frontiers | Frontiers in Computational Neuroscience

TYPE Original Research
PUBLISHED 10 December 2025
DOI 10.3389/fncom.2025.1699179

Check for updates

OPEN ACCESS

EDITED BY
Jian K. Liu,
University of Birmingham, United Kingdom

REVIEWED BY
Shengdun Wu,
Zhejiang Lab, China
Jizhao Liu,
Lanzhou University, China

*CORRESPONDENCE
Wei P. Dai
✉ weidai@fudan.edu.cn
Yaoyu Zhang
✉ zhyy.sjtu@sjtu.edu.cn

RECEIVED 04 September 2025
REVISED 12 November 2025
ACCEPTED 20 November 2025
PUBLISHED 10 December 2025

CITATION
Yan J, Feng Y, Dai WP and Zhang Y (2025) State-dependent filtering as a mechanism toward visual robustness. *Front. Comput. Neurosci.* 19:1699179. doi: 10.3389/fncom.2025.1699179

COPYRIGHT
© 2025 Yan, Feng, Dai and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

State-dependent filtering as a mechanism toward visual robustness

Jing Yan[1], Yunxuan Feng[1], Wei P. Dai[2,3]* and Yaoyu Zhang[1]*

[1]School of Mathematical Sciences, Institute of Natural Sciences and MOE-LSC, Shanghai Jiao Tong University, Shanghai, China, [2]Research Institute of Intelligent Complex Systems, Fudan University, Shanghai, China, [3]Shanghai Artificial Intelligence Laboratory, Shanghai, China

Robustness, defined as a system's ability to maintain functional reliability in the face of perturbations, is achieved through its capacity to filter external disturbances using internal priors encoded in its structure and states. While biophysical neural networks are widely recognized for their robustness, the precise mechanisms underlying this resilience remain poorly understood. In this study, we explore how orientation-selective neurons arranged in a one-dimensional ring network respond to perturbations, with the aim of uncovering insights into the robustness of visual subsystems in the brain. By analyzing the steady-state dynamics of a rate-based network, we characterize how the activation state of neurons influences the network's response to disturbances. Our results demonstrate that the activation state of neurons, rather than their firing rates alone, governs the network's sensitivity to perturbations. We further show that lateral connectivity modulates this effect by shaping the response profile across spatial frequency components. These findings suggest a state-dependent filtering mechanism that contributes to the robustness of visual circuits, offering theoretical insight into how different components of perturbations are selectively modulated within the network.

KEYWORDS

ring model, robustness, lateral connections, state-dependent filtering, visual processing

1 Introduction

Robustness—the ability of a system to maintain its functionality in the face of perturbations—is a critical property of many complex systems, including neural networks (Kitano, 2004; Alderson and Doyle, 2010). The visual system, in particular, exhibits remarkable robustness, accurately perceiving and recognizing objects despite variations in lighting, viewpoints, and other distortions in the visual scene (DiCarlo et al., 2012). Understanding the mechanisms that enable such stability is a central goal in computational neuroscience and has direct implications for the development of reliable artificial visual systems. This raises a fundamental question: How does robustness develop from the architecture and dynamics of cortical circuits?

A growing body of research suggests that cortical computations are dynamically shaped by internal states and recurrent interactions. Normalization provides a canonical account of gain control and robustness (Carandini and Heeger, 2012), predictive–processing frameworks emphasize adaptive feedback and contextual modulation (Keller and Mrsic-Flogel, 2018; Mante et al., 2013), and circuit-level studies link connectivity structure to emergent, state-dependent computations (Mastrogiuseppe and Ostojic, 2018; Stringer et al., 2019; Rubin et al., 2015). Building on these advances, we focus on the specific contribution of lateral connectivity to robustness, independent of higher-area feedback. Our contribution is to provide a compact, quantitative framework that formalizes state-dependent filtering by analyzing the system's Jacobian through singular value decomposition (SVD) as an analytical tool for characterizing sensitivity. This analysis identifies maximally amplified perturbation modes and demonstrates how they correspond to perceptually interpretable transformations—such as contrast modulation, small rotations, and elongation—via a Gabor mapping. Together, these results clarify how lateral connectivity and activation state jointly implement selective robustness within a simplified cortical model.

To isolate the role of lateral connectivity, simplified recurrent architectures such as the ring model have proven particularly useful. The ring model, consisting of orientation-selective neurons with Gaussian-shaped connectivity on a one-dimensional ring, has been widely used to study fundamental aspects of visual processing, including orientation selectivity (Ben-Yishai et al., 1995), contrast invariance (Carandini and Heeger, 2012), surround suppression (Somers et al., 1995; Sompolinsky and Shapley, 1997; Rubin et al., 2015), and binocular rivalry and fusion (Said and Heeger, 2013; Wilson, 2017; Wang et al., 2020). Despite its simplicity, the ring model can incorporate effective single-neuron nonlinearities and experimentally derived connectivity profiles, enabling tractable analysis of how neuronal states and recurrent interactions shape network responses.

In this study, we investigate the state-dependent response of the ring model to structured perturbations and its implications for visual robustness. We first analyze a steady-state rate-based ring model and derive analytical expressions relating perturbation responses to activation states and connectivity. We then identify the perturbations that elicit the largest state-dependent responses and examine their functional properties. To validate these results, we extend the analysis to a more biologically plausible spiking version of the ring model. Finally, by mapping the network's orientation-domain responses into image space through Gabor filters, we demonstrate that structured perturbations—such as contrast or aspect-ratio modulation—induce far stronger responses than random noise. By elucidating the model's state-dependent responses to perturbations, we instantiate the priors that filter external perturbations and pave the way for developing more robust and biologically inspired artificial vision systems.

2 Results

2.1 Robustness as selective filtering

In this study, we define robustness not merely as insensitivity to external noise, but as the capacity to selectively filter perturbations based on their semantic relevance—that is, the extent to which a perturbation modifies perceptually or behaviorally meaningful aspects of the input.

As discussed in Goodfellow et al. (2014), small adversarial perturbations—imperceptible to human observers—can cause deep neural networks to produce drastically incorrect outputs. For example, a slight input modification may cause a model to misclassify a panda as a gibbon, even though there is no significant semantic change in the image. In this case, the system is not robust because it reacts disproportionately to irrelevant noise.

In contrast, a biologically inspired visual system should instead prioritize perturbations that correspond to meaningful changes in the input—such as modifications in shape, orientation, or contrast—while suppressing perturbations that do not alter perceptual semantics. These types of perturbations are illustrated in Figure 1, where we show example modifications to a Gabor stimulus that are perceptually salient yet small in magnitude.

This perspective motivates our study: we hypothesize that robustness in visual processing relies on an internal filtering mechanism that adapts based on the input. In particular, the model should exhibit state-dependent filtering, in which the response to a perturbation depends on the current activation pattern induced by the input.

To explore whether biologically inspired systems could support such selective filtering, we study a simplified model of the primary visual cortex—the ring model. We provide a method to compute the perturbation direction that maximizes the change in the model response across different activation patterns. Remarkably, when these perturbations are mapped back into the image domain, they often resemble structured patterns such as rotation, elongation, or contrast change. In contrast, random noise-like perturbations of the same energy are consistently attenuated by the model, indicating that the system selectively filters out semantically irrelevant inputs.

This finding supports the hypothesis that robust computation may emerge from state-dependent filtering, in which the structure of meaningful perturbations aligns with the system's internal sensitivity.

Formally, the sensitivity of the ring model to small perturbations can be described by the Jacobian operator:

$$DF(\mathbf{r}_0) := \frac{\partial \mathbf{r}}{\partial \mathbf{I}}\bigg|_{\mathbf{r}=\mathbf{r}_0}, \tag{1}$$

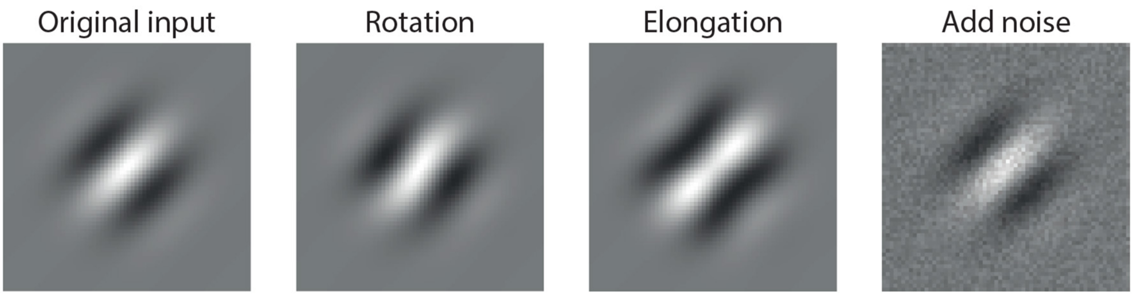

FIGURE 1
Example of perturbations applied to a standard Gabor stimulus: (1) original input, (2) rotated version (orientation change), (3) elongated version (spatial shape change), and (4) noise-added version (random perturbation). While all three perturbations are small in magnitude, only the first two induce semantically meaningful changes in the stimulus, which a robust visual system should prioritize. The last perturbation, although visually subtle, is semantically irrelevant and ideally should be suppressed.

which maps infinitesimal changes in input $\Delta\mathbf{I}$ to changes in system response $\Delta\mathbf{r}$ around a fixed operating point $\mathbf{r}_0$. Applying singular value decomposition (SVD) to $DF(\mathbf{r}_0)$ yields orthogonal perturbation directions, called singular vectors, each associated with an amplification factor, the corresponding singular value. The right singular vectors indicate orthogonal directions in the input space that produce maximal response changes, while the singular values quantify the corresponding frequency response magnitude. Intuitively, singular vectors represent the most "effective perturbations" of the system: those aligned with large singular values cause strong changes in network activity, whereas those aligned with small singular values are effectively filtered out. This provides a principled mathematical framework linking our semantic notion of robustness to structured perturbation patterns, such as orientation shifts or elongation, that we will analyze in subsequent sections.

