \nu^{\mathrm{S2}}) = -\int_{\mathcal{C}} \boldsymbol{F}\, ds$, where we parameterized the line integral along the path $\nu : [0,1] \to \mathcal{C}, t \mapsto t \cdot (\nu^{\mathrm{S1}}, \nu^{\mathrm{S2}})$, which yields

$$U(\nu^{\mathrm{S1}}, \nu^{\mathrm{S2}}) = \frac{1}{2}(\nu^{\mathrm{S1}})^2 + \frac{1}{2}(\nu^{\mathrm{S2}})^2 - \int_0^{\nu^{\mathrm{S1}}} \Phi_{\mathrm{S1}}\left(\nu, \nu\frac{\nu^{\mathrm{S2}}}{\nu^{\mathrm{S1}}}, f(\nu, \nu\frac{\nu^{\mathrm{S2}}}{\nu^{\mathrm{S1}}})\right) - \int_0^{\nu^{\mathrm{S2}}} \Phi_{\mathrm{S2}}\left(\nu\frac{\nu^{\mathrm{S1}}}{\nu^{\mathrm{S2}}}, \nu, f(\nu\frac{\nu^{\mathrm{S1}}}{\nu^{\mathrm{S2}}}, \nu)\right). \tag{30}$$

The numerical evaluation of this two-dimensional potential is shown in Figure S9 - figure supplement 2, whereas sketches in Fig. 9e show a one-dimensional section (grey lines in Figure S9 - figure supplement 2) that goes anti-diagonal through the two minima corresponding to one population being in the high-activity state and the other one being in the low-activity state.

Critical modularity for piecewise linear activation function

To obtain a closed-form analytic solution for the critical modularity, in the following we consider a neuron model with piecewise linear activation function

$$\nu(\mu) = \nu_{\max}\frac{\mu - \mu_{\min}}{\mu_{\max} - \mu_{\min}} \tag{31}$$

for $\mu \in [\mu_{\min}, \mu_{\max}]$, $\nu(\mu) = 0$ for $\mu < \mu_{\min}$ and $\nu(\mu) = \nu_{\max}$ for $\mu > \mu_{\max}$ (Fig. 8a). Successful denoising requires the non-stimulated sub-populations to be silent, $\nu^{NS} = 0$, and the stimulated sub-populations to be active, $\nu^{S} > 0$. We first study solutions where $0 < \nu^{S} < \nu_{\max}$ and afterwards the case where $\nu^{S} = \nu_{\max}$. Inserting Eq. 31 into Eq. 9 and Eq. 10, we obtain

$$\mu^{\mathrm{S}} = \alpha \mathcal{J} \nu_x + \kappa_{\mathrm{S,S}}(m)\, \nu_{\max} \frac{\mu_S - \mu_{\min}}{\mu_{\max} - \mu_{\min}} \,,$$
$$\mu^{\mathrm{NS}} = \alpha \mathcal{J} \nu_x + \kappa_{\mathrm{NS,S}}(m)\, \nu_{\max} \frac{\mu_S - \mu_{\min}}{\mu_{\max} - \mu_{\min}} \,.$$

The first equation can be solved for μ^{S}

$$\frac{\mu^{\mathrm{S}}}{\mu_{\min}} = 1 + \frac{\alpha \mathcal{J} \nu_x - \mu_{\min}}{\mu_{\min} - \kappa_{\mathrm{S,S}}(m)\, \nu_{\max} \frac{\mu_{\min}}{\mu_{\max} - \mu_{\min}}} \,, \tag{32}$$

which holds for

$$\mu_{\min} \leq \mu^{\mathrm{S}} \leq \mu_{\max} \,, \tag{33}$$
$$\mu^{\mathrm{NS}} \leq \mu_{\min} \,. \tag{34}$$

