Emergence of robust global modules from local interactions and smooth gradients

Mikail Khona*
Physics, MIT and
Brain and Cognitive Sciences, and McGovern Institute, MIT

Sarthak Chandra*
Brain and Cognitive Sciences, and McGovern Institute, MIT

Ila Fiete
Brain and Cognitive Sciences, and McGovern Institute, MIT
(Dated: June 6, 2023)

Modular structure and function are ubiquitous in biology, from the scale of ecosystems to the organization of animal bodies and brains. However, the mechanisms of modularity emergence over development remain unclear. Here we introduce the principle of *peak selection*, a process in which two local interactions self-organize discontinuous module boundaries from a smooth global gradient, unifying the positional hypothesis and the Turing pattern formation hypothesis for morphogenesis. Applied to the brain's grid cell networks, peak selection results in the spontaneous emergence of functionally distinct modules with discretely spaced spatial periods. Applied to ecological systems, a generalization of the process results in discrete systems-level niches. The dynamics exhibits emergent self-scaling to variations in system size and "topological robustness" [1] that renders module emergence and module properties insensitive to most parameters. Peak selection substantially ameliorates the fine-tuning requirement of continuous attractor dynamics even within single modules. It makes a detail-independent prediction that grid module period ratios should approximate adjacent integer ratios, furnishing the most accurate match to data to date, with additional predictions to connect physiology, connectomics, and transcriptomics data. In sum, our results indicate that local competitive interactions combined with low-information global gradients can lead to robust global module emergence.

INTRODUCTION

Modular structures are ubiquitous in natural systems, from ecological niches to human communities, and from body structures to circuits in the brain. This is probably so because they are robust to localized perturbations [2, 3], faster to adapt if the world requires sparse or modular changes [4], and permit flexible, high-capacity computation through compositionality [5–10]. In this sense, modularity is the crux of biological organization.

The prevalence of modularity raises critical questions about its evolutionary and developmental origins: From an evolutionary perspective, modular solutions to a given problem form a vanishingly small subset of all possible solutions, raising the question of how and why they are selected. From the perspective of development, which we take here, the question is how modular structures form, and whether module features such as size, number, and boundary locations need to be genetically instructed or spontaneously emerge through unfolding physical processes such as symmetry breaking.

One hypothesis for the developmental emergence of structure is the positional information hypothesis espoused by Lewis Wolpert: Gene expression generates spatial morphogen concentration gradients, and different downstream genes become activated by thresholding the morphogen concentration [11, 12]. In line with this hypothesis, body segmentation in Drosophila [13], Fig. 1a-c, is controlled by a family of genes that become activated at different concentrations of maternally deposited Bicoid RNA. This gradient precedes and causally guides spatial bands of gene expression whose boundaries demarcate segment boundaries. A distinct hypothesis by Alan Turing is the idea of spontaneous pattern emergence from local competitive interactions, minimizing the use of detailed genetic information [12, 14]. Supporting this hypothesis is evidence of digit formation via pattern formation in hand morphogenesis [15]. A clear example of Turing pattern formation comes from the grid cell system in the medial entorhinal cortex (MEC) of mammalian brains, Fig. 1d-f. MEC neurons form a triangular grid-like pattern of activity as a function of space when animals navigate. Underlying these spatially periodic responses are periodic activity patterns in the cortex [16, 17]. Continuous attractor neural network (CAN) models of grid cells are based on Turing patterning [18] and

* Equal contribution

make a number of stringent predictions that are supported by evidence from many studies [19–23]. However, it is unclear whether and how these CAN models for single grid cell modules translate to the formation of multiple discrete modules of grid cells with distinct periods [24], Fig. 1i.

The positional and Turing hypotheses individually have weaknesses [12, 25–27]: The positional mechanism is susceptible to noise in copy number [12, 28–30], and requires different downstream genetic cascades for each facet of structure that specify how and where that structure forms. Its implication that modular structure or function are driven by modularity in gene expression runs counter to at least some experimental studies that find continuous gradients in gene expression can underlie modular function [28, 31–44]. The pattern forming mechanism typically produces structure at a single scale given by the width of the local lateral interactions. These models do not possess scale invariance to (self-scaling with) the size of the system undergoing patterning.

Here, we hypothesize that positional and pattern forming mechanisms can be unified in a way such that modularity emerges through self-organization from local interactions without requiring modularity in gene expression, and the resulting process is scale-invariant, Fig. 1g. We show that such a process exists and is robust to most parametric variation and noise.

We focus first on the mammalian grid cell system, which exhibits structure on two scales: locally periodic patterns and global modules. The model produces strikingly accurate predictions about the sequence of successive spatial period ratios in grid cells, improving substantially on existing models. The process exhibits a "topological robustness" property that substantially eases the usual fine-tuning requirements of continuous attractor models of grid cells [45]. It also generates numerous predictions for future physiology, transcriptomics, and connectomics experiments.

Analyzing the underlying dynamical mechanisms of the process allows us to extract a general principle for global module emergence with smooth gradients and local interactions, which we call the *peak selection* principle. We then apply the peak selection principle to a very different problem, showing the emergence of modular multi-species niches in an interacting ecological system.

GENERALIZATION OF CONTINUOUS ATTRACTOR MODELS FOR SINGLE GRID MODULES

Grid cells in the mammalian medial entorhinal cortex (MEC) of mammals exhibit spatially periodic response patterns as animals explore open spaces [16]. Moving ventrally along the long (dorsoventral or DV) axis of MEC, Fig. 1h, the grid cells form modular and functionally independent subnetworks, with discrete jumps in spatial periodicity, Fig. 1i. First, we extend models of grid cells within single modules before considering mechanisms for the formation of multiple discrete modules of different scales, which remain unknown.

The dynamics of grid cells within a module are consistent with predictions of continuous attractor neural network (CAN) models [45]. CAN models are based on pattern formation through a linear (Turing) instability, driven by strong local interactions between neurons. Realized predictions include cell-cell relationships that are preserved across environments [19], and from waking to sleep [20, 21], and the direct demonstration of a toroidal set of stable states in the dynamics of single modules, Fig. 1e-f [22].

Existing CAN models show that grid tuning can arise from either of two specific interaction profiles: a center excitation-surround inhibition (Mexican hat) or purely inhibitory surround shape [18], or a uniform local inhibition shape (which we term a "Lincoln hat") [48]. We derive a set of conditions (SI A) that we hypothesize are sufficient for grid-like patterning, and show numerically that an infinite set of randomly sampled interaction profiles (kernels) under these conditions do so (Fig. 2a). The conditions are simple constraints on the local interaction kernel W: that it is radially symmetric, that its integral is negative (inhibition dominated: $\int W(x,x')dx' < 0$), and that it is strong enough such that that there is some value of $k > 0$ such that its Fourier transform satisfies $\tilde{W}(k) > 1/\tau$. The latter can almost always be made to hold for interactions that satisfy the former by scaling W (SI Sec. A). This result significantly expands the generality of CAN models for single grid cell modules.