2.2 The perturbed system: determined by lateral connections and activation pattern

The ring model provides a simplified framework for studying orientation columns in the primary visual cortex (V1), where neurons' preferred orientations are arranged on a ring. It contains excitatory (E) and inhibitory (I) populations, connected through Gaussian-shaped kernels. We begin our analysis with a steady-state rate-based version of the ring model (see Section 4), defined by the following system of equations:

$$\begin{aligned} r_E &= g(I_E + k_{EE} * r_E - k_{EI} * r_I), \\ r_I &= g(I_I + k_{IE} * r_E - k_{II} * r_I), \end{aligned} \tag{2}$$

where r_X denotes the firing rate vector of population $X \in \{E, I\}$, I_X the external input, and k_{XY} the lateral connectivity kernel from population Y to X, where $Y \in \{E, I\}$ as well. The convolution operation $(*)$ is circular, enforcing the ring topology over preferred orientations (see Section 4). The activation function g is ReLU, chosen for its biological plausibility and analytical tractability.

Linearizing around a fixed operating point results in the perturbed system

$$\begin{aligned} \delta r_E &= g'_E \odot (\delta I_E + k_{EE} * \delta r_E - k_{EI} * \delta r_I), \\ \delta r_I &= g'_I \odot (\delta I_I + k_{IE} * \delta r_E - k_{II} * \delta r_I). \end{aligned} \tag{3}$$

where g'_X is binary (0 or 1), depending on whether the neuron is active, and $\odot$ denotes element-wise multiplication. In matrix form, with diagonal matrices $G_X = \mathrm{diag}(g'_X)$ and circulant matrices K_{XY}, (see Section 4): Equation 2

$$\begin{aligned} \delta r_E &= G_E(\delta I_E + K_{EE}\delta r_E - K_{EI}\delta r_I), \\ \delta r_I &= G_I(\delta I_I + K_{IE}\delta r_E - K_{II}\delta r_I), \end{aligned} \tag{4}$$

The formal solution is as follows:

$$\begin{aligned} \delta r_E &= (I - G_E K_{EE} + G_E K_{EI}(I + G_I K_{II})^{-1} G_I K_{IE})^{-1} \\ &\quad (G_E \delta I_E - G_E K_{EI}(I + G_I K_{II})^{-1} G_I \delta I_I), \\ \delta r_I &= (I + G_I K_{II})^{-1} G_I(\delta I_I + K_{IE}\delta r_E), \end{aligned} \tag{5}$$

This solution reveals how perturbations to the input propagate through the network, depending on both the active set of neurons and the structure of lateral connectivity. Since G_X encodes the current activation pattern and K_{XY} the connection topology, the system effectively implements a state-dependent linear filter. Different activation states modulate the frequency response function and the direction of input perturbations, thereby enabling the network to selectively amplify structured perturbations aligned with the activation state while suppressing irrelevant ones.

In the following sections, we analyze this linearized system from multiple perspectives to uncover how state-dependent filtering emerges and contributes to robust computation.

2.2.1 Effects of lateral connections on model response with fully active neurons

We first consider the case that all neurons are active. In this case, the perturbed system can be solved in spatial frequency space, implying that sinusoids are eigenvectors: they retain their shape but only change in intensity as they pass through the system. The change in intensity, which we call frequency response from now on, is determined by the lateral connections. The specific relation is given by the following equation:

$$\hat{\delta r}_E = \left(1 - \hat{k}_{EE} + \frac{\hat{k}_{EI}\hat{k}_{IE}}{1 + \hat{k}_{II}}\right)^{-1}\left(\hat{\delta I}_E - \frac{\hat{k}_{EI}}{1 + \hat{k}_{II}}\hat{\delta I}_I\right), \tag{6}$$

where $\hat{v}$ denotes the DFT (Discrete Fourier Transform) of a vector v. Details are shown in Section 4. Here, we have suppressed the element-wise multiplication symbol $(\odot)$ for notational simplicity, as all multiplications in frequency space are understood to be element-wise. We focus on the excitatory population since it provides the main output of the ring model and is most relevant for the downstream readout, and we will keep this focus throughout the following analyses. We denote

$$\hat{h}_{-1} := \hat{h}_0^{-1} := \left(1 - \hat{k}_{EE} + \frac{\hat{k}_{EI}\hat{k}_{IE}}{1 + \hat{k}_{II}}\right)^{-1}, \tag{7}$$

which we call the frequency response function from now on.

We assume that inhibitory-to-inhibitory (I–I) connections are absent, i.e., $\hat{k}_{II} = 0$, for analytical simplicity and interpretability. Although such connections exist in biological circuits and may contribute to overall inhibitory modulation, they are not essential for the frequency-selective filtering we focus on. Omitting them allows us to highlight the balance between excitation and inhibition in shaping the model's robustness.

Lateral connections can be divided into two parts: the excitatory term $\hat{k}_{EE}$, which enhances signals through recurrent excitation, and the recurrent inhibitory term $\hat{k}_{EI}\hat{k}_{IE}$, which reduces signals through the E→I→E pathway. Assuming Gaussian kernels,

$$k_{XY}(x) = \alpha_{XY} \exp\left(-\frac{x^2}{2\sigma_{XY}^2}\right), \quad X, Y \in \{E, I\}, \tag{8}$$

where α_{XY} sets the connection strength and σ_{XY} the spatial spread.

Because Gaussian kernels remain Gaussian in frequency space, both excitation and inhibition act as low-pass filters: excitation boosts low frequencies, while recurrent inhibition suppresses them.

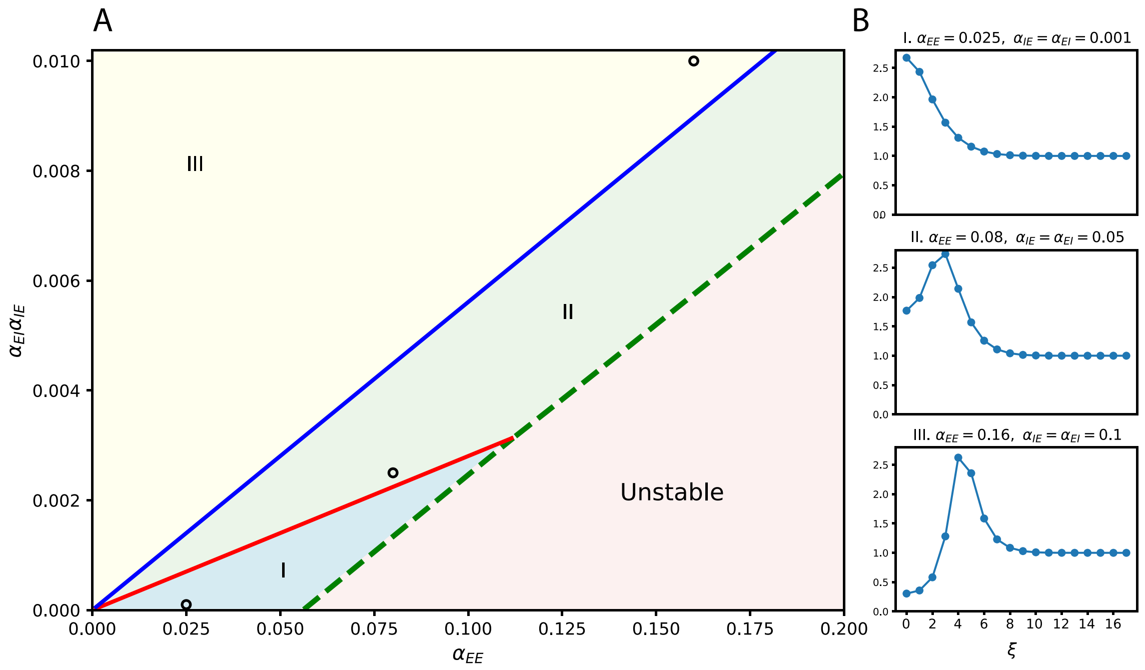

FIGURE 2

Frequency-selective effects of lateral connectivity. **(A)** Phase diagram over connection strengths, with α_{EE} on the horizontal axis and $\alpha_{EI}\alpha_{IE}$ on the vertical axis. The diagram is divided into three stable regions (I–III) and one unstable region, separated by critical lines. Example parameter choices are marked by circles. **(B)** Frequency (ξ) response curves corresponding to the marked parameters in **(A)**. Different regions of the phase diagram produce qualitatively distinct filtering profiles. Here, the widths of lateral connections are fixed and equal ($\sigma_{EE} = \sigma_{EI} = \sigma_{IE} = 10$).

High-frequency perturbations, in contrast, pass almost unchanged. The range of affected frequencies scales inversely with σ_{XY}: broader connections suppress or enhance lower frequencies. Since k_{IE} and k_{EE} often share similar widths, the inhibitory term ($k_{EI} * k_{IE}$) typically spans a broader frequency range, reducing lower frequencies more strongly than excitation enhances them.

The resulting frequency-dependent amplification profiles can be summarized in a phase diagram (Figure 2A), parameterized by the widths $\sigma_{EE}, \sigma_{EI}, \sigma_{IE}$. Distinct regions of stability (I, II, and III) and instability are determined by the ratios α_{EE} and α_{EI}/α_{IE}. Each stable region corresponds to a characteristic shape of the frequency response curve (Figure 2B), where we plot only the first several dominant frequencies associated with the largest singular values. These already capture the main trend of frequency selectivity. As parameters vary or as the system transitions between regions, the frequency selectivity shifts accordingly.

In short, lateral connectivity controls which perturbation frequencies are amplified or suppressed, acting as a tunable filter that balances excitation and inhibition. This frequency-selective mechanism lays the foundation for robustness: it boosts structured perturbations while suppressing noise-like ones. In later sections, we will further show that these frequency preferences correspond to concrete image-level patterns, such as rotation or elongation.

2.2.2 Effect of activation patterns on model response

We next discuss how the system's behavior changes with different activation patterns, focusing on perturbations that can lead to strong responses.

For a fixed activation pattern $G = (G_E, G_I)$, the perturbed system remains linear. We therefore analyze the excitatory population's sensitivity using the singular value decomposition (SVD) of the corresponding linear operator. Singular vectors of this operator identify perturbations that are maximally amplified; their singular values s quantify the amplification.

