Featured in Physics

Mechanisms for Spontaneous Symmetry Breaking in Developing Visual Cortex

Francesco Fumarola*
Laboratory for Neural Computation and Adaptation, RIKEN Center for Brain Science, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

Bettina Hein
Center for Theoretical Neuroscience, College of Physicians and Surgeons and Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, New York 10027, USA

Kenneth D. Miller†
Center for Theoretical Neuroscience, Department of Neuroscience, Swartz Program in Theoretical Neuroscience, Kavli Institute for Brain Science, College of Physicians and Surgeons and Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, New York, USA

(Received 8 October 2021; revised 3 May 2022; accepted 28 June 2022; published 11 August 2022)

For the brain to recognize local orientations within images, neurons must spontaneously break the translation and rotation symmetry of their response functions—an archetypal example of unsupervised learning. The dominant framework for unsupervised learning in biology is Hebb's principle, but how Hebbian learning could break such symmetries is a longstanding biophysical riddle. Theoretical studies argue that this requires inputs to the visual cortex to invert the relative magnitude of their correlations at long distances. Empirical measurements have searched in vain for such an inversion and report the opposite to be true. We formally approach the question through the Hermitianization of a multilayer model, which maps it into a problem of zero-temperature phase transitions. In the emerging phase diagram, both symmetries break spontaneously as long as (i) recurrent interactions are sufficiently long range and (ii) Hebbian competition is duly accounted for. A key ingredient for symmetry breaking is competition among connections sprouting from the same afferent cell. Such a competition, along with simple monotonic falloff of input correlations with distance, is capable of triggering the broken-symmetry phase required by image processing. We provide analytic predictions on the relative magnitudes of the relevant length scales needed for this novel mechanism to occur. These results reconcile experimental observations to the Hebbian paradigm, shed light on a new mechanism for visual cortex development, and contribute to our growing understanding of the relationship between learning and symmetry breaking.

DOI: 10.1103/PhysRevX.12.031024

Subject Areas: Biological Physics, Interdisciplinary Physics, Statistical Physics

I. INTRODUCTION

The primary visual cortex (V1)—the first receiving area in the cerebral cortex for visual sensory information—receives signals from the lateral geniculate nucleus of the thalamus (LGN), which in turn receives signals directly from the eyes. Both LGN and V1 extend in two dimensions so as to embody a continuous map of the world as seen through the two eyes. In other words, these brain regions are arranged "retinotopically," each neuron responding to input in the vicinity of a certain point on the retina, with neighboring areas of the retina represented by neighboring neural areas. Moreover, cells in the LGN can be excited either by light onset or by light offset on the corresponding spot of the retina, and cells of these two types are known, respectively, as ON-center and OFF-center cells.

*francesco.fumarola@riken.jp
†kdm2103@columbia.edu

We think of the instantaneous visual stimulus as pixel values for each position in the two-dimensional retinotopic space. The "receptive field" (RF) of a visual neuron is defined as the linear kernel that best determines its activity as a function of the visual stimulus. LGN cells, like retinal cells [1,2], have RFs that are roughly circularly symmetric [3], while V1 neurons best respond to a particular orientation of a light or dark edge [4,5]. Furthermore, in response to a drifting periodic luminance grating, the temporal mean of the total LGN input to a V1 cell is untuned for orientation, while the size of the temporally periodic modulation about the mean of this input is orientation tuned [6,7], which indicates that deviations from circular symmetry in the RFs of individual LGN cells do not contribute appreciably to the orientation selectivity of V1 cells [8]. Instead, rotational symmetry is broken by the spatial arrangement of the set of LGN cells that make synaptic connections onto a given V1 neuron. As originally postulated by Hubel and Wiesel [5], this spatial arrangement appears to involve spatially adjacent subregions alternating between ON-center and OFF-center inputs [6,9] representing adjacent retinotopic subregions in which light or dark stimuli, respectively, best drive the V1 cell. Here we focus on understanding the origin of this spatial arrangement that breaks rotational symmetry and endows V1 cells with selectivity for the orientation of a light or dark edge. We note that this orientation selectivity is distinct from direction selectivity—selectivity for the direction of motion of such an edge or of other stimuli. The requirements for development of direction selectivity are quite different from those for development of orientation selectivity [10,11].

The orientation preferences of V1 cells are locally continuous and rotate roughly periodically with movement along the two dimensions representing retinotopy in all species that have been studied outside of rodents and lagomorphs (Ref. [12]; reviewed in Refs. [13,14]). This arrangement of orientations over cortical space is known as an "orientation map." We also address the breaking of translational symmetry that is necessary for such a map.

The sensitivity to local orientation develops in V1 cells even when the animal is denied visual experience by rearing in darkness [12,15–17] but depends on normal patterns of spontaneous activity (activity without vision, e.g., in the dark) in the LGN and V1 [18,19]; reviewed in Ref. [20]. For this and other reasons, V1 orientation selectivity is thought to arise from a process of activity-dependent self-organization, most likely instructed by the spontaneous activity patterns [13,21]. The problem is thus placed within the general framework of symmetry breaking during learning, a branch of theoretical physics that recently achieved substantial progress [22–26].

Activity-dependent self-organization of neural connections is commonly thought to be guided by "Hebbian" rules of synaptic plasticity—proposed independently by Hebb [27] and Konorski [28]—in which synapses are strengthened by temporal correlation between their pre- and postsynaptic patterns of activity ("neurons that fire together, wire together") [29–31]. This mechanism provides a concrete biophysical basis for the notion dating back to the 19th century [32–34], that associating information is the quintessential brain function. The development of orientation selectivity and other related selectivity and organization in V1 was one of the earliest testing grounds for Hebbian learning [35–38].

Many authors have proposed models of the origin of V1 orientation selectivity guided by various forms of Hebbian plasticity given various forms of structured input. Many relied on visual input from natural scenes [39–45] or oriented bars [46,47], which cannot explain development in the absence of visual experience. Others considered retinal waves [48] or various combinations of these different input ensembles [49–51]. Waves of activity travel across retina in early spontaneous activity before vision [52], and the boundaries of these waves are postulated to serve as oriented edges. However, propagating retinal waves very largely [53] or entirely [54] disappear before the major development of V1 orientation selectivity. For example, in ferrets, which are often used for developmental studies because they are born relatively early in development [18], the waves have disappeared before postnatal day 30 (P30) [54], whereas the major development of orientation selectivity occurs after P30 [18,55].

Another study proposed that early cortical activity may occur in large elongated "coarse" patterns, at least several millimeters long and $\tilde{1}$ mm wide (which patterns have not been seen, but see Ref. [56]), while on a fine scale (within 1 mm), activity is confined to periodic patches [57]. Then, under a Hebbian mechanism, intracortical connections will come to connect fine patches along a long line, with different local fine patches connected along differently oriented lines. This could create orientation selectivity for long stimuli that extend several millimeters across the cortex and would provide a scaffolding for the development of finer-scale orientation selectivity, but no mechanism was proposed for the latter development except possibly learning from natural scenes.

Others have shown how normative principles such as sparse coding or independent components analysis can yield orientation selectivity, again assuming that these principles act on an ensemble of natural scenes, oriented bars, or retinal waves [47,58–62]. Such normative principles do not address the biological mechanism by which orientation selectivity develops, although several groups have shown that biologically plausible local learning rules could implement such normative principles and produce orientation selectivity from such input ensembles [40–43,45,47,49]. But, like the other models discussed above, these approaches have not addressed the development of orientation selectivity from the spontaneous activity that exists after retinal waves disappear.

There is weak orientation selectivity in a minority of cells as early as has been studied, which could result simply from random connectivity, but their number and degree of selectivity do not change, and no organization into orientation maps is observed, until after P30 in ferrets [18,55]. It is not impossible that retinal waves guide the development of these early orientation-selective cells and that they somehow seed the rest of the map once development is biologically opened, but this gives no explanation for how this guidance occurs. The most straightforward interpretation is that the development of orientation selectivity is guided by the spontaneous activity that exists after P30 and that this development is independent of, and overwrites, any weak initial orientation selectivity. Here we address the question of how orientation selectivity could develop under this interpretation.

One of us proposed an answer several decades ago, namely, that orientation selectivity and its local continuity across the cortex could arise through Hebbian dynamics from a so-called Mexican-hat profile of input correlations, in which same-center-type inputs are more correlated than opposite-center type at short separations and the converse is true at longer separations [63,64]. This correlation structure was expected because the RFs of ON cells consist of a circular ON (light-preferring) center and an OFF (dark-preferring) surround that forms a ring about the center [1–3], and similarly, OFF cells have an OFF center and an ON surround. Thus, at small separations where centers strongly overlap, cells of a given type should be best correlated with others of their own type, while at larger separations where the center of one cell overlaps the surround of the other cell, cells of a given type should be best correlated with others of the opposite type. Related models include that of Ref. [36] analyzed in more detail by Refs. [65–67] (see Ref. [64] for a discussion of the relationships between these models).

With this premise, the model of Ref. [64] predicts that orientation selectivity will arise via activity-dependent competition between ON- and OFF-center inputs. Hebbian plasticity leads an individual cell to receive a well-correlated set of inputs. This yields a set of inputs to each cell that alternates between inputs of one center type and the other, with a spatial period corresponding to the alternation between same-type and opposite-type pairs being best correlated. In addition, local excitatory connections between cortical cells lead nearby cells to develop similar preferred orientation and similar absolute spatial phase (the location in visual space of light-preferring or dark-preferring subregions), eliminating high spatial frequencies in the map of orientations. At larger distances, different orientations arise from the random initial conditions, so that a low-pass map arises. The actual maps are bandpass (periodic) rather than low pass [13]. Thus, the model does not produce realistic map structure but does show the breaking of translational symmetry and elimination of high spatial frequencies seen in real maps.

This and similar theories of Hebbian development of orientation selectivity were further studied mathematically by Refs. [68,69]. In Ref. [69], a rigorous proof was worked out that, indeed, "in order to get orientation-selective receptive fields, the spatial correlation function of the inputs that drive the development must have a zero crossing."

This scenario, however, was called into question by direct measurements of correlations in LGN activity as a function of retinotopic distance [70]. Experiments on young ferrets, at the ages over which orientation selectivity develops, found no Mexican-hat correlation structure in the LGN. Instead, same-type pairs were more strongly correlated than opposite-type pairs at all retinotopic separations, and the decay with distance was monotonic. Those authors suggested that such a correlation function could yield development of orientation selectivity if there were a constraint ensuring that a postsynaptic cell received equal numbers of ON-center and OFF-center input (a similar constraint played an important role in Ref. [36] as analyzed in Refs. [65,66]). Without this constraint, a best-correlated set of inputs would be all of a single-center type, but given the constraint, it instead consists of ON and OFF inputs each filling one of two adjacent subregions of visual space, yielding orientation selectivity. However, no molecular cue distinguishing ON-center and OFF-center LGN inputs has ever been demonstrated, so that it seems likely that the two center types are distinguished only by their patterns of activity. If so, then there is no biological basis for such a constraint; instead, the percentages of ON or OFF inputs received by a cell will be the outcome of activity-dependent plasticity, rather than being determined *a priori*.

In sum, despite many points of uncertainty in our knowledge, we believe that the key outstanding question for understanding the development of V1 orientation selectivity is how it can arise from local rules of synaptic plasticity guided by spontaneous activity with the monotonically decaying correlation functions measured by Ohshiro and Weliky [70], without any constraint on the percentage of ON or OFF inputs received by individual V1 cells. Here we provide an answer to this question, including a complete analysis of the model's phase space to show where rotational symmetry is broken, yielding orientation selectivity, and where translational symmetry is broken, a necessary condition for orientation maps to form.

II. RESULTS

We address the problem through the model illustrated in Fig. 1, in which V1 is represented by a single postsynaptic layer and the LGN by two overlapping layers containing ON-center and OFF-center cells, respectively. The layers are modeled as infinite planes endowed with a retinotopic metric and inhabited by a continuum of cells.

Synapses from a given afferent cell of either type are said to belong to an "arbor" [71], and the number of them

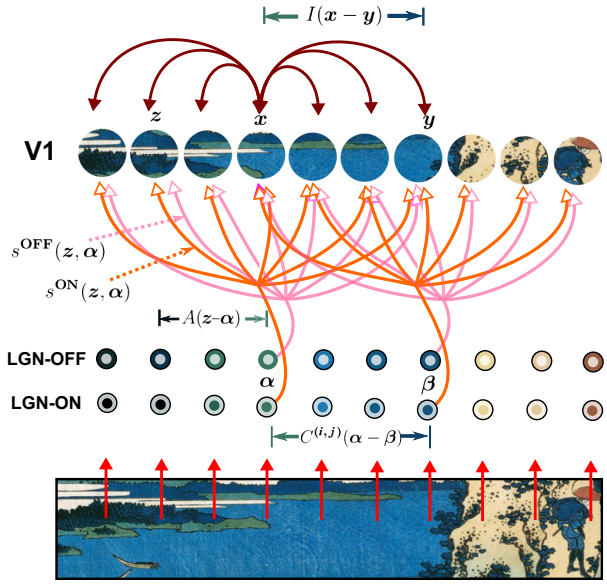

FIG. 1. Schematic depiction of the model. Primary visual cortex (V1) is the first receiving area in the cerebral cortex for visual sensory information. Here it is represented by a single postsynaptic layer of cells depicted by the upper row of larger circles. Inputs to V1 come from the lateral geniculate nucleus of the thalamus (LGN), which in turns takes input from the eyes. LGN cells are excited either by light onset or by light offset on the corresponding spot of the retina, and these two types of cells are represented by the two rows of smaller circles. The quantity of synapses from a LGN location $\boldsymbol{\alpha}$ to a V1 location $\boldsymbol{z}$ is described by an "arbor density" $A(\boldsymbol{z}-\boldsymbol{\alpha})$. Their total synaptic strength is defined as either of the functions $s^{\mathrm{ON}}(\boldsymbol{z},\boldsymbol{\alpha})$ or $s^{\mathrm{OFF}}(\boldsymbol{z},\boldsymbol{\alpha})$ depending on whether the presynaptic cell is of the ON or OFF type. Correlations between the activities of two presynaptic cells of types $(i)\in\{\mathrm{ON},\mathrm{OFF}\}$ and $(j)\in\{\mathrm{ON},\mathrm{OFF}\}$ located at retinotopic positions $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are described by a set of functions $C^{(i,j)}(\boldsymbol{\alpha}-\boldsymbol{\beta})$. The effect via lateral connections (brown arrows) of activity at cortical position $\boldsymbol{x}$ on activity at cortical position $\boldsymbol{y}$ is characterized by the function $I(\boldsymbol{x}-\boldsymbol{y})$.

targeting a given cortical cell is a function $A(\boldsymbol{r})$ of the retinotopic distance vector $\boldsymbol{r}$ between the two cells, decaying over a length scale ρ ("arbor width"). In addition, cells internal to the cortical layer are connected by recurrent synapses whose connection strength decreases with retinotopic distance over a length scale η. [Note that, for both analytical and conceptual simplicity, we take the intracortical interactions to be purely excitatory and monotonically decaying with distance, although their nature biologically is not determined. Anatomically, connections of excitatory cells are more wide ranging than those of inhibitory cells (reviewed in Ref. [72]), but biologically plausible means by which disynaptic inhibition could create functionally "Mexican-hat" interactions have been proposed [73].]

Correlations between the activities of two presynaptic cells of types $(i)=\mathrm{ON}/\mathrm{OFF}$ and $(j)=\mathrm{ON}/\mathrm{OFF}$ are described by a function $C^{(i,j)}(\boldsymbol{r})$ of the retinotopic distance vector $\boldsymbol{r}$. This function must be monotonically decreasing by the experimental results discussed in the Introduction, and may be thus assumed to be a Gaussian with a given correlation length ζ. A similarly Gaussian ansatz is adopted for the arbor density function (with length scale ρ) and for the strength of lateral connections between cortical neurons (with length scale η), as detailed in Appendix A 1. The phase diagram can thus be plotted using for coordinates any pair of dimensionless ratios among the three parameters ρ, η, ζ.

Let s^{ON} and s^{OFF} be the synaptic strength of connections from ON and OFF cells to the cortex, as per Fig. 1. Taking synaptic plasticity to be slow on the timescale of neural activity, we obtain independent equations for their sum $s^{(S)}\equiv s^{\mathrm{ON}}+s^{\mathrm{OFF}}$ and difference $s^{(D)}\equiv s^{\mathrm{ON}}-s^{\mathrm{OFF}}$ (see Appendix A 1). Our interest is in the development of orientation selectivity via the formation of alternating RF subregions in which ON or OFF LGN inputs, respectively, are dominant. Hence, we are interested in the development of a pattern in $s^{(D)}$, while $s^{(S)}$ is not expected to form interesting structure. Thus, we focus on computing $s\equiv s^{(D)}$ and the symmetry properties of this function.

The possible outcomes are easily inferred from first principles. A fundamental invariance law for models of brain vision is that a simultaneous translation or rotation of the animal and of the image it views will not affect brain activity [74]. More specifically, a simultaneous translation or rotation of the primary visual cortex and of its input source should not affect the resulting RFs. These two symmetries (translation and rotation) lead to three relevant symmetry classes of the solution detailed in Table I. It remains to be understood which phases are allowed by biological mechanisms.

In the model, we like to incorporate the fact that correlation-based development is competitive, meaning that when some synapses grow stronger, others grow weaker. Beginning with the first model of activity-dependent development of von der Malsburg [35], theorists have often modeled this competition as a constraint that the total synaptic weight received by a neuron is conserved. Alternatively, it can be modeled as a homeostatic process that maintains the average postsynaptic firing rate of neurons about some set point, as is seen experimentally [75]. Such a constraint upon the summed postsynaptic strength will alter the equation for the sum $s^{(S)}$ but not for the difference $s^{(D)}$.

TABLE I. Symmetry classes of the solution. The labels N, R, T are defined in the legend of Fig. 2.

Phase label	Receptive field	Orientation selectivity
N	Uniform across the cortex	Nonselective
R	Uniform across the cortex	Selective
T	Varies across the cortex	Selective

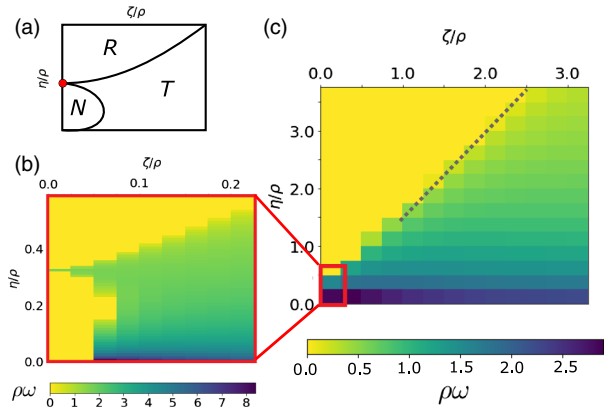

FIG. 2. Phase diagram of the model. (a) Structure of the phase diagram resulting from analytical investigation of the time evolution. The model's parameters are the correlation length ζ of the inputs from the LGN, the radius of outgoing connections ρ, and the length scale η of lateral connections in V1. A phase diagram may use for coordinates any pair of dimensionless ratios among these parameters; here, ζ/ρ and η/ρ are used. The labels N, R, T correspond, respectively, to no symmetry breaking, breaking of rotational symmetry only, and breaking of both translational and rotational symmetry (cf. Table I). The red dot marks the inferred location of a triple point. (b) Dominant wave number of map modulation across the cortex in units of the inverse arbor radius. The preservation of cortically uniform states is confirmed in the regions corresponding to the R and N phases. Clearly visible are the sharp boundaries of the two cortically uniform phases N and R separated by the nonuniform phase T, where orientation maps arise. The plots are obtained by diagonalizing a discretized version of the operator L^p defined by Eq. (A36), where RFs are confined to a square of side equal to six arbor radii represented by a 15×15 grid. (c) Larger-scale view of the phase diagram displaying as a gray dotted line the asymptotic phase boundary $\eta_c \sim \sqrt{2}\zeta$ obtained from an analytic solution of the model.

Since we are concerned with the development of $s^{(D)}$, we ignore such a constraint here.

Another form of competition is that presynaptic axonal arbors compete for postsynaptic connections. In many neural systems, it has been shown that if an arbor loses overall synaptic strength, it competes more effectively to retain it, while if it has too much, it competes less effectively, so that arbor retraction takes place (for reviews, see Ref. [71] Chap. 8 and Ref. [76]). While the physics of competition among LGN arbors for cortical innervation is still unclear, it seems reasonable to assume that, given statistically equal activity of LGN cells, it cannot be the case that some arbors lose most of their innervation to the cortex while others take over. Rather, all arbors should retain roughly equal innervation. This can also be thought of as a homeostatic process, maintaining the overall projection strength of each neuron about some set point.

Since axonal competition will separately constrain the overall strength of ON innervation and of OFF innervation, it will constrain the development of $s^{(D)}$. We model this constraint as a precise conservation of the total strength of synapses projected by each arbor (see Appendix A 1). The introduction of homeostatic constraints, which is demanded by biophysical considerations, considerably complicates the problem. Indeed, such constrained models have been predominantly an object of numerical investigations (for the case of a single V1 cell see also Ref. [77]), and these happened to miss their most relevant potentiality—the emergence of a fully symmetry-breaking and hence orientation-selective and map-forming phase without Mexican-hat-like functions (which drive development of periodic patterns) in either correlations or intracortical interactions.

As detailed in Appendixes A 2 and A 3, after incorporating the above constraints the time-evolution operator can be Hermitianized and thus mapped into a quantum-mechanical Hamiltonian whose low-energy states describe the long-term relaxation of the system. Quantum-mechanical tools allow us to see that all three phases listed in Table I enter the phase diagram, whose structure is illustrated by Fig. 2(a).

For low values of both ζ/ρ and η/ρ (in a lunette extending from the lowest stretch of the η axis), the cortex is in a symmetry-preserving state that we term the "N phase," where the RF is identical and unoriented at all points in retinotopic space. This would mean that, upon eye opening, a given cell's responses could indicate the location of an object in visual space but could not indicate its orientation. In terms of the cortical position $\boldsymbol{x}$ and of the difference between pre- and postsynaptic positions $\boldsymbol{r}$, a variational approximation to the RF in this phase is given by

$$s_N(\boldsymbol{x},\boldsymbol{r}) \propto (R^2 - \boldsymbol{r}^2)\exp\left(-\frac{\boldsymbol{r}^2}{2\hat{\rho}^2}\right), \tag{1}$$

where $\hat{\rho} = \sqrt{2}\rho[1+\sqrt{1+(4\rho^2/\eta^2+\zeta^2)}]^{-1/2}$ and $R \sim (4+\sqrt{10}/3)^{1/2}\rho^{1/2}(\eta^2+\zeta^2)^{1/4}$ (see Fig. 3).

For values of η above a critical value which increases as a function of ζ, the cortex is in a state where rotation symmetry is spontaneously broken at every point ("R phase"). Since by contrast, translation symmetry is not broken, the RFs (including their orientation preferences) are the same everywhere. To an animal with this visual cortex, a pencil slanting at the right angle would be perceivable as such; even shifting it in front of its eyes would pose no hindrance to vision, but a major handicap would emerge if the pencil tilted at a different angle, as all cortical cells would be poorly responsive or unresponsive to this stimulus.