FIXED AND GRADED LOCAL INTERACTIONS FOR GLOBAL MODULARITY EMERGENCE

Physiological experiments reveal biophysical gradients along the DV axis[32, 38, 42], Fig. 1h, raising the possibility that combining such a gradient with CAN pattern formation mechanisms might lead to the emergence of multiple discrete modules. We replaced the translation-invariant interaction kernels in CAN grid cell models by kernels with a slowly increasing width $\sigma(n_{DV})$ along the DV axis of the model neural sheet, SI Fig.11a, where n_{DV} refers to DV location (while $\mathbf{x}$ refers to the general 2-dimensional position on the neural strip):

$$W^g(|\mathbf{x}-\mathbf{x}'|) \to W(|\mathbf{x}-\mathbf{x}'|;\sigma(n_{DV})) \equiv W^g_{n_{DV}}(\mathbf{\Delta}).$$

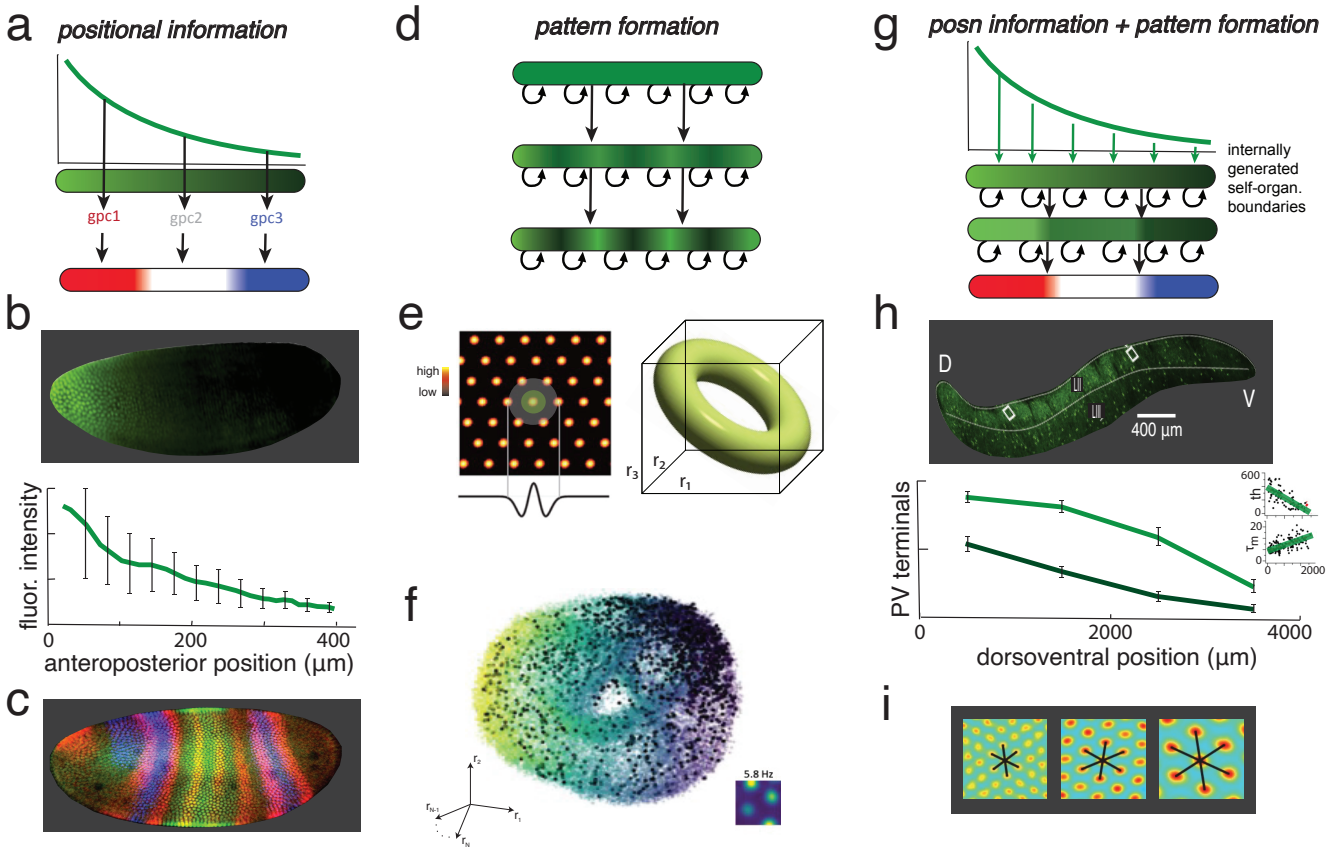


FIG. 1. **Positional versus pattern-forming mechanisms for structure formation and our hypothesis.** (a) The positional hypothesis: global gradients are thresholded by different downstream gene expression cascades to generate structure [11]. (b) Fluorescence image of maternally deposited protein *bcd* RNA (based on maternal bicoid RNA deposition) early in development of the Drosophila embryo [46] sets up a polarity gradient. (c) A downstream gene-protein expression cascade, including gap and pair-rule genes, sets up body segment-defining bands by thresholding the bicoid gradient (immunofluorescence image adapted from [47]; segmentation figure adapted from [13].) (d) Spontaneous self-organized structure emergence (pattern formation) through competitive lateral interactions [14]. (e-f) The continuous attractor neural network (CAN) model for single grid cell modules [18] is based on Turing patterning based on local interactions and its predictions are consistent with the experimental data [19–21, 23]. These include the prediction of a continuous set of stable states with toroidal geometry across waking and sleep, and its recent confirmation [22]. (g) Our hypothesis: Positional and pattern forming mechanisms can interact to lead to structure emergence that exhibits the strengths of both mechanisms. (h-i) The long dorsoventral (DV) axis of medial entorhinal cortex (MEC; image of layers II and III) [32] exhibits smooth-seeming gradients in multiple cellular properties, while along the same axis, grid cells are organized into discrete modules with discontinuous jumps in their spatial periods (adapted from [24]).

In numerical simulations, this slow DV variation of the local interaction width in the CAN model produces hexagonally arranged activity bumps with a growing period (SI Fig. 11b). However, the variation in pattern period is smooth, reflecting the smooth width gradient: there is no emergent modularization (SI Fig. 11c).

Given that global modularity and local patterning together cover two spatial scales, it might be necessary to include two scales in the lateral interactions. However, just as local interactions lead to globally periodic structure, we hypothesize that adding a second but still *local* interaction might be sufficient to induce globally modular structure.