Using Gaussian-shaped lateral connections, we compute the singular vectors under three representative activation conditions (Figure 3). Figures 3A–C show the case with all neurons active; Figures 3D–F show a case where half of the neurons are consecutively active; Figures 3G–I introduce random neuron loss, where each neuron is independently silenced with probability 0.1. Across all conditions, the singular vectors in orientation space are sinusoidal or resemble sinusoidal patterns. When the same vectors are reordered by their dominant frequency ξ (middle column) and plotted in the frequency domain (right column), their spectra exhibit clear peaks, indicating frequency-dominant structure.

These observations allow us to summarize the system's selectivity using frequency response curves (Figure 4). As the active set shrinks, the gain profile becomes smoother and less sharply tuned: peaks broaden, and their energy redistributes into neighboring frequencies. Peak locations are largely stable across conditions; when shifts occur, they are slight and infrequent. For clarity, Figure 4 displays only the first few dominant frequencies, which already capture the main trend. In terms of magnitudes, the largest singular values s associated with the most selective components often decrease modestly as activation is reduced, whereas nearby components can remain comparable or increase, reflecting the redistribution of energy across adjacent frequencies.

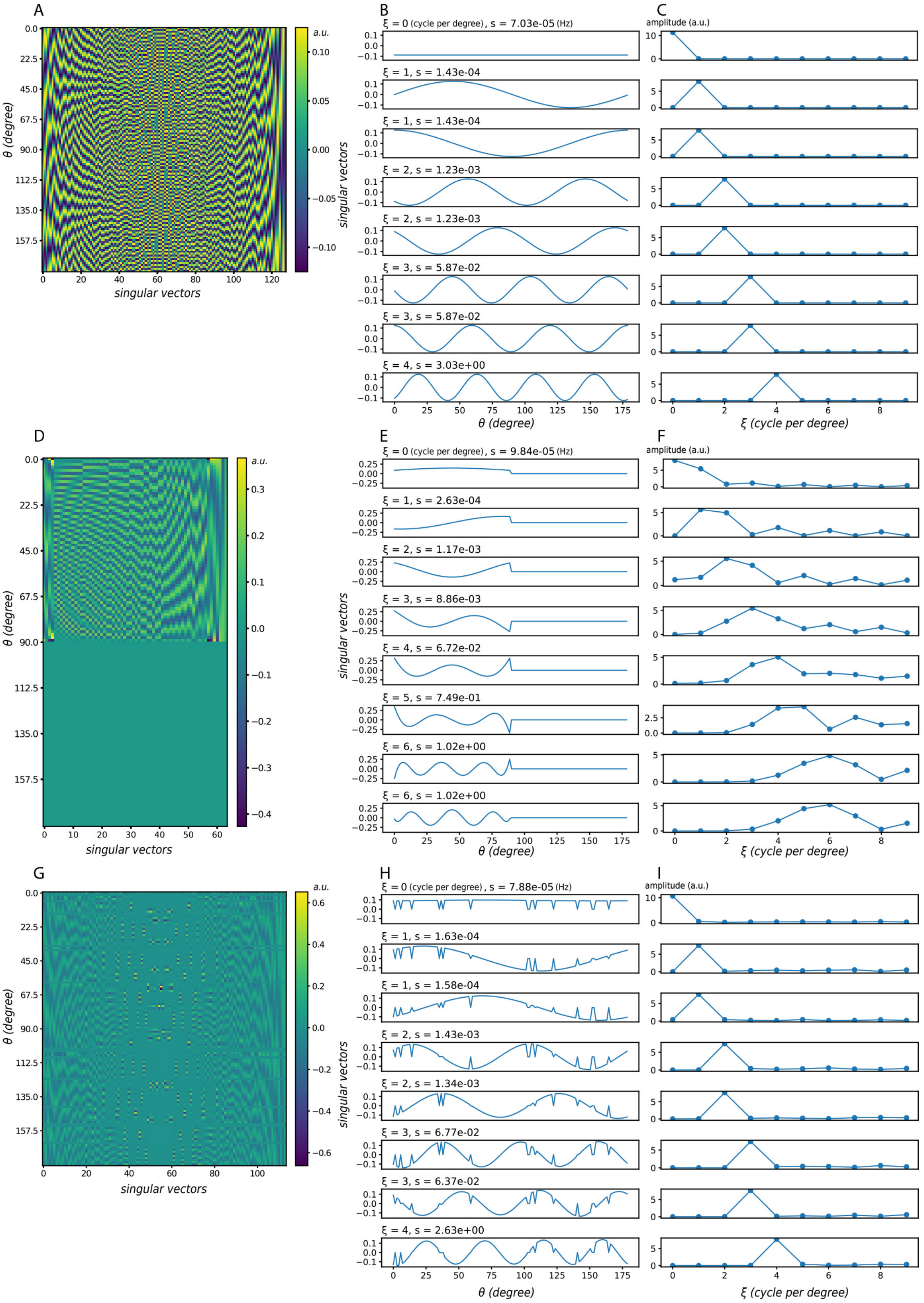

FIGURE 3
Singular vectors under different activation patterns. **(A–C)** All neurons are active. **(D–F)** Half of the neurons are active in succession. **(G–I)** Random neuron loss, where each neuron is silenced independently with probability 0.1. In each case: **(left)** singular vector matrices, where each column corresponds to one singular vector, ordered by decreasing singular value s; **(middle)** the same singular vectors plotted in orientation space, but reordered according to their dominant frequency ξ; **(right)** corresponding frequency-domain representations, where the peak indicates the dominant frequency and its height reflects the strength of that component. Here ξ denotes the dominant frequency of each singular vector, and s the corresponding singular value (amplification factor). Despite partial deactivation, the dominant frequency components remain visible, showing that the system's preference for frequency-dominant perturbations is robust to neuron loss. Model parameters: $\sigma_{EE} = \sigma_{EI} = \sigma_{IE} = 24$, $\alpha_{EE} = 4$, $\alpha_{EI} = \alpha_{IE} = 2$.

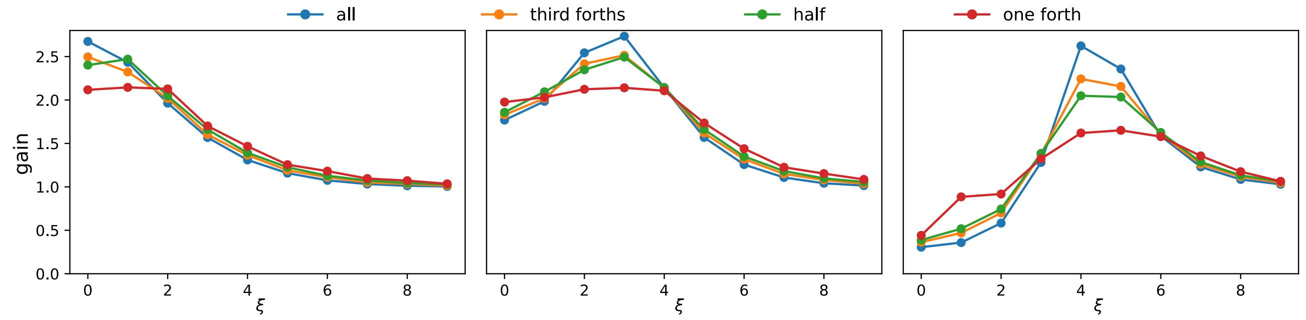

FIGURE 4
Frequency response curves under different activation patterns. Each subpanel corresponds to one parameter setting (same as Figure 2B). Colors indicate different fractions of consecutively active neurons. The x-axis denotes the dominant frequency ξ of the singular vectors, while the y-axis shows the corresponding singular value s. As the proportion of active neurons decreases, the frequency response curve becomes smoother, reflecting weaker frequency selectivity.

In summary, activation patterns strongly shape the system's filtering properties. By changing which frequencies are emphasized, they determine the orientation and directional preferences of the effective filter. Robustness thus arises from this state-dependent filtering: even when only part of the population is active, the network continues to favor structured, frequency-dominant perturbations.

2.3 Results for spiking ring models

To validate the generality of our theoretical findings, we conducted experiments using a more biologically realistic conductance-based spiking neuron model. Although the steady-state rate model and the spiking neuron model differ significantly in their implementations, important terms such as 'firing rates' and 'lateral connections' are preserved across both models (see Section 4).

Based on previous study with this model, we consider the mean-driven regime (Cai et al., 2004). The mean-driven regime is more compatible with our theoretical analysis while still capturing biological characteristics. We obtained results in the mean-driven regime that fully align with the steady-state rate model.

First, we demonstrate the response of different frequency perturbations in orientation space and frequency space for unconnected and fully connected networks, as shown in Figures 5A, B. Then, we present experimental results in the mean-driven regime. Consistent with the steady-state rate model, the spiking neuron model in the mean-driven regime also demonstrates a preference for different frequencies. This quantitative consistency could be achieved by considering a dimensionful mapping from the dimensionless steady-state rate model to the dimensionful spiking neuron model. As shown in Figure 5C, when no connections were present, the ring model responded uniformly to all frequencies. With only recurrent excitatory connections, the ring model enhanced the low-frequency response. In contrast, when only recurrent inhibitory connections were present, the ring model suppressed the low-frequency response. And when recurrent excitatory and inhibitory connections were present, the model demonstrated a preference for a specific frequency.

2.4 From orientation-domain frequencies to image-space perturbations

2.4.1 Gabor operator modes: full vs. partial activation

To link orientation-domain analysis with image space, we examined the singular value decomposition (SVD) of the Gabor operator $\mathcal{F}_{\mathcal{G}}$, which maps an image to orientation responses (see Section 4). With all orientations active, $\mathcal{F}_{\mathcal{G}}$ is approximately block-circulant along the orientation dimension. Its singular vectors therefore align with discrete Fourier modes: the orientation-domain vectors are sinusoids, and their paired image-domain vectors are structured stripe patterns (Figure 6A). These modes are arranged from left to right and top to bottom in order of decreasing singular value s. Due to the circular symmetry in this fully active case, no directional preference emerges; the organizing index is solely the orientation frequency ξ.