The receptive field in this phase has a degeneracy of order 2, which can be represented with the exact basis (see Fig. 3)

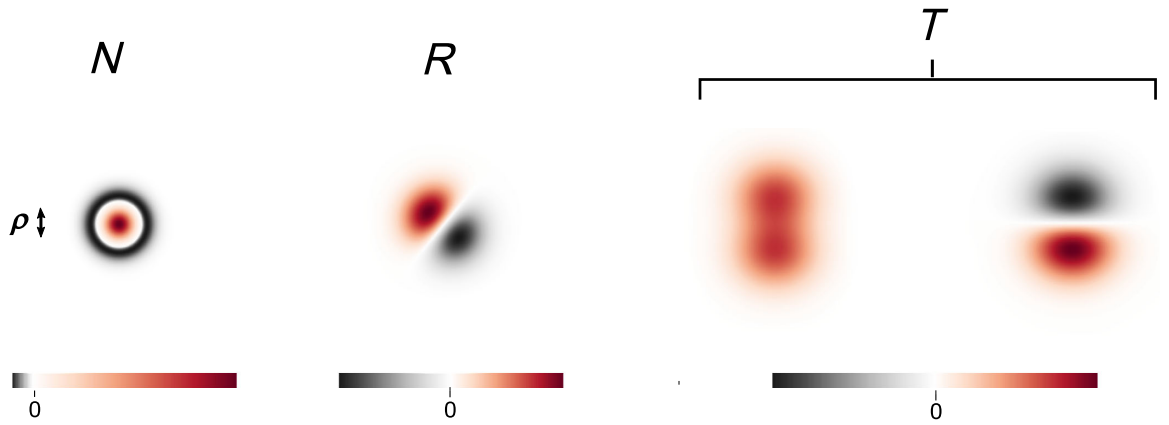

FIG. 3. Receptive fields by symmetry class. Plots of the receptive fields of Eqs. (1), (2), and (3), respectively, for the N, R, and T phases. Hue scale ranges between minimum and maximum values, allowing for an arbitrary scaling factor. Functions are evaluated at representative points $P = (\zeta/\rho, \eta/\rho)$ in parameter space given by $P_N = (0.02, 0.2)$, $P_R = (0.05, 0.7)$, $P_T = (5, 3)$ (see Appendix C for numerical results with the same parameters). The relative length of an arbor radius ρ for all plots is shown to the left of the figure. The receptive field for the R phase is plotted with a randomly chosen orientation. For the T phase, the function s_T of Eq. (3) is rotated by the complex angle $\phi_0 = \arctan(-\int \Im s_T / \int \Re s_T)$ so as to make the imaginary part odd under inversion of the cortical modulation axis (the real part becomes symmetric as a consequence; see Appendix Sec. B 2), and we separately display its real (left) and imaginary (right) components. The fastest-growing mode of time evolution is obtained by multiplying receptive field by a modulating phase factor $e^{i\omega x}$, with x being the coordinate for the degenerate direction of modulation in cortical space.

$$s_R^{x,y}(\boldsymbol{x}, \boldsymbol{r}) \propto \left\{ \begin{matrix} r_x \\ r_y \end{matrix} \right\} \exp\left(-\frac{r^2}{2\hat{\rho}^2}\right). \tag{2}$$

The qualitatively novel phase reported in this paper ("T phase") appears for sufficiently large values of ζ. In the T phase, a double spontaneous breaking of rotation and translation symmetries allows the animal to perceive, in principle, both shifts and rotations of an elongated object. This biologically plausible phase had previously been found only in the presence of finely tailored choices of the input's correlation function Eq. (A13), such as differences between Gaussians of different widths [64,69] or similar [36], which the experiments discussed previously [70] argue against.

A dipole approximation on the relevant time-evolution operator detailed in the Appendix C 2 reveals that the main curve partitioning the diagram (RT boundary) is asymptotically linear far from the origin [$\max(\zeta, \eta) \gg \rho$], and approximated by $\eta_c \sim \sqrt{2}\zeta$ [gray dotted line in Fig. 2(b)]. For interaction lengths above that boundary, the system breaks only rotation symmetry; beneath it, phenomenologically relevant orientation maps emerge.

On the T side of the transition line, the RF is approximated by a linear combination of the real and imaginary parts of the function

$$s_T(\boldsymbol{x}, \boldsymbol{r}) \propto e^{-\frac{r^2}{2\rho^2} - i\frac{\eta^2}{\mu^2}\boldsymbol{\omega}\boldsymbol{r} + i\boldsymbol{\omega}\boldsymbol{x}} \times \left(1 - e^{-\frac{1\zeta^4}{2\mu^4}\rho^2\omega^2 - i\frac{\zeta^2}{\mu^2}\boldsymbol{\omega}\boldsymbol{r}}\right), \tag{3}$$

where $\mu = \sqrt{\eta^2 + \zeta^2}$ and $\omega \equiv |\boldsymbol{\omega}| \sim (\mu/\zeta^2)\sqrt{(2\zeta^2/\eta^2) - 1}$ in the vicinity of the phase boundary.

The long-term behavior described by Eq. (3) is degenerate in the direction of the $\boldsymbol{\omega}$ vector, and will be summed over directions made available by the initial condition. Notice that, because $\hat{\rho} \sim \rho$ for large μ/ρ, at the phase boundary the T-phase functions (3) transform continuously into the projection of the R-phase eigenfunction (2) on the axis parallel to $\boldsymbol{\omega}$. The point where this phase boundary hits the $\zeta = 0$ axis is a triple point at which all three phases coexist.

Finally, the above results are confirmed by numerical investigation. Figure 2(b) shows the results for the dominant wave number of map modulation across the cortex and confirms the preservation of cortically uniform states in the regions corresponding to the R and N phases. The different symmetries of the RFs for the two cortically uniform regions [predicted as in Fig. 2(a)] are readily confirmed by inspection (see Figs. 6, 9, and 10).

We further perform numerical simulations of the equations of motion [Eq. (A6)] both in the absence of any constraint and while imposing the constraint [Eq. (A9)]. Simulations are carried out by evolving random initial conditions and waiting for the configurations to stabilize. The unconstrained case displayed in Fig. 4(e) leads to a fully symmetric solution, with all RFs identical and lacking orientation tuning. For parameters picked in the T phase, dynamical simulations in the presence of the constraint yield symmetry breaking in both s^{ON} [Fig. 4(a)] and s^{OFF} [Fig. 4(b)] and ultimately in the quantity of interest, their difference [Figs. 4(c) and 4(d)].

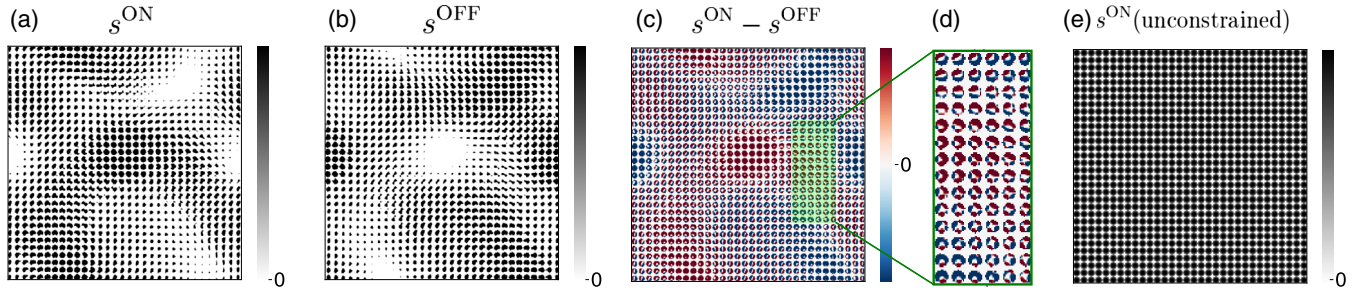

FIG. 4. Results of dynamical simulations. Simulations of the equations of motion [Eqs. (A7)] are performed on discretized receptive fields evolving from random initial conditions while waiting for the configurations to stabilize. We use 32×32 grids of cortical cells and of ON and OFF LGN cells, with periodic boundary conditions. Each LGN cell projects an arbor of inputs to the cortex that is nonzero over a circle of diameter 13 centered on its corresponding point in the cortical grid. (a) Large-time configuration of s^{ON} after evolving the constraint coming from Hebbian competition. Each 13×13 square contains the diameter-13 (or smaller) set of ON weights to one cortical cell. A 32×32 grid of cortical cells is illustrated. (b) As in panel (a) but for s^{OFF}, which yields analogous breaking of the symmetries. (c) Difference between s^{ON} and s^{OFF} of panels (a) and (b) with a portion expanded to highlight cell-by-cell structure [panel (d)]. (e) Large-time state of s^{ON} evolved *without* the competitive constraint [Eq. (A6)] and displaying indeed no trace of either symmetry breaking. In this fully symmetric solution, all RFs are identical and lacking orientation tuning. Parameters used for all these simulations: $\eta/\rho = 0.75, \zeta/\rho = 0.25, \rho = 6.5$.

III. DISCUSSION

Our understanding of unsupervised learning in biological systems relies crucially on Hebb's principle that coactive cells strengthen mutual connections. Visual processing is a classical testing ground for this hypothesis. As we discuss in the Introduction, the development of the quintessential feature of V1, orientation selectivity, depends on neural activity, occurs in the absence of visual experience, and very largely occurs at a time after "retinal waves" have disappeared (see, e.g., Refs. [20,54]). This motivates the hypothesis that this development occurs via Hebbian plasticity based on the structure of spontaneous activity found in LGN and V1 at this time. Existing theories explained this development by postulating that correlations in spontaneous LGN activity have a sign switch: At small separations, cells of a given center type (ON or OFF) are best correlated with other cells of the same type, while at larger distances they are best correlated with cells of the opposite type. However, experimental studies showed no trace of such a switch [70]. This motivated us to reexamine the problem with a focus on the role of competition among LGN inputs to the cortex. We build and solve analytically a model of receptive field development in which the total projection strength from each ON or OFF cell in the LGN is constrained to remain constant. We show that this constrained dynamics is able to produce orientation-selective RFs that vary smoothly across the cortex, even with data-compatible input correlations that decay monotonically with distance, without a sign change.

The multilayer model we employ consists of two bottom layers representing ON or OFF cells in the LGN and a top layer representing V1. The key dimensionless parameters are the widths of LGN correlations and of cortical lateral interactions, in units of the arbor radius. As the model is translationally and rotationally invariant, possible solutions can break either symmetry, or both, or none. Orientation selectivity requires breaking rotation symmetry, while variation of preferred orientations across the cortex requires also the breaking of translation symmetry. We ignore the fact that intracortical connections develop in a pattern that breaks rotational symmetry [78,79], which yields other symmetry classes [74].

The uniformly nonselective phase (where neither symmetry is broken) prevails for sufficiently small values of the width parameters. Rotational invariance is broken for sufficiently long-range cortical interactions, and translational invariance, for sufficiently long-range LGN correlations. Besides calculating the phase boundaries, we estimate the functional form of receptive fields in the various regimes and, for the T phase, which has nonzero cortical wave number, we evaluate explicitly the preferred value of this wave number across parameter space.

Why does orientation selectivity develop without zero-crossing correlations in the inputs to the cortex? The dynamical equations maximize, subject to the constraint, the sum of correlations between synapse pairs, weighted by their synaptic strengths and the cortical interaction between their postsynaptic cells. Previous analyses [64] focused on maximizing correlations within one RF, and implicitly considered interactions between RFs as a perturbation that coordinates the developing orientations and spatial phases between cells. For a maximally correlated RF to be oriented, the sign change in correlations is required. However, when cortical interactions or input correlations are sufficiently long range, interactions between RFs can become equal to or dominant over interactions within RFs. Our results show that, in this case, correlations are maximized, given the constraint, by segregating ON and OFF subregions within RFs so that RF pairs at many

distances can achieve overlaps of same-type subregions. This is favored over the alternative, which is to have a periodic alternation across the cortex of all-ON and all-OFF RFs (with the period as large as possible while satisfying the constraint, as discussed for competition between inputs from the two eyes in Ref. [38]); this would have greater within-RF correlations but smaller overall summed correlations.

The relevance of the study to the current understanding of cortical development lies in the demonstration that orientation selectivity and its smooth variation across the cortex can develop without zero-crossing correlations. This reinstates the applicability of Hebbian development, bridges it with advances in our experimental knowledge of V1, and casts light on the function of competition among axonal arbors for innervation, the biophysical mechanisms of which are still far from being properly understood [80].

We do not consider here proposals that orientation selectivity may arise without instruction by activity from random LGN-to-V1 connectivity [81,82]. We consider these proposals unlikely given the evidence that orientation selectivity and maps depend on and are instructed by activity [20,21] and arise through self-organization [13]. In addition, for one proposal based on the spatial pattern of retinal ganglion cells [81], strong arguments have been given that it cannot account for observations [83,84]. The other proposal aimed specifically at mouse V1 [82], argues that a tight "balancing" of excitation and inhibition cancels a large untuned component of input, leaving a small, random, tuned component. However, there is strong evidence against the existence of such tight balancing [85], and the excitatory input to mouse V1 cells is well tuned for orientation [7], rather than showing a small-tuned component on a large-untuned background.

Our model shares with previous models the nonbiological feature that the developed orientation maps are low pass rather than bandpass (periodic). This is because lateral interactions that maximize correlation between RFs of cells that excite one another, and anticorrelation when one cell inhibits another (e.g., if intracortical interactions have a Mexican-hat structure), lead to periodic changes in the spatial phase of RFs rather than in their preferred orientation [64]. These periodic phase changes maximize the spatial overlap of subregions of the same ON or OFF type to maximize correlations, and of opposite types to maximize anticorrelations. We argue [86] that, if orientation selectivity develops in phase-selective cells (simple cells), periodic maps might develop through interactions between phase-nonselective cells (complex cells) receiving input from simple cells of multiple preferred phases; the influence of the complex cells would propagate back to the phase-selective cells to organize maps even as the phase-selective cells are responsible for the development of orientation selectivity itself. Antolík and Bednar [50] demonstrated such a scenario, but using some nonbiological assumptions such as no interactions among developing simple cells.

One direction for future studies lies in the possible nonlinear complications of Hebbian models of cortical development #1 (e.g., Ref. [45]), including simultaneous development of intracortical and input synapses. Nonlinear variants of Hebbian dynamics may be key to creating bottom-up models of orientation development that are able to reproduce band-pass, periodic orientation maps like those measured in the cortex (e.g., Ref. [87]). This may further serve to bridge Hebbian theory to Landau-type models of universal behavior such as those of Ref. [13].

By addressing the riddle of orientation selectivity in biological networks, our work suggests some possible new avenues for research on artificial networks, which have been seen to perform unsupervised learning in a strikingly brainlike fashion [88]. Future research in this sense may require two directions: (i) a comparison of the multilayer dynamics that we demonstrate to the learning trajectories of units in convolutional networks notably during pretraining (along the lines of Refs. [89–91]) to shed light on the degree of universality of the mechanisms discussed, and (ii) the deployment of the competitive constraints that we demonstrate toward engineering convolutional networks for unsupervised learning such as those of Refs. [92–96], in analogy with performance boosting via cortexlike bottlenecks in Ref. [97].

ACKNOWLEDGMENTS

F. F. worked on the analytics, B. H. on the numerics, and K. D. M. supervised the project. All authors have reviewed the manuscript. We benefited from conversations with Tadashi Yamazaki, Tomokazu Ohshiro, and Hideaki Shimazaki. This research has been supported by NSF (Grant No. DBI-1707398) and by the Gatsby Charitable Foundation (Grant No. GAT3708). F. F. has also been supported by RIKEN Center for Brain Science, Brain/MINDS from Japan's Agency for Medical Research and Development under Grant No. JP20dm020700, and JSPS KAKENHI Grant No. JP18H05432.

APPENDIX A: FORMULATION OF THE MODEL EQUATIONS

1. Model setup

Here we derive the dynamical equations of the model. We label the locations in the presynaptic layers with greek indices $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \ldots)$ and the locations in the postsynaptic layer with letters from the end of the latin alphabet $(\boldsymbol{x}, \boldsymbol{y}, \ldots)$. We call $r(\boldsymbol{x}, \tau)$ the firing rate of neurons in the cortical layer, and $r^{\mathrm{ON}}(\boldsymbol{\alpha}, \tau)$ and $r^{\mathrm{OFF}}(\boldsymbol{\alpha}, \tau)$ the firing rates of ON or OFF neurons in the LGN layers at time τ.

Correlations between the activities of two presynaptic cells of types $(i) \in \{\mathrm{ON}, \mathrm{OFF}\}$ and $(j) \in \{\mathrm{ON}, \mathrm{OFF}\}$ located at retinotopic positions $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are described by

a function $C^{(i,j)}(\boldsymbol{\alpha},\boldsymbol{\beta}) \equiv C^{(i,j)}(\boldsymbol{\alpha}-\boldsymbol{\beta})$. Lateral connections between cortical neurons at positions $\boldsymbol{x}$ and $\boldsymbol{y}$ have synaptic strength given by a function $W(\boldsymbol{x},\boldsymbol{y}) \equiv W(\boldsymbol{x}-\boldsymbol{y})$. We assume both of these to be time independent over the timescale of orientation map development.

The quantity of synapses from a presynaptic location $\boldsymbol{\alpha}$ to a postsynaptic location $\boldsymbol{z}$ is described by an arbor density $A(\boldsymbol{z},\boldsymbol{\alpha}) \equiv A(\boldsymbol{z}-\boldsymbol{\alpha})$, which is also assumed to be time independent. The synaptic strength of the kth individual synapse from the afferent at location $\boldsymbol{\alpha}$ to the cortical cell at location $\boldsymbol{x}$ at time τ is called $s_k^{\mathrm{ON}}(\boldsymbol{x},\boldsymbol{\alpha},\tau)$ or $s_k^{\mathrm{OFF}}(\boldsymbol{x},\boldsymbol{\alpha},\tau)$ depending on whether the presynaptic cell is of the ON or OFF type. The index k runs from 1 to $A(\boldsymbol{x}-\boldsymbol{\alpha})$.

Finally, let $s^{\mathrm{ON}}(\boldsymbol{x},\boldsymbol{\alpha},\tau) = \sum_{k=1}^{A(\boldsymbol{x}-\boldsymbol{\alpha})} s_k^{\mathrm{ON}}(\boldsymbol{z},\boldsymbol{\alpha},\tau)$ and $s^{\mathrm{OFF}}(\boldsymbol{x},\boldsymbol{\alpha},\tau) = \sum_{k=1}^{A(\boldsymbol{x}-\boldsymbol{\alpha})} s_k^{\mathrm{OFF}}(\boldsymbol{z},\boldsymbol{\alpha},\tau)$ be the total synaptic strengths at time τ from the afferent at location $\boldsymbol{\alpha}$ of type ON or OFF to the cortical cell at $\boldsymbol{x}$. For ease of notation, we omit the time variable from the arguments of the functions.

Applying standard rate dynamics to this model [98], we can write that the cortical firing rates evolve according to

$$T(\boldsymbol{x})\frac{dr(\boldsymbol{x})}{d\tau} = -r(\boldsymbol{x}) + \int \boldsymbol{dy}\, W(\boldsymbol{x}-\boldsymbol{y})r(\boldsymbol{y}) + \sum_{(i)=\mathrm{ON,OFF}} \int \boldsymbol{d\alpha}\, s^{(i)}(\boldsymbol{x},\boldsymbol{\alpha})r^{(i)}(\boldsymbol{\alpha}), \quad \text{(A1)}$$

where $T(\boldsymbol{x})$ is a diagonal matrix whose entries are the timescales of neural activity.

According to Hebb's rule, synapses are strengthened or stabilized if there is temporal correlation between their pre- and postsynaptic patterns of activity. For sufficiently small variations, this principle can be linearized into the statement that for all $k = 1,\ldots,A(\boldsymbol{x},\boldsymbol{\alpha})$, we have

$$T_{\mathrm{pl}}\frac{ds_k^{(i)}(\boldsymbol{x},\boldsymbol{\alpha})}{d\tau} \propto r(\boldsymbol{x})r^{(i)}(\boldsymbol{\alpha}), \quad \text{(A2)}$$

where the index (i) distinguishes ON and OFF cells, while T_{pl} is the timescale for synaptic plasticity. Notice that we are also omitting possible constant terms included by Ref. [36] but not essential to the development of selectivity. Assuming symmetry of the two center types, any such constants will disappear when we focus below on the development of the *difference* between s^{ON} and s^{OFF}.

Once Eq. (A2) is summed over all synapses sharing the same afferent and target cell, we obtain

$$T_{\mathrm{pl}}\frac{ds^{(i)}(\boldsymbol{x},\boldsymbol{\alpha})}{d\tau} \propto A(\boldsymbol{x}-\boldsymbol{\alpha})r(\boldsymbol{x})r^{(i)}(\boldsymbol{\alpha}). \quad \text{(A3)}$$

The values of $A(\boldsymbol{x}-\boldsymbol{\alpha})$ are continuous, as they represent the local spatial density of arborization from LGN position $\boldsymbol{\alpha}$ at cortical position $\boldsymbol{x}$.

We take the timescale T_{pl} of synaptic plasticity that figures in Eq. (A2) to be much slower than the timescale of neural activity as given by the entries of T in Eq. (A1), which seems consistent with experiments [99]. This allows us to model synaptic development by relying on the steady state of the fast dynamics from Eq. (A1), which is given by

$$r(\boldsymbol{x}) = \sum_{(i)=\mathrm{ON,OFF}} \int \boldsymbol{dy d\alpha}\, I(\boldsymbol{x}-\boldsymbol{y})s^{(i)}(\boldsymbol{y},\boldsymbol{\alpha})r^{(i)}(\boldsymbol{\alpha}), \quad \text{(A4)}$$

where $I(\boldsymbol{x}-\boldsymbol{y}) - \int d\boldsymbol{z} I(\boldsymbol{x}-\boldsymbol{z})W(\boldsymbol{z}-\boldsymbol{y}) = \delta(\boldsymbol{x}-\boldsymbol{y})$.

Replacing Eq. (A4) with Eq. (A3) yields

$$T_{\mathrm{pl}}\frac{ds^{(i)}(\boldsymbol{x},\boldsymbol{\alpha})}{d\tau} \sim A(\boldsymbol{x}-\boldsymbol{\alpha}) \sum_{(j)=\mathrm{ON,OFF}} \int \boldsymbol{dyd\beta}\, I(\boldsymbol{x}-\boldsymbol{y})s^{(j)}(\boldsymbol{y},\boldsymbol{\beta})r^{(i)}(\boldsymbol{\alpha})r^{(j)}(\boldsymbol{\beta}). \quad \text{(A5)}$$

We can now average Eq. (A5) over a timescale sufficiently longer than the typical timescale of firing-rate dynamics, yet shorter than the typical timescale of synaptic evolution. The averaging leads to the equation

$$\frac{ds^{(i)}(\boldsymbol{x},\boldsymbol{\alpha})}{d\tau} \sim A(\boldsymbol{x}-\boldsymbol{\alpha}) \sum_{(j)=\mathrm{ON,OFF}} \int \boldsymbol{dyd\beta}\, I(\boldsymbol{x}-\boldsymbol{y})s^{(j)}(\boldsymbol{y},\boldsymbol{\beta})C^{(ij)}(\boldsymbol{\alpha}-\boldsymbol{\beta}), \quad \text{(A6)}$$

where $C^{(ij)}_{\boldsymbol{\alpha}-\boldsymbol{\beta}} = \langle r^{(i)}_{\boldsymbol{\alpha}} r^{(j)}_{\boldsymbol{\beta}} \rangle$, and we set the units of time so that $T_{\mathrm{pl}} = 1$.