The meaning of a second interaction must be distinct from merely a shape difference relative to the first, because the sum of two local kernels is otherwise just another local kernel, and the graded period result of SI Fig. 11b-c will still hold. Therefore, we explored adding a second type of local lateral interaction, W^f, which remains fixed across the DV axis while the first is graded. The combined interaction is:

$$W_{n_{DV}}(\mathbf{\Delta x}) = W^g_{n_{DV}}(\mathbf{\Delta x}) + W^f(\mathbf{\Delta x}). \tag{1}$$

We assume that the fixed interaction is wider than the graded interaction, but both interactions are local and much smaller than the length of the cortical sheet: $\sigma_{\max} \leq d \ll L$, where d is the width of W^f and L is the DV length

of the neural sheet. Remarkably, the addition of such a fixed-scale interaction leads the network to spontaneously decompose into a few discrete modules, with coherent periodic activity patterns locally and discontinuous jumps in period globally, Fig. 2b-d. As with single module grid patterning, there is broad latitude in the shapes of the interaction kernels, W^g and W^f, so long as one is graded and the other is wider but fixed in width along the DV axis, Fig. 2f-h. The same combination of a graded-width and a fixed-scale interaction produces robust and spontaneous decomposition of dynamics into discrete modules across 1D and 2D network models, Fig. 2e,i. (See SI Sec. D 8 for additional model results for 2D networks.)

The modules that emerge are much larger than both local interaction scales, cf. Fig. 2b, f (where the interaction widths are rendered to scale), versus the formed modules in Fig. 2c-e, g-i. Given this surprising result of an emergent scale that is not directly related to the underlying physical interaction scales, we seek its causes below.

Formed modules are functionally independent

The emergent modules, which are distinct in the periodicity of their patterning, are also functionally independent units. In single-module CAN models, providing cells with a velocity input causes the pattern to flow in a direction and with a speed proportional to the velocity input — the network can thus track current location by integrating a velocity signal [49–52], as grid cells are believed to do [18]. For the distinct modules to independently perform velocity integration means that their patterns flow independently, even though connectivity is as continuous across module boundaries as it is within modules, which seems intuitively implausible. However, when we drive all modules with a common velocity input, we find that the patterns in each module flow independently. The boundaries between modules remain sharp and fixed at the same cortical locations, with a spatially discontinuous phase dislocation between boundaries with module phases separately updating and shifting relative to each other in time, Fig. 2j. This dynamics results in veridical, independent velocity integration within and across modules, so that all cells (even those close to the module boundaries) have periodic spatial tuning curves, Fig. 2k.

ANALYTICAL THEORY OF MODULARIZATON: PEAK SELECTION, TOPOLOGICAL ROBUSTNESS, AND SELF-SCALING

The generality and robustness with which discrete modules emerge from the combination of a fixed-scale and a graded-scale local interaction — with very different kernel shapes as in Fig.2b,f — suggests a general principle at work. Starting from an initial condition of uniform activity, the network exhibits nearly immediate (within 1-2 biophysical time-constants τ) signs of modularization, Fig. 3a. Modularization begins before most neurons have crossed their nonlinear thresholds, and unfolds concurrently with local periodic patterning (Fig. 3a,e). The system also exhibits localized eigenvectors (SI sec G; similar to the Anderson localization phase transition in condensed matter physics [53]). Both phenomena point to a linear instability-driven modularization mechanism, hinting at the possibility of a unified theory for patterning and modularity. We derive such a theory, summarizing it below with details in SI (Sec.B). Besides establishing how, why, and when modularity emerges, the theory accurately predicts the discrete pattern periods of all modules, the number and sizes of modules, and the locations of module boundaries (explored below).

The theory examines how infinitesimal perturbations evolve from an unstable initial fixed point $s_0(n_{DV})$. In spatially local windows that are much larger than the local interaction scale $\sigma(n_{DV}), d$ but much smaller than the full system size L, the dynamics can be decomposed into local Fourier modes (SI sectionD 8) and solved because the spatial variation in the interaction $W_{n_{DV}}$ of Eq. (1) is slow. Under the same simple conditions on the interaction kernels given above (in Generalization of continuous attractor models for single grid cell modules), the network forms a *spatially varying* patterned state with the following inverse periods (wave vectors) as a function of the DV location n_{DV}:

$$\lambda(n_{DV})^{-1} = k^*(n_{DV}) = \arg\max_k \{\tilde{W}^f(k) + \tilde{W}^g_{n_{DV}}(k)\}, \tag{2}$$

where $\tilde{W}$ denotes the Fourier transform of W.

The spatial structure of this patterned solution can be gleaned from the interplay between the two terms. Suppose the interaction kernel W^f is *simple*: primarily characterized by one length-scale $d > \sigma_{\max}$ and such that other length-scales are not prominent and much smaller than d (Fig. 3b, top, orange). Then (except for a set of hypothesized measure zero in the space of functions W^f; SI Sec. D for details), its Fourier transform has an approximate form ($\sim \cos(kd - \phi)$), with closely spaced peaks every $\sim 1/d$ (Fig. 3b, bottom, orange) and a phase ϕ. These local maxima remain the same across the neural strip because W^f is not graded.