When $\mathcal{F}_{\mathcal{G}}$ is restricted to a contiguous subset of active orientations, producing a sub-matrix $\tilde{\mathcal{F}}_{\mathcal{G}}$, the circular symmetry is broken. In this scenario, we follow the idea underlying singular value decomposition: we use an optimization procedure to extract patterns in the image space that are mutually orthogonal and satisfy the constraint

$$\|\tilde{\mathcal{F}}^{c}_{\mathcal{G}}p\|_\infty \leq \|\tilde{\mathcal{F}}_{\mathcal{G}}p\|_\infty,$$

where p denotes a candidate pattern. The goal is to identify those patterns that retain the most "energy," meaning that they produce outputs with the largest norm under these constraints. Although this is not a literal SVD, the construction parallels its logic, so we continue to refer to the resulting modes as singular vectors for consistency. As Figure 6B shows, the leading singular vectors remain frequency-dominant but are biased toward the active sector. In this setting, low-order frequencies correspond to interpretable image-level perturbations: $\xi = 0$ corresponds to contrast modulation, $\xi = 1$ to small rotations of oriented content,

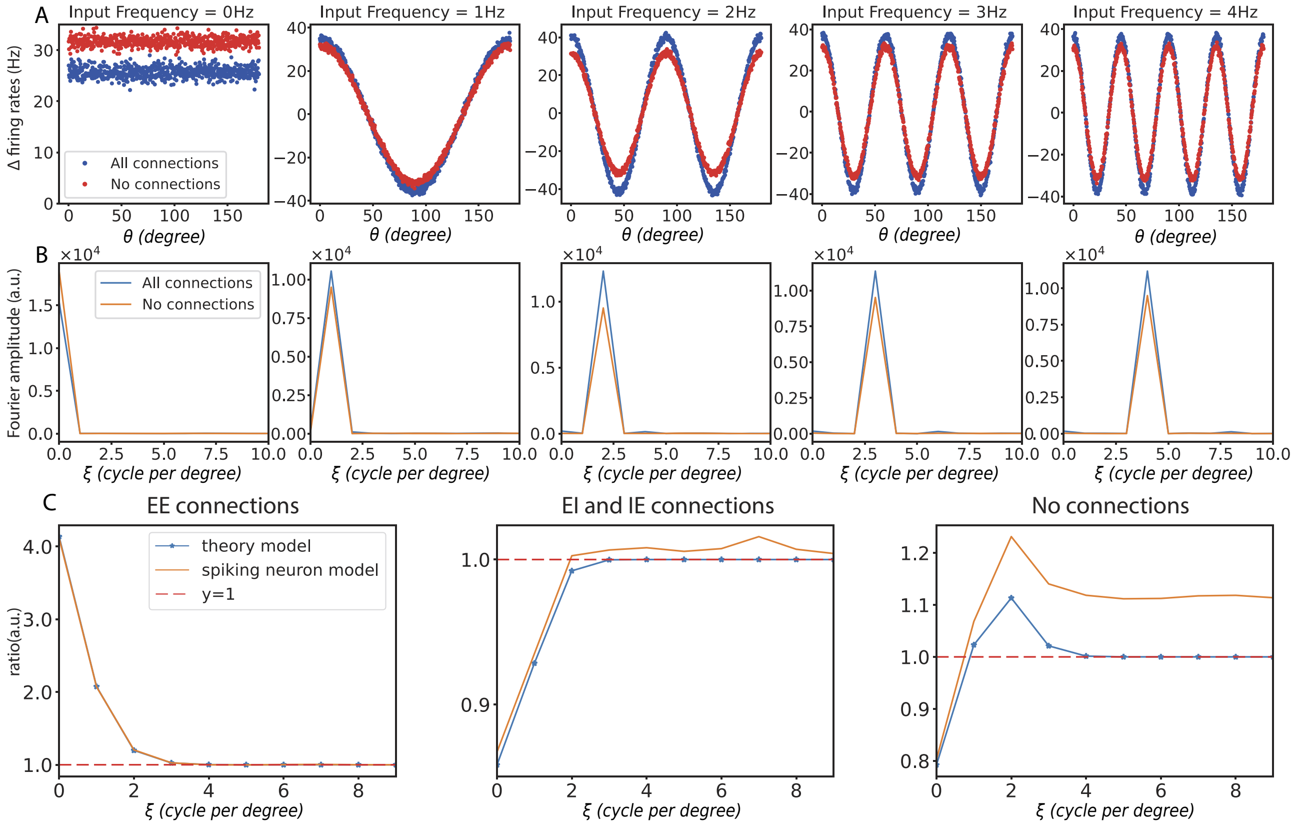

FIGURE 5
Results with spiking neuron model. **(A)** The changes in firing rates caused by different input frequencies under the condition of full connectivity. **(B)** The corresponding changes in firing rates are shown in frequency space under the same conditions. **(C)** Frequency curves under different connection conditions: the left panel shows the frequency response curve with only recurrent excitation (EE) connections, the middle panel shows the curve with both recurrent excitation (EI) and inhibition (IE) connections, and the right panel shows the frequency response curve without any connections. In all subfigures, ξ represents input frequency.

and $\xi = 2$ to elongation or aspect-ratio changes. These patterns are also arranged in decreasing order of singular value. This parallel between Gabor SVD modes and ring-model frequency modes shows that both systems naturally produce frequency-dominant eigenmodes that can be aligned.

2.4.2 Perturbation analysis with Gabor inputs

Next, we analyzed the complete mapping from the image space through the Gabor operator $\mathcal{F}_G$ into the ring model. For different regimes of lateral connectivity, we identified the image perturbations that maximize the excitatory response under a fixed norm constraint (Figure 7). The maximally effective perturbations fall into three families—contrast, rotation-like, and elongation—consistent with the orientation frequencies $\xi = 0, 1, 2$ identified earlier.

To quantify these effects, we examined the gain curves (ΔR vs. ΔG) for three canonical perturbation families: elongation, contrast scaling, and Gaussian noise (Figure 7). The gain curves show that structured perturbations elicit much stronger responses than noise. In regime (A), contrast consistently dominates across amplitudes. In regime (B), elongation initially dominates at small to moderate ΔG, but its gain decreases as elongation increases further. Along this path, the maximizing perturbation pattern can shift from elongation-like to rotation-like. This arises because the Gabor filter bank attenuates energy from the center to the periphery; once the base image is highly elongated, further elongation yields diminishing returns, making a small rotation more effective. By contrast, along the contrast and noise paths, elongation remains the strongest perturbation.

The side images illustrate this behavior: the left column shows the input stimulus at different amplitudes, while the right column shows the maximizing perturbation pattern for that input. Elongation corresponds to the stretching of the Gabor envelope, in contrast to global intensity modulation, and noise remains ineffective. The observed preferences are not fixed; they can be systematically altered by tuning the lateral connectivity, as summarized in the phase diagram (Figure 2). In principle, these behaviors are encoded in the frequency response function $\hat{h}_{-1}$, but here we highlight their manifestation empirically through gain curves.

To further compare the impacts of elongation and contrast change, we plot the gain curves in Figure 8. The effectiveness of a perturbation can be assessed from the slope of its curve. It is evident that both patterns produce a larger response than the noise. Furthermore, when we apply all types of lateral connections, as

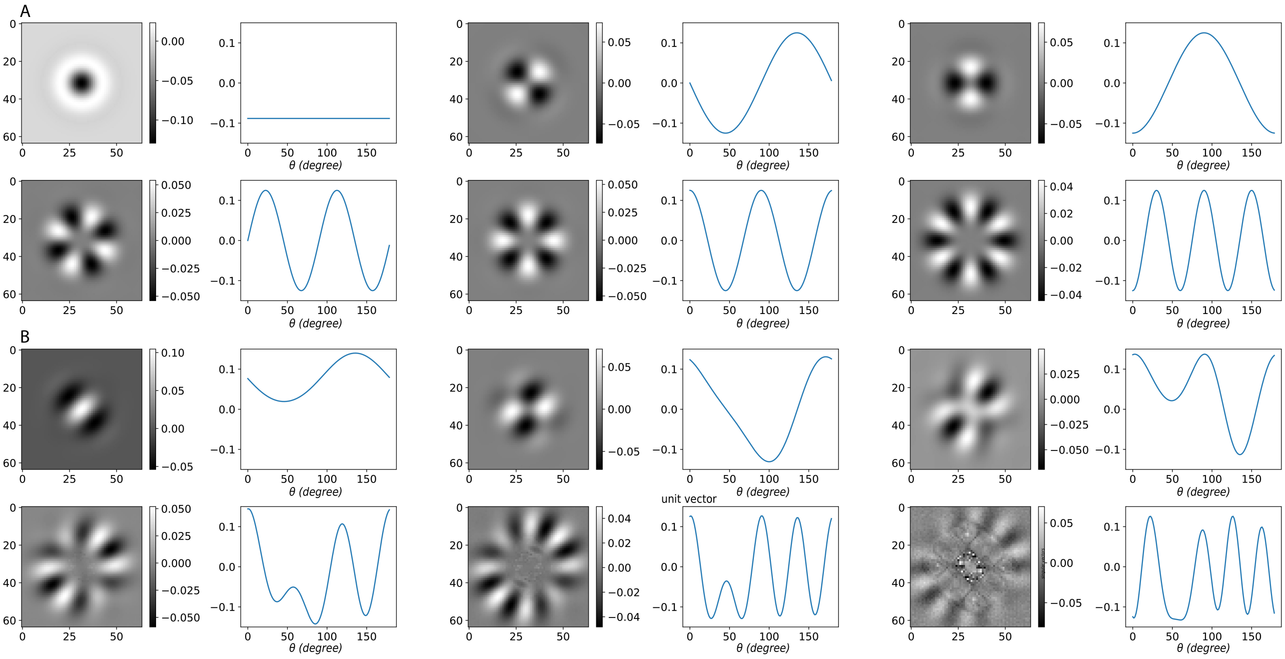

FIGURE 6
SVD of the Gabor operator under different activation conditions. **(A)** Fully active case. Because each orientation channel is a rotated copy of the same kernel, the Gabor operator $\mathcal{F}_G$ is approximately block-circulant along the orientation. Its left singular vectors U (orientation domain) align with discrete Fourier modes (sinusoids over orientation), shown as waveforms, while the paired right singular vectors V (image domain) are reshaped into grayscale image patterns. The x-axis of these waveforms denotes the orientation angle θ, and the y-axis indicates the relative weight of each singular vector across orientation channels. In this symmetric setting, no directional bias appears; the organizing index is the orientation frequency ξ. **(B)** Partially active case with a contiguous active sector. Restricting $\mathcal{F}_G$ to a subset of active orientations breaks circular symmetry and introduces directional preference. The leading singular vectors remain frequency-dominant but become biased toward the active sector. Low-order frequencies correspond to familiar image-level perturbations: $\xi = 0$ represents contrast modulation; $\xi = 1$ corresponds to a small rotation (odd, shift-like); and $\xi = 2$ corresponds to elongation or aspect-ratio change (even, two-lobe). Each image tile shows the right singular vector V (image-space pattern), with the paired left singular vector U plotted as a sinusoid over orientation. Here, ξ denotes the dominant orientation frequency of the singular vector (DFT index).

commonly assumed in ring model studies to reproduce orientation selectivity and contrast invariance, the elongation perturbation exerts a stronger influence on the output. In contrast, in the absence of such lateral connections, the contrast change perturbation is more effective.