Assuming for simplicity symmetry under interchange of ON and OFF, so that $C^{(\mathrm{ON,ON})} = C^{(\mathrm{OFF,OFF})}$, we can transform coordinates to obtain independent equations for $s^{(S)} \equiv s^{\mathrm{ON}} + s^{\mathrm{OFF}}$ and $s^{(D)} \equiv s^{\mathrm{ON}} - s^{\mathrm{OFF}}$ (superscripts S and D stand for "sum" and "difference," respectively). Our interest is in the development of orientation selectivity via the formation of alternating RF subregions in which ON or OFF LGN inputs, respectively, are dominant. Hence, we are interested in the development of a pattern in $s^{(D)}$, while $s^{(S)}$ is not expected to form interesting structure. Thus, we focus on the equation for $s^{(D)}$, which is

$$\frac{ds^{(D)}(\boldsymbol{x},\boldsymbol{\alpha})}{d\tau} \sim A(\boldsymbol{x}-\boldsymbol{\alpha}) \int \boldsymbol{dyd\beta}\, I(\boldsymbol{x}-\boldsymbol{y})s^{(D)}(\boldsymbol{y},\boldsymbol{\beta})C^{(D)}(\boldsymbol{\alpha}-\boldsymbol{\beta}), \quad \text{(A7)}$$

where $C^{(D)} = C^{(\mathrm{ON,ON})} - C^{(\mathrm{ON,OFF})}$ is the difference between same-center-type and opposite-center-type correlations.

It is characteristic of Hebbian rules that synaptic strengths tend to increase without limit [98]. Reference [64] modeled the biological mechanisms for saturation by including an upper bound and a zero lower bound for all synaptic strengths ($0 \leq s^{(i)} \leq s_{\max}$), which becomes a limit on $s^{(D)}$ of $-s_{\max} \leq s^{(D)} \leq s_{\max}$. This turns Eq. (A7) into a nonlinear equation. However, we imagine development starting from an initial condition in which there are roughly equal strengths of ON and OFF innervation throughout the receptive field, and thus in which the values of $s^{(D)}$ are small random perturbations about 0. We can always assume the synaptic weight bounds large enough so that the principal features of the $s^{(D)}$ dynamics are established before the bounds are saturated (as in Refs. [36,64]). Once the bounds are reached, they will simply capture and preserve the existing weight structure with little subsequent change.

We can therefore extract the long-term behavior of the synaptic weights simply by analyzing the properties of the time-evolution operator in the linear regime. As per Eq. (A7), this will be, in a first approximation, the integral operator characterized by the kernel

$$K^{(D)}(\boldsymbol{x},\boldsymbol{\alpha};\boldsymbol{y},\boldsymbol{\beta}) = A(\boldsymbol{x}-\boldsymbol{\alpha})I(\boldsymbol{x}-\boldsymbol{y})C^{(D)}(\boldsymbol{\alpha}-\boldsymbol{\beta}). \quad \text{(A8)}$$

We now incorporate the fact that correlation-based development is competitive. In this framework, indeed, a mechanism of the order of competition is necessary for neurons to become selective for certain features. Without competition, all synapses onto a cell could grow to their maximum possible value, eliminating all selectivity save the retinotopic selectivity embodied in the arbor density. As we discuss in the main text, we model competition by conserving the total strength of synapses projected by each arbor:

$$\frac{d}{d\tau}\int \boldsymbol{dx}\, s^{\mathrm{ON}}_{\boldsymbol{x},\boldsymbol{\alpha}} = \frac{d}{d\tau}\int \boldsymbol{dx}\, s^{\mathrm{OFF}}_{\boldsymbol{x},\boldsymbol{\alpha}} = 0 \quad \forall\ \boldsymbol{\alpha}, \quad \text{(A9)}$$

where the arguments of functions are written as subscripts for the sake of compactness.

Together, these imply

$$\frac{d}{d\tau}\int \boldsymbol{dx}\, s^{D}_{\boldsymbol{x},\boldsymbol{\alpha}} = \frac{d}{d\tau}\int \boldsymbol{dx}\, s^{S}_{\boldsymbol{x},\boldsymbol{\alpha}} = 0 \quad \forall\ \boldsymbol{\alpha}. \quad \text{(A10)}$$

Henceforth, we focus only on the development of $s^{(D)}$. We drop the "D" superscript, simply writing s for $s^{(D)}$, C for $C^{(D)}$, and K for $K^{(D)}$ [Eq. (A8)]. Including the constraint Eq. (A10), the equation we study is

$$\frac{ds(\boldsymbol{x},\boldsymbol{\alpha})}{d\tau} = \int \boldsymbol{dyd\beta}\, K(\boldsymbol{x},\boldsymbol{\alpha};\boldsymbol{y},\boldsymbol{\beta})s(\boldsymbol{y},\boldsymbol{\beta}) - \frac{A(\boldsymbol{x}-\boldsymbol{\alpha})}{\int \boldsymbol{dz}A(\boldsymbol{z}-\boldsymbol{\alpha})}\int \boldsymbol{dydqd\beta}\, K(\boldsymbol{q},\boldsymbol{\alpha};\boldsymbol{y},\boldsymbol{\beta})s(\boldsymbol{y},\boldsymbol{\beta}). \quad \text{(A11)}$$

To completely specify the model, it remains only to choose a form for the functions A, I, and C. The assumption of monotonically decreasing functions reflects a principle of modeling economy for I and is suggested by the general decay of arborizations with distance for A and by the experimental results discussed in the Introduction for C. We take them for definiteness to have Gaussian dependences on distances:

$$A(\boldsymbol{x}-\boldsymbol{\alpha}) \propto e^{-\frac{(x-\alpha)^2}{2\rho^2}}, \quad \text{(A12)}$$

$$C(\boldsymbol{\alpha}-\boldsymbol{\beta}) \propto e^{-\frac{(\alpha-\beta)^2}{2\zeta^2}}, \quad \text{(A13)}$$

$$I(\boldsymbol{x}-\boldsymbol{y}) \propto e^{-\frac{(x-y)^2}{2\eta^2}}, \quad \text{(A14)}$$

where the arbor radius ρ, the interaction length η, and the correlation length ζ are the three characteristic length scales of the model.

2. Hermitian formulation of the constrained problem

The homeostatic constraint mechanism modeled by Eq. (A9) conserves the total projection strength from each presynaptic cell. We start by illustrating how this constraint can be incorporated into the theory, namely, by adding to the Hebbian law Eq. (A2) a suitable leak term $\epsilon^{\mathrm{ON}}_{\boldsymbol{\alpha}}$. This yields

$$\frac{ds^{(i)}_k(\boldsymbol{x},\boldsymbol{\alpha})}{d\tau} \propto -\epsilon^{(i)}_{\boldsymbol{\alpha}} + r(\boldsymbol{x})r^{(i)}(\boldsymbol{\alpha}). \quad \text{(A15)}$$

Here, $\epsilon^{\mathrm{ON}}_{\boldsymbol{\alpha}}$ and $\epsilon^{\mathrm{OFF}}_{\boldsymbol{\alpha}}$ are unspecified quantities that are defined in such a way as to implement the conservation constraints Eq. (A9) of the model.

Summing Eq. (A15) over all synapses with the same afferent and performing the averaging over time as done for Eq. (A5), we obtain

$$\frac{d}{d\tau}s^{\mathrm{ON}}_{\boldsymbol{x},\boldsymbol{\alpha}} = -\epsilon^{\mathrm{ON}}_{\boldsymbol{\alpha}}A_{\boldsymbol{x}-\boldsymbol{\alpha}} + A_{\boldsymbol{x}-\boldsymbol{\alpha}}\sum_{\boldsymbol{y},\boldsymbol{\beta}} I_{\boldsymbol{x}-\boldsymbol{y}}[C^{\mathrm{ON,ON}}_{\boldsymbol{\alpha}-\boldsymbol{\beta}}s^{\mathrm{ON}}_{\boldsymbol{y},\boldsymbol{\beta}} + C^{\mathrm{ON,OFF}}_{\boldsymbol{\alpha}-\boldsymbol{\beta}}s^{\mathrm{OFF}}_{\boldsymbol{y},\boldsymbol{\beta}}], \quad \text{(A16)}$$

$$\frac{d}{d\tau}s^{\mathrm{OFF}}_{\boldsymbol{x},\boldsymbol{\alpha}} = -\epsilon^{\mathrm{OFF}}_{\boldsymbol{\alpha}}A_{\boldsymbol{x}-\boldsymbol{\alpha}} + A_{\boldsymbol{x}-\boldsymbol{\alpha}}\sum_{\boldsymbol{y},\boldsymbol{\beta}} I_{\boldsymbol{x}-\boldsymbol{y}}[C^{\mathrm{OFF,OFF}}_{\boldsymbol{\alpha}-\boldsymbol{\beta}}s^{\mathrm{OFF}}_{\boldsymbol{y},\boldsymbol{\beta}} + C^{\mathrm{OFF,ON}}_{\boldsymbol{\alpha}-\boldsymbol{\beta}}s^{\mathrm{ON}}_{\boldsymbol{y},\boldsymbol{\beta}}]. \quad \text{(A17)}$$

The crucial step for all that follows is to render the operation in Eqs. (A16) and (A17) symmetric, which redefines the time evolution in terms of Hermitian operators. We do so by defining

$$t^{\mathrm{ON}}_{\boldsymbol{x},\boldsymbol{\alpha}} = \frac{s^{\mathrm{ON}}_{\boldsymbol{x},\boldsymbol{\alpha}}}{\sqrt{A_{\boldsymbol{x}-\boldsymbol{\alpha}}}}, \qquad t^{\mathrm{OFF}}_{\boldsymbol{x},\boldsymbol{\alpha}} = \frac{s^{\mathrm{OFF}}_{\boldsymbol{x},\boldsymbol{\alpha}}}{\sqrt{A_{\boldsymbol{x}-\boldsymbol{\alpha}}}}, \tag{A18}$$

so that Eqs. (A16) and (A17) become

$$\frac{d}{d\tau} t^{\mathrm{ON}}_{\boldsymbol{x},\boldsymbol{\alpha}} = \sqrt{A_{\boldsymbol{x}-\boldsymbol{\alpha}}}\Big\{-\epsilon^{\mathrm{ON}}_{\boldsymbol{\alpha}} + \sum_{\boldsymbol{y},\boldsymbol{\beta}} I_{\boldsymbol{x}-\boldsymbol{y}}[C^{\mathrm{ON,ON}}_{\boldsymbol{\alpha}-\boldsymbol{\beta}} \sqrt{A_{\boldsymbol{y}-\boldsymbol{\beta}}} t^{\mathrm{ON}}_{\boldsymbol{y},\boldsymbol{\beta}} + C^{\mathrm{ON,OFF}}_{\boldsymbol{\alpha}-\boldsymbol{\beta}} \sqrt{A_{\boldsymbol{y}-\boldsymbol{\beta}}} t^{\mathrm{OFF}}_{\boldsymbol{y},\boldsymbol{\beta}}]\Big\}, \tag{A19}$$

$$\frac{d}{d\tau} t^{\mathrm{OFF}}_{\boldsymbol{x},\boldsymbol{\alpha}} = \sqrt{A_{\boldsymbol{x}-\boldsymbol{\alpha}}}\Big\{-\epsilon^{\mathrm{OFF}}_{\boldsymbol{\alpha}} + \sum_{\boldsymbol{y},\boldsymbol{\beta}} I_{\boldsymbol{x}-\boldsymbol{y}}[C^{\mathrm{OFF,OFF}}_{\boldsymbol{\alpha}-\boldsymbol{\beta}} \sqrt{A_{\boldsymbol{y}-\boldsymbol{\beta}}} t^{\mathrm{OFF}}_{\boldsymbol{y},\boldsymbol{\beta}} + C^{\mathrm{OFF,ON}}_{\boldsymbol{\alpha}-\boldsymbol{\beta}} \sqrt{A_{\boldsymbol{y}-\boldsymbol{\beta}}} t^{\mathrm{ON}}_{\boldsymbol{y},\boldsymbol{\beta}}]\Big\}. \tag{A20}$$

It will be convenient to regard the functions t^{ON} and t^{OFF} as vectors in Hilbert space, so that (in bracket notation) $t^{\mathrm{ON}}_{\boldsymbol{x},\boldsymbol{\alpha}} = \langle \boldsymbol{x},\boldsymbol{\alpha}|t^{\mathrm{ON}}\rangle$ and $t^{\mathrm{OFF}}_{\boldsymbol{x},\boldsymbol{\alpha}} = \langle \boldsymbol{x},\boldsymbol{\alpha}|t^{\mathrm{OFF}}\rangle$.

Since ON and OFF cells are subjected to the same inputs from the retina, we may assume $C^{\mathrm{ON,ON}} = C^{\mathrm{OFF,OFF}}$. Using this and the fact that $C^{\mathrm{ON,OFF}} = C^{\mathrm{OFF,ON}}$, we rewrite Eqs. (A19) and (A20) in terms of the sole operators $\hat{L}^s$ and $\hat{L}^d$ defined as follows (hats on the names of operators distinguish them from ordinary variables):

$$\langle \boldsymbol{x},\boldsymbol{\alpha}|\hat{L}^s|\boldsymbol{y},\boldsymbol{\beta}\rangle = \sqrt{A_{\boldsymbol{x}-\boldsymbol{\alpha}}} I_{\boldsymbol{x}-\boldsymbol{y}} C^{\mathrm{ON,ON}}_{\boldsymbol{\alpha}-\boldsymbol{\beta}} \sqrt{A_{\boldsymbol{y}-\boldsymbol{\beta}}}, \tag{A21}$$

$$\langle \boldsymbol{x},\boldsymbol{\alpha}|\hat{L}^d|\boldsymbol{y},\boldsymbol{\beta}\rangle = \sqrt{A_{\boldsymbol{x}-\boldsymbol{\alpha}}} I_{\boldsymbol{x}-\boldsymbol{y}} C^{\mathrm{ON,OFF}}_{\boldsymbol{\alpha}-\boldsymbol{\beta}} \sqrt{A_{\boldsymbol{y}-\boldsymbol{\beta}}}, \tag{A22}$$

yielding

$$\frac{d}{d\tau}|t^{\mathrm{ON}}\rangle = \hat{L}^s|t^{\mathrm{ON}}\rangle + \hat{L}^d|t^{\mathrm{OFF}}\rangle - \hat{\mathcal{E}}^{\mathrm{ON}}|a\rangle, \tag{A23}$$

$$\frac{d}{d\tau}|t^{\mathrm{OFF}}\rangle = \hat{L}^s|t^{\mathrm{OFF}}\rangle + \hat{L}^d|t^{\mathrm{ON}}\rangle - \hat{\mathcal{E}}^{\mathrm{OFF}}|a\rangle. \tag{A24}$$

Here the vector $|a\rangle$ is defined by $\langle \boldsymbol{x},\boldsymbol{\alpha}|a\rangle = \sqrt{A_{\boldsymbol{x}-\boldsymbol{\alpha}}}$, while the operators $\hat{\mathcal{E}}^{\mathrm{ON}}$ and $\hat{\mathcal{E}}^{\mathrm{OFF}}$ have the form

$$\hat{\mathcal{E}}^{\mathrm{ON}} = \sum_{\boldsymbol{\beta}} \epsilon^{\mathrm{ON}}_{\boldsymbol{\beta}} \hat{P}_{\boldsymbol{\beta}}, \qquad \hat{\mathcal{E}}^{\mathrm{OFF}} = \sum_{\boldsymbol{\beta}} \epsilon^{\mathrm{OFF}}_{\boldsymbol{\beta}} \hat{P}_{\boldsymbol{\beta}}, \tag{A25}$$

where $\hat{P}_{\boldsymbol{\beta}}$ is the operator that effects projection into the subspace with basis $\{|\boldsymbol{y},\boldsymbol{\beta}\rangle\}_{\boldsymbol{y}}$, that is,

$$\hat{P}_{\boldsymbol{\beta}} \equiv \sum_{\boldsymbol{y}} |\boldsymbol{y},\boldsymbol{\beta}\rangle\langle \boldsymbol{y},\boldsymbol{\beta}|. \tag{A26}$$

The expressions for ϵ^{ON}, ϵ^{OFF} can be found from the conservation laws Eq. (A9), which may be rewritten in the form

$$\frac{d}{d\tau} \sum_{\boldsymbol{y}} \sqrt{A_{\boldsymbol{y}-\boldsymbol{\beta}}}\, t^{\mathrm{ON,OFF}}_{\boldsymbol{y},\boldsymbol{\beta}} = 0 \quad \forall\ \boldsymbol{\beta}, \tag{A27}$$

or rather

$$\frac{d}{d\tau}\langle a|\hat{P}_{\boldsymbol{\beta}}|t^{\mathrm{ON}}\rangle = 0 \quad \forall\ \boldsymbol{\beta}, \tag{A28}$$

$$\frac{d}{d\tau}\langle a|\hat{P}_{\boldsymbol{\beta}}|t^{\mathrm{OFF}}\rangle = 0 \quad \forall\ \boldsymbol{\beta}. \tag{A29}$$

Substituting Eqs. (A23) and (A24) into the two constraints (A28) and (A29), and using the expressions (A25) for $\mathcal{E}^{\mathrm{ON}}$, $\mathcal{E}^{\mathrm{OFF}}$, we finally arrive at

$$\epsilon^{\mathrm{ON}}_{\boldsymbol{\beta}} = \frac{\langle a|\hat{P}_{\boldsymbol{\beta}} L^s|t^{\mathrm{ON}}\rangle + \langle a|\hat{P}_{\boldsymbol{\beta}} L^d|t^{\mathrm{OFF}}\rangle}{\langle a|\hat{P}_{\boldsymbol{\beta}}|a\rangle}, \tag{A30}$$

$$\epsilon^{\mathrm{OFF}}_{\boldsymbol{\beta}} = \frac{\langle a|\hat{P}_{\boldsymbol{\beta}} L^s|t^{\mathrm{OFF}}\rangle + \langle a|\hat{P}_{\boldsymbol{\beta}} L^d|t^{\mathrm{ON}}\rangle}{\langle a|\hat{P}_{\boldsymbol{\beta}}|a\rangle}. \tag{A31}$$

For the difference $s_{\boldsymbol{x},\boldsymbol{\alpha}} \equiv s^{\mathrm{ON}}_{\boldsymbol{x},\boldsymbol{\alpha}} - s^{\mathrm{OFF}}_{\boldsymbol{x},\boldsymbol{\alpha}}$, we have

$$s_{\boldsymbol{x},\boldsymbol{\alpha}} = \sqrt{A_{\boldsymbol{x}-\boldsymbol{\alpha}}}(t^{\mathrm{ON}}_{\boldsymbol{x},\boldsymbol{\alpha}} - t^{\mathrm{OFF}}_{\boldsymbol{x},\boldsymbol{\alpha}}), \tag{A32}$$

so that we must proceed to compute the time evolution of $|t\rangle \equiv |t^{\mathrm{ON}} - t^{\mathrm{OFF}}\rangle$. From Eqs. (A23) and (A24), we find

$$\frac{d}{d\tau}|t\rangle = \left[\mathbf{1} - \sum_{\boldsymbol{\beta}} \frac{\hat{P}_{\boldsymbol{\beta}}|a\rangle\langle a|\hat{P}_{\boldsymbol{\beta}}}{\langle a|\hat{P}_{\boldsymbol{\beta}}|a\rangle}\right]\hat{L}|t\rangle, \tag{A33}$$

where 1 stands for the identity operator, and we define $\hat{L} \equiv \hat{L}^s - \hat{L}^d$.

3. Projection operators

To rewrite Eq. (A33) in a more transparent form, we define the single-arbor ket $|a_{\boldsymbol{\beta}}\rangle = \hat{P}_{\boldsymbol{\beta}}|a\rangle$, with elements $\langle \boldsymbol{x},\boldsymbol{\alpha}|a_{\boldsymbol{\beta}}\rangle = \delta_{\boldsymbol{\alpha},\boldsymbol{\beta}}\sqrt{A_{\boldsymbol{x}-\boldsymbol{\alpha}}}$. Notice that the operator $\hat{P}_{\boldsymbol{\beta}}$ of Eq. (A26) is orthogonal, and therefore, being also a projection, it is self-adjoint. Using this fact, as well as the idempotence of $\hat{P}_{\boldsymbol{\beta}}$, we obtain

$$\frac{d}{d\tau}|t\rangle = \left[\mathbf{1} - \sum_{\boldsymbol{\beta}} \frac{|a_{\boldsymbol{\beta}}\rangle\langle a_{\boldsymbol{\beta}}|}{\langle a_{\boldsymbol{\beta}}|a_{\boldsymbol{\beta}}\rangle}\right]\hat{L}|t\rangle. \tag{A34}$$

Defining $\hat{P} = \mathbf{1} - \sum_{\boldsymbol{\beta}}(|a_{\boldsymbol{\beta}}\rangle\langle a_{\boldsymbol{\beta}}|/\langle a_{\boldsymbol{\beta}}|a_{\boldsymbol{\beta}}\rangle)$, we have from Eq. (A34)

$$\frac{d}{d\tau}|t\rangle = \hat{P}\,\hat{L}\,|t\rangle = \hat{P}\,\hat{L}\,\hat{P}\,|t\rangle + \hat{P}\,\hat{L}(\mathbf{1}-\hat{P})|t\rangle. \tag{A35}$$

Since we are interested in calculating the final outcome of development, we must focus on the long-term behavior

of this dynamics. To do so, we notice that the components of $|t\rangle$ projected away by $\hat{P}$ cannot be made to grow by the constrained Hebbian dynamics of Eq. (A35). Therefore, positive eigenvalues leading to exponential growth must be found in the space in which $\hat{P}$ is projecting, and after waiting a sufficiently long time, one may always approximate the state of the system as contained in that space. Noting this, we can drop the last term in Eq. (A35) and write simply

$$\frac{d}{d\tau}|t\rangle = \hat{P}\,\hat{L}\,\hat{P}\,|t\rangle \equiv \hat{L}^p|t\rangle. \tag{A36}$$

Finally, notice that, since the arbor density $A(\boldsymbol{r})$ has the meaning of a density, we can define it as being properly normalized so that $\int d\boldsymbol{r}A(\boldsymbol{r}) = 1$, i.e., $\langle a_{\boldsymbol{\beta}}||a_{\boldsymbol{\beta}}\rangle = 1$. This allows us to remove the denominators in the definition of the projection operator $\hat{P}$, which becomes simply

$$\hat{P} = \mathbf{1} - \sum_{\boldsymbol{\beta}} |a_{\boldsymbol{\beta}}\rangle\langle a_{\boldsymbol{\beta}}|. \tag{A37}$$

The principal eigenspace of the operator L^p of Eq. (A36) defined in terms of Eq. (A37) is thus the object of our interest, as it will determine the fastest-growing modes of the system.