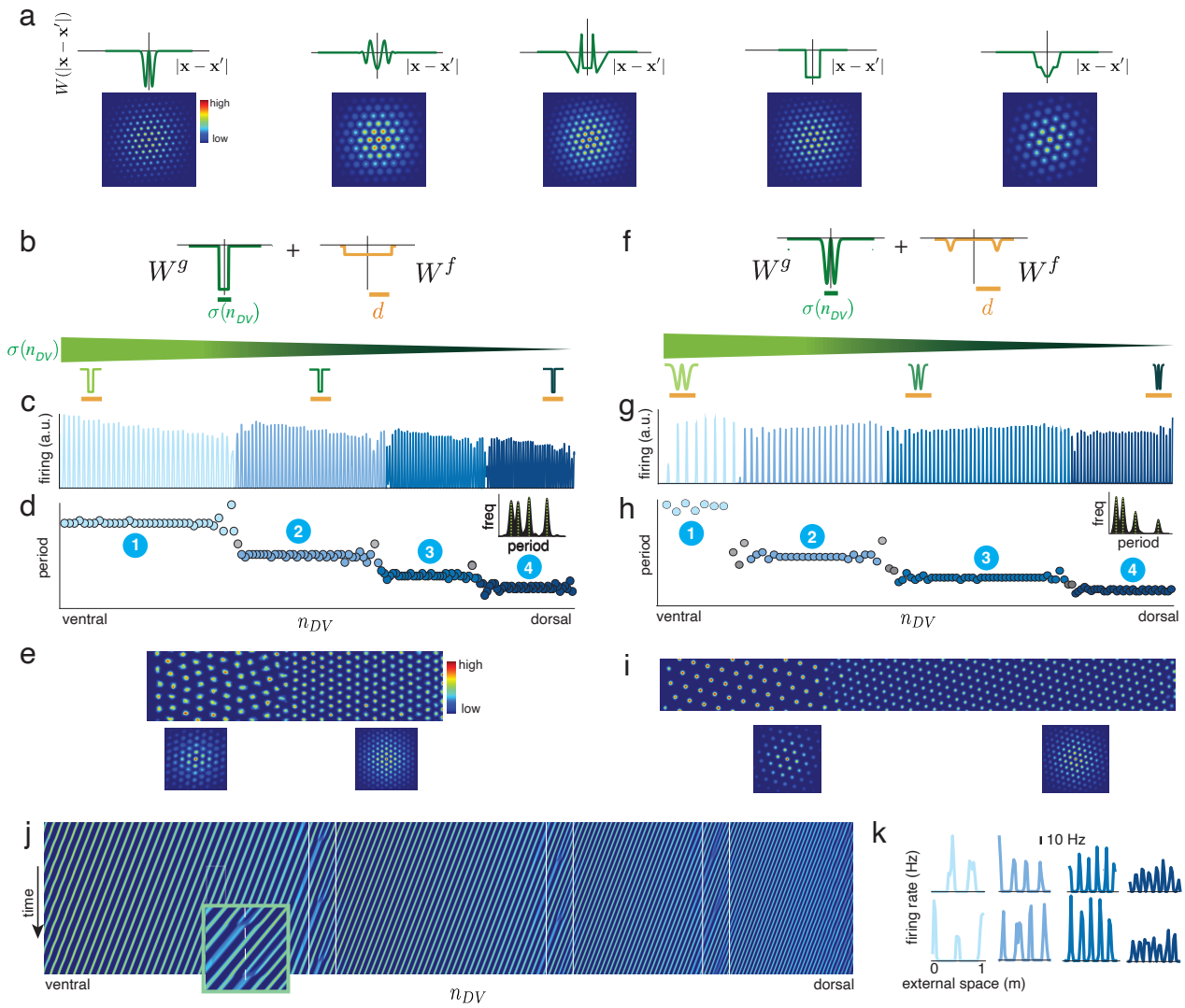


FIG. 2. **Two *local* interactions, with graded and fixed widths, respectively, lead to global module emergence.** (a) Generalization of CAN grid cell models: 5 examples from an infinite set of distinct local interaction kernel shapes that can lead to grid-like patterning. (b) Combining two local interactions, one whose width ($\sigma(n_{DV})$) scales smoothly along the DV axis ($W^g_{n_{DV}}$, green) and a broader but still-local one whose width (d) remains fixed along the neural strip. Interaction widths indicated below the gradient are drawn to scale relative to the activity shown in (c). (c-e) The two interactions from (b) lead to spontaneous emergence of modules with distinct periods in 1-dimensional (c) neural strip, with extracted periods shown in (d). The same kernels applied to a 2-dimensional neural sheet (e), with the 2d autocorrelation function of the local (single-module) patterns in the neural sheet (bottom). (f-i) Same as (c-e), but for a different pair of interaction kernels $W^g_{n_{DV}}, W^f$ with distinct gradient shape $\sigma(n_{DV})$ and endpoints $(\sigma_{\min}, \sigma_{\max})$ from (c-e). (j) The response of the 1-dimensional neural strip shown over time when the network is driven by a smoothly graded velocity input, white lines highlighting the temporal evolution of dynamics at the module boundary (inset: magnification of the first boundary). (k) The independent velocity-driven pattern dynamics in each module result in regular periodic spatial tuning curves (shown are 2 cells per module). *See Methods for parameter and simulation details.*

By contrast, $\tilde{W}^g_{n_{DV}}$ exhibits a broad Fourier peak (of scale $\sim 1/\sigma(n_{DV}) \gg 1/d$). The width and location of this peak smoothly contracts along the long axis (as $\sigma(n_{DV})$ increases; Fig. 3b, bottom, green). This interaction drives spatial patterning and its graded variation is ultimately responsible for changes in period. However, the narrow peaks of $\tilde{W}^f(k)$ determine the set of potential locations of the global maximum, while the smoothly moving peak from $\tilde{W}^g_{n_{DV}}(k)$ performs "peak selection" on these possibilities (Fig. 3c and SI Movie 1). It sweeps through the narrow local maxima of $\tilde{W}^f$, and as it does so one of the narrowly spaced peaks becomes the global maximum. As it sweeps on, the global maximum abruptly and discontinuously jumps to the next peak of $\tilde{W}^f(k)$, defining the abrupt changes