3 Discussion

3.1 Advantages and limitations

This study provides insights into the robustness mechanisms of the visual system using the ring model, which represents a simplified network of orientation-selective neurons in the primary visual cortex (V1). One of the key findings is that the system's response to perturbations is strongly shaped by the state of neuronal activation and the pattern of lateral connectivity, which together act as the network's internal priors. Using both a steady-state rate model and a spiking network implementation, we bridge theoretical analysis with biological plausibility.

Our results reveal that distinct activation patterns lead to qualitatively different filtering properties, suggesting that the visual system may adaptively respond to perturbations depending on its current state. This phenomenon echoes the adaptive nature of biological organisms, which continuously adjust to environmental variability (DiCarlo et al., 2012). Moreover, frequency response analysis shows that the strength and spatial structure of lateral connections dictate the degree of selectivity in filtering, supporting the view that biological networks are fine-tuned to enhance relevant inputs while suppressing noise (Carandini and Heeger, 2012). Such findings may inspire more adaptive artificial neural networks.

While omitting inhibitory–inhibitory (I–I) connections and adopting a ReLU activation function simplifies the model, these choices maintain analytical tractability and interpretability. Including I–I connections would mainly shift quantitative gain profiles leftward without altering the core frequency-selective, state-dependent filtering mechanism. Concerning the activation function, we note that softplus can be viewed as a two-segment approximation of ReLU—with slopes 0 and 1, and only a narrow, smooth transition—so our results naturally extend to the softplus case. Similarly, the sigmoid activation introduces an additional saturating segment; however, within the regime of interest in our study (inputs near the threshold), saturation rarely occurs, and the network operates in the quasi-linear region equivalent to ReLU or softplus. Therefore, the qualitative mechanism of state-dependent filtering remains unchanged across these monotonic rectifying nonlinearities.

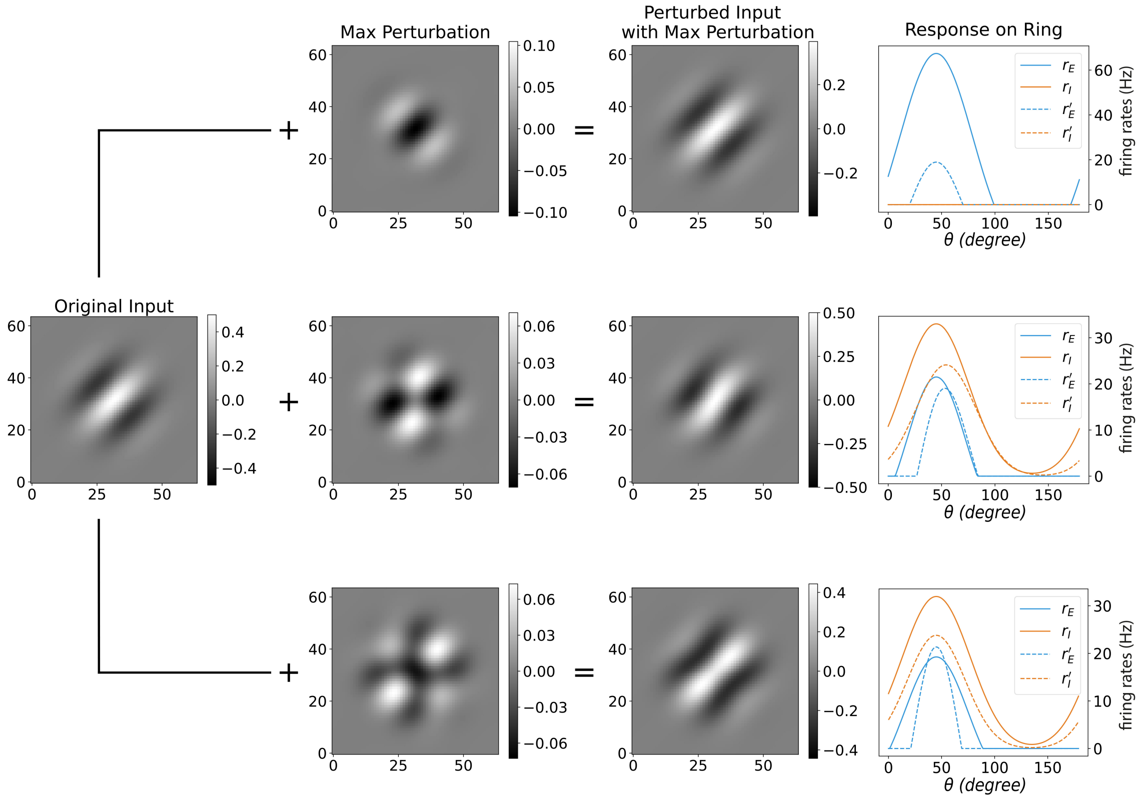

FIGURE 7
Maximally effective image perturbations for the full pipeline (image → $\mathcal{F}_{\mathcal{G}}$ → ring model). Rows correspond to three lateral–connection regimes around the same base input: **(A)** no lateral connections; **(B)** $g_{EE} = 0.03$, $g_{EI} = 0.06$, $g_{IE} = 0.04$; **(C)** $g_{EE} = 0.01$, $g_{EI} = 0.05$, $g_{IE} = 0.04$. Columns (left→right): original Gabor input; the maximizing image perturbation;the perturbed input; and the ring responses before/after perturbation (excitatory r_E vs. r'_E, inhibitory r_I vs. r'_I; dashed = before, solid = after). The maximizing pattern depends on connectivity: in **(A)** it is *contrast* modulation; in **(B)** it is *rotation*-like; in **(C)** it is *elongation* (envelope aspect ratio). These classes match the orientation–frequency mapping established earlier (contrast ↔ $\xi = 0$, rotation ↔ $\xi = 1$, elongation ↔ $\xi = 2$), showing how frequency-dominant modes translate into image-space perturbations under different lateral profiles.

Despite these advantages, the study also has several limitations. The analytical tractability of the ring model comes at the cost of biological completeness: complex cell-type heterogeneity, long-range feedback, and top-down modulation are not represented, which may restrict the generalization of our findings (Somers et al., 1995).

Although this study does not include a direct comparison between the 1-D ring model and 2-D topology, we postulate that a more realistic local connectivity in a cortical sheet will not make a large difference in the basic mechanism of state-dependent filtering. As other aspects of the two topologies have been studied extensively, there is no qualitative difference. Among them, the notable recent one is the extensive study on response normalization across orientation preferences, contextual (surround suppression), and contrast modulation studies (Ben-Yishai et al., 1995; Somers et al., 1995; Bressloff and Cowan, 2002; Rubin et al., 2015). In our case, the corresponding mechanism for state-dependent filtering can be simply extended to 2-D, but with a caveat of some difficulty in dealing with neurons at the pinwheel center for an ordered orientation map in non-rodents. Specifically, their state-dependence could be very different, and future studies that do employ such geometry can help elucidate such differences.

The present framework can also be interpreted from an adversarial perspective. By locally linearizing the network around a fixed activation state, the system becomes a linear operator represented by the Jacobian $DF(\mathbf{r}_0)$. Within this local approximation, identifying perturbations of fixed norm that induce the largest change in response is mathematically analogous to constructing adversarial perturbations in machine learning. The leading singular vectors of $DF(\mathbf{r}_0)$, corresponding to the largest singular values, define a small number of orthogonal directions along which the system is most sensitive. These directions represent the principal modes of effective perturbations—comparable to the dominant adversarial directions in artificial networks. In contrast to conventional adversarial analysis, which exploits such directions to reveal model vulnerabilities, the present study uses them to examine how biological circuits selectively suppress or amplify structured perturbations through state-dependent filtering.