4. Matrix elements of the Hermitianized operator

The matrix elements of the unconstrained operator $\hat{L}$ appearing in Eq. (A33) are, in the $|\boldsymbol{x},\boldsymbol{\alpha}\rangle$ basis,

$$\begin{aligned} L(\boldsymbol{x},\boldsymbol{\alpha};\boldsymbol{y},\boldsymbol{\beta}) &\equiv \langle \boldsymbol{x},\boldsymbol{\alpha}|L|\boldsymbol{y},\boldsymbol{\beta}\rangle \\ &= \sqrt{A_{\boldsymbol{x}-\boldsymbol{\alpha}}} I_{\boldsymbol{x}-\boldsymbol{y}} C_{\boldsymbol{\alpha}-\boldsymbol{\beta}} \sqrt{A_{\boldsymbol{y}-\boldsymbol{\beta}}}, \end{aligned} \tag{A38}$$

where $C_{\boldsymbol{\alpha}-\boldsymbol{\beta}} = C^{\mathrm{ON,ON}}_{\boldsymbol{\alpha}-\boldsymbol{\beta}} - C^{\mathrm{ON,OFF}}_{\boldsymbol{\alpha}-\boldsymbol{\beta}}$.

The generic matrix element of the operator $\hat{L}^p$ of Eq. (A36) is written, using Eq. (A37), as

$$\begin{aligned} &L^p(\boldsymbol{x},\boldsymbol{\alpha};\boldsymbol{y},\boldsymbol{\beta}) = \langle \boldsymbol{x},\boldsymbol{\alpha}|\hat{L}^p|\boldsymbol{y},\boldsymbol{\beta}\rangle \\ &= \int d\boldsymbol{x}_1 d\boldsymbol{\alpha}_1 d\boldsymbol{x}_2 d\boldsymbol{\alpha}_2 [\delta(\boldsymbol{x}-\boldsymbol{x}_1)\delta(\boldsymbol{\alpha}-\boldsymbol{\alpha}_1) \\ &- \sqrt{A(\boldsymbol{x}-\boldsymbol{\alpha})}\delta(\boldsymbol{\alpha}-\boldsymbol{\alpha}_1)\sqrt{A(\boldsymbol{x}_1-\boldsymbol{\alpha}_1)}] L(\boldsymbol{x}_1,\boldsymbol{\alpha}_1;\boldsymbol{x}_2,\boldsymbol{\alpha}_2) \\ &\times [\delta(\boldsymbol{x}_2-\boldsymbol{y})\delta(\boldsymbol{\alpha}_2-\boldsymbol{\beta}) - \sqrt{A(\boldsymbol{x}_2-\boldsymbol{\alpha}_2)}\delta(\boldsymbol{\alpha}_2-\boldsymbol{\beta})\sqrt{A(\boldsymbol{y}-\boldsymbol{\beta})}]. \end{aligned} \tag{A39}$$

The natural variables in which to express a RF are the relative coordinates $\boldsymbol{r} = \boldsymbol{\alpha} - \boldsymbol{x}$, and we abuse the notation by writing $L^p(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s}) \equiv L^p(\boldsymbol{x},\boldsymbol{\alpha}-\boldsymbol{x};\boldsymbol{y},\boldsymbol{\beta}-\boldsymbol{y})$ and $L(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s}) \equiv L(\boldsymbol{x},\boldsymbol{\alpha}-\boldsymbol{x};\boldsymbol{y},\boldsymbol{\beta}-\boldsymbol{y})$. Integrating out the delta functions in Eq. (A39), we obtain

$$\begin{aligned} L^p(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s}) &= L(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s}) + S(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s}) \\ &+ T(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s}) + \tilde{T}(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s}). \end{aligned} \tag{A40}$$

On the rhs of this equation, the first term is obtained with a change of variables in the arguments of Eq. (A38), leading to

$$L(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s}) = \sqrt{A(r)}\sqrt{A(s)} I(\boldsymbol{x}-\boldsymbol{y}) C(\boldsymbol{x}-\boldsymbol{y}+\boldsymbol{r}-\boldsymbol{s}), \tag{A41}$$

while the last three terms are given by

$$\begin{aligned} \frac{S(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s})}{\sqrt{A(r)A(s)}} &= \int d\boldsymbol{s}_1 d\boldsymbol{s}_2 \sqrt{A(s_1)A(s_2)} \\ &\times L(\boldsymbol{x}+\boldsymbol{r}-\boldsymbol{s}_1,\boldsymbol{s}_1;\boldsymbol{y}+\boldsymbol{s}-\boldsymbol{s}_2,\boldsymbol{s}_2), \end{aligned} \tag{A42}$$

$$T(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s}) = -\sqrt{A(r)} \int d\boldsymbol{u} \sqrt{A(u)} L(\boldsymbol{x}+\boldsymbol{r}-\boldsymbol{u},\boldsymbol{u};\boldsymbol{y},\boldsymbol{s}), \tag{A43}$$

$$\tilde{T}(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s}) = -\sqrt{A(s)} \int d\boldsymbol{u} \sqrt{A(u)} L(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y}+\boldsymbol{s}-\boldsymbol{u},\boldsymbol{u}). \tag{A44}$$

In the special case where the functions $I(\boldsymbol{x}-\boldsymbol{y})$ and $C(\boldsymbol{\alpha}-\boldsymbol{\beta})$ are even under parity, it is seen from Eq. (A41) that $\hat{L}$ becomes symmetric under swapping of the LGN and cortical coordinates, and it follows that $\hat{S} = \hat{S}^T$, $\hat{\tilde{T}} = \hat{T}^T$.

5. Cortical Fourier transform

Noting the translation invariance of $\hat{L}$ in Eq. (A41) with respect to the cortical location variable, we can define

$$L(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s}) \equiv \int \frac{d\boldsymbol{\omega}}{(2\pi)^2} e^{-i\boldsymbol{\omega}(\boldsymbol{x}-\boldsymbol{y})} L(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\omega}), \tag{A45}$$

and similar for $\hat{L}^p$, from the translation invariance seen in Eq. (A39). This means that the eigenfunctions of $\hat{L}^p$ in Eq. (A40) will be of the form $e^{i\boldsymbol{\omega}\boldsymbol{x}}\psi_{\omega}(\boldsymbol{r})$, where $e^{i\boldsymbol{\omega}\boldsymbol{x}}$ describes an oscillation across the cortical coordinate $\boldsymbol{x}$, and $\psi_{\omega}(\boldsymbol{r})$ describes the RF as a function of the LGN position $\boldsymbol{\alpha}$ relative to $\boldsymbol{x}$, i.e., $\boldsymbol{r} = \boldsymbol{\alpha} - \boldsymbol{x}$.

The function $\psi_{\omega}(\boldsymbol{r})$ is complex, and we write it in the form of real functions as $\psi_{\omega}(\boldsymbol{r}) = \psi^R_{\omega}(\boldsymbol{r}) + i\psi^L_{\omega}(\boldsymbol{r})$. Then, this eigenfunction corresponds to the real functions $\cos(\boldsymbol{\omega}\boldsymbol{x}+\phi)\psi^R_{\omega}(\boldsymbol{r}) + \sin(\boldsymbol{\omega}\boldsymbol{x}+\phi)\psi^R_{\omega}(\boldsymbol{r})$ for arbitrary phase ϕ.

We refer to the spatial frequency vector $\boldsymbol{\omega}$ as the cortical wave vector and to its modulus as the cortical wave number. The Fourier transform of the constrained operator Eq. (A39), which determines the RF eigenfunctions $\psi_{\omega}(\boldsymbol{r})$, is then

$$L^p(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\omega}) = L(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\omega}) - \sqrt{A(r)}\int \boldsymbol{dr_1}\sqrt{A(r_1)}L(\boldsymbol{r_1},\boldsymbol{s};\boldsymbol{\omega})e^{-i\boldsymbol{\omega}(\boldsymbol{r}-\boldsymbol{r_1})} - \sqrt{A(s)}\int \boldsymbol{ds_1}\sqrt{A(s_1)}L(\boldsymbol{r},\boldsymbol{s_1};\boldsymbol{\omega})e^{-i\boldsymbol{\omega}(\boldsymbol{s_1}-\boldsymbol{s})} + \sqrt{A(r)A(s)}\int dr_1 ds_1\sqrt{A(r_1)A(s_1)}L(\boldsymbol{r_1},\boldsymbol{s_1};\boldsymbol{\omega})e^{-i\boldsymbol{\omega}(\boldsymbol{r}-\boldsymbol{s}-\boldsymbol{r_1}+\boldsymbol{s_1})}. \tag{A46}$$

The transform of Eq. (A41) reads

$$L(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\omega}) \sim L^u(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\mu})\exp\left[-\frac{\omega^2}{2\Omega^2} - i\frac{\eta^2}{\mu^2}\boldsymbol{\omega}(\boldsymbol{r}-\boldsymbol{s})\right], \tag{A47}$$

as long as we adopt Eqs. (A13) and (A14) for the interaction and correlation functions. We neglect here an overall prefactor that can be absorbed in the definition of time. The "effective length" μ and "cutoff wave number" Ω in Eq. (A47) are given by

$$\mu^2 = \eta^2 + \zeta^2, \qquad \Omega^2 = \frac{1}{\eta^2} + \frac{1}{\zeta^2}, \tag{A48}$$

and $L^u(r,s;\mu) = \sqrt{A(r)A(s)}e^{-[(\boldsymbol{r}-\boldsymbol{s})^2/2\mu^2]}$, with the apex u standing for "unconstrained." We may refer to the parameter μ as the correlation-interaction length, as it is a Pythagorean combination of the two "intralayer" length scales of the problem.

We now (i) substitute the Fourier-transformed matrix element of $\hat{L}$ as per Eq. (A47) into the expression (A46) for the matrix element of $\hat{L}^p$, (ii) insert a specific (Gaussian) assumption for the arbor density function $A(r) = (1/2\pi\rho^2)\exp[-(r^2/2\rho^2)]$, and (iii) perform the integration over all intermediate space variables.

We thus arrive at decomposing the constrained operator of Eq. (A39) into

$$\hat{L}^p = \hat{L} + \hat{S} + \hat{T} + \hat{T}^\dagger, \tag{A49}$$

where in the coordinate representation, these are

$$L(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\omega}) = \exp\left[-\frac{\omega^2}{2\Omega^2} - i\frac{\eta^2}{\mu^2}\boldsymbol{\omega}(\boldsymbol{r}-\boldsymbol{s}) - \frac{r^2+s^2}{4\rho^2} - \frac{(\boldsymbol{r}-\boldsymbol{s})^2}{2\mu^2}\right], \tag{A50}$$

$$S(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\omega}) = \frac{\mu^2}{\mu^2+2\rho^2}\exp\left[-\frac{\omega^2}{2\Omega^2} - \frac{\rho^2\zeta^4\omega^2}{\mu^2(\mu^2+2\rho^2)} - i\boldsymbol{\omega}(\boldsymbol{r}-\boldsymbol{s}) - \frac{r^2+s^2}{4\rho^2}\right], \tag{A51}$$

$$T(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\omega}) = -\frac{\mu^2}{\mu^2+\rho^2}\exp\left[-\frac{\omega^2}{2\Omega^2} - \frac{\rho^2\zeta^4\omega^2}{2\mu^2(\mu^2+\rho^2)} - \frac{1}{2}\left(\frac{1}{2\rho^2} + \frac{1}{\rho^2+\mu^2}\right)r^2 - \frac{s^2}{4\rho^2} - i\boldsymbol{\omega}\left(\frac{\rho^2+\eta^2}{\rho^2+\mu^2}\boldsymbol{r} - \boldsymbol{s}\right)\right],$$

$$T^\dagger(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\omega}) = T^*(\boldsymbol{s},\boldsymbol{r};\boldsymbol{\omega}), \tag{A52}$$

and the additive constraint operators $\hat{S}$ and $\hat{T}$ have fully separable matrix elements.

APPENDIX B: PROPERTIES OF THE HERMITIANIZED OPERATOR

We analyze here the basic properties of the time-evolution operator $\hat{L}^p$ of Eq. (A36), including (i) its positive semidefiniteness, (ii) the general structure of its spectrum, (iii) its commutation properties with translation and rotation operators, (iv) its symmetry with respect to parity, complex conjugation, and their combination, and (v) we finally write down the exact diagonalization of the unconstrained operator L, which will serve as the starting point for studying the properties of $\hat{L}^p$ in greater detail.

1. Positive semidefiniteness

It will be useful to rely on the positive semidefiniteness of the Fourier-projected operator $\hat{L}^p(\omega)$ for any given wave number ω.

Consider Eq. (A46), and suppose we regard the direction of the wave vector $\boldsymbol{\omega}$ as fixed (in the following, with no loss of generality, we take it to be parallel to the x axis). We can then write

$$\hat{L}^p(\omega) = (1 - |a_\omega\rangle\langle a_\omega|)\hat{L}(\omega)(1 - |a_\omega\rangle\langle a_\omega|), \tag{B1}$$

where $\langle \boldsymbol{r}|a_\omega\rangle = \sqrt{A(r)}e^{-i\omega r_x}$. Thus, even for a given wave number ω, the constrained operator is nothing but the unconstrained operator sandwiched between two identical projection operators.

The unconstrained operator $\hat{L}$ is clearly positive definite. This follows from the fact that it is Hermitian with an all-positive kernel; we see explicitly that its eigenvalues are all positive but dense in a neighborhood of zero. Recall now the following lemma: If $\hat{O}$ is a positive-definite operator in a linear space and $\hat{P}$ a projection operator on some subspace, then $\hat{P}\,\hat{O}\,\hat{P}$ is positive semidefinite—from which it follows that $\hat{L}^p$ is positive semidefinite.

2. Long-term dynamics

Once the operator is diagonalized for an arbitrary cortical wave number, we expect to find the eigenvalues from a series of possibly overlapping bands, where each given band corresponds to a set of eigenfunctions with varying ω. Different bands may come from different sets of eigenstates characterized by discrete numbers that we may term quantum numbers. As we show, for low wave numbers

these bands correspond to different rotational eigenstates. The "principal band" of the spectrum is the one that contains the principal mode, which drives the long-term dynamics, $\Lambda_M(\omega)$ is the wave-number-dependent eigenvalues of the principal band, and ω_M is the position of its (possibly broad) maximum as a function of ω corresponding to the fastest-growing eigenspace. In the following, we are interested in the principal eigenspace of the operator L^p, as this determines the fastest-growing modes. We can gain insight into long-term behavior by focusing on the principal band and on the RFs it represents.

Indeed, the developmental process for a given band has a similar effect as filtering with a spatially isotropic bandpass with respect to cortical wave number. The function $\Lambda_M(\omega)$ can be interpreted as the corresponding filter profile, and the location of the maximum of this filter may depend in nontrivial ways on the model's parameters.

If the system lies in parameter space at a point such that $\omega_M = 0$, the dynamics will tend to flatten out any spatial inhomogeneity in the initial condition. If $\omega_M > 0$, on the other hand, the long-term RF will vary spatially on a scale of approximately $1/\omega_M$. Since, as will be seen, we have a broad maximum of a nearly optimal wave vector, we may expect local but no long-range periodicity. Anisotropies in the initial conditions can also be magnified by the dynamics.

The evolution of the RF at any given point in the cortex, finally, may cancel or emphasize whatever degree of orientation selectivity is possessed by the initial condition, depending on the structure of the eigenspace associated with the principal mode.

3. Symmetries of the system: Translations and rotations

Since LGN activity reflects retinal input and we average at the outset over an isotropic input ensemble, we expect no change in the dynamics from simultaneously rotating both the cortical layer and the two LGN sheets by the same angle. The same is true, as already noted, if we consider simultaneous translations of the three layers (see Fig. 5).

The time-evolution operator $\hat{L}^p(\boldsymbol{\omega})$ thus has two symmetries: (i) Simultaneous shifts of the cortex and of the LGN do not affect the matrix elements. (ii) If the wave vector $\boldsymbol{\omega}$ is rotated, and the relative coordinates $(\boldsymbol{r}, \boldsymbol{s})$ are rotated by the same angle, the matrix elements are also unchanged. If we consider the null wave number $\omega = 0$, the latter operation reduces to rotating the $\boldsymbol{r}$ coordinates only, which therefore does not affect the matrix elements. Since the time-evolution operator at zero wave number can be diagonalized simultaneously with the rotation operator, we occasionally follow [100] in referring to its eigenstates through the language of atomic orbitals ($1s$, nodeless; $2s$, one radial node; $2p$, one angular node, etc.).

From the discussion in Sec. B 2, it follows that the problem can be treated analogously to the study of zero-temperature phase transitions, in which different phases are often entirely characterized by changes in symmetry.

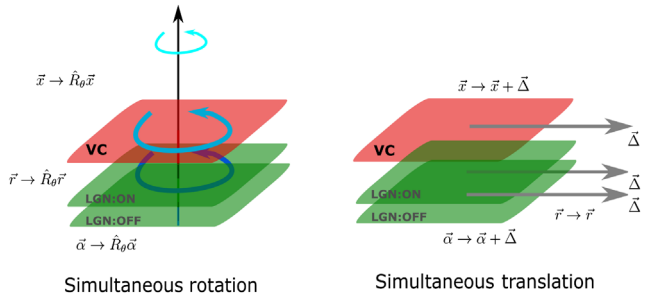

FIG. 5. Invariances of the theory. Depiction of the two main transformations under which the model is invariant: simultaneous rotation and simultaneous translation of the three neuron layers. Notice that the receptive field coordinate $\boldsymbol{r}$ is left untransformed by translations.

Translation symmetry is broken if the principal eigenstate of the system corresponds to a nonzero wave number. Rotation symmetry is broken if the principal eigenstate is not invariant under simultaneous rotations of the wave vector and of the radial coordinate. For instance, this happens if the wave number under consideration is $\omega = 0$ and the eigenfunction has angular momentum $l = 1$ (a "p wave") or any other angle-dependent (hence, orientation-selective) functional form.

It follows that there are multiple symmetry classes for the solution, and it is convenient to introduce shorthand labels for the phases that emerge from the analysis. We call the phase in which no invariance is broken the N phase (nonselective), the R phase is the phase where rotation symmetry is broken but translation symmetry is not, and the T phase is the phase where translation symmetry is broken, and so is rotation symmetry. A summary of these phases is given in Table I.

4. Symmetries of the system: Parity and *CP* symmetry

An important property of the eigenfunctions of $\hat{L}^p$ concerns their behavior under the action of the operators $\not{P}_x$ and $\not{P}_y$ defined by

$$\not{P}_x \psi(r_x, r_y) = \psi(-r_x, r_y),$$
$$\not{P}_y \psi(r_x, r_y) = \psi(r_x, -r_y). \tag{B2}$$

As we take the wave vector $\boldsymbol{\omega}$ to be aligned with the x axis, the commutation rule $[\hat{L}^p, \not{P}_y] = 0$ is immediately verified from Eq. (A39); hence, $\hat{L}^p$ and $\not{P}_x$ can be diagonalized simultaneously, and the eigenfunctions of $\hat{L}^p$ may be chosen as either symmetric or antisymmetric under inversion of the r_y coordinate.

On the other hand, the operators $\hat{L}^p$ and $\not{P}_x$ do not commute, as can be seen from Eq. (A39). However, $\hat{L}^p$

does commute with the product $\mathcal{C}\not{P}_x$, where $\mathcal{C}$ is the antilinear operator such that $\mathcal{C}\psi(r) = \psi^*(r)$.

Writing the complex RF $\psi(r) = u(r) + iv(r)$ as the real-valued vector function $\psi(r) = \binom{u(r)}{v(r)}$, we have that

$$\mathcal{C}\not{P}_x = \begin{pmatrix} \not{P}_x & 0 \\ 0 & -\not{P}_x \end{pmatrix}, \qquad \mathcal{C}\not{P}_y = \begin{pmatrix} \not{P}_y & 0 \\ 0 & -\not{P}_y \end{pmatrix}, \tag{B3}$$

which is a Hermitian operator, so that its eigenvalues must be real. Since $(\mathcal{C}\not{P}_x)^2 = (\mathcal{C}\not{P}_y)^2 = 1$, it follows that the eigenvalues are ± 1.

In this representation, a generic integral operator $\hat{O}$ takes the matrix form $\hat{O} = \left(\begin{smallmatrix} \hat{A} & -\hat{B} \\ \hat{B} & \hat{A} \end{smallmatrix}\right)$, where the kernels $A(r,s)$ and $B(r,s)$ of $\hat{A}$ and $\hat{B}$ are the real and imaginary parts of the integral kernel $O(r,s)$ of $\hat{O}$. Such an operator clearly commutes with multiplications of the wave functions by an arbitrary "gauge factor" $e^{i\theta}$. Indeed, such a gauge transformation is represented by the rotation of the complex plane

$$\hat{R}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{B4}$$

and we have $[\hat{O}, \hat{R}] = 0$.

If $[\hat{O}, \mathcal{C}\not{P}] = 0$ for the parity operator $\not{P}$ corresponding to a given coordinate r, it follows that the eigenfunctions $[u(r), w(r)]$ of $\hat{O}$ can be chosen to be eigenvectors of the operator $\mathcal{C}\not{P}$, whose eigenvalues we discuss after Eq. (B3). That is, they can be chosen to obey the constraint

$$\begin{aligned} \begin{pmatrix} u(-r) \\ -w(-r) \end{pmatrix} &= \mathcal{C}\not{P} \begin{pmatrix} u(r) \\ w(r) \end{pmatrix} \\ &= \lambda_{CP} \begin{pmatrix} u(r) \\ w(r) \end{pmatrix} = \pm \begin{pmatrix} u(r) \\ w(r) \end{pmatrix}, \end{aligned} \tag{B5}$$

from which we can see that either $u(r)$ is symmetric and $w(r)$ is antisymmetric, or vice versa. In both cases, the symmetric and antisymmetric parts of the function are separated by a phase shift of magnitude π.

Applying this to the constrained time-evolution operator $\hat{L}^p$, we conclude that its eigenfunctions will consist of a component ψ_S that is symmetric in $\hat{P}_x$ and a component ψ_S that is antisymmetric, the two components being separated by a phase shift π.

We can thus write

$$\psi(\boldsymbol{r}) \propto \psi_S(\boldsymbol{r}) \pm i\psi_A(\boldsymbol{r}), \tag{B6}$$

where ψ_S and ψ_A are real, and ψ_S (ψ_A) an even (odd) function in r_x.

Notice that the operator $\mathcal{C}\not{P}$ does not commute with the gauge operator $\hat{R}_\theta$ defined by Eq. (B4). This means that by diagonalizing $\mathcal{C}\not{P}$ we effectively fix the gauge of the wave functions. Thus, we show that it is *possible* to write the eigenfunctions of $\hat{L}^p$ in the form $\psi_S(\boldsymbol{r}) + i\psi_A(\boldsymbol{r})$. If we back transform to real space in the cortical coordinates $\boldsymbol{x}$, this means that symmetric and antisymmetric RFs will alternate along the direction of cortical modulation.

5. Diagonalization of the unconstrained dynamics

In Fourier space, the unconstrained two-layer model is given by Eq. (A50), which can be diagonalized exactly.