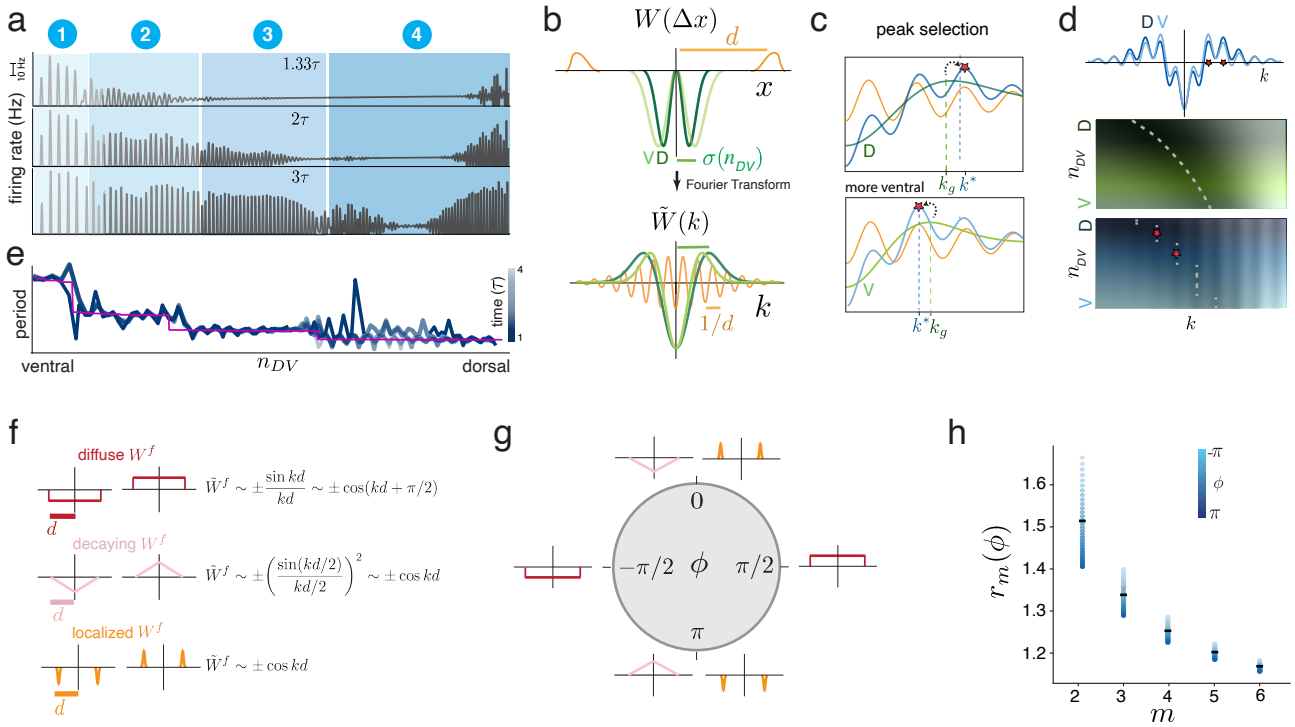


FIG. 3. **Theory of module emergence: multi-scale linear instability and topological peak selection** (a) Snapshots of population activity within a few neural time-constants (τ) of initializing the dynamics at a uniform state. Modules appear in situ at the same time as local patterning, before most neurons have hit their nonlinear thresholds. (b) Top: Schematics of fixed-width (orange) and graded (green) interaction kernels. Kernels from two different DV locations (designated D for dorsal and V for ventral). Bottom: their Fourier transforms. (c) Peak selection process: The global maxima in Fourier space (blue) are based on combining the graded interaction (green) with the fixed interaction (orange). As the green peak slides across, the global maximum (marked by red star) jumps abruptly from the position of one orange peak to the next. (d) Top: Summed Fourier transform of the two local interactions (darker (lighter) blue: more dorsal (ventral)). Middle: The location of the maximum of the graded interaction varies smoothly as a function of DV location. Bottom: the maximum of the summed interaction jumps discontinuously (bottom). (e) Dark to light blue curves: activity pattern periods from (a) for early to late times after initialization. Module boundaries and periods remain unchanged from the earliest time-points. Pink: theoretical prediction of periods and module boundaries from Eq. 2. (f) Left: Example simple fixed-scale interaction profiles that produce modularization: profiles can be roughly categorized as diffuse, decaying, or localized. Right: the dominant terms in their Fourier transforms. (g) The Fourier phases of the interactions in (a). (h) Theoretically predicted sequence of period ratios for any value of ϕ (blue circles), for module numbers 2-6. Any dependence on ϕ and thus the shape of the fixed-scale interaction is weaker for higher module numbers (smaller period/dorsal modules). *See Methods for parameter and simulation details for (a) and (e).*

in period between modules, Fig. 3c-d. Thus the spatial periods along the DV axis form a discrete set determined by the maxima of $\tilde{W}^f(k)$, which occur at

$$\lambda_m^{-1} = k_m^* \approx \left\{ \frac{2\pi m + \phi}{d} \middle| m \in \mathbb{Z}^+ \right\}. \tag{3}$$

The module periods are proportional to the width d of the fixed-scale interaction W^f, and are given by this scale divided by a set of integers m.

Permitted values of the integers m are determined by the set of local maxima $k^*(n_{DV})$ of the fixed-scale interaction (in Eq. 2) that fall within the range $[\eta/\sigma_{\max}, \eta/\sigma_{\min}]$, where η is a proportionality constant (SI Sec. D 7). It follows that the number of allowed periods (and thus modules) is determined by the set of integers m that fit in the following interval:

$$\frac{\eta d/\sigma_{\max} - \phi}{2\pi} \leq m \leq \frac{\eta d/\sigma_{\min} - \phi}{2\pi} \tag{4}$$

The phase $\phi \in [-\pi, \pi]$ is a simple scalar, determined by coarse-grained features of W^f (Fig. 3f-g): A diffuse W^f out to scale d leads to $\phi \approx \pm\pi/2$ for inhibitory or excitatory interactions respectively; a localized W^f at scale d leads

to a value of ϕ close to 0 or π for excitatory or inhibitory interactions respectively; a decaying W^f also leads to ϕ close to 0 or π for inhibitory or excitatory interactions, respectively. Intermediate values of ϕ can be obtained by interpolating between these fixed-scale lateral interaction shapes (See Fig. 12 for examples of numerical simulations for several cases).

The analytical expression for module periods obtained by evaluating Eq. 2 on the Fourier transform of W exactly predicts the values from numerical simulation (Figs. 3e, 4b, SI Fig. 12). Similarly, the even simpler analytical expression for period in Eq. 3 with ϕ computed from W^f and without free parameters, also exactly predicts module periods from numerical simulation across diverse lateral interaction shapes (Figs. 3a, 4b, SI Fig. 12).

Period ratio prediction and parameter invariance through *topological protection*

The (inverse) module period expression of Eq. 3 supplies a quantitative prediction about module period ratios, which have been characterized experimentally [24] and are the subject of several theoretical models [54, 55]. In contrast to results in which period ratios have been described with one value regardless of the module number, our prediction varies with module: the period ratio of the mth module to the $m+1$th module is given by:

$$\frac{\lambda_m}{\lambda_{m+1}} \equiv r(m,\phi) = \frac{(m+1+\phi/2\pi)}{(m+\phi/2\pi)}. \tag{5}$$

Note that while the module periods are determined by the scale d of the fixed-scale interaction, the module period ratios are strikingly and completely independent of any scale: neither $d, \sigma_{n_{DV}}$, nor L. For $\phi = 0$, The ratios of adjacent modules are simply successive integer ratios, with the integer indexing the module number (Fig. 3h). It is also independent of the functional form of gradation in $\sigma_{n_{DV}}$.

The only parameter dependence in the period ratios is through the scalar phase ϕ, which itself depends only on coarse properties of the lateral interaction W^f (Fig. 3f-h). This extreme invariance of the period ratios to almost all parameters is a signature of a topologically protected[1] process (SI Sec.D 7), and is due to the discrete topology of the set of successive integers which are generated by the peak selection process. Thus, the modularization process is robust to nearly all parameteric variation.

Emergent self-scaling of modules

Above we saw that within-module properties like pattern periods within each module are independent of details of the lateral interaction kernel shapes, the size L of the neural sheet, the width of the graded interaction and the functional form of the gradient. However, total independence of structure emergence from system parameters can be a weaknesses of pattern forming dynamics, if structures do not scale with quantities like system size [12, 56].