Requirement (Eq. 33) is equivalent to an inequality for m

$$0 \leq \frac{\alpha \mathcal{J} \nu_x - \mu_{\min}}{\mu_{\max} - \frac{\mathcal{J}}{N_{\mathrm{C}}} (1 + \gamma g)\, \nu_{\max} - \frac{(1-\alpha)\mathcal{J}\nu_{\max}}{(N_{\mathrm{C}}-1)(1-m)+1} - \mu_{\min}} \leq 1$$

that, depending on the dynamic range of the neuron, the strength of the external background input and the recurrence, yields

$$m = \frac{N_{\mathrm{C}}}{N_{\mathrm{C}} - 1} - \frac{1}{N_{\mathrm{C}} - 1} \frac{(1-\alpha)\mathcal{J}\nu_{\max}}{\mu_{\max} - \alpha \mathcal{J} \nu_x - \frac{\mathcal{J}}{N_{\mathrm{C}}} (1 + \gamma g)\, \nu_{\max}} \tag{35}$$

as an upper or lower bound for the modularity (Fig. 8). Requirement (Eq. 34) with the solution (Eq. 32) for μ^{S} inserted yields a further lower bound

$$m \geq \frac{(\mu_{\max} - \mu_{\min}) N_{\mathrm{C}}}{(1-\alpha)\mathcal{J}\nu_{\max} + (\mu_{\max} - \mu_{\min})(N_{\mathrm{C}} - 1)} \tag{36}$$

for the modularity that is required for denoising. This criterion is independent of the external background input and the recurrence of the SSN.

Now we turn to the saturated scenario $\nu^S = \nu_{\max}$ and $\nu^{NS} = 0$ and obtain

$$\mu^{\mathrm{S}} = \alpha \mathcal{J} \nu_x + \kappa_{\mathrm{S,S}}(m)\, \nu_{\max} \,,$$
$$\mu^{\mathrm{NS}} = \alpha \mathcal{J} \nu_x + \kappa_{\mathrm{NS,S}}(m)\, \nu_{\max} \,,$$

with the criteria

$$\mu^{\mathrm{S}} \geq \mu_{\max} \,, \tag{37}$$
$$\mu^{\mathrm{NS}} \leq \mu_{\min} \,. \tag{38}$$

The first criterion (Eq. 37) yields the same critical value (Eq. 35) that for $\mu_{\max} - \alpha \mathcal{J} \nu_x - \frac{\mathcal{J}}{N_{\mathrm{C}}} (1 + \gamma g)\, \nu_{\max} \geq 0$ is a lower bound and otherwise an upper bound. The second criterion (Eq. 38) yields an additional lower bound for $\mathcal{J}(1-\alpha)\nu_{\max} - (N_{\mathrm{C}} - 1) \left(\mu_{\min} - \alpha \mathcal{J} \nu_x - \frac{\mathcal{J}}{N_{\mathrm{C}}} (1 + \gamma g)\, \nu_{\max} \right) \geq 0$ (Fig. 8):

$$m \geq 1 - \frac{\left(\mu_{\min} - \alpha \mathcal{J} \nu_x - \frac{\mathcal{J}}{N_{\mathrm{C}}} (1 + \gamma g) \nu_{\max}\right)}{\mathcal{J}(1-\alpha)\nu_{\max} - (N_{\mathrm{C}} - 1)\left(\mu_{\min} - \alpha \mathcal{J} \nu_x - \frac{\mathcal{J}}{N_{\mathrm{C}}} (1 + \gamma g) \nu_{\max}\right)} \,. \tag{39}$$

The above criteria yield necessary conditions for the existence of a fixed point with $\nu^S > 0$ and $\nu^{NS} = 0$. Next we study the stability of such solutions. This works analogous to the stability in the spiking models discussed in Section "Effective connectivity and stability analysis" by studying the spectrum of the effective connectivity matrix. For the model Eq. 31, the effective connectivity is given by

$$w_{ij} = \frac{\partial \nu_i}{\partial \nu_j} = \nu'(\mu_i) \frac{\partial \mu_i}{\partial \nu_j} = \nu'(\mu_i) \mathcal{J}_{ij} \tag{40}$$

with $\nu'(\mu) = \frac{d\nu}{d\mu}(\mu)$ and $\mathcal{J}_{ij} = \tau_x J_{ij}$. On the level of stimulated and non-stimulated sub-populations across layers, the effective connectivity becomes

$$W = \begin{pmatrix} \kappa_{\mathrm{S,S}}(m)\nu'(\mu^{\mathrm{S}}) & \kappa_{\mathrm{S,NS}}(m)\nu'(\mu^{\mathrm{NS}}) \\ \kappa_{\mathrm{NS,S}}(m)\nu'(\mu^{\mathrm{S}}) & \kappa_{\mathrm{NS,NS}}(m)\nu'(\mu^{\mathrm{NS}}) \end{pmatrix} \tag{41}$$