Indeed, if we define the basis transformation $\Psi(\boldsymbol{r}) = \exp[i(\eta^2/\mu^2)\omega r_x]\chi(r_x, r_y)$, it is clear that $\Psi(\boldsymbol{r})$ is an eigenfunction of $\hat{L}$ if and only if $\chi(\boldsymbol{r})$ is an eigenfunction with the same eigenvalue of the integral operator with kernel

$$L^\chi(\boldsymbol{r}, \boldsymbol{s}) = \exp\left[-\frac{\omega^2}{2\Omega^2} - \frac{r^2 + s^2}{4\rho^2} - \frac{(\boldsymbol{r} - \boldsymbol{s})^2}{2\mu^2}\right]. \tag{B7}$$

The full diagonalization of this operator was first accomplished in Cartesian coordinates by Wimbauer *et al.* [101]. They found that the normalized eigenfunctions have the form

$$\chi_{n_x n_y}(\boldsymbol{r}) = A_{n_x,n_y} e^{-(r^2/2\gamma^2)} H_{n_x}\left(\frac{r_x}{\gamma}\right) H_{n_y}\left(\frac{r_y}{\gamma}\right), \tag{B8}$$

with $A_{n_x,n_y} = (2^{(n_x+n_y/2)}\sqrt{\pi n_x! n_y!}\gamma)^{-1}$, the numbers n_x and n_y being non-negative integers, and the functions H_n Hermite polynomials. The corresponding eigenvalues are

$$\Lambda_{n_x n_y} = 2\pi\mu^2 e^{-\frac{\omega^2}{2\Omega^2}} \beta^{-n_x - n_y - 1}, \tag{B9}$$

and the two parameters entering these formulas are

$$\gamma = \sqrt{2}\rho\left(1 + \frac{4\rho^2}{\mu^2}\right)^{-1/4}, \quad \beta = 1 + \frac{\mu^2}{2\rho^2} + \frac{\mu}{\rho}\sqrt{1 + \frac{\mu^2}{4\rho^2}}. \tag{B10}$$

The parameter γ is the width of the receptive fields, which quantifies how the arbor radius ρ is renormalized by recurrence and input correlations. For fixed ρ, the width γ of the eigenfunctions is a monotonically decreasing function of the ratio ρ/μ. In this representation [having divided the RFs by $\sqrt{A(r)}$ at the outset in Sec. A 2], the unrenormalized arbor radius is represented by $\sqrt{2}\rho$. It follows that, if the ratio ρ/μ is very small, no renormalization occurs: $\gamma = \sqrt{2}\rho$.

If the ratio ρ/μ tends to infinity (i.e., if the arbors are comparatively wide, asymptotically extending over all the cortex), the range of the eigenfunction will be restricted by the correlation-interaction length scale, becoming equal to

the geometric mean of the two length scales, namely, $\gamma \sim \sqrt{\rho\mu}$.

For future reference, we note the three highest-lying eigenfunctions Ψ_{n_x,n_y} of $\hat{L}$:

$$\Psi_{0,0}(\boldsymbol{r}) = \frac{1}{\sqrt{\pi}\gamma}\exp\left(i\frac{\eta^2}{\mu^2}\omega r_x - \frac{r^2}{2\gamma^2}\right), \tag{B11}$$

$$\Psi_{0,1}(\boldsymbol{r}) = \frac{r_x}{\sqrt{\pi}\gamma^2}\exp\left(i\frac{\eta^2}{\mu^2}\omega r_x - \frac{r^2}{2\gamma^2}\right), \tag{B12}$$

$$\Psi_{1,0}(\boldsymbol{r}) = \frac{r_y}{\sqrt{\pi}\gamma^2}\exp\left(i\frac{\eta^2}{\mu^2}\omega r_x - \frac{r^2}{2\gamma^2}\right). \tag{B13}$$

Since the operator $\hat{L}^{\chi}$ is symmetric with respect to rotations of the vector $\boldsymbol{r}$, it can also be diagonalized simultaneously with the generator of rotations for the vector $\boldsymbol{r}$, as done more recently by Davey *et al.* [102]. This leads to writing the eigenvalues of $\hat{L}$ in the equivalent angular form

$$\lambda_{N,m} = 2\pi\mu^2 e^{-(\omega^2/2\Omega^2)}\beta^{-2N-m-1}, \tag{B14}$$

where the integer m is the angular momentum, or the number of angular nodes in the eigenfunctions, of $\hat{L}^{\chi}$ while N is their number of radial nodes, and we define

$$\beta \equiv 1/q = 1 + \frac{\eta^2}{2\rho^2} + \frac{\eta}{\rho}\sqrt{1+\frac{\eta^2}{4\rho^2}}.$$

The corresponding eigenfunctions $\Phi_{N,m}$ of $\hat{L}$ are best written as functions of polar coordinates (r,ϕ). The highest such eigenfunctions are

$$\Phi_{0,0}(\boldsymbol{r}) = \Psi_{0,0}(\boldsymbol{r}), \tag{B15}$$

$$\Phi^{\pm}_{0,1}(\boldsymbol{r}) = \frac{r}{\sqrt{\pi}\gamma^2}\exp\left(i\frac{\eta^2}{\mu^2}\omega r\cos\phi - \frac{r^2}{2\gamma^2} \pm i\phi\right), \tag{B16}$$

$$\Phi_{1,0}(\boldsymbol{r}) = \frac{(\gamma^2 - r^2)}{\sqrt{\pi}\gamma^3}\exp\left(i\frac{\eta^2}{\mu^2}\omega r\cos\phi - \frac{r^2}{2\gamma^2}\right). \tag{B17}$$

The eigenfunctions of the angular-momentum representation with an even (odd) number of angular nodes are built with appropriate Clebsch-Gordan coefficients from eigenfunctions of the Cartesian representation where n_x and n_y have the same (different) parity.

From Eqs. (B9) and (B14), it is seen that the dependence of the eigenvalue on the wave number lies entirely in the exponential prefactor. Hence, the optimal wave number is always $\omega = 0$. Translation symmetry is never broken in the absence of homeostatic constraints.

Since $\hat{L}$ is diagonalizable, and the other operators summed into $\hat{L}^p$ have separable matrix elements, *each* of the four operators summing up to $\hat{L}^p$ in Eq. (A49) is diagonalizable exactly. Unfortunately, the sum of the four is not. But while no closed-form solution is available in general, it will be possible to study the operator separately in various regions of parameter space.

APPENDIX C: COMPUTATION OF THE PHASE DIAGRAM

1. Cortically uniform phases

Let us assume that, for some given values of ζ and η, the principal eigenfunction has the form $\Psi(\boldsymbol{x},\boldsymbol{r}) \equiv \Psi(\boldsymbol{r})$, which is uniform over the cortex, or in other words, that the principle eigenfunction in that point of parameter space is $\omega = 0$. We refer to such regions as "uniform phases." We would like to know, given a point in parameter space where such a phase is dominant, whether it will be of the R or N type.

It can be seen that the operator $\hat{L}^p$ of Eq. (A49), acting in such a case on functions of the single variable $\boldsymbol{r}$, becomes equal to the operator $\hat{L}^f$ with matrix elements

$$\frac{L^f(\boldsymbol{r},\boldsymbol{s})}{\sqrt{A(r)A(s)}} = I^{\mu}(\boldsymbol{r},\boldsymbol{s}) + \int d\boldsymbol{r}_1 d\boldsymbol{s}_1 A(r_1)A(s_1)I^{\mu}(\boldsymbol{r}_1,\boldsymbol{s}_1) - \int d\boldsymbol{u}A(u)I\mu(\boldsymbol{r},\boldsymbol{u}) - \int d\boldsymbol{u}A(u)I\mu(\boldsymbol{u},\boldsymbol{s}), \tag{C1}$$

where $I^{\mu} \equiv = \exp\{-[(\boldsymbol{x}-\boldsymbol{y})^2/2\mu^2]\}$ is a version of the interaction function in Eq. (A14) corrected by the input.

We can also rewrite Eq. (C1) compactly as

$$\hat{L}^f = \hat{L}^{\mu} + |a_0\rangle\langle a_0|\hat{L}^{\mu}|a_0\rangle\langle a_0| - 2\mathrm{HP}[\hat{L}^{\mu}|a_0\rangle\langle a_0|], \tag{C2}$$

where "HP" is the Hermitian part of an operator, and the unconstrained part of the $\hat{L}^f$ operator has matrix elements

$$L^{\mu}(\boldsymbol{r},\boldsymbol{s};\mu) = \sqrt{A(r)A(s)}I^{\mu}(\boldsymbol{r},\boldsymbol{s}). \tag{C3}$$

The rest of this section is devoted to the diagonalization of $\hat{L}^f$, which we perform by treating separately the regimes with large and small values of μ/ρ.

a. Long effective length ($\mu \gg \rho$)

Although the operator $\hat{L}^f$ is not amenable to exact diagonalization, it is easy to show that, in the regime of long effective length ($\mu \gg \rho$), rotation symmetry is broken, leading to the development of orientation selectivity.

To see this, assume self-consistently that all the radial variables in the eigenvalue equation for $\hat{L}^f$ will be confined to a region of order ρ. Hence, expanding the unconstrained operator in Eq. (C62) can be expanded as

$$L^{\mu}(\boldsymbol{r},\boldsymbol{s};\mu) \approx e^{-\frac{r^2+s^2}{4\rho^2}}\left[1-\frac{(\boldsymbol{r}-\boldsymbol{s})^2}{2\mu^2}\right], \tag{C4}$$

where further corrections inside the square brackets are of order $(\rho/\mu)^4$.

If we substitute Eq. (C4) into Eq. (C61), we find that in the asymptotic matrix element the terms of order $(\rho/\mu)^0$ vanish exactly. The terms of order $(\rho/\mu)^2$ cancel each other leaving only the following:

$$\begin{aligned} L^{f}(\boldsymbol{r},\boldsymbol{s};\mu) &\approx \frac{\boldsymbol{r}\cdot\boldsymbol{s}}{\mu^2}\sqrt{A(r)A(s)} \\ &= \frac{rs}{\mu^2}\sqrt{A(r)A(s)}\cos(\phi_r-\phi_s), \end{aligned} \tag{C5}$$

while further corrections are again of order $(\rho/\mu)^4$.

It is clear that the only positive eigenvalue of the operator defined by the kernel (C5) corresponds to the eigenfunction

$$\psi(\boldsymbol{r}) \propto r\sqrt{A(r)}\cos(\phi-\phi_0), \tag{C6}$$

the corresponding eigenvalue being just the p-wave eigenvalue $\lambda_{0,1}$ of the unconstrained model. The expansion is self-consistent because indeed the function (C6) vanishes for $r \gg \rho$.

Following a convention in the literature [65,66], we refer to eigenfunctions $\psi(\boldsymbol{r}) = f(r)$ as "s-wave" states. We call p-wave states eigenfunctions having angular momentum $m = 1$, i.e., with angular dependence $\cos[m(\phi-\phi_0)]$ with $m = 1$. From Eq. (C5), we see that all s-wave eigenstates have zero eigenvalue to this order in the expansion. For sufficiently long effective length μ, the principal eigenspace is thus composed by the p-wave functions described in Eq. (C6).

b. Short effective length ($\mu \ll \rho$)

We argue that p waves dominate the uniform phases in the limit of long effective length $\mu \gg \rho$. We would like to inquire whether there exist regions of parameters where this is not the case, i.e., where rotation symmetry is not broken and s waves dominate the uniform phases. These s waves would describe RFs that are unable to discriminate among the possible orientations of visual input.

If that is the case, there can be no smooth crossover between the two regimes. A linear combination of an s wave ($m = 0$) and of a p wave ($m = 1$) could not be an eigenfunction of $\hat{L}^f$ other than at special points of degeneracy. Let us tentatively call $\Theta_c = \mu_c/\rho$ the largest value of $h = \mu/\rho$ where the principal eigenfunction is nonselective. We would like to find if $\Theta_c > 0$ and, if so, compute the structure of the receptive field for $h < \Theta_c$.

A natural tool to address this question is the variational method for linear operators. We assume a functional form (trial function) for the principal eigenfunction; we normalize it, we find the expectation value of our operator in that state, and we maximize it with respect to variational parameters. This leads to the best available approximation of the principal eigenvalue within the given Hilbert subspace.

The expectation value of the operator $\hat{L}^f$ in the trial state $|\psi\rangle$ is defined as

$$E[\psi] = \frac{\langle\psi|\hat{L}^f|\psi\rangle}{\langle\psi|\psi\rangle}. \tag{C7}$$

It can be shown by the same arguments as in Refs. [65,66] that the principal eigenfunction of $\hat{L}^f$ in the s sector must be of the $2s$ type, i.e., with one radial node. We thus choose our variational trial function to be a RF with the same functional form as the $2s$ eigenfunction of the unconstrained model, only with the position of the node unspecified.

The unconstrained $2s$ wave function is, as per Eq. (B17), a Gaussian RF of width γ multiplied by the polynomial $(\gamma^2 - r^2)$, so that the radial node is located at $r = \gamma$. We now replace the nodal radius γ with an unspecified radius R, obtaining a trial function that is a generalization of Eq. (B17), and we optimize the expectation value of $\hat{L}^f$ with respect to R over all Hilbert space. Our "movable-node" trial function is therefore,

$$\psi_T^{(R)}(\boldsymbol{r}) = \frac{N}{\sqrt{\pi}\gamma}(R^2-r^2)\exp\left(-\frac{r^2}{2\gamma^2}\right), \tag{C8}$$

where the value of γ is given by Eq. (B10), and we introduce the normalization factor $N = [\gamma^4 + (\gamma^2 - R^2)^2]^{-1/2}$.

Let us consider the expectation value Eq. (C7) of the unconstrained operator of Eq. (C62) in the state (C8). This is given by

$$\begin{aligned} \mathcal{E}_0(R) &\equiv \langle\psi_T^{(R)}|\hat{L}^{\mu}|\psi_T^{(R)}\rangle \\ &= \left(\frac{N^2}{\pi\gamma^2}\right)4\pi^2\int_0^{\infty} dr\, r(R^2-r^2) \\ &\quad\times\int_0^{\infty} ds\, s(R^2-s^2)I_0 \\ &\quad\times\left(\frac{rs}{\mu^2}\right)e^{-\left(\frac{1}{2\rho^2}+\frac{1}{\mu^2}+\frac{1}{\gamma^2}\right)\frac{r^2+s^2}{2}}, \end{aligned} \tag{C9}$$

which, after integration, yields

$$\mathcal{E}_0(R) = \frac{2\pi\mu^2}{\beta^3}\frac{1+\beta^2(1-R^2/\gamma^2)^2}{1+(1-R^2/\gamma^2)^2} \tag{C10}$$

with β defined in Eq. (B10).

To optimize this expectation value, we need to maximize the function $f(x) = [(1+\beta^2 x/1+x)$, where $x = (R^2/\gamma^2-1)^2$. The derivative is $f'(x) = [\beta^2 - 1/(1+x)^2]$,

always non-negative because $\beta \geq 1$; hence, it is sufficient to maximize x, which is done by choosing the limit $R \to \infty$. The result is unsurprising: In the limit $R \to \infty$, the moveable-node function becomes in fact nodeless, and it is nothing but the $1s$ Gaussian of width γ which we know as the principal eigenfunction of $\hat{L}^{\mu}$.

Let us now consider the expectation value Eq. (C7) of the full operator $\hat{L}^{f}$, as described by Eq. (C64), calculated in the moveable-node state of Eq. (C8). This can be written as

$$\mathcal{E}(R) = \mathcal{E}_0(R) + \frac{\mu^2 A^2}{\mu^2 + 2\rho^2} - \frac{2\mu^2 AB}{\mu^2 + \rho^2}, \tag{C11}$$

where

$$A = \frac{N}{\sqrt{\pi}\gamma} \times 2\pi \int_0^{\infty} (R^2 - r^2) \exp\left[-\left(\frac{1}{\gamma^2} + \frac{1}{2\rho^2}\right)\frac{r^2}{2}\right] r\, dr, \tag{C12}$$

$$B = \frac{N}{\sqrt{\pi}\gamma} \times 2\pi \int_0^{\infty} (R^2 - r^2) \times \exp\left[-\left(\frac{1}{\gamma^2} + \frac{1}{2\rho^2} + \frac{1}{\rho^2 + \mu^2}\right)\frac{r^2}{2}\right] r\, dr, \tag{C13}$$

or, upon integration,

$$A = 4\sqrt{\pi} N \gamma \rho^2 \frac{(2\rho^2 + \gamma^2)R^2 - 4\gamma^2\rho^2}{(2\rho^2 + \gamma^2)^2}, \tag{C14}$$

$$B = 4\sqrt{\pi} N \gamma \rho^2 (\rho^2 + \mu^2) \times \frac{2\rho^2(R^2 - 2\gamma^2)(\rho^2 + \mu^2) + R^2\gamma^2(3\rho^2 + \mu^2)}{[(\rho^2 + \mu^2)(2\rho^2 + \gamma^2) + 2\gamma^2\rho^2]^2}. \tag{C15}$$

While expression (C11) with the substitution of Eqs. (C14) and (C15) is somewhat intricate, we are ultimately interested only in its maximal value over all the range of nodal radii R. We thus expand $\mathcal{E}$ in $h = \mu/\rho$ with the ansatz $R^2 = \rho^2(k^2 h + O(h^2))$, yielding

$$\frac{\mathcal{E}(\rho k\sqrt{h})}{2\pi\mu^2} = 1 - 2f(k)h + O(h^2), \tag{C16}$$

where $f(k) = (8 - 7k^2 + 2k^4/2 - 2k^2 + k^4)$. The requirements $f'(\bar{k}) = 0, f''(\bar{k}) > 0$ lead to

$$\bar{k} = \sqrt{\frac{4 + \sqrt{10}}{3}}, \tag{C17}$$

which means that the node behaves as $R \sim [(4+\sqrt{10}/3)\mu\rho]^{1/2}$. Inserting this into Eq. (C8) and applying Eq. (A32) leads straight to Eq. (1) in the main text (for a comparison with numerics, see Fig. 6).

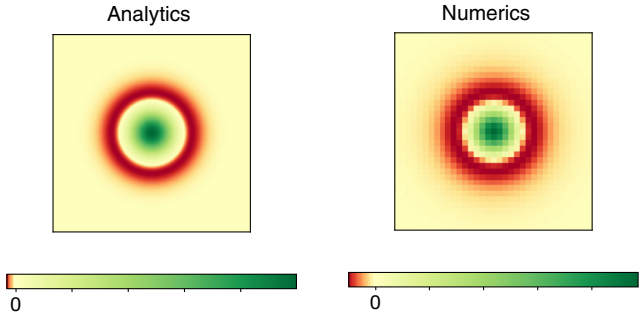

FIG. 6. Receptive fields for the N phase. Example of the variational RF of Eq. (1) in the main text compared to the result of numerically diagonalizing the full operator $\hat{L}^{p}$ and rescaling the eigenfunction by Eq. (A32). The parameters used here are $\zeta/\rho = 0.02$, $\eta\rho = 0.2$. A side of the grid has length equal to 5ρ; color scale ranges between min and values.

Further substituting into Eq. (C11), we find that the optimal expectation value is

$$\mathcal{E} \equiv \mathcal{E}(\bar{k}\sqrt{h}) = 2\pi\mu^2[1 - (5 - \sqrt{10})h]. \tag{C18}$$

We can now compare $\mathcal{E}$ with the exact eigenvalue of the dominant p wave, which is given by Eq. (B14) as $\lambda_{0,1} = 2\pi\mu^2/\beta^2 \sim 1\text{–}2h$. Since $(5 - \sqrt{10}) \sim 1.83 < 2$, we conclude that the principal s-wave eigenvalue

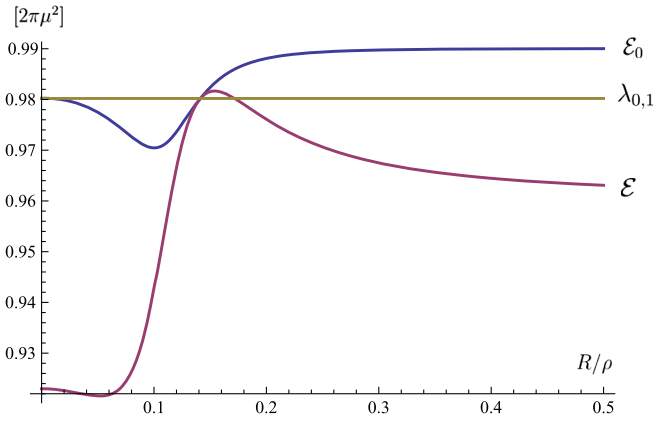

FIG. 7. Variational landscape for cortically uniform phases. Expectation values of the constrained time-evolution operator $\hat{L}^{f}$ (in units of $2\pi\mu^2$) plotted as a function of R/ρ. The three curves refer to (i) $\lambda_{0,1}$ (expectation value of $\hat{L}^{\mu}$ or $\hat{L}^{f}$ in the exact eigenfunction $\Psi_{0,1}$) plotted in green; (ii) $\mathcal{E}_0$, expectation value of the unconstrained operator $\hat{L}^{\mu}$ in the moveable-node state ψ_T plotted in blue; (iii) $\mathcal{E}$, expectation value of the constrained operator $\hat{L}^{f}$ in the state ψ_T plotted in purple. The figure refers to $h = (\mu/\rho) = .01$. Values of R for which the moveable-node state is preferred to the orientation-selective state are different for the two operators $\hat{L}^{\mu}$ and $\hat{L}^{f}$. Namely, there exists a minimal value $\tilde{R}$, in this example being approximately equal to 0.14, such that $\hat{L}^{\mu}$ opts for ψ_T at sufficiently high values of the node radius $R > \tilde{R}$, while $\hat{L}^{f}$ does so for values of R in a narrow window $R \gtrsim \tilde{R}$.

approximated by Eq. (C8) lies *higher*. Therefore, the s waves do indeed dominate for small $h = (\mu/\rho)$.

The eigenvalue landscape leading to dominance of s waves is displayed in full in Fig. 7 for a fixed (sufficiently low) value of $h = (\mu/\rho)$. As can be seen, choices of the movable node below a certain threshold $\tilde{R}(h)$ would lead to s dominance, but the optimal R in the presence of constraints (purple curve) lies beyond that threshold.

c. Phase boundary of the uniform phases

We now like to have a lower bound on the critical value of the interaction length $\mu_c = \Theta_c \rho$ at which rotation symmetry is first broken. We can define Θ_c as the largest value of μ/ρ where the s mode dominates. We proceed by finding the h for which the expectation value of Eq. (C18) is equal to the exact $2p$ eigenvalue. The fact that this will indeed yield a lower bound on the actual value of the transition point can be understood as follows.

If the variational method reveals the transition at a point $h = \theta_c$, it means that we have found an s-wave state whose expectation value is larger than the exact eigenvalue of the principal p wave for all $h < \theta_c$. Suppose *ad absurdum* that the actual critical point Θ_c is $\Theta_c < \theta_c$. That means in the region $\Theta_c < h < \theta_c$ the actual principal state of the operator is a orientation selective, i.e., $m > 0$. And since the $m > 0$ sector is exactly diagonalizable, this principal p wave must be the one we already calculated, with eigenvalue $\lambda_{0,1}$.

But if that was true, all the s-wave functions would yield expectation values lower than that eigenvalue. Then it would not be possible to create a linear combination of them (our trial function) that yields an expectation value $> \lambda_{0,1}$, as we have done. We deduce that we must have $\Theta_c \geq \theta_c$. That is, the variational method provides a lower bound on the actual critical point.