Interestingly and surprisingly, if the minimum and maximum widths ($\sigma(0) = \sigma_{\max}$, $\sigma(L) = \sigma_{\min}$) of the graded interaction remain fixed at the original absolute sizes, as does the absolute width of the fixed-scale interaction, while the network size L is varied, the number of formed modules remains unchanged, meaning that the module formation process has self-scaled to the system size (Fig.4a-b). Each module's internal periodicity remains unchanged (Fig.4a-b), but the modules are now larger in width so that they each occupy the same fraction of the length of the neural sheet (Fig.4b). In other words, the dynamics of module formation exhibit an emergent invariance or self-scaling property with brain size, automatically adjusting to the size of the substrate without requiring any change in the biological and genetic parameters that control connectivity width.

We can understand this analytically within our theoretical framework: The number of formed modules N_{mod} is given by the number of integers satisfying Eq. (4), and thus is given by:

$$N_{mod} = \left\lfloor \frac{\eta d}{2\pi\sigma_{\min}} - \frac{\phi}{2\pi} \right\rfloor - \left\lceil \frac{\eta d}{2\pi\sigma_{\max}} - \frac{\phi}{2\pi} \right\rceil$$

where $\lfloor\ \rfloor$, $\lceil\ \rceil$ indicate the floor and ceiling operations, respectively. In words, the number of modules is determined by the (integer) difference in the ratios of d with $\sigma_{\min}, \sigma_{\max}$: it is based on the interplay (difference) between the width ratios of the two local kernels, without depending on their specific widths or even the values of their ratios. Consistent with the numerical simulation results (Fig.4a-b), the number of formed modules is predicted to independent of the size of L if $\sigma_{\min}, \sigma_{\max}, d$ are held fixed, which gives the dynamics its self-scaling property.

Even more so than for the within-module periods (cf. Eq. 3), the number of modules depends on the shapes of W^f, W^g only slightly, through the scalar factor ϕ. Large variations in the shape result in the same number of modules

(Fig. 2 and additional profiles in SI Fig.12). The number of modules is also notably predicted to not depend on the shape of the monotonically varying width function $\sigma(n_{DV})$ (Fig. 4c).

We can make a more approximate but still fairly accurate prediction for where module boundaries will form (Fig. 3e, pink, and SI D 7):

$$\text{boundary locations} \approx \sigma^{-1}\left(\frac{\eta d}{(2m+1)\pi}\right). \tag{6}$$

where σ^{-1} is the inverse function of $\sigma(n_{DV})$, such that $\sigma^{-1} \circ \sigma(n_{DV}) = n_{DV}$. Unlike the module number and period, the module boundary locations are predicted to depend on the shape of the gradient $\sigma(n_{DV})$ (Fig. 3c). Numerical simulations confirm the independence of module number and period to changes in the shape of $\sigma(n_{DV})$ when $\sigma_{\min}, \sigma_{\max}$ are fixed, but that boundary locations and therefore relative module sizes can change (Fig. 4d), see SI Sec D 7 for more details.

Finally, even if $\sigma_{\min}, \sigma_{\max}$ are varied smoothly (while d is held fixed), or if d is varied smoothly (while $\sigma_{\min}, \sigma_{\max}$ are held fixed), the number of modules will remain fixed until the change becomes large enough to accommodate one additional or one less module. All these features are properties of the topologically robust process of module emergence, and the number of modules is a topological invariant of the system [1].

A corollary of the independence of module number to system size is that the average module size will scale in proportion to system size, as L/N_{mod}. Thus, if the neural sheet is large, the module sizes can be orders of magnitude larger than any of the lateral interaction scales, $\sigma_{\min}, \sigma_{\max}, d$, resolving the mystery of what sets the scale of individual modules, why they are unrelated to the local interaction scales, and why they are global in size, Fig. 2b-d,f-h.

Self-scaling occurs when holding the endpoints of the global gradient fixed as system size is varied, and is a property of positional information models. Thus, similar results could have been obtained by applying a set of thresholds to the function $\sigma_{n_{DV}}$, which then activated different downstream process to specify the boundaries and period for the corresponding module, which would correspond to the positional or french-flag model [11]. The difference here is that the modules are *self-generated* and do not require externally imposed thresholds or control. The self-organization process allows for mathematical prediction of the periods. An additional consequence for evolutionary dynamics is that changing the number of modules in the peak selection mechanism involves simply changing the value of one endpoint ($\sigma_{\max}$ or $\sigma_{\min}$) of the graded interaction function $\sigma(n_{DV})$, rather than creating new gene expression cascades for each added module (Fig. 4e).

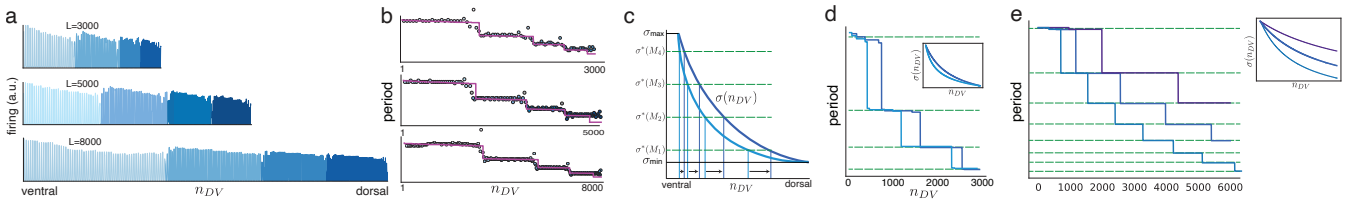


FIG. 4. **Emergent self-scaling.** (a) Increasing the size of the neural sheet while holding constant the minimum and maximum graded interaction widths and the fixed interaction width, the within-module periods remain the same but module sizes expand so that the system has the same number of modules regardless of system size. (b) Extracted periods from results in (a). The neural axis is scaled (normalized) by network size to compare relative module sizes; the period axis is the same across plots (preserved periods in each module). Pink: analytical predictions from Eq. 2. (c) Different functions (shapes) for the monotonically graded interaction width $\sigma(n_{DV})$ are predicted theoretically to result in the same number of modules if the minimum and maximum values of the width ($\sigma_{\min}, \sigma_{\max}$) remain unchanged. Shape changes only affect the detailed positions of module boundaries. (d) k^* calculated from numerical Fourier transform of interaction matrix with two different gradient shapes, holding $\sigma_{\min}$ and $\sigma_{\max}$ fixed. Module number and periods remain unchanged, while boundaries shift. (inset) The shapes of the gradient in the width of the primary pattern-forming interaction for the two choices of gradient shapes. Green dashed lines are scales corresponding to each local maxima of the secondary interaction. (e) k^* calculated from numerical Fourier transform of interaction matrix with three significantly different values of $\sigma_{\min}$ while holding the spatial extent of the system fixed. The number of formed modules changes from 3 to 5 to 8, while the periods of the first few modules (that are common across all three simulations) remain unchanged. (inset) The shapes of the gradient $\sigma(x)$ in the primary pattern-forming interaction for the three choices of gradient shapes. *See Methods and SI Sec D 7 for parameter and simulation details.*

The simple analytical expressions for module period and period ratios, their parameter (in)dependence, and self-scaling properties accurately match the results of full numerical simulations, Fig. 5a.