3.2 Role of lateral connection to visual robustness

Our analysis reveals that lateral connectivity patterns play a crucial role in determining how the ring model filters out perturbations to an input. Specifically, we find that, with a biologically realistic lateral connection, the network exhibits a

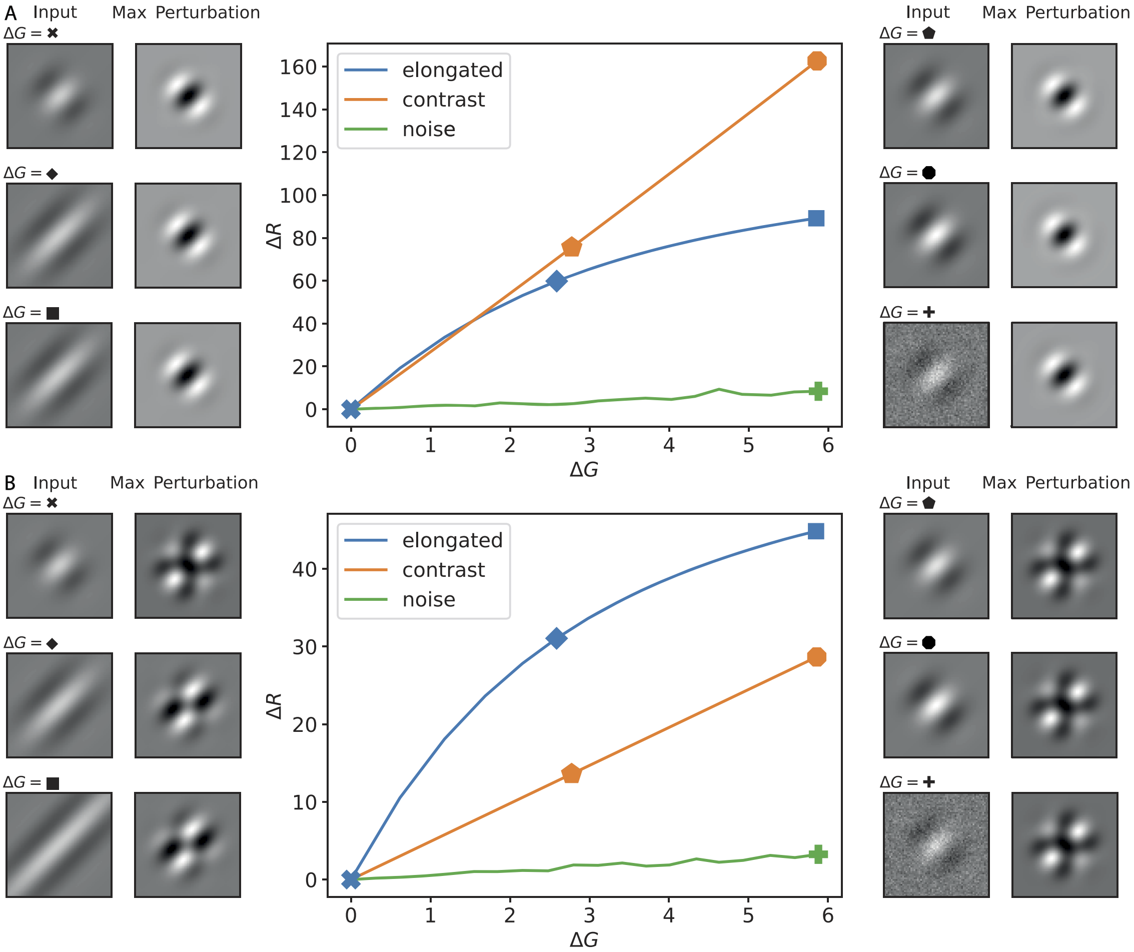

FIGURE 8
Response functions (ΔR vs. ΔG) for canonical image perturbations, with examples at selected amplitudes. **(A)** Disconnected lateral connections; **(B)** fully connected regime (cf. Figure 2). Center plots show the change in ring response ΔR versus input amplitude ΔG for three perturbation families: elongation (blue), contrast scaling (orange), and Gaussian noise (green). Side thumbnails illustrate, for selected amplitudes (indicated by symbols on the curves), the input image **(left)** and the maximizing perturbation pattern for that input **(right)**.

stronger response to the elongation of a standard Gabor input compared to changes in contrast. This behavior, which extends beyond linear receptive field modeling, contributes significantly to the system's robustness by prioritizing changes in features relevant to the input over irrelevant signals, such as very low or high-frequency spatial variations or perturbations in the orthogonal orientation.

Furthermore, since lateral connections are shaped by correlations in input statistics, we hypothesize that they selectively prioritize perturbations that align with the input pattern and natural input statistics while filtering out those that deviate from them. This mechanism enhances the robustness of visual information extraction and processing. Given that lateral connectivity is a canonical feature of cortical circuits across layers and brain regions, we propose that it serves as a potential mechanism for filtering out irrelevant perturbations while amplifying relevant ones in response to each input. This results in a system that exhibits remarkable resilience to atypical external disturbances, including adversarial inputs, even in the absence of adversarial training.

Building on this interpretation, our analysis suggests that similar principles could inform architectural designs in artificial networks. For example, lateral recurrent modules with input-dependent gating could emulate state-dependent filtering within convolutional or transformer-based architectures. Such modules would allow feature selectivity to adapt dynamically to input statistics, thereby improving robustness against structured perturbations.

While the ring model simplifies V1 by omitting feedback from higher cortical areas and other hierarchical processing stages (possibly mediated by disinhibition from other inhibitory neurons), it retains the essential feature of lateral connectivity, which serves as the primary mechanism underlying state-dependent filtering. We hypothesize that this lateral connectivity alone, even without complex feedback structures, is sufficient to induce the observed robustness phenomena. The presence of lateral connections creates a system in which perturbations aligned with the network's current state are selectively amplified and irrelevant noise is suppressed, providing a basic form of robustness. This simplified model does not preclude the role of higher-level feedback, but we posit that lateral connectivity is the critical factor driving the robustness observed in our analysis.

In contrast, most current deep-learning architectures—except for transformers—lack mechanisms that incorporate lateral computations analogous to those in biological visual pathways. As a result, these models remain highly vulnerable to small, unseen adversarial perturbations, highlighting a significant gap in robustness between biological and artificial systems.

3.3 Future research

Moving forward, several avenues of investigation could further enhance our understanding and application of robustness in visual systems:

1. Hierarchical models: Extending the analysis to hierarchical models of visual processing, including multiple layers that mimic the entire visual pathway, would provide a more comprehensive understanding of robustness across different stages of visual processing (Felleman and Van Essen, 1991).
2. Comparative studies: Conducting comparative studies between ring models and networks without lateral connections will elucidate the advantages of this connectivity in filtering out abnormal patterns and noise (Rubin et al., 2015).
3. Enhanced biological models: Incorporating more detailed biological data into the spiking network models could improve their realism. This could involve complex neurotransmitter dynamics, dendritic processing, and more accurate replication of neural circuitry (Said and Heeger, 2013).
4. Artificial neural networks optimization: Applying insights from biological systems to optimize artificial neural networks could lead to more robust machine learning algorithms. The focus should be on minimizing the sample sizes required for training and improving the network's resilience to input perturbations (Goodfellow et al., 2016).

In summary, this study provides novel insights into the role of state-dependent filtering in shaping robust visual processing. We demonstrate how activation states and lateral connectivity influence neural responses to infinitesimal perturbations and suggest that integrating state-dependent priors into artificial models may improve their adaptability and resilience in complex environments. This perspective bridges the gap between biological and artificial neural networks, offering new directions for both neuroscience and AI research.

4 Materials and methods

4.1 Steady-state ring model

We use the steady-state rate ring model, which can be viewed as the solution to the steady state of a dynamic system.

$$\begin{aligned} r_E &= g(I_E + k_{EE} * r_E - k_{EI} * r_I), \\ r_I &= g(I_I + k_{IE} * r_E - k_{II} * r_I). \end{aligned} \tag{9}$$

r_E (r_I) is a vector denoting the firing rates of the excitatory (inhibitory) neuron population. I_E (I_I) represents the excitatory (inhibitory) external input, and k_{XY} ($X, Y \in \{E, I\}$) is the connectivity kernel from population Y to X, implementing the Gaussian profile.

The activation function g is chosen as the rectified linear unit (ReLU), and $*$ denotes the circular convolution operation, enforcing the ring topology. The mathematical formula is as follows:

$$x * h = \sum_{k=0}^{N-1} x[k]h[(n-k) \bmod N], \tag{10}$$

$x[n]$ and $h[n]$ here are of length N.

It should be noted that k_{XY} has a Gaussian profile, which is symmetric about $\theta = 0$. The depiction of k_{XY} shows a Gaussian profile over the angle range from -90 degrees to 90 degrees. However, when performing the convolution, k_{XY} should be adjusted from 0 to -180 in the inverse direction. Therefore, the k here needs to be shifted. There is some symbol abuse here, and the meaning should be interpreted in context.

With small perturbations δI_E and δI_I, we obtain the perturbed system:

$$\begin{aligned} \delta r_E &= g'_E \odot (\delta I_E + k_{EE} * \delta r_E - k_{EI} * \delta r_I), \\ \delta r_I &= g'_I \odot (\delta I_I + k_{IE} * \delta r_E), \end{aligned} \tag{11}$$

$\odot$ stands for the Hadamard (element-wise) product. Here, g is the ReLU function, and each element of g' takes the value 0 or 1, depending on whether the corresponding neuron is active. For convenience, we encode the information of g'_X in a matrix G_X and derive the matrix-vector form of the perturbed system.

$$\begin{aligned} \delta r_E &= G_E(\delta I_E + K_{EE}\delta r_E - K_{EI}\delta r_I), \\ \delta r_I &= G_I(\delta I_I + K_{IE}\delta r_E - K_{II}\delta r_I), \end{aligned} \tag{12}$$

K_{XY} is a circulant matrix of the lateral connection k_{XY}, with the first row taken from k_{XY}, discretized from angle 0 to -180. Since the kernel k_{XY} is symmetric about 0, the generated matrix is symmetric.

For clarity, we provide explicit definitions of the matrices G_X and K_{XY} used in Equation 11.

The matrix $G_X \in \mathbb{R}^{N \times N}$ is a diagonal matrix encoding the derivative of the activation function for each neuron in population $X \in \{E, I\}$. For ReLU activation, it is defined as follows:

$$(G_X)_{ij} = \begin{cases} 1, & \text{if } i = j \text{ and } (r_X)_i > 0 \\ 0, & \text{otherwise} \end{cases} \quad \text{or equivalently,} \quad G_X = \mathrm{diag}(g'_X), \tag{13}$$

,

where $g'_X \in \mathbb{R}^N$ is a binary vector with $(g'_X)_i = 1$ if the i-th neuron is active, and 0 otherwise.

The matrix $K_{XY} \in \mathbb{R}^{N \times N}$ is a circulant matrix generated from the lateral connectivity kernel k_{XY}. The first row of K_{XY} is obtained by discretizing the kernel as follows:

$$(K_{XY})_{1j} = k_{XY}(\theta_j), \quad \theta_j = -\frac{180}{N}(j-1), \quad j = 1, \ldots, N, \tag{14}$$

and each subsequent row is a right circular shift of the previous one. Due to the even symmetry of k_{XY}, the resulting K_{XY} is both symmetric and circulant:

$$K_{XY} = \text{circ}\left(k_{XY}(\theta_1), k_{XY}(\theta_2), \ldots, k_{XY}(\theta_N)\right). \tag{15}$$

The solution to the equation can be given in a formal expression:

$$\begin{aligned} \delta r_E &= (I - G_E K_{EE} + G_E K_{EI}(I + G_I K_{II})^{-1} G_I K_{IE})^{-1}(G_E \delta I_E \\ &- G_E K_{EI}(I + G_I K_{II})^{-1} G_I \delta I_I), \quad (16) \\ \delta r_I &= (I + G_I K_{II})^{-1} G_I (\delta I_I + K_{IE} \delta r_E), \quad (17) \end{aligned}$$

Replacing ReLU with softplus or sigmoid preserves the qualitative state-dependent filtering while only smoothing out the active-inactive transition, confirming robustness to the choice of activation function.