Let us proceed with the calculation. We first expand R to a higher order as $R^2/\rho^2 = c_1 h + c_2 h^2 + O(h^3)$. The coefficient c_1 can be determined by maximizing Eq. (C11) to the order $O(h)$, which gives $c_1 = \bar{k}^2 = (4 + \sqrt{10}/3)$. Then we calculate the second term in the expansion of $\mathcal{E}$, plug in the value of c_1 we found, and maximize with respect to c_2. This second-order correction computed at the optimal value of c_2 is then included in the expectation value, and the whole thing is compared to the eigenvalue of the leading p waves to see which is dominating. One obtains

$$\frac{\mathcal{E}}{2\pi\mu^2} \sim 1 - (5 - \sqrt{10})h + \left(\frac{33}{2} - \frac{51}{\sqrt{10}}\right)h^2. \tag{C19}$$

The critical point θ_c is found where this s-wave expectation value intersects the p-wave eigenvalue given by Eq. (B14), that is,

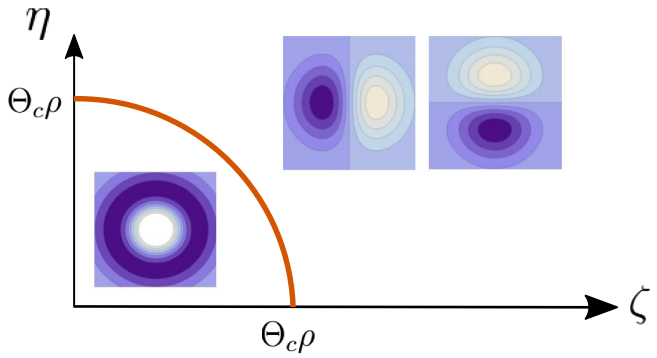

FIG. 8. Structure of cortically uniform phases over the phase diagram. Insets show plots of the principal eigenfunctions, respectively, the movable-node approximation for the N phase and the exact eigenfunctions with longitudinal and transverse alignments for the R phase.

$$\frac{\lambda_{0,1}}{2\pi\mu^2} \sim 1 - 2h + 2h^2; \tag{C20}$$

setting $\mathcal{E} = \lambda_{0,1}$ yields

$$\theta_c = \frac{2(75 - 8\sqrt{10})}{997} \approx 0.1, \tag{C21}$$

which is a rigorous lower bound to the critical point. The actual value is easiest to find numerically by projecting the operator $\hat{L}^f$ into the $m = 0$ subspace and thus turned into an operator $L^{(s)}$ acting on functions of the sole radial variable, whose matrix element is

$$\begin{aligned} L^{(s)}(r,s) = {} & I_0\left(\frac{rs}{\mu^2}\right) e^{-\left(\frac{1}{2\rho^2}+\frac{1}{\mu^2}\right)\frac{r^2+s^2}{2}} \\ & + \frac{\mu^2}{\mu^2 + 2\rho^2} e^{-\frac{r^2+s^2}{4\rho^2}} \\ & - \frac{\mu^2}{\mu^2 + \rho^2}\left(e^{-\frac{r^2}{4\tilde{\rho}^2}-\frac{s^2}{4\rho^2}} + e^{-\frac{r^2}{4\rho^2}-\frac{s^2}{4\tilde{\rho}^2}}\right), \end{aligned} \tag{C22}$$

where $\tilde{\rho} = [(1/\rho^2) + (2/\mu^2 + \rho^2)]^{-1/2}$, and I_0 is the modified Bessel function of the first kind. The principal eigenvalue of $L^{(s)}$ must be compared to the exact p-wave eigenvalue so as to obtain the transition point yielding $\Theta_c \approx 0.34$.

The resulting phase diagram for uniform phases is illustrated in Fig. 8. In terms of the general model (which allows modulation across the cortex as well), the regions within and outside the quarter circle of Fig. 8 can be taken to identify forbidden regions for the R and N phases. Within the quarter circle ($\mu < \Theta_c \rho$), the R phase is forbidden because, if at any point the optimal wave number happens to be zero, it must yield an N phase instead. Outside the quarter circle ($\mu > \Theta_c \rho$), the N phase is

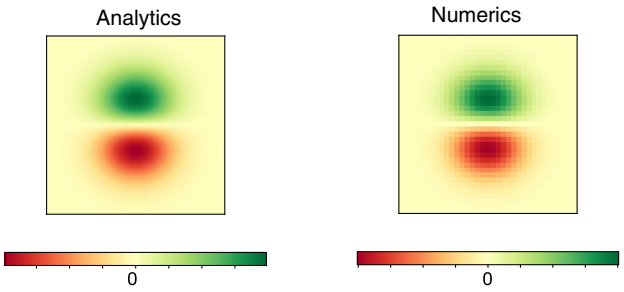

FIG. 9. Receptive fields for the R phase. Example of the RF of Eq. (2) compared to the result of numerically diagonalizing the full operator $\hat{L}^p$ and rescaling the eigenfunction by Eq. (A32). The parameters used here are $\zeta/\rho = 0.05, \eta\rho = 0.7$. A side of the grid has length equal to 5ρ; color scale ranges between min and values.

forbidden because, if the optimal wave number is zero, it must yield an R phase instead.

d. Degeneracies in the uniform phases

To sum up, we find that if the uniform phase is dominant ($\omega = 0$), the principal eigenfunction of the operator is orientation selective if $\mu > \Theta_c\rho$ and nonselective if $\mu < \Theta_c\rho$, with the critical ratio Θ_c bounded from below by the value θ_c of Eq. (C21).

For $\mu > \Theta_c\rho$, there is a two-dimensional degeneracy in the orientation of the symmetry breaking. The exact principal eigenfunction as per Eqs. (A16) and (A17) is given by

$$\Psi(\boldsymbol{x},\boldsymbol{r}) = (k_+ e^{+i\phi} + k_- e^{-i\phi}) \exp\left[-\frac{r^2}{2\gamma^2}\right] \tag{C23}$$

for arbitrary coefficients k_+ and k_-. We refer in particular to the combinations

$$\begin{pmatrix} \Psi^x(\boldsymbol{x},\boldsymbol{r}) \\ \Psi^y(\boldsymbol{x},\boldsymbol{r}) \end{pmatrix} = \begin{pmatrix} r_x \\ r_y \end{pmatrix} \exp\left(-\frac{r_x^2 + r_y^2}{2\gamma^2}\right), \tag{C24}$$

which are the instances of the $2p$ waves $\Phi_{0,1}$ and $\Phi_{1,0}$ corresponding to uncorrelated input.

Since the cortical wave vector is aligned along the x axis, Ψ^x describes RFs aligned parallel to the cortical wave vector, and Ψ^y describes the RFs aligned orthogonally to it. Accordingly, we call Ψ^x the longitudinal eigenfunction and Ψ^y the transverse eigenfunction. Equation (2) of the main text is obtained from Eq. (C24) by applying Eq. (A32), and is compared to the numerics in Fig. 9.

These two eigenfunctions share, as per Eq. (B14), the exact eigenvalue

$$\Lambda^{x,y} = 2\pi\mu^2 \left(1 + \frac{\mu^2}{2\rho^2} + \frac{\mu}{\rho}\sqrt{1 + \frac{\mu^2}{4\rho^2}}\right)^{-2}. \tag{C25}$$

2. The long-range limit

a. The long-range limit: Derivation

We call the long-range limit ($\mu \gg \rho$) the case where either cortical interactions are long range ($\eta \gg \rho$) or LGN interactions are ($\zeta \gg \rho$), or both.

We begin by Taylor expanding Eq. (A50) into

$$\hat{L}^p \sim \hat{L}_0 + \hat{R} + \hat{S} + \hat{T} + \hat{T}^\dagger \tag{C26}$$

with

$$L_0(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\omega}) = \exp\left[-\frac{\omega^2}{2\Omega^2} - i\frac{\eta^2}{\mu^2}\boldsymbol{\omega}(\boldsymbol{r}-\boldsymbol{s}) - \frac{r^2+s^2}{4\hat{\rho}^2}\right], \tag{C27}$$

$$R(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\omega}) = \left(\frac{r_i s_i}{\mu^2} + \frac{r_i r_j s_i s_j}{2\mu^4} + \ldots\right) \times \exp\left[-\frac{\omega^2}{2\Omega^2} - i\frac{\eta^2}{\mu^2}\boldsymbol{\omega}(\boldsymbol{r}-\boldsymbol{s}) - \frac{r^2+s^2}{4\hat{\rho}^2}\right], \tag{C28}$$

where summation over repeated indices is implied. The first terms in R are the dipole and quadrupole components, whereas the quantity $\hat{\rho} = [(1/\rho^2) + (2/\mu^2)]^{-1/2}$ plays the role of a renormalized "mass."

We are interested in the structure of the phase diagram in the leading order in the small parameter of ρ/μ. Since we are interested in the leading order, it appears from the equations that we may replace the renormalized $\hat{\rho}$ with the bare ρ. Moreover, within expressions (A51) and (A52) for $\hat{S}$ and $\hat{T}$, denominators of the form $\mu^2 + \rho^2$ and $\mu^2 + 2\rho^2$ can be approximated with μ^2.

In every integration where this kernel would play a role, variables representing relative LGN-cortex coordinates are confined by the arbor densities to a radius of order ρ, and if we rely on the smallness of ρ/μ, we can also rely on the smallness of the variables r/μ, s/μ in their absolute values. However, these variables are also associated with angular directions which can lead to annihilating whole terms of an operator, no matter how large ρ/μ, when integrated over the orthogonal angular component. Thus, the smallness of r/μ, s/μ cannot be used to discard Eq. (C28) by comparison with Eq. (C27).

The approximation we pursue is to discard all but the dipole term in R [first term in expression (C28)]. This is indeed the simplest restriction of Hilbert space that allows us to explore whether, anywhere in the phase diagram, the system breaks out of circular symmetry. Doing so transforms Eq. (C26) into

$$L^p(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\omega}) = e^{-\frac{\omega^2}{2\Omega^2}}\{a_c^*(\boldsymbol{r})a_c(\boldsymbol{s}) + \boldsymbol{A}^*(\boldsymbol{r})\boldsymbol{A}(\boldsymbol{s}) + q^2 a_1^*(\boldsymbol{r})a_1(\boldsymbol{s}) - q[a_1^*(\boldsymbol{r})a_d(\boldsymbol{s}) + a_d^*(\boldsymbol{r})a_1(\boldsymbol{s})]\}, \tag{C29}$$

where $c=(\eta/\mu)^2$, $d=(\eta^2+\rho^2/\mu^2)$, $q=\exp[-(\zeta^4\rho^2\omega^2/2\mu^4)]$, and we define the functions

$$a_v(r) = \exp\left(iv\omega r_x - \frac{r^2}{4\rho^2}\right),$$
$$\boldsymbol{A}(\boldsymbol{r}) = \frac{\boldsymbol{r}}{\mu}\exp\left[-\frac{r^2}{4\rho^2} + i\frac{\eta^2}{\mu^2}\omega r_x\right], \tag{C30}$$

where the index v takes the values 1, c, and d.

We now treat the $\eta \gg \rho$ and $\zeta \gg \rho$ cases separately, even if these assumptions lead to similar results.

The regime $\eta \gg \rho$.—If $\eta \gg \rho$, we have $d \sim c$, so the operator Eq. (C26) becomes

$$L^p(\boldsymbol{r},\boldsymbol{s};\boldsymbol{\omega}) = e^{-\frac{\omega^2}{2\Omega^2}}[a_c^*(\boldsymbol{r})a_c(\boldsymbol{s}) + \boldsymbol{A}^*(\boldsymbol{r})\boldsymbol{A}(\boldsymbol{s}) + q^2 a_1^*(\boldsymbol{r})a_1(\boldsymbol{s}) - q(a_1^*(\boldsymbol{r})a_c(\boldsymbol{s}) + a_c^*(\boldsymbol{r})a_1(\boldsymbol{s}))]. \tag{C31}$$

Given one eigenfunction $\psi(r)$, let us now define the two unknowns $I_v = \int a_v(\boldsymbol{r})\psi(\boldsymbol{r})d\boldsymbol{r}$ for $v = 1, c$, and the third unknown $K = \int A_x(\boldsymbol{r})\psi(\boldsymbol{r})d\boldsymbol{r}$, and use the self-consistent assumption that $\int A_y(\boldsymbol{r})\psi(\boldsymbol{r})d\boldsymbol{r} = 0$ (which is checked below in Sec. C 2 b). The eigenvalue equation for the operator of Eq. (C31) becomes

$$\lambda e^{\frac{\omega^2}{2\Omega^2}}\psi(\boldsymbol{r}) = I_1[q^2 a_1^*(\boldsymbol{r}) - q a_c^*(\boldsymbol{r})] + I_c[-q a_1^*(\boldsymbol{r}) + a_c^*(\boldsymbol{r})] + K A_x^*(\boldsymbol{r}), \tag{C32}$$

and computing the three unknown integrals from Eq. (C32) itself, one obtains

$$\frac{\lambda}{2\pi\rho^2}e^{\frac{\omega^2}{2\Omega^2}}I_1 = i\frac{\rho^2\zeta^2}{\mu^3}\omega q K, \tag{C33}$$

$$\frac{\lambda}{2\pi\rho^2}e^{\frac{\omega^2}{2\Omega^2}}I_c = (q^3 - q)I_1 + (1-q^2)I_c, \tag{C34}$$

$$\frac{\lambda}{2\pi\rho^2}e^{\frac{\omega^2}{2\Omega^2}}K = -q^3\frac{i\rho^2\zeta^2}{\mu^3}\omega I_1 + q^2\frac{i\rho^2\zeta^2}{\mu^3}\omega I_c + \frac{\rho^2}{\mu^2}K, \tag{C35}$$

from which it follows that we can replace the infinite-dimensional operator of Eq. (A49) with the 3×3 matrix $\hat{L} = 2\pi\rho^2 e^{-(\omega^2/2\Omega^2)}\hat{M}$, where

$$\hat{M} = \begin{pmatrix} 0 & 0 & Jq \\ q^3 - q & 1-q^2 & 0 \\ -q^3 J & q^2 J & (\rho/\mu)^2 \end{pmatrix} \tag{C36}$$

for $J = i\rho^2\zeta^2\omega/\mu^3$.

This matrix has only two nonzero eigenvalues, both positive as we may expect from the discussion in Sec. B 1. The larger one is

$$\lambda = \pi\rho^2 e^{-\frac{\omega^2}{2\Omega^2}}\left(1 + \rho^2/\mu^2 - e^{-\omega^2\zeta^4\rho^2/\mu^4} + \sqrt{(1-\rho^2/\mu^2 - e^{-\omega^2\zeta^4\rho^2/\mu^4})^2 + \frac{4\rho^4\zeta^4\omega^2}{\mu^6}e^{-2\omega^2\zeta^4\rho^2/\mu^4}}\right). \tag{C37}$$

The corresponding eigenfunction is obtained from Eq. (C32) through the principal eigenvector of the matrix $\hat{M}$. This is found from Eq. (C36) to be, before normalization,

$$I_1 = 2J^2 q^3, \tag{C38}$$

$$I_c = (1-q^2)[1 - s - q^2 + \sqrt{(1-s-q^2)^2 - 4J^2q^4}], \tag{C39}$$

$$K = Jq^2[1 + s - q^2 + \sqrt{(1-s-q^2)^2 - 4J^2q^4}] \tag{C40}$$

with $s = (\rho/\mu)^2$. We now take the long-range limit as $s \to 0$ while keeping q fixed, which yields

$$(I_1, I_c, K) \to (0, 1, 0); \tag{C41}$$

hence, the principal eigenfunction for the kernel Eq. (C31) is found to be

$$\psi(\boldsymbol{r}) = \psi_1(\boldsymbol{r};\omega_M\hat{\omega}) \propto -q a_1^*(\boldsymbol{r}) + a_c^*(\boldsymbol{r}) = e^{-\frac{r^2}{4\rho^2}}[e^{-i\frac{\eta^2}{\mu^2}\omega_M\hat{\omega}\boldsymbol{r}} - e^{-i\omega_M\hat{\omega}\boldsymbol{r} - \frac{1}{2}(\frac{\zeta^2\rho\omega_M}{\mu^2})^2}], \tag{C42}$$

where $\hat{\omega}$ is an arbitrary unit vector.

The regime $\zeta \gg \rho$.—For $\zeta \gg \rho$, the eigenvalue equation of Eq. (C29) can be written as

$$\lambda e^{\frac{\omega^2}{2\Omega^2}}\psi(\boldsymbol{r}) = I_1[q^2 a_1^*(\boldsymbol{r}) - q a_d^*(\boldsymbol{r})] + I_c a_c^*(\boldsymbol{r}) - I_d q a_1^*(\boldsymbol{r}) + K A_x^*(\boldsymbol{r}), \tag{C43}$$

where we define the three unknown quantities $I_v = \int a_v(\boldsymbol{r})\psi(\boldsymbol{r})d\boldsymbol{r}$ (for $v = 1, c, d$) and the fourth unknown $K = \int A_x(\boldsymbol{r})\psi(\boldsymbol{r})d\boldsymbol{r}$. Again, we are using the self-consistent assumption that $\int A_y(\boldsymbol{r})\psi(\boldsymbol{r})d\boldsymbol{r} = 0$, which is duly checked in Sec. C 2 b.

Define $J_x = \exp\left(-\frac{\rho^2 x^2 \omega^2}{2}\right)$, so that

$$q = J_{1-c}, \tag{C44}$$

$$\int a_\alpha(\boldsymbol{r}) a_\beta^*(\boldsymbol{r}) d\boldsymbol{r} = 2\pi\rho^2 J_{\alpha-\beta}, \tag{C45}$$

$$\int a_\beta(\boldsymbol{r}) A_x^*(\boldsymbol{r}) d\boldsymbol{r} = (2\pi\rho^2) i \frac{\rho^2\omega}{\mu}(\beta - c) J_{\beta-c}, \tag{C46}$$

while $\int d\boldsymbol{r} A_x(\boldsymbol{r}) A_x^*(\boldsymbol{r}) = 2\pi\rho^4/\mu^2$.

From Eq. (C43), we obtain

$$\frac{\lambda}{2\pi\rho^2} e^{\frac{\omega^2}{2\Omega^2}} I_1 = (q^2 - qJ_{1-d}) I_1 + qI_c - qI_d + i\frac{\rho^2\omega}{\mu}(1-c) qK, \tag{C47}$$

$$\frac{\lambda}{2\pi\rho^2} e^{\frac{\omega^2}{2\Omega^2}} I_c = (q^3 - qJ_{d-c}) I_1 + I_c - q^2 I_d, \tag{C48}$$

$$\frac{\lambda}{2\pi\rho^2} e^{\frac{\omega^2}{2\Omega^2}} I_d = (q^2 J_{1-d} - q) I_1 + J_{d-c} - qJ_{1-d} I_d + \frac{i\omega\rho^4}{\mu^3} qK, \tag{C49}$$

$$\frac{\lambda}{2\pi\rho^2} e^{\frac{\omega^2}{2\Omega^2}} K = \left(-i\frac{\omega\zeta^2\rho^2 q^3}{\mu^3} + iq\frac{\omega\rho^4}{\mu^3} J_{d-c}\right) I_1 + i\frac{\omega\zeta^2\rho^2}{\mu^3} q^2 I_d + \frac{\rho^2}{\mu^2} K, \tag{C50}$$

from which it follows that, in this limit, we can replace our infinite-dimensional operator with the 4×4 matrix $\hat{L} = 2\pi\rho^2 e^{-(\omega^2/2\Omega^2)} \hat{M}$, where

$$\hat{M} = \begin{pmatrix} q^2 - qJ_{1-d} & q & -q & \frac{i\omega\zeta^2\rho^2}{\mu^3} q \\ q^3 - qJ_{d-c} & 1 & -q^2 & 0 \\ q^2 J_{1-d} - q & J_{d-c} & -qJ_{1-d} & \frac{i\omega\rho^4}{\mu^3} q \\ -i\frac{\omega\zeta^2\rho^2 q^3}{\mu^3} + iq\frac{\omega\rho^4}{\mu^3} J_{d-c} & 0 & i\frac{\omega\zeta^2\rho^2}{\mu^3} q^2 & \frac{\rho^2}{\mu^2} \end{pmatrix}. \tag{C51}$$

Now, we have $d - c = (\rho^2/\mu^2)$ and $1 - d = (\zeta^2 - \rho^2/\mu^2)$. Since we are considering the regime where $\zeta \gg \rho$, we can write $1 - d \sim (\zeta^2/\mu^2)$, so that $J_{1-d} \sim q$. Notice that we are making no assumption on the magnitude of η. The matrix thus simplifies to

$$\hat{M} = \begin{pmatrix} 0 & q & -q & \frac{i\omega\rho^2\zeta^2}{\mu^3} q \\ q^3 - qJ_{d-c} & 1 & -q^2 & 0 \\ q^3 - q & J_{d-c} & -q^2 & \frac{i\omega\rho^4}{\mu^3} q \\ -i\frac{\omega\zeta^2\rho^2}{\mu^3} q^3 + i\frac{\omega\rho^4}{\mu^3} qJ_{d-c} & 0 & i\frac{\omega\zeta^2\rho^2}{\mu^3} q^2 & \frac{\rho^2}{\mu^2} \end{pmatrix}. \tag{C52}$$

Let us adopt one more self-consistent assumption concerning the optimal wave number, which is immediately verified once the optimal wave number is computed from the resulting eigenvalue. Namely, we assume $\omega \ll (\mu/\rho^2)$, so that we can write $J_{d-c} \sim 1$ and neglect the terms in $\omega\rho^4/\mu^3$. The matrix Eq. (C52) becomes

$$\hat{M} = \begin{pmatrix} 0 & q & -q & \frac{i\omega\rho^2\zeta^2}{\mu^3} q \\ q^3 - q & 1 & -q^2 & 0 \\ q^3 - q & 1 & -q^2 & 0 \\ -i\frac{\omega\zeta^2\rho^2}{\mu^3} q^3 & 0 & i\frac{\omega\zeta^2\rho^2}{\mu^3} q^2 & \frac{\rho^2}{\mu^2} \end{pmatrix}. \tag{C53}$$

We reduce an infinite-dimensional problem to a four-dimensional problem, which we can solve exactly. From Eq. (C53), we see that

$$\det(\hat{M} - \lambda) = \lambda^2 \left[\frac{\rho^2}{\mu^2} - \frac{\omega^2\rho^4\zeta^4}{\mu^6} q^4 - \frac{\rho^2}{\mu^2} q^2 + \lambda\left(q^2 - 1 - \frac{\rho^2}{\mu^2}\right) + \lambda^2\right], \tag{C54}$$

and from Eq. (C54), we find that the two non-null eigenvalues correspond to those of Eq. (C36). Hence, formula (C37) for the eigenvalue still holds true and, in particular, the optimal wave number will be the same in the two regimes.

b. The long-range limit: Analysis of results

Phase boundary and critical behavior.—The system is in the T phase if the wave number maximizing the principal eigenvalue is positive, while it is in either the R or N phase if that optimal wave number is null. In terms of the dimensionless variable $x = \omega^2\zeta^4\rho^2/\mu^4$ (such that $q = e^{-x/2}$), we can write the principal eigenvalue (C37) as

$$\lambda = \pi\rho^2 f(x), \tag{C55}$$

$$f(x) = e^{-\alpha x/2}\left(1 + s - e^{-x} + \sqrt{(1 - s - e^{-x})^2 + 4sxe^{-2x}}\right) \tag{C56}$$

with $\alpha = (\mu\eta/\zeta\rho)^2$.