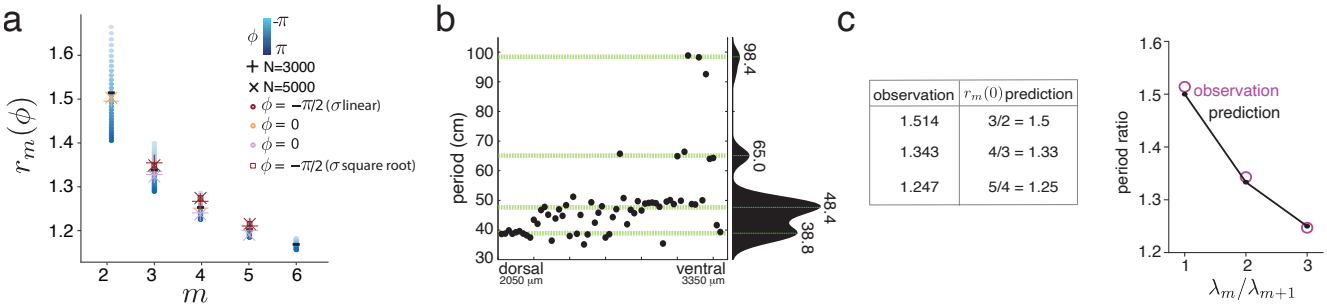


FIG. 5. **Comparison of precise period ratio predictions with data** (a) Period ratio predictions from 3h together with numerical simulation (other symbols) of neural circuit models with the set of fixed-scale interaction profiles shown in 3f, same color code. Numerical simulations with all combinations of network size and weight profiles are shown. (b) Observed periods of grid cells from multiple modules [24] (c) Successive period ratios computed from the observation (left column), and predicted period ratios for $\phi = 0$ (middle column). Ratios match predicted values with $R^2 = 0.999$ (right column). *See Methods for parameter and simulation details.*

Neural data matches detailed predicted sequence of period ratios

We compare our predictions to existing results, which estimate a single ratio across adjacent module pairs [24], by averaging the predicted ratios across 4 modules and across different phases ϕ (SI Sec. D 6). This yields a predicted value of 1.37, in good agreement with experimental results of an average ratio of ~ 1.42 across animals as reported in [24] and 1.368 for the animal shown in Fig. 5b.

Further, we compare our more fine-grained successive period ratio predictions with published per-module period values. Our prediction with $\phi = 0$ matches the sequence of observed period ratios [24] strikingly well, Fig. 5c, also for other datasets in which multiple grid periods ratios are available from single individuals (SI Sec. D 10).

PEAK SELECTION ENHANCES ROBUSTNESS WITHIN INDIVIDUAL GRID MODULE ATTRACTOR NETWORKS

We have seen that peak selection-based emergence leads to module formation via patterning on two scales, and that the results are invariant to variations in parameters, function shapes, and the form of the global gradients. We next investigate and discover that the two-scale interaction and topological structure emergence process also confers substantial additional robustness to traditional continuous attractor models. Specifically, this mechanism makes continuous attractor dynamics resistant to several forms of weight heterogeneity and activity perturbation.

Robustness within individual continuous attractor networks

Standard attractor models [18, 23, 57, 58] require weight homogeneity to generate regular patterns and continuous attractor states for representing continuous variables. Heterogeneities degrade the pattern and the continuity of attractor states. If the heterogeneities are sufficiently large [59, 60], pattern formation can entirely fail (Fig. 6a, left and SI Sec. D 9 for details). This susceptibility of continuous attractor dynamics to noise is one of the fundamental open problems for continuous attractor models [23].

We hypothesize that the second wider-scale local interaction that leads to multi-module formation might enforce pattern homogeneity even in the absence of homogeneity in the pattern-forming weights because it imposes an additional set of constraints on local pattern formation.

We numerically simulated the standard CAN model for one grid module (i.e. no gradient in the pattern forming interaction), and added noise, both in the shared radial structure of the interaction weights across neurons and i.i.d. noise in every weight, SI Fig. 20 and SI sec D 9 for details and visualizations). With the same level of added noise that largely destroys patterning in conventional attractor networks (Fig. 6a, left), inclusion of a second wider-scale local interaction subject to same added noise results in robust and homogeneous pattern formation (Fig. 6a, right). We quantified this effect, finding that within-module homogeneity despite weight heterogeneity is substantially higher with peak selection than without, Fig. 6b.

The robustness enhancement from the addition of a second local interaction for single modules also holds in the formation of multiple modules (Fig. 6c). Without a second interaction, weight inhomogeneities drive variations in pattern period, whose magnitude scales like $1/\sigma$ (the width of the minimum of $\tilde{W}$) as the interaction profile is linearly stretched by σ, Fig. 6d (left). When a broader second scale is added, its narrower Fourier peaks confine the permitted solutions, enforcing a more homogeneous period despite weight inhomogeneity, Fig. 6d (right). Variations in period scale as $\sim 1/d$, a relative reduction of σ/d. Further, since module periods scale as d/m, the smaller period modules should be more robust than larger-period modules. (Details in SI D 9).

Conceptually, the peaks of the wider local interaction "focus" the dynamics to a narrower region than specified by the narrower local interaction alone (Fig. 6d).