4.2 Frequency response curve for all neurons active

Perform a Fourier transformation on both sides of the equation with respect to neuronal positions on the ring.

$$\begin{aligned} \hat{\delta r}_E &= \hat{\delta I}_E + \hat{k}_{EE} \odot \hat{\delta r}_E - \hat{k}_{EI} \odot \hat{\delta r}_I, \\ \hat{\delta r}_I &= \hat{\delta I}_I + \hat{k}_{IE} \odot \hat{\delta r}_E, \end{aligned} \tag{18}$$

Here, the Fourier transformation is in the form

$$\hat{x} = \sum_{n=0}^{N-1} x[n] \exp(-ik\omega_0 n), \omega_0 = \frac{2\pi}{N}. \tag{19}$$

The perturbed system can be solved now directly in the frequency space, and the solution is

$$\begin{aligned} \hat{\delta r}_E &= (\hat{\delta I}_E - \hat{k}_{EI} \odot \hat{\delta I}_I)(1 - \hat{k}_{EE} + \hat{k}_{EI} \odot \hat{k}_{IE})^{-1}, \\ \hat{\delta r}_I &= (\hat{k}_{IE} \odot \hat{\delta I}_E + \hat{\delta I}_I - \hat{k}_{EE} \odot \hat{\delta I}_I)(1 - \hat{k}_{EE} + \hat{k}_{EI} \odot \hat{k}_{IE})^{-1}. \end{aligned} \tag{20}$$

The singular values are given by $(1 - \hat{k}_{EE} + \hat{k}_{EI} \odot \hat{k}_{IE})^{-1}$, with the frequency order. As we have mentioned before, the singular values against frequency can be classified into several cases. Here, we give the specific classification basis when the kernels are given as follows:

$$k_{XY} = \alpha_{XY} e^{-x^2/2\sigma_{XY}^2}, \tag{21}$$

$X, Y \in \{E, I\}$. With these Gaussian kernels, we have them in frequency space as follows:

$$\hat{k}_{XY}(\xi) = \frac{N}{T} \alpha_{XY} \sqrt{2\pi} \sigma_{XY} e^{-2\pi^2 \sigma_{XY}^2 \xi^2 / T^2}, \tag{22}$$

where N is the number of neurons, and T is the period, which takes 180 here.

Denote $\tilde{\alpha}_{XY} = \frac{N}{T}\alpha_{XY}$, we care about the change of eigenvalues related to the kernel's parameters. We have

$$\begin{aligned} \hat{h}_0 &= 1 - \hat{k}_{EE} + \hat{k}_{EI}\hat{k}_{IE} \\ &= 1 - \tilde{\alpha}_{EE}\sqrt{2\pi}\sigma_{EE} e^{-2\pi^2\sigma_{EE}^2\xi^2/T^2} \\ &+ 2\pi\tilde{\alpha}_{EI}\tilde{\alpha}_{IE}\sigma_{EI}\sigma_{IE} e^{-2\pi^2(\sigma_{EI}^2+\sigma_{IE}^2)\xi^2/T^2}, \end{aligned} \tag{23}$$

There are three key quantities here.

- $(\sigma_{EI}^2 + \sigma_{IE}^2)/\sigma_{EE}^2$. We can view $\sigma_{EI}^2 + \sigma_{IE}^2$ as the scope for the recurrent inhibitory lateral connection and σ_{EE} for the excitatory one. Thus, this quantity determines which part has a wider scope.
- $\tilde{\alpha}_{EE}\sigma_{EE}/(\tilde{\alpha}_{EI}\tilde{\alpha}_{IE}\sigma_{EI}\sigma_{IE}$ determines whether all the frequencies are uniformly enhanced or suppressed.
- $(\tilde{\alpha}_{EE}\sigma_{EE}^3)/(\tilde{\alpha}_{EI}\tilde{\alpha}_{IE}\sigma_{EI}\sigma_{IE}(\sigma_{IE}^2 + \sigma_{EI}^2))$ determines whether the zero frequency is enhanced/suppressed the most.

It appears that a condition is missing here. Additionally, more illustrative diagrams could be added to facilitate a thorough discussion of different scenarios. Moreover, references to this section should be incorporated into the earlier content.

4.3 Frequency response curve for part of neurons active

Perform a Fourier transformation on both sides of the equation with respect to neuronal positions on the ring

$$\begin{aligned} \hat{\delta r}_E &= \hat{g}_E' * (\hat{\delta I}_E + \hat{k}_{EE} \odot \hat{\delta r}_E - \hat{k}_{EI} \odot \hat{\delta r}_I), \\ \hat{\delta r}_I &= \hat{g}_I' * (\hat{\delta I}_I + \hat{k}_{IE} \odot \hat{\delta r}_E), \end{aligned} \tag{24}$$

4.4 Gabor filters

We are interested in how the model is sensitive to changes in images and in the gap between a signal and an image, or vice versa. We bridge the gap with Gabor filters. A Gabor filter is constructed as

$$\mathcal{F}_{\mathcal{G}} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix}, \tag{25}$$

where $g_i = g^{(i-1)\omega}, i = 1, 2 \cdots n$, $(\omega = 2\pi/n)$ are row vectors, generated by flattening discretized 2-d Gabor functions with the expression

$$\begin{aligned} g^\theta &= g(x, y; A, \lambda, \theta, \psi, \sigma, \gamma) \\ &= A \exp(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}) \cos(2\pi f x' + \psi), \end{aligned} \tag{26}$$

where

$$\begin{aligned} x' &= x\cos(\theta) + y\sin(\theta), \\ y' &= -x\sin(\theta) + y\cos(\theta). \end{aligned} \tag{27}$$

4.5 Steady-state ring model for given input

In the above discussion, we focused mainly on perturbations; that is, we did not consider the inputs. In this section, we will describe how to extend the previous steady-state ring model to create a mapping from image input to input. We used the following model:

$$\begin{aligned} I_E &= \mathcal{F}_{\mathcal{G}} x_{\text{image}}, \\ r_E &= g(I_E + k_{EE} * r_E - k_{EI} * r_I + b_E) \\ r_I &= g(k_{IE} * r_E) \end{aligned} \tag{28}$$

where $\mathcal{F}_{\mathcal{G}}$ is the Gabor filter and b_E is the bias of the model. Here, we use a model with added bias, compared to the model from the theoretical analysis. However, this change does not affect the theoretical analysis, as it is based on the amount of variation; the fixed constant is eliminated after the actual input. At the same time, the bias values are hyperparameters that control how many neurons can be activated by an input. This is similar to the fact in biology that neurons are not activated by tiny stimuli, but only when the stimulus reaches a certain intensity.

For an input, we use numerical methods to solve the above equations. After completing the solution, we can obtain the neuron's activation, and with the fixed activation, the perturbation analysis can be performed as described before.

4.6 I&F neuronal ring model

In this study, we consider a fully connected network with conductance-based, integrate-and-fire neuron (Obeid and Miller, 2021). The population consists of $N = 600$ neurons, each labeled by its orientation preference, $\theta_k = 0.3k$ degrees, which forms a ring. The dynamics of each neuron in the network are modeled by the leaky integrate-and-fire equation:

$$\tau_m \frac{\mathrm{d}V}{\mathrm{d}t} = -(V - R_L) + \frac{g_E}{g_L}(R_E - V) + \frac{g_I}{g_L}(R_I - V), \tag{29}$$

where τ_m denotes the time constant, g_L is the leak conductance, g_E and g_I are the time-dependent excitatory and inhibitory conductances, and R_L, R_E, R_I are reversal potentials. When the neuron's membrane potential reaches threshold V_{th}, the neuron generates a spike, and the membrane potential returns to the resting potentialV_{rest} and remains at the resting potential until the end of refractory period τ_{ref}. Biophysical parameters are used: $g_L = 10nS, R_L = -70mV, R_E = 0mV, R_I = -80mV, V_{th} = -50mV, V_{\text{rest}} = -56mV, \tau_m = 15ms, \tau_{ref} = 0ms$. For any neuron n of type $X \in \{E, I\}, g_E, g_I \geq 0$ are its excitatory and inhibitory conductances governed by

$$\frac{\mathrm{d}g_E}{\mathrm{d}t} = -g_E/\tau_E + \sum_{b=1}^{N_E} \sum_j W_{XE}\delta(t - t_{Ej}) + \sum_k g_{X,\text{ext}}\delta(t - t_{\text{ext},k}), \tag{30}$$

$$\frac{\mathrm{d}g_I}{\mathrm{d}t} = -g_I/\tau_I + \sum_{b=1}^{N_I} \sum_j W_{XI}\delta(t - t_{Ij}), \tag{31}$$

where $\tau_E = 3$ms and $\tau_I = 3$ms are decay rates for excitatory and inhibitory conductances, respectively. And $W_{XY} = w * g_{XY}$, where $w = A + B * \exp\left(-(\theta_{\text{pre}} - \theta_{\text{post}})^2/(2\sigma_{ori}^2)\right)$, and θ_{pre} is the reference angle for pre-synaptic neurons, θ_{post} is the reference angle for post-synaptic neurons, $\sigma_{ori}$is the width of the Gaussian kernel, and A, B are constants satisfying $A + B = 1$. Synaptic inputs from other neurons within the network are described in the second terms on the right sides of Equations 29, 30 $t_{Ej}, t_{Ij}$are the spike times of all the E- and I-neurons pre-synaptic to neuron n. And external synaptic input is described by a Poisson sequence with rates r_{ext} that arrives at neuron n in $t_{\text{ext},k}$. And $\delta(\cdot)$ is the Dirac delta function indicating an instantaneous jump of conductance g_E or g_I upon the arrival of an E- or I- or external spike, with amplitude equal to g_E, g_I, g_{ext} respectively. In addition, our Poisson groups and E- and I-neurons are all one-to-one connected (i.e., we have $2N$ Poisson neurons), and the rates of the Poisson groups of neurons with corresponding preferential angles satisfy the following equation:

$$r_{\text{ext}} = A * cos(2\pi\theta f) + C, \tag{32}$$

where A represents the amplitude of the fluctuation, f represents the frequency of the fluctuation, θ is the preference angle of the neuron, and C is the strength of the base frequency inputs. This equation describes how the input changes spatially, i.e., we can generate stimuli of different spatial frequencies. At the same time, we ensure that the input mean is constant and its variance can be adjusted, i.e., the product of r_{ext} and g_{ext} should be constant.