For η/ζ of order 1 and $\mu \gg \rho$, α diverges, so the exponential prefactor in Eq. (C56) confines x to values

of order $1/\alpha$. We can thus expand the expression in parentheses in x without any assumption on the magnitude of s, yielding

$$f(x) \sim 2e^{-\alpha x/2}(s+x)$$

with derivative $f'(x) \sim e^{-\alpha x/2}(2-\alpha s-\alpha x)$. Since this corresponds to a single maximum, the condition for T-phase dominance is simply $f'(0) > 0$, i.e.,

$$\zeta > \zeta_c(\eta) = \sqrt{\frac{1}{2}}\eta, \tag{C57}$$

whereas the wave number near the phase boundary is given by

$$\omega \sim \frac{\mu}{\zeta^2}\sqrt{\frac{2\zeta^2}{\eta^2}-1}.$$

Form of the eigenfunction.—Separately pursuing as above the assumptions $\eta \gg \rho$ and $\zeta \gg \rho$ leads, as we see, to the same eigenfunction (C42). This can be written as

$$\psi(\boldsymbol{r}) \propto e^{-\frac{r^2}{4\rho^2}-i\frac{\eta^2}{\mu^2}\boldsymbol{\omega}\boldsymbol{r}}\left(1-e^{-i\frac{\zeta^2}{\mu^2}\boldsymbol{\omega}\boldsymbol{r}-\frac{1}{2}\frac{\zeta^4}{\mu^4}\rho^2\omega^2}\right). \tag{C58}$$

Notice that the value of ω to be plugged into Eq. (C58) is the value that maximizes the eigenvalue (C37). In regimes where the optimal wave number is null, we must take the $\omega \to 0$ limit in Eq. (C58). Expanding the two complex exponentials to the first order in ω and keeping only the lowest order in the result yields the unnormalized eigenfunction

$$\psi(\boldsymbol{r}) \propto r_x \exp\left(-\frac{r^2}{4\rho^2}\right) \tag{C59}$$

equal to the orientation-selective eigenfunction we find for the zero-wave-number region, that is, to an R phase. (In those parts of the phase diagram, therefore, the homeostatic constraint is satisfied through the individual selectivity of cells, and does not need to be satisfied through variations over cortical space; that is why translation symmetry can be restored.)

With $\psi(\boldsymbol{r};\boldsymbol{\omega})$ given by Eq. (C58), the eigenfunctions $\psi(\boldsymbol{x},\boldsymbol{r}) = \psi(\boldsymbol{r};\boldsymbol{\omega})e^{i\boldsymbol{\omega}\boldsymbol{x}}$ for wave vectors $\boldsymbol{\omega}$ and $-\boldsymbol{\omega}$ are degenerate and complex conjugates of each other. A real linear combination of the two is obtained by taking either the real or imaginary part. From this, via Eq. (A32), Eq. (3) of the main text is obtained (see Fig. 10).

Normally oriented eigenfunctions.—In order to obtain Eq. (C42), we make at the very outset [below (Eq. (C31)] the self-consistent assumption $\int A_y^*(\boldsymbol{r})\psi(\boldsymbol{r})d\boldsymbol{R} = 0$, which we use to write both Eqs. (C32) and (C43). The subspace we focus on is indeed orthogonal to A_y, and we find this subspace to be an asymptotic eigenspace of the system; see Eq. (C58). Nonetheless, the same system may also possess eigenfunctions having a nonzero overlap with $A_y(\boldsymbol{r})$. Do these eigenfunctions correspond to an eigenvalue higher than those we calculate?

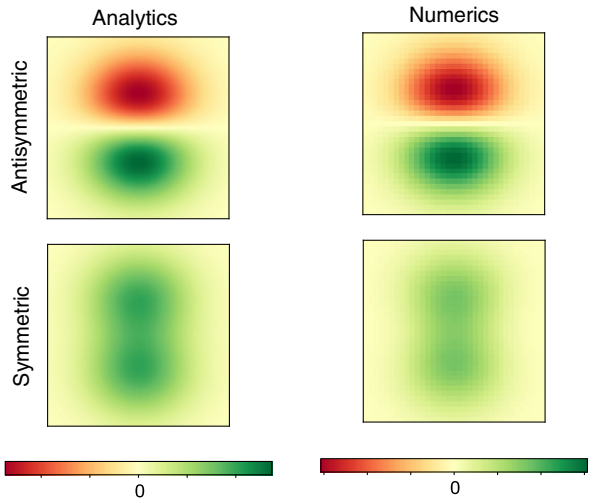

FIG. 10. Receptive fields for the T phase. Plots of the symmetric and antisymmetric components of Eq. (3) in the main text to the Fourier-transformed eigenfunction compared to the result of numerically diagonalizing the full operator $\hat{L}^p$ and rescaling the principal eigenfunction by Eq. (A32). Receptive fields $\tilde{s}$ are rotated by the complex angle $\phi_0 = \arctan\left(-\int \Im\tilde{s}/\int \Re\tilde{s}\right)$ so as to make the imaginary part odd under inversion of the cortical modulation axis (the real part becomes symmetric as a consequence; see Sec. B 2). The parameters used here are $\zeta = 5\rho$ and $\eta = 3\rho$ corresponding to a ground state at $\omega \approx 0.48/\rho$. A side of the grid has length equal to 5ρ; color scale ranges between min and values.

The self-consistency of our initial assumption is straightforward to check. If we repeat the above by relaxing the assumption $\int A_y^*(\boldsymbol{r})\psi(\boldsymbol{r})d\boldsymbol{R} = 0$, we have to diagonalize a 5×5 matrix instead of a 4×4 one. However, this matrix is diagonal in its A_y sector. The resulting extra eigenvalue is a strictly decreasing function of the wave number; hence, it must be computed at $\omega = 0$, where we find $\Lambda_y = 2\pi\rho^4/\mu^2$.

Let us compare this eigenvalue with the eigenvalue of the cortically modulated solution Λ_M which we find above. In the regions of the phase diagram where $\omega_M > 0$, we have $\Lambda_y < \Lambda_M$; hence, the normally oriented solution is suppressed at long times. In the regions where the optimal wave number is $\omega_M = 0$, on the other hand, it can be seen that $\Lambda_y = \Lambda_M$.

We thus find that, when the RF varies across the cortex, it tends to vary from negative to positive along the direction of cortical modulation, so that the orientation is orthogonal to that direction. When it is uniform across the cortex, its direction becomes immaterial; hence, we have degeneracy in the orientation of the RF.

3. The uncorrelated regime

a. The operator with uncorrelated input ($\zeta \sim 0$)

In order to infer the other main feature of the phase diagram, i.e., the existence of a triple point, we must focus on the uncorrelated limit $\zeta \ll \min(\rho, \eta)$. In this limit, it follows from Eq. (A41) that the matrix element of $\hat{L}^p$ in real space takes the form

$$L^p(\boldsymbol{x},\boldsymbol{r};\boldsymbol{y},\boldsymbol{s}) = \delta(\boldsymbol{x}-\boldsymbol{y}+\boldsymbol{r}-\boldsymbol{s})L^c(\boldsymbol{r},\boldsymbol{s}), \quad \text{(C60)}$$

where the operator $\hat{L}^c$ has the kernel

$$L^c(\boldsymbol{r},\boldsymbol{s}) = \int d\boldsymbol{r}_1 d\boldsymbol{s}_1 [\delta(\boldsymbol{r}-\boldsymbol{r}_1) - \sqrt{A(\boldsymbol{r})A(\boldsymbol{r}_1)}] L^0(\boldsymbol{r}_1,\boldsymbol{s}_1) \times [\delta(\boldsymbol{s}_1-\boldsymbol{s}) - \sqrt{A(\boldsymbol{s}_1)A(\boldsymbol{s})}] \quad \text{(C61)}$$

with

$$L^0(\boldsymbol{r},\boldsymbol{s};\eta) = \sqrt{A(r)A(s)} I(\boldsymbol{r},\boldsymbol{s}), \quad \text{(C62)}$$

where to write the last equality we apply the delta function of Eq. (C60) to infer $I(\boldsymbol{x},\boldsymbol{y}) = I(\boldsymbol{x}-\boldsymbol{y}) = I(\boldsymbol{r}-\boldsymbol{s})$.

Implementing the delta functions in Eq. (C61), we have

$$\frac{L^c(\boldsymbol{r},\boldsymbol{s})}{\sqrt{A(r)A(s)}} = I(\boldsymbol{r},\boldsymbol{s}) + \int d\boldsymbol{r}_1 d\boldsymbol{s}_1 A(r_1)A(s_1) I(\boldsymbol{r}_1,\boldsymbol{s}_1) - \int d\boldsymbol{u} A(u) I(\boldsymbol{r},\boldsymbol{u}) - \int d\boldsymbol{u} A(u) I(\boldsymbol{u},\boldsymbol{s}). \quad \text{(C63)}$$

We can also rewrite Eq. (C61) compactly as

$$\hat{L}^c = \hat{L}^0 + |a_0\rangle\langle a_0|\hat{L}^0|a_0\rangle\langle a_0| - 2\mathrm{HP}[\hat{L}^0|a_0\rangle\langle a_0|], \quad \text{(C64)}$$

where HP is the Hermitian part of an operator.

If we look for eigenfunctions of $\hat{L}^p$ in the form

$$\Psi(\boldsymbol{x},\boldsymbol{r}) \equiv \langle \boldsymbol{x},\boldsymbol{r}|\Psi\rangle = \psi(\boldsymbol{r}) e^{-i\omega(x+r)}, \quad \text{(C65)}$$

the characteristic equation $\Lambda\Psi = \hat{L}^p\Psi$ reduces to $\Lambda\psi(\boldsymbol{r}) = \int \boldsymbol{ds} L^c(\boldsymbol{r},\boldsymbol{s})\psi(\boldsymbol{s})$, which means that $\psi(\boldsymbol{r})$ is the corresponding eigenfunction of $\hat{L}^c$, and the eigenvalue is independent of the cortical wave number. Hence, in the limit of uncorrelated inputs there is complete degeneracy in the wave number.

Because of this degeneracy, the principal eigenfunction could be calculated by focusing solely on the zero-wave-number sector. The results of Sec. C 1 apply and can be used to compute $\psi(\boldsymbol{r})$ which is then replaced for $\psi(\boldsymbol{r})$ in Eq. (C65), yielding the principal eigenfunction for all wave numbers at $\zeta = 0$. Thus, we find that if the input is completely uncorrelated ($\zeta = 0$), the principal eigenfunction of the operator is orientation selective if $\eta > \Theta_c\rho$ and nonselective if $\eta < \Theta_c\rho$, with the critical ratio Θ_c bounded from below by the value θ_c of Eq. (C21).

For $\zeta = 0$ and $\eta < \theta_c\rho$, in the variational approximation of Eq. (C8), the principal eigenfunction is

$$\Psi_0(\boldsymbol{x},\boldsymbol{r}) \propto (R^2 - r^2)\exp\left(-\frac{r^2}{2\sigma^2} + i\omega(x + r_x)\right), \quad \text{(C66)}$$

where σ is the value of γ as given by Eq. (B10) evaluated at $\zeta = 0$.

The nodal radius $R \propto \sqrt{\eta\rho}$ is as calculated in Sec. C 1 and the eigenvalue is $\lambda \sim (2\pi\eta^3/\rho)[\rho - (5-\sqrt{10})\eta]$. While the eigenfunction depends parametrically on the wave number ω, the eigenvalue is entirely degenerate in it, as follows from the divergence of the cutoff wave number Ω in Eq. (A48).

For $\zeta = 0$ and $\eta > \Theta_c\rho$, the exact principal eigenfunction as per Eqs. (B16) and (B17) is given by

$$\Psi(\boldsymbol{x},\boldsymbol{r}) = (k_+ e^{+i\phi} + k_- e^{-i\phi})\exp\left[-i\boldsymbol{\omega}(\boldsymbol{x}+\boldsymbol{r}) - \frac{r^2}{2\sigma^2}\right] \quad \text{(C67)}$$

for any vector $\boldsymbol{\omega}$ and arbitrary coefficients k_+ and k_-. Again we refer to the longitudinal and orthogonal combinations

$$\begin{pmatrix}\Psi^x(\boldsymbol{x},\boldsymbol{r})\\ \Psi^y(\boldsymbol{x},\boldsymbol{r})\end{pmatrix} = \begin{pmatrix} r_x \\ r_y\end{pmatrix}\exp\left(-\frac{r_x^2+r_y^2}{2\sigma^2} + i\boldsymbol{\omega}(\boldsymbol{x}+\boldsymbol{r})\right), \quad \text{(C68)}$$

which are the instances of the $2p$ waves $\Phi_{0,1}$ and $\Phi_{1,0}$ corresponding to uncorrelated input.

These two eigenfunctions share, as per Eq. (B14), the exact eigenvalue

$$\Lambda^{x,y} = 2\pi\eta^2\left(1 + \frac{\eta^2}{2\rho^2} + \frac{\eta}{\rho}\sqrt{1+\frac{\eta^2}{4\rho^2}}\right)^{-2}, \quad \text{(C69)}$$

which is independent of the wave number. Thus, the $x-y$ degeneracy we have for $\eta > \Theta_c\rho$ adds up to the overall degeneracy in the cortical wave number that exists for any value of η.

b. Perturbative input correlations

We show that the point $P_0 = (\zeta_0 = 0, \eta_0 = \Theta_c\rho)$ where the R and N phases meet is a point of nonanalyticity for the principal eigenvalue regarded as a function of the parameters and thus belongs to a phase boundary. Moreover, this phase boundary cannot stop there, because it is a boundary between two phases that have different symmetries—one that displays orientation selectivity and one that does not. How is this phase boundary continued for $\zeta > 0$? Will it curve up or down in the (ζ,η) space?

Since we possess the exact solution for $\zeta = 0, \eta > \Theta_c\rho$, perturbation theory is an ideal tool to address this question. We build a perturbation theory in the small parameter ζ/η. As we see in Sec. C 3 a, our starting point for perturbation theory is a highly degenerate set of eigenfunctions, mainly due to the degeneracy in the wave number. But since the full operator for $\zeta > 0$ commutes with cortical translations, different translational eigenstates are not coupled by the perturbation, and nondegenerate perturbation theory with respect to wave numbers may be applied.

The theory will prove the following three facts:

(i) The ω degeneracy is removed by an infinitesimal $\zeta > 0$ for any η, and this happens in such a way that $\omega = 0$ is always the principal eigenstate.
(ii) The phase boundary starting at the point $P_0 = (0, \Theta_c\rho)$ has a flat slope at that point in the ζ/η plane.
(iii) The $x - y$ degeneracy of the p-wave eigenfunctions survives at finite ζ.

c. Optimal wave number for oriented eigenfunctions: Transverse orientations

We mention that the perturbation does not couple degenerate wave numbers. The same is true with the additional degeneracy in the orientation of selectivity, and it is possible to study the two $2p$ eigenfunctions separately because the full operator $\hat{L}^p$ does not couple them for any value of the parameters. Indeed, we have it by symmetry that $\langle \Psi^x | \hat{L}^p | \Psi^y \rangle = 0$ for any ζ and ω. This means that we can study the effect of a small but finite ζ separately on the two eigenfunctions (applying nondegenerate perturbation theory).

We begin with the y-oriented wave (transverse orientation). While Ψ^y is an eigenfunction of $\hat{L}$ only for $\zeta = 0$, it can be checked that its generalization

$$\chi^y_{0,1}(\boldsymbol{r}) = \frac{r_y}{\sqrt{\pi}\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) e^{-i(\eta^2/\mu^2)\boldsymbol{\omega} \boldsymbol{r}} \tag{C70}$$

is an exact eigenfunction of the full operator $\hat{L}^p$ over the whole phase diagram. Indeed, it is an exact eigenfunction of $\hat{L}$ and, being orthogonal to the constraint ket $|a_\omega\rangle$, it belongs to the null space of the constraint operators $\hat{S}$ and $\hat{T}$:

$$\langle \chi^y_{0,1} | \hat{S} | \chi^y_{0,1} \rangle = \langle \chi^y_{0,1} | \hat{T} | \chi^y_{0,1} \rangle = 0. \tag{C71}$$

The corresponding eigenvalue of $\hat{L}^p$ is given, for every point in (ζ, η) space, by

$$\Lambda^y = \frac{2\pi\mu^2}{\beta^2} \exp\left(-\frac{\omega^2}{2\Omega^2}\right) \tag{C72}$$

with β and Ω defined according to Eqs. (B10) and (A48), respectively.

For $\zeta = 0$, as we know, this eigenvalue is independent of the wave number. However, for any $\zeta > 0$, Eq. (C72) describes an eigenvalue that decreases monotonically with the wave number; hence, the degeneracy is removed. We can conclude that, in the limit of small ζ, the principal y-oriented eigenfunction is uniform over the cortex; i.e., translation symmetry is not broken.

d. Optimal wave number for oriented eigenfunctions: Longitudinal orientations

We now turn to considering the x-oriented function Ψ^x. It can be checked that Ψ^x is orthogonal to the constraint state $|a_\omega\rangle$, which entails

$$\langle \Psi^x | \hat{S} | \Psi^x \rangle = \langle \Psi^x | \hat{T} | \Psi^x \rangle = 0. \tag{C73}$$

This holds true for any value of ζ. However, Ψ^x is an eigenstate only for $\zeta = 0$ and, different from the case of Ψ^y seen above, it is not straightforward to build a generalization of Ψ^x that will be an eigenstate of $\hat{L}^p$ at any point in parameter space. Therefore, here we restrict our attention to sufficiently small nonzero values of ζ and build a perturbation theory in the parameter $\epsilon = \zeta^2/\eta^2$. Hence, we write the operator $\hat{L}^p$ as $\hat{L}^p = \hat{L}^p(\epsilon = 0) + \hat{\Delta} + O(\epsilon^2)$, where $\hat{\Delta}$ includes the first order in ϵ, and we treat $\hat{\Delta}$ as a perturbation. In the shift operator $\hat{\Delta} = \Delta\hat{L} + \Delta\hat{S} + \Delta\hat{T} + \Delta\hat{T}^\dagger$, because of Eq. (C73), we have to compute only the $\hat{L}$ term.

We begin by expanding to the first order in ϵ Eq. (A50), which yields

$$\Delta L(\boldsymbol{r}, \boldsymbol{s}; \omega) = \epsilon \left[-\frac{\omega^2\eta^2}{2} + i\omega(r_x - s_x) + \frac{(\boldsymbol{r} - \boldsymbol{s})^2}{2\eta^2} \right] \times e^{-i\omega(r_x - s_x) - \frac{r^2 + s^2}{4\rho^2} - \frac{(\boldsymbol{r}-\boldsymbol{s})^2}{2\eta^2}}. \tag{C74}$$

We keep only the terms that have a nonvanishing expectation value in Ψ^x; in particular, we neglect terms that change sign if we swap the two variables r_x and s_x, because the integral would be zero. In addition, we may ignore terms whose expectation value in Ψ^x (i.e., whose contribution to $\langle \Psi^x | \Delta\hat{L} | \Psi^x \rangle$) will bear no dependence on the wave number. After some algebra, this leaves a single first-order term in Eq. (C74) that obeys all these requirements, namely,

$$\Delta L(\boldsymbol{r}, \boldsymbol{s}; \omega) \sim -\frac{\omega^2\zeta^2}{2} e^{-i\omega(r_x - s_x) - \frac{r^2 + s^2}{4\rho^2} - \frac{(\boldsymbol{r}-\boldsymbol{s})^2}{2\eta^2}}. \tag{C75}$$

The corresponding expectation value is

$$\Delta = \langle \Psi^x | \Delta\hat{L} | \Psi^x \rangle = -\frac{\omega^2\zeta^2}{2} \Lambda_{2p}(\zeta = 0), \tag{C76}$$

a negative shift in the eigenvalue that is minimized by setting $\omega = 0$.

We thus prove that, for sufficiently small ζ and given η, the principal eigenstate is always cortically uniform ($\omega = 0$), as long as the principal eigenstate for the given η and $\zeta = 0$ is an R phase. This entails that in the limit $\zeta \to 0$, the principal eigenstate of the system has a zero wave number for any $\eta > \Theta_c \rho$. Hence, the slope of the phase boundary at $(0, \Theta_c \rho)$ cannot be positive.

While the degeneracy in the wave number is removed by first-order perturbation theory, the degeneracy between the x and y orientations is not removed, as seen by comparing Eqs. (C72) and (C76) to the second order in ω and using $\Omega \sim 1/\zeta$.

e. Optimal wave number for nonoriented eigenfunctions

The s wave (that is, nonoriented) eigenfunctions of $\hat{L}^p$ are also degenerate in the cortical wave number for $\zeta = 0$. To see which wave number effectively prevails, we must build a perturbation theory in $\epsilon = \zeta^2/\eta^2$ starting from the (unknown) principal eigenstate of the zero-wave-number operator, which we call $|s\rangle$ because of it being an s mode.

From the discussion of Sec. C 3 [see Eq. (C65)], we know that the principal eigenstate at $\zeta = 0$, when an s wave, must have the form $\langle \omega, \boldsymbol{r}|s\rangle = \psi_s(r) e^{-i\omega r_x}$. We can thus use Eq. (B1) to write the level shift as

$$\Delta\hat{L}^p(\omega) = \langle s|(1 - |a_\omega\rangle\langle a_\omega|)\Delta\hat{L}_\omega(1 - |a_\omega\rangle\langle a_\omega|)|s\rangle, \quad \text{(C77)}$$

where the matrix elements of $\Delta\hat{L}_\omega$ have the form given in Eq. (C74).

We notice now that factors of the type $e^{-i\omega r_x}$ will cancel in the integrands of all scalar products that appear in Eq. (C77). As a consequence, the third term in the square brackets of Eq. (C74) may be ignored, as it adds no dependence on the cortical wave number.

The first term in the square brackets of Eq. (C74), on the other hand, yields the level shift

$$\begin{aligned}\Delta\hat{L}_e^{(1)} &= -\frac{\omega^2\zeta^2}{2}\langle s|\hat{L}^p(\zeta = 0)|s\rangle \\ &= -\frac{\omega^2\zeta^2}{2}\int \psi_s(\boldsymbol{r}) L^c(\boldsymbol{r}, \boldsymbol{s})\psi_s(\boldsymbol{s})\boldsymbol{dr} \\ &= -\frac{\omega^2\zeta^2\Lambda_s}{2}, \end{aligned} \quad \text{(C78)}$$

where to write the last equality, we use the fact that ψ_s is, by definition, an eigenfunction of $\hat{L}^c$ with a positive eigenvalue Λ_s. The resulting shift is a monotonically decreasing function of the wave number.