with eigenvalues

$$\begin{aligned} \lambda_{\pm} = & \frac{\kappa_{\mathrm{S,S}}(m)\nu'(\mu^{\mathrm{S}}) + \kappa_{\mathrm{NS,NS}}(m)\nu'(\mu^{\mathrm{NS}})}{2} \\ & \pm \sqrt{\left(\frac{\kappa_{\mathrm{S,S}}(m)\nu'(\mu^{\mathrm{S}}) + \kappa_{\mathrm{NS,NS}}(m)\nu'(\mu^{\mathrm{NS}})}{2}\right)^2 - \left(\kappa_{\mathrm{S,S}}(m)\nu'(\mu^{\mathrm{S}})\kappa_{\mathrm{NS,NS}}(m)\nu'(\mu^{\mathrm{NS}}) - \kappa_{\mathrm{S,NS}}(m)\nu'(\mu^{\mathrm{NS}})\kappa_{\mathrm{NS,S}}(m)\nu'(\mu^{\mathrm{S}})\right)} \,. \end{aligned} \tag{42}$$

The saturated fixed point $\nu^S = \nu_{\max}$ and $\nu^{NS} = 0$ has $\nu'(\mu^{\mathrm{S}}) = \nu'(\mu^{\mathrm{NS}}) = 0$, leading to $\lambda_{\pm} = 0$. This fixed point is always stable. The non-saturated fixed point also has $\nu'(\mu^{\mathrm{NS}}) = 0$. Consequently, Eq. 42 simplifies to $\lambda_{-} = 0$ and

$$\lambda_{+} = \frac{\nu_{\max}}{\mu_{\max} - \mu_{\min}} \kappa_{\mathrm{S,S}}(m) \,. \tag{43}$$

For $\lambda > 1$ fluctuations in the stimulated sub-population are being amplified. These fluctuations also drive fluctuations of the non-stimulated sub-population via the recurrent coupling. The fixed point thus becomes unstable and the necessary distinction between the stimulated and non-stimulated sub-populations vanishes. For inhibition-dominated recurrence, $\kappa_{\mathrm{S,S}}(m)$ is small enough to obtain stable fixed points at non-saturated rates (Fig. 8c). In the case of no recurrence or excitation-dominated recurrence, $\kappa_{\mathrm{S,S}}(m)$ is much larger, typically driving λ_{+} across the line of instability and preventing non-saturated fixed points to be stable. In such networks, only the saturated fixed point at $\nu^S = \nu_{\max}$ is stable and reachable (Fig. 8d,e).

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] Rufin VanRullen and Christof Koch. Is perception discrete or continuous? *Trends in Cognitive Sciences*, 7(5):207–213, May 2003.

[44] Christian Klos, Daniel Miner, and Jochen Triesch. Bridging structure and function: A model of sequence learning and prediction in primary visual cortex. *PLOS Computational Biology*, 14(6):e1006187, Jun. 2018.

[45] Carl van Vreeswijk and Haim Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274:1724–1726, December, 6 1996.

[46] 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.

[47] 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.

[48] 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.

[49] Bart Borghuis, Duje Tadin, Martin Lankheet, Joseph Lappin, and Wim van de Grind. Temporal limits of visual motion processing: Psychophysics and neurophysiology. *Vision*, 3(1):5, Jan. 2019.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.

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

[58] 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.

[59] 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.

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

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

[62] 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.

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

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

[65] 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.

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

[67] 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.

[68] Baktash Babadi and Haim Sompolinsky. Sparseness and expansion in sensory representations. *Neuron*, 83(5):1213–1226, Sep. 2014.

[69] 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.

[70] 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.

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

[72] 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.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] 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.

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

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

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

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

[82] 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.

[83] 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.

[84] 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.

[85] 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.

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

[87] 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.

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

[89] 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.

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

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

[92] 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.

[93] 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.

[94] Eilen Nordlie, Marc-Oliver Gewaltig, and Hans Ekkehard Plesser. Towards reproducible descriptions of neuronal network models. *PLoS Computational Biology*, 5(8):e1000456, Aug. 2009.

[95] Moritz Layer, Johanna Senk, Simon Essink, Karolína Korvasová, Alexander van Meegen, Hannah Bos, Jannis Schuecker, and Moritz Helias. Lif meanfield tools, Feb. 2020.

[96] Nicolas Fourcaud and Nicolas Brunel. Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Comput.*, 14:2057–2110, 2002.

[97] Massimo Mascaro and Daniel J Amit. Effective neural response function for collective population states. *Network: Computation in Neural Systems*, 10(4):351–373, 1999.