If we consider the mean-driven regime mentioned above, we need:

$$N \to +\infty, \ \ g_{XY} \to 0, \ \ N_{input} \to +\infty,$$

which means we tune the network primarily based on the mean of the inputs rather than their fluctuations, i.e., the input variance tends to 0. With these parameters, the network will experience even less randomness, reducing fluctuations and bringing it closer to our theoretical analysis.

Finally, we verified how well the model parameters match biological phenomena, as shown in Figure 9. We demonstrated that the model can reproduce biological observations, such as lateral inhibition and winner-take-all behavior, which are essential for tasks like orientation selectivity.

4.7 Measurement of spiking neuron ring model at different input frequencies

Our goal is to test how the ring model's response varies with different frequency inputs under different connectivity conditions (i.e., connection strengths and spatial extents between excitatory and inhibitory neurons). Therefore, we chose to benchmark the ring model's response at different frequencies, without any connections, to test the variation across different connections. Various measurements are accomplished through the following process:

1. Vary the amplitude of the fluctuation A when there is no connection and test the change in the issuance rate.

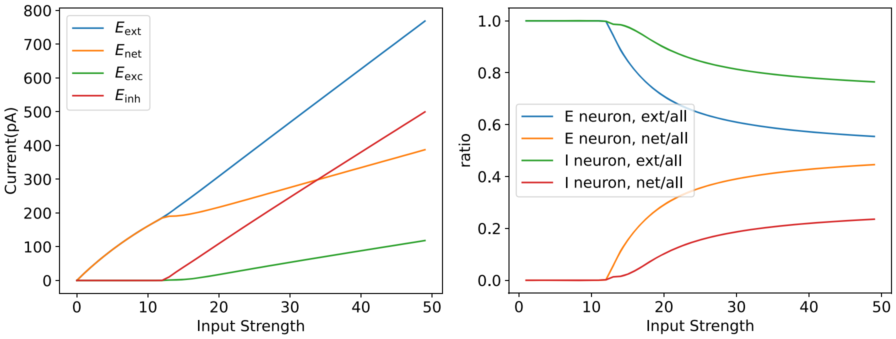

FIGURE 9
Validation of parameter rationality. The figure on the **left** illustrates the variation of the network current as the input strength varies. The figure on the **right** shows the internal network current versus the external input current as a percentage of the total network current.

2. At the same time, take the discrete Fourier transform of the change in response and the change in the input current to obtain $\widehat{\Delta r}/\widehat{\Delta I}$.
3. Change the amplitude of the fluctuation A again, but with the corresponding connection, and test the change in the rate of issuance.
4. Take a discrete Fourier transform of the new change in the issuance rate versus the change in input current to get $\widehat{\Delta r'}/\widehat{\Delta I'}$.
5. Comparing the two ratios, that is, we end up with $\widehat{\Delta r'}/\widehat{\Delta r}$.

Therefore, in our experiments, we mainly examine the relationship between A and $\widehat{\Delta r'}/\widehat{\Delta r}$ as a subject of analysis, which is similar to that analyzed by the steady-state rate model.

4.8 Parameter correspondence in mean-driven regime

Since the steady-state rate model is a dimensionless model, we need to consider its parametric correspondence to the actual model with the following equations:

$$r_E = g(I_E + k_{EE} * r_E - k_{EI} * r_I), \tag{33}$$

$$r_I = g(I_I + k_{IE} * r_E), \tag{34}$$

where $k_{XY} = \alpha_{XY} \exp(-x^2/(2\sigma^2))$. We fix $g = 1$ and set r_E, r_I to match the experimental results in units of Hz. We then calculate I_E, I_I, α_{XY} in the steady-state rate model.

To simplify the description, we denote I_E and I_I as the current of the theoretical model, denote I'_E and I'_I as the current of the spiking neuron model, and denote $\bar{r_E}$ and $\bar{r_I}$ as the mean of all neuron firing rates (only the spiking neuron model needs to be averaged here, and this is due to the randomness it introduces at the time of the experiment). We will complete the correspondence of the parameters by the following process:

1. Determine the correspondence of I_E, I_I in the absence of any connection: $r_E = I_E = kI'_E + b$.
2. We compute α_{EE} when there is only an E-to-E coupling. We consider $r_E = I_E + k_{EE} * r_E$, and we have
$$\alpha_{EE} = \frac{\bar{r_E} - (kI'_E + b)}{\bar{r_E} \int_0^{180} e^{-x^2/(2\sigma^2)} dx}.$$
3. We compute α_{IE} when there is only an I to E coupling. We consider $r_E = I_E - k_{IE} * r_I$, and we have
$$\alpha_{IE} = \frac{\bar{r_E} - (kI'_E + b)}{\bar{r_I} \int_0^{180} e^{-x^2/(2\sigma^2)} dx}.$$
4. We compute α_{EI} when there is only an E to I coupling. We consider $r_I = I_I + k_{EI} * r_E$, and we have
$$\alpha_{EI} = \frac{\bar{r_I} - (kI'_I + b)}{\bar{r_E} \int_0^{180} e^{-x^2/(2\sigma^2)} dx}.$$

Finally, we can reasonably compare the steady-state rate model with the distribution model.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

JY: Conceptualization, Methodology, Investigation, Software, Formal analysis, Visualization, Writing – original draft. YF: Investigation, Software, Formal analysis, Visualization, Writing – original draft. WD: Formal analysis, Supervision, Writing – review & editing, Funding acquisition. YZ: Conceptualization, Methodology, Supervision, Writing – review & editing, Funding acquisition.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Science and Technology Innovation 2030–Brain Science and Brain-Inspired Intelligence Project Grant No. 2021ZD0201301 (WD), the National Natural Science Foundation of China Grant No. 12201125 (WD), the Natural Science Foundation of China Grant No. 12571567 (YZ), the Natural Science Foundation of Shanghai Grant No. 25ZR1402280 (YZ), and the Science and Technology Committee of Shanghai Municipality Grant No. 22YF1403300 (WD).

Conflict of interest

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.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. Generative AI was used in the preparation of this manuscript solely for language polishing and grammar refinement. All scientific content, data analysis, and conclusions were conceived, developed, and verified entirely by the author(s). The author(s) take full responsibility for the accuracy and integrity of the manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncom.2025.1699179/full#supplementary-material

References

Alderson, D. L., and Doyle, J. C. (2010). Contrasting views of complexity and their implications for network-centric infrastructures. *IEEE Trans. Syst. Man Cybern.-Part A* 40, 839–852. doi: 10.1109/TSMCA.2010.2048027

Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc. Nat. Acad. Sci.* 92, 3844–3848. doi: 10.1073/pnas.92.9.3844

Bressloff, P. C., and Cowan, J. D. (2002). The visual cortex as a crystal. *Physica D* 173, 226–258. doi: 10.1016/S0167-2789(02)00677-2

Cai, D., Tao, L., Shelley, M., and McLaughlin, D. W. (2004). An effective kinetic representation of fluctuation-driven neuronal networks with application to simple and complex cells in visual cortex. *Proc. Nat. Acad. Sci.* 101, 7757–7762. doi: 10.1073/pnas.0401906101

Carandini, M., and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62. doi: 10.1038/nrn3136

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi: 10.1016/j.neuron.2012.01.010

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

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. London: MIT press.

Goodfellow, I. J., Shlens, J., and Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

Keller, G. B., and Mrsic-Flogel, T. D. (2018). Predictive processing: a canonical cortical computation. *Neuron* 100, 424–435. doi: 10.1016/j.neuron.2018.10.003

Kitano, H. (2004). Biological robustness. *Nat. Rev. Genet.* 5, 826–837. doi: 10.1038/nrg1471

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

Mastrogiuseppe, F., and Ostojic, S. (2018). Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron* 99, 609–623.e29. doi: 10.1016/j.neuron.2018.07.003

Obeid, D., and Miller, K. D. (2021). Stabilized supralinear network: Model of layer 2/3 of the primary visual cortex. *BioRxiv, 2020-12*. doi: 10.1101/2020.12.30.424892

Rubin, D. B., Van Hooser, S. D., and Miller, K. D. (2015). The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* 85, 402–417. doi: 10.1016/j.neuron.2014.12.026

Said, C. P., and Heeger, D. J. (2013). A model of binocular rivalry and cross-orientation suppression. *PLoS Comput. Biol.* 9:e1002991. doi: 10.1371/journal.pcbi.1002991

Somers, D. C., Nelson, S. B., and Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.* 15, 5448–5465. doi: 10.1523/JNEUROSCI.15-08-05448.1995

Sompolinsky, H., and Shapley, R. (1997). New perspectives on the mechanisms for orientation selectivity. *Curr. Opin. Neurobiol.* 7, 514–522. doi: 10.1016/S0959-4388(97)80031-1

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature* 571, 361–365. doi: 10.1038/s41586-019-1346-5

Wang, Z., Dai, W., and McLaughlin, D. W. (2020). Ring models of binocular rivalry and fusion. *J. Comput. Neurosci.* 48, 193–211. doi: 10.1007/s10827-020-00744-7

Wilson, H. R. (2017). Binocular contrast, stereopsis, and rivalry: toward a dynamical synthesis. *Vision Res.* 140, 89–95. doi: 10.1016/j.visres.2017.07.016