The only remaining term is the second one in the square brackets of Eq. (C74), namely,

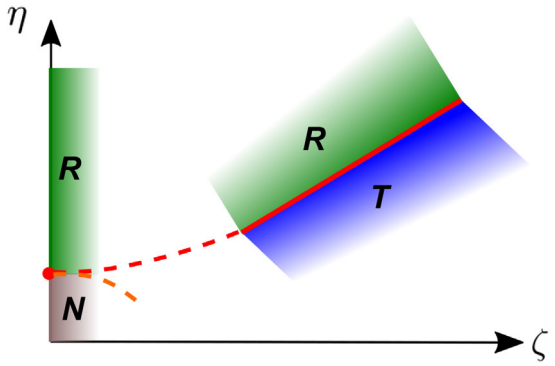

FIG. 11. Patching together of phase boundaries. For uncorrelated inputs (Sec. C 3), the η axis contains a transition point where rotation symmetry is broken (red dot). The perturbation theory for short-range input correlations (Sec. C 3 b) shows that this transition point is continued by a flat phase boundary. The asymptotic rank reduction used for the long-range limit (Sec. C 2) reveals an R-T boundary far away from the origin. We also know from Sec. C 1 c that the N phase is forbidden outside a quarter circle containing the red dot on its contour, and the R phase is forbidden inside it. This leads to predicting an N-T boundary (orange dashed line) contained within the quarter circle, and an R-T boundary stretching from the red dot into the long-range regime. The red dot is a triple point of the system.

$$\Delta L^{(2)}(\boldsymbol{r}, \boldsymbol{s}; \omega) = i\epsilon\omega(r_x - s_x) e^{-i\omega(r_x - s_x) - \frac{r^2 + s^2}{4\rho^2} - \frac{(r-s)^2}{2\eta^2}}, \quad \text{(C79)}$$

which also yields a level shift of the form

$$\begin{aligned}\Delta\hat{L}_e^{(2)} &= \langle s|\Delta\hat{L}^{(2)}|s\rangle + |\langle s|a_\omega\rangle|^2\langle a_\omega|\Delta\hat{L}_\omega|a_\omega\rangle \\ &\quad - 2\Re[\langle s|a_\omega\rangle\langle a_\omega|\Delta\hat{L}^{(2)}|s\rangle]. \end{aligned} \quad \text{(C80)}$$

The ensuing integrals are quickly estimated by symmetry considerations. The factors $e^{-i\omega(r_x - s_x)}$ in Eq. (C79) cancel everywhere in Eq. (C80). While only the real part of these imaginary exponentials would contribute, the resulting matrix element is effectively antisymmetric in the swapping of the $\boldsymbol{r}$ and $\boldsymbol{s}$ coordinates. This makes the first two terms in Eq. (C80) vanish by symmetry; since the final square bracket is purely imaginary, the third term is also zero. Hence, the full eigenvalue shift is given by Eq. (C78) and decreases monotonically as a function of the wave number. We conclude that the wave-number degeneracy at $\zeta = 0$ is removed even by an infinitesimal range of presynaptic correlations, and the uniform cortical mode is favored for any η.

Building upon this, we conclude that, to the lowest order in ζ/η, the principal eigenvalues of the s and p modes are unchanged from those at $\zeta = 0$, and therefore, the phase boundary starting from $P_0 = (0, \Theta_c\rho)$ has a flat slope at that point (Fig. 11).

4. The triple point

We point out the existence of a point P on the $\zeta = 0$ axis where the N phase transitions into the R phase, and show that this phase boundary continues parallel to the ζ axis for perturbatively small values of ζ. This finding must be matched with what is shown about the long-range limit, the existence of a linear phase boundary between the R and T phases. These boundary lines cannot terminate, but can continue only into each other, the reason being that beyond the termination point of a phase boundary, two different symmetries would have to merge. The simplest diagram adhering to this requirement is one where the R region extending above the $T - R$ boundary for long ranges connects, at short ranges, to the R region that extends above the $N - R$ boundary. The missing stretch of phase boundary is sketched as a dashed red curve in Fig. 11.

The immediate consequence of this scenario is that an extra boundary located at values of η below the lower limit of the R phase, must separate the short-range N region from the T region that begins for longer ranges (orange dashed curve of Fig. 11). In principle, one could expect this second boundary to start from any point along the lower contour of the R phase, i.e., at an arbitrary value of ζ. However, it must be remembered the discontinuity between the R and N eigenfunctions is only necessary for $\zeta = 0$, where projective fields from the LGN are decoupled from each other and there is degeneracy in the wave number. For $\zeta > 0$, a suitable path through the T phase can always bridge the N and R eigenfunctions continuously. Variational reasoning is sufficient to infer that, where such a transition exists, it is advantageous over a sharp transition.

It follows that we expect the T region to taper all the way to point P_0, which consequently is a triple point of the system.

[1] S. Kuffler, *Discharge Patterns and Functional Organization of Mammalian Retina*, J. Neurophysiol. **16**, 37 (1953).

[2] H. B. Barlow, *Summation and Inhibition in the Frog's Retina*, J. Physiol. **119**, 69 (1953).

[3] D. Hubel and T. Wiesel, *Integrative Action in the Cat's Lateral Geniculate Body*, J. Physiol. **155**, 385 (1961).

[4] D. Hubel and T. Wiesel, *Receptive Fields of Single Neurones in the Cat's Striate Cortex*, J. Physiol. **148**, 574 (1959).

[5] D. H. Hubel and T. N. Wiesel, *Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex*, J. Physiol. **160**, 106 (1962).

[6] D. Ferster and K. D. Miller, *Neural Mechanisms of Orientation Selectivity in the Visual Cortex*, Annu. Rev. Neurosci. **23**, 441 (2000).

[7] A. D. Lien and M. Scanziani, *Tuned Thalamic Excitation Is Amplified by Visual Cortical Circuits*, Nat. Neurosci. **16**, 1315 (2013).

[8] T. W. Troyer, A. E. Krukowski, N. J. Priebe, and K. D. Miller, *Contrast-Invariant Orientation Tuning in Cat Visual Cortex: Feedforward Tuning and Correlation-Based Intracortical Connectivity*, J. Neurosci. **18**, 5908 (1998).

[9] R. C. Reid and J. M. Alonso, *Specificity of Monosynaptic Connections from Thalamus to Visual Cortex*, Nature (London) **378**, 281 (1995).

[10] Y. Li, D. Fitzpatrick, and L. E. White, *The Development of Direction Selectivity in Ferret Visual Cortex Requires Early Visual Experience*, Nat. Neurosci. **9**, 676 (2006).

[11] X. Ge, K. Zhang, A. Gribizis, A. S. Hamodi, A. M. Sabino, and M. C. Crair, *Retinal Waves Prime Visual Motion Detection by Simulating Future Optic Flow*, Science **373**, eabd0830 (2021).

[12] D. Hubel and T. Wiesel, *Ordered Arrangement of Orientation Columns in Monkeys Lacking Visual Experience*, J. Comp. Neurol. **158**, 307 (1974).

[13] M. Kaschube, M. Schnabel, S. Löwel, D. M. Coppola, L. E. White, and F. Wolf, *Universality in the Evolution of Orientation Columns in the Visual Cortex*, Science **330**, 1113 (2010).

[14] K. E. Schmidt and F. Wolf, *Punctuated Evolution of Visual Cortical Circuits? Evidence from the Large Rodent Dasyprocta leporina, and the Tiny Primate Microcebus murinus*, Curr. Opin. Neurobiol. **71**, 110 (2021).

[15] D. H. Hubel and T. N. Wiesel, *Receptive Fields of Cells in Striate Cortex of Very Young, Visually Inexperienced Kittens*, J. Neurophysiol. **26**, 994 (1963).

[16] M. C. Crair, D. C. Gillespie, and M. P. Stryker, *The Role of Visual Experience in the Development of Columns in Cat Visual Cortex*, Science **279**, 566 (1998).

[17] L. E. White, D. M. Coppola, and D. Fitzpatrick, *The Contribution of Sensory Experience to the Maturation of Orientation Selectivity in Ferret Visual Cortex*, Nature (London) **411**, 1049 (2001).

[18] B. Chapman and M. P. Stryker, *Development of Orientation Selectivity in Ferret Visual Cortex and Effects of Deprivation*, J. Neurosci. **13**, 5251 (1993).

[19] B. Chapman and I. Gödecke, *Cortical Cell Orientation Selectivity Fails to Develop in the Absence of On-Center Retinal Ganglion Cell Activity*, J. Neurosci. **20**, 1922 (2000).

[20] A. D. Huberman, M. B. Feller, and B. Chapman, *Mechanisms Underlying Development of Visual Maps and Receptive Fields*, Annu. Rev. Neurosci. **31**, 479 (2008).

[21] K. D. Miller, E. Erwin, and A. Kayser, *Is the Development of Orientation Selectivity Instructed by Activity?*, J. Neurobiol. **41**, 44 (1999).

[22] R. Fok, A. An, and X. Wang, *Spontaneous Symmetry Breaking in Neural Networks*, arXiv:1710.06096.

[23] T. Wu and I. Fischer, *Phase Transitions for the Information Bottleneck in Representation Learning*, arXiv:2001.01878.

[24] A. Gordon, A. Banerjee, M. Koch-Janusz, and Z. Ringel, *Relevance in the Renormalization Group and in Information Theory*, Phys. Rev. Lett. **126**, 240601 (2021).

[25] H. Tanaka and D. Kunin, *Noether's Learning Dynamics: The Role of Kinetic Symmetry Breaking in Deep Learning*, arXiv:2105.02716.

[26] A. G. Kline and S. E. Palmer, *Gaussian Information Bottleneck and the Non-Perturbative Renormalization Group*, arXiv:2107.13700.

[27] D. Hebb, *The Organization of Behavior* (Wiley & Sons, New York, 1949).
[28] J. Konorski, *Conditioned Reflexes and Neuron Organization* (Cambridge University Press, Cambridge, England, 1948).
[29] C. J. Shatz, *The Developing Brain*, Sci. Am. **267**, 60 (1992).
[30] R. Lamprecht and J. LeDoux, *Structural Plasticity and Memory*, Nat. Rev. Neurosci. **5**, 45 (2004).
[31] H. Markram, W. Gerstner, and P. J. Sjöström, *A History of Spike-Timing-Dependent Plasticity*, Front. Synaptic Neurosci. **3**, 4 (2011).
[32] A. Bain, *The Senses and the Intellect* (JW Parker & Son, London, 1855).
[33] T. Ribot, *La Psychologie Anglaise Contemporaine* (Librairie philosophique de Ladrange, Paris, 1870).
[34] W. James, *The Principles of Psychology* (Henry Holt and Company, New York, 1890), Vol. 2.
[35] C. von der Malsburg, *Self-Organization of Orientation Sensitive Cells in the Striate Cortex*, Kybernetik **14**, 85 (1973).
[36] R. Linsker, *From Basic Network Principles to Neural Architecture (Series)*, Proc. Natl. Acad. Sci. U.S.A. **83**, 7508 (1986).
[37] K. D. Miller, J. B. Keller, and M. P. Stryker, *Models for the Formation of Ocular Dominance Columns Solved by Linear Stability Analysis*, Soc. Neurosci. Abstr. **12**, 1373 (1986), https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Models+for+the+2059+Formation+of+Ocular+Dominance+Columns+Solved+by+Linear+2060+Stability+Analysis&btnG=.
[38] K. D. Miller, J. B. Keller, and M. P. Stryker, *Ocular Dominance Column Development: Analysis and Simulation*, Science **245**, 605 (1989).
[39] H. Shouval, N. Intrator, C. C. Law, and L. N. Cooper, *Effect of Binocular Cortical Misalignment on Ocular Dominance and Orientation Selectivity*, Neural Comput. **8**, 1021 (1996).
[40] B. S. Blais, N. Intrator, H. Z. Shouval, and L. N. Cooper, *Receptive Field Formation in Natural Scene Environments. Comparison of Single-Cell Learning Rules*, Neural Comput. **10**, 1797 (1998).
[41] M. S. Falconbridge, R. L. Stamps, and D. R. Badcock, *A Simple Hebbian/Anti-Hebbian Network Learns the Sparse, Independent Components of Matural Images*, Neural Comput. **18**, 415 (2006).
[42] J. Zylberberg, J. T. Murphy, and M. R. DeWeese, *A Sparse Coding Model with Synaptically Local Plasticity and Spiking Neurons Can Account for the Diverse Shapes of V1 Simple Cell Receptive Fields*, PLoS Comput. Biol. **7**, e1002250 (2011).
[43] P. D. King, J. Zylberberg, and M. R. DeWeese, *Inhibitory Interneurons Decorrelate Excitatory Cells to Drive Sparse Code Formation in a Spiking Model of V1*, J. Neurosci. **33**, 5475 (2013).
[44] T. Miconi, J. L. McKinstry, and G. M. Edelman, *Spontaneous Emergence of Fast Attractor Dynamics in a Model of Developing Primary Visual Cortex*, Nat. Commun. **7**, 13208 (2016).
[45] C. S. Brito and W. Gerstner, *Nonlinear Hebbian Learning as a Unifying Principle in Receptive Field Formation*, PLoS Comput. Biol. **12**, e1005070 (2016).
[46] C. von der Malsburg, *Self-Organization of Orientation Selective Cells in the Striate Cortex*, Kybernetik **14**, 85 (1973).
[47] P. Földiák, *Forming Sparse Representations by Local Anti-Hebbian Learning*, Biol. Cybern. **64**, 165 (1990).
[48] J. J. Hunt, M. Ibbotson, and G. J. Goodhill, *Sparse Coding on the Spot: Spontaneous Retinal Waves Suffice for Orientation Selectivity*, Neural Comput. **24**, 2422 (2012).
[49] C. Savin, P. Joshi, and J. Triesch, *Independent Component Analysis in Spiking Neurons*, PLoS Comput. Biol. **6**, e1000757 (2010).
[50] J. Antolík and J. A. Bednar, *Development of Maps of Smple and Complex Cells in the Primary Visual Cortex*, Front. Comput. Neurosci. **5**, 17 (2011).
[51] J. L. Stevens, J. S. Law, J. Antolík, and J. A. Bednar, *Mechanisms for Stable, Robust, and Adaptive Development of Orientation Maps in the Primary Visual Cortex*, J. Neurosci. **33**, 15747 (2013).
[52] R. O. Wong, *Retinal Waves and Visual System Development*, Annu. Rev. Neurosci. **22**, 29 (1999).
[53] J. Demas, S. J. Eglen, and R. O. Wong, *Developmental Loss of Synchronous Spontaneous Activity in the Mouse Retina Is Independent of Visual Experience*, J. Neurosci. **23**, 2851 (2003).
[54] R. O. Wong, M. Meister, and C. J. Shatz, *Transient Period of Correlated Bursting Activity During Development of the Mammalian Retina*, Neuron **11**, 923 (1993).
[55] B. Chapman, M. P. Stryker, and T. Bonhoeffer, *Development of Orientation Preference Maps in Ferret Primary Visual Cortex*, J. Neurosci. **16**, 6443 (1996).
[56] G. B. Smith, B. Hein, D. E. Whitney, D. Fitzpatrick, and M. Kaschube, *Distributed Network Interactions and Their Emergence in Developing Neocortex*, Nat. Neurosci. **21**, 1600 (2018).
[57] A. Grabska-Barwinska and C. von der Malsburg, *Establishment of a Scaffold for Orientation Maps in Primary Visual Cortex of Higher Mammals*, J. Neurosci. **28**, 249 (2008).
[58] B. A. Olshausen and D. J. Field, *Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images*, Nature (London) **381**, 607 (1996).
[59] A. J. Bell and T. J. Sejnowski, *The "Independent Components" of Natural Scenes Are Edge Filters*, Vision Res. **37**, 3327 (1997).
[60] J. H. van Hateren and D. L. Ruderman, *Independent Component Analysis of Natural Image Sequences Yields Spatio-Temporal Filters Similar to Simple Cells in Primary Visual Cortex*, Proc. Biol. Sci. **265**, 2315 (1998).
[61] M. V. Albert, A. Schnabel, and D. J. Field, *Innate Visual Learning through Spontaneous Activity Patterns*, PLoS Comput. Biol. **4**, e1000137 (2008).
[62] S. Behpour, D. J. Field, and M. V. Albert, *On the Role of LGN/V1 Spontaneous Activity as an Innate Learning Pattern for Visual Development*, Frontiers of oral physiology **12**, 695431 (2021).

[63] K. D. Miller, *Development of Orientation Columns via Competition Between ON- and OFF-Center Inputs*, NeuroReport **3**, 73 (1992).
[64] K. D. Miller, *A Model for the Development of Simple Cell Receptive Fields and the Ordered Arrangement of Orientation Columns through Activity-Dependent Competition between On-and Off-Center Inputs*, J. Neurosci. **14**, 409 (1994).
[65] D. MacKay and K. D. Miller, *Analysis of Linsker's Simulations of Hebbian Rules*, Neural Comput. **2**, 173 (1990).
[66] D. J. C. MacKay and K. D. Miller, *Analysis of Linsker's Applications of Hebbian Rules to Linear Networks*, Network **1**, 257 (1990).
[67] M. Miyashita and S. Tanaka, *A Mathematical Model for the Self-Organization of Orientation Columns in Visual Cortex*, NeuroReport **3**, 69 (1992).
[68] T. Yamazaki, *Mathematical Analysis of a Correlation-Based Model for Orientation Map Formation*, Neural Netw. **16**, 47 (2003).
[69] S. Wimbauer, W. Gerstner, and J. van Hemmen, *Analysis of a Correlation-Based Model for the Development of Orientation-Selective Receptive Fields in the Visual Cortex*, Network **9**, 449 (1998).
[70] T. Ohshiro and M. Weliky, *Simple Fall-Off Pattern of Correlated Neural Activity in the Developing Lateral Geniculate Nucleus*, Nat. Neurosci. **9**, 1541 (2006).
[71] M. Jacobson, *Developmental Neurobiology* (Springer Science & Business Media, New York, 2013).
[72] D. R. Muir and M. Cook, *Anatomical Constraints on Lateral Competition in Columnar Cortical Architectures*, Neural Comput. **26**, 1624 (2014).
[73] J. Antolík, *Rapid Long-Range Disynaptic Inhibition Explains the Formation of Cortical Orientation Maps*, Front. Neural Circuits **11**, 21 (2017).
[74] P. C. Bressloff, J. D. Cowan, M. Golubitsky, P. J. Thomas, and M. C. Wiener, *Geometric Visual Hallucinations, Euclidean Symmetry and the Functional Architecture of Striate Cortex*, Phil. Trans. R. Soc. B **356**, 299 (2001).
[75] G. Turrigiano, *Too Many Cooks? Intrinsic and Synaptic Homeostatic Mechanisms in Cortical Circuit Refinement*, Annu. Rev. Neurosci. **34**, 89 (2011).
[76] L. Luo and D. O'Leary, *Axon Retraction and Degeneration in Development and Disease*, Annu. Rev. Neurosci. **28**, 127 (2005).
[77] K. D. Miller and D. J. C. MacKay, *The Role of Constraints in Hebbian Learning*, Neural Comput. **6**, 100 (1994).
[78] W. H. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick, *Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex*, J. Neurosci. **17**, 2112 (1997).
[79] K. E. Schmidt, R. Goebel, S. Löwel, and W. Singer, *The Perceptual Grouping Criterion of Colinearity is Reflected by Anisotropies of Connections in the Primary Visual Cortex*, Eur. J. Neurosci. **9**, 1083 (1997).
[80] T. Keck, T. Toyoizumi, L. Chen, B. Doiron, D. E. Feldman, K. Fox, W. Gerstner, P. G. Haydon, M. Hübener, H.-K. Lee *et al.*, *Integrating Hebbian and Homeostatic Plasticity: The Current State of the Field and Future Research Directions*, Phil. Trans. R. Soc. B **372**, 20160158 (2017).
[81] S. B. Paik and D. L. Ringach, *Retinal Origin of Orientation Maps in Visual Cortex*, Nat. Neurosci. **14**, 919 (2011).
[82] J. J. Pattadkal, G. Mato, C. van Vreeswijk, N. J. Priebe, and D. Hansel, *Emergent Orientation Selectivity from Random Networks in Mouse Visual Cortex*, Cell Rep. **24**, 2042 (2018).
[83] M. Schottdorf, S. J. Eglen, F. Wolf, and W. Keil, *Can Retinal Ganglion Cell Dipoles Seed Iso-Orientation Domains in the Visual Cortex?*, PLoS One **9**, e86139 (2014).
[84] M. Schottdorf, W. Keil, D. Coppola, L. E. White, and F. Wolf, *Random Wiring, Ganglion Cell Mosaics, and the Functional Architecture of the Visual Cortex*, PLoS Comput. Biol. **11**, e1004602 (2015).
[85] Y. Ahmadian and K. D. Miller, *What Is the Dynamical Regime of Cerebral Cortex?*, Neuron **109**, 3373 (2021).
[86] A. S. Kayser and K. D. Miller, *Opponent Inhibition: A Developmental Model of Layer 4 of the Neocortical Circuit*, Neuron **33**, 131 (2002).
[87] M. Kaschube, F. Wolf, T. Geisel, and S. Löwel, *Genetic Influence on Quantitative Features of Neocortical Architecture*, J. Neurosci. **22**, 7206 (2002).
[88] C. Zhuang, S. Yan, A. Nayebi, M. Schrimpf, M. C. Frank, J. J. DiCarlo, and D. L. Yamins, *Unsupervised Neural Network Models of the Ventral Visual Stream*, Proc. Natl. Acad. Sci. U.S.A. **118**, e2014196118 (2021).
[89] D. L. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo, *Performance-Optimized Hierarchical Models Predict Neural Responses in Higher Visual Cortex*, Proc. Natl. Acad. Sci. U.S.A. **111**, 8619 (2014).
[90] G. W. Lindsay and K. D. Miller, *How Biological Attention Mechanisms Improve Task Performance in a Large-Scale Visual System Model*, eLife **7**, e38105 (2018).
[91] M. Schrimpf, J. Kubilius, H. Hong, N. J. Majaj, R. Rajalingham, E. B. Issa, K. Kar, P. Bashivan, J. Prescott-Roy, F. Geiger *et al.*, *Brain-Score: Which Artificial Neural Network for Object Recognition Is Most Brain-like?*, bioRxiv.407007.
[92] B. Miclut, *Committees of Deep Feedforward Networks Trained with Few Data*, in *Proceedings of the German Conference on Pattern Recognition* (Springer, New York, 2014), pp. 736–742.
[93] A. Dundar, J. Jin, and E. Culurciello, *Convolutional Clustering for Unsupervised Learning*, arXiv:1511.06241.
[94] B. Knyazev, E. Barth, and T. Martinetz, *Recursive Autoconvolution for Unsupervised Learning of Convolutional Neural Networks*, in *Proceedings of 2017 IJCNN* (IEEE, New York, 2017), pp. 2486–2493.
[95] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, *Barlow Twins: Self-Supervised Learning via Redundancy Reduction*, arXiv:2103.03230.
[96] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, *A Simple Framework for Contrastive Learning of Visual Representations*, in *Proceedings of the International Conference on Machine Learning* (PMLR, 2020), pp. 1597–1607, https://arxiv.org/abs/2002.05709.
[97] J. Dapello, T. Marques, M. Schrimpf, F. Geiger, D. D. Cox, and J. J. DiCarlo, *Simulating a Primary Visual Cortex at the Front of CNNS Improves Robustness to Image*

Perturbations, bioRxiv, http://proceedings.mlr.press/v119/chen20j/chen20j.pdf.

[98] P. Dayan and L. F. Abbott, *Theoretical Neuroscience* (MIT Press, Cambridge, MA, 2001).

[99] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and A. J. Hudspeth, *Principles of Neural Science*, 6th ed. (McGraw-Hill Education, New York, 2012).

[100] D. J. C. MacKay and K. D. Miller, *Analysis of Linsker's Application of Hebbian Rules to Linear Networks*, Network: Computation in Neural Systems **1**, 257 (1990).

[101] S. Wimbauer, W. Gerstner, and J. L. Van Hemmen, *Analysis of a Correlation-Based Model for the Development of Orientation-Selective Receptive Fields in the Visual Cortex*, Network: Computation in Neural Systems **9**, 449 (1998).

[102] C. E. Davey, D. B. Grayden, and A. N. Burkitt, *Emergence of Radial Orientation Selectivity: Effect of Cell Density Changes and Eccentricity in a Layered Network*, arXiv: 1805.03749.