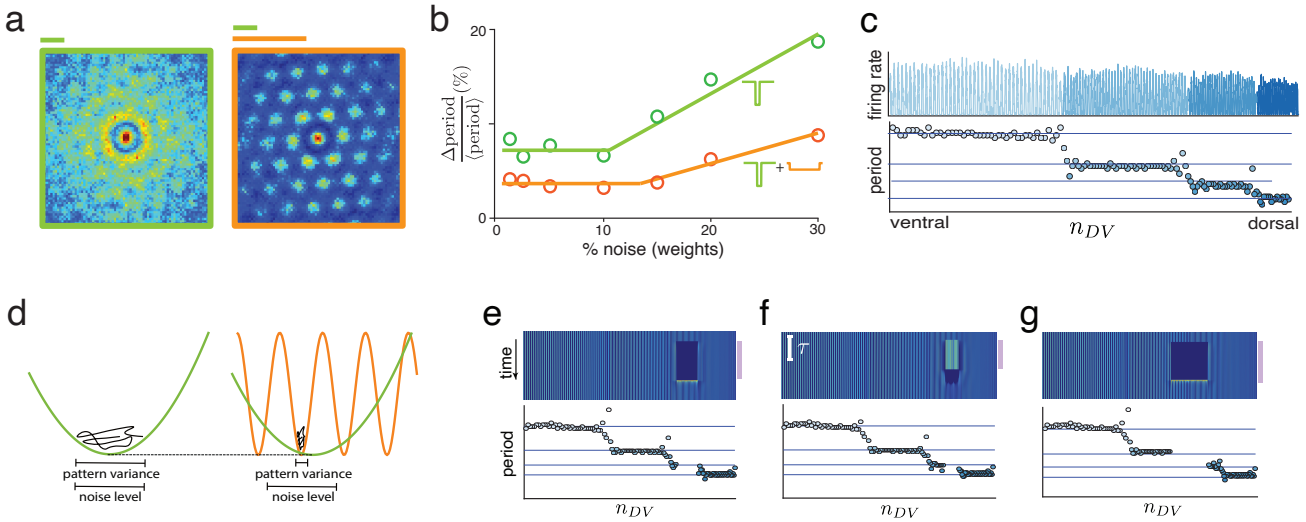


FIG. 6. **Enhanced robustness to weight heterogeneity, noise, and activity perturbation by peak selection.** (a) Left: Weight heterogeneity (here, radial asymmetry and i.i.d. noise) quickly destroys discernable pattern structure in standard continuous attractor models [18]). (Simulation of a 100x100 neuron network with kernel W_g as in Fig. 2d-e; scale shown by green bar.) Right: Addition of a secondary wider local interaction (scale shown by orange bar), with noise in both sets of weights, rescues patterning. (b) Variability in 1-dimensional patterning versus the magnitude of added noise in the weights, for single-scale weights (green), and for networks with two local interaction scales (orange). Pattern variation is the ratio of the standard deviation to the mean of the pattern period. (c) The same as (a), showing regularity in period despite the addition of noise in a 1-dimensional setting. (d) The mechanism for enhanced within-module robustness: the broader local interaction scale enforces a narrower set of solutions in the energy landscape than possible with the pattern-forming interaction alone. (e-g) Inter-module dynamical independence: (e) An entire module is transiently silenced for 50 ms; (f) a large fraction of a module is externally driven by large-amplitude fixed, random, independent perturbations; (g) a continuous region that spans two modules is transiently silenced. In all cases, the perturbation remains local so neighboring regions and modules are unaffected, and the perturbed module recovers within one neural time-constant after removal of the perturbation. *See Methods for parameter and simulation details.*

Robustness to module-wide and across-module perturbations in activity

Given that the lateral neural interactions are not discontinuous across modules (weights span modules), and given that global modules emerge from local interactions and local dynamics, we wondered whether erasing patterning within entire modules or large across-module regions of the network will affect patterning in the rest of the modules.

Upon silencing activity in one module (mimicking optogenetic inactivation) the other modules, module periods, and even adjacent module boundaries remain stable, Fig.6e. When a subset of cells in one module is driven to maintain randomly chosen values between 0 and 20Hz, patterned activity is disrupted in adjacent regions of the module, Fig.6f, but when the perturbation is lifted, the module period at the perturbation site is restored. If a region spanning the boundary between two modules is silenced, the dynamics and periods in spared parts of the two modules remain unchanged, and the boundary re-emerges at its pre-perturbation position after removal of the suppressive input (Fig.6g). In all cases, pre-perturbation states are restored within one cellular time-constant ($\approx \tau$). These findings contrast with existing models of module formation in which modules interact in a stacked architecture [61]: these

models exhibit cascading dependencies between modules, so that perturbation of one module will have propagating effects across all downstream modules.

GENERALIZED ENERGY LANDSCAPE VIEW OF MODULE EMERGENCE

Motivated by the peak-selection process of grid module emergence, we hypothesize that the same principle could supply a general mechanism for module emergence even in the absence of periodic pattern formation. The theory can be generalized in two steps. First, by translating the linear dynamics of Fourier modes into nonlinear dynamics on an general energy landscape, and next translating the Fourier peaks and troughs into multiple rugged local optima in the energy landscape (SI Sec. E).

Consider a generalized energy landscape (Lyapunov function) constructed as the sum of two terms (Fig. 7b): a spatially independent term $f(\theta)$ that is rugged with multiple similar-depth minima (constructed here as a sample from a Gaussian process with a radial basis function kernel, shown in orange in Fig. 7b), and another that has a single broad minimum (constructed here as a quadratic term, shown in green in Fig. 7b). The location of the broad minimum moves as a function $\theta_0(x)$ of some smoothly varying parameter x (Fig. 7a). This results in a combined energy landscape

$$L(\theta, \theta_0(x)) = (1-\alpha)f(\theta) + \alpha\|\theta - \theta_0(x)\|^2 \tag{7}$$

The problem can be viewed as regularized optimization on a rugged loss landscape, with a spatially-dependent regularizer $||\theta - \theta_0(x)||^2$ acting as a prior that selects one of the local minima at each location. The dynamics of θ are defined as downhill flow on the Lyapunov function landscape: $\dot{\theta} = -\partial L/\partial\theta$. The landscape is first governed by the broad quadratic term, then sculpted by the rugged landscape (with α gradually decreasing with time starting from $\alpha = 1$, SI Sec. F). Smoothly varying the parameter x results in a set of modular solutions with steps along the smoothly varying parameter dimension (Fig. 7c), corresponding to the state settling into the local minimum closest to the quadratic minimum at that x. This analytical formulation with numerical validation provides a simple mathematical framework of peak selection that shows how smooth gradients can lead to discrete patterning and modular structure emergence [31, 40, 62] beyond linear instability, Fourier modes, and periodic solutions.

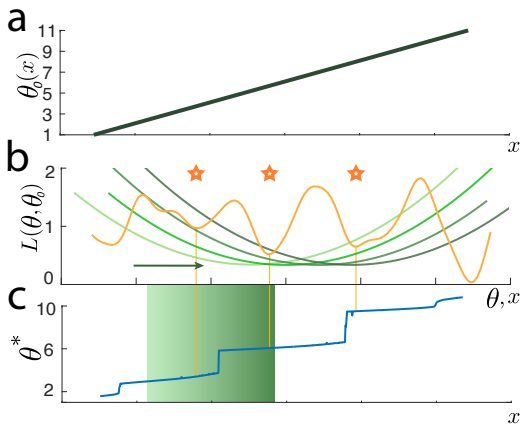


FIG. 7. **Generalized peak-selection mechanism for module emergence.** (a) Parametric gradient of parameter θ_0 as a function of position x. (b) The Lyapunov (generalized energy) function is a sum of two functions: one static function with multiple local minima, and one with a single broad minimum whose location smoothly varies with x according to $\theta_0(x)$. (c) Fixed points of the dynamics upon simulation of dynamics on this Lyapunov function, as a function of the location x. *See SI Sec. F for simulation details.*

SELF-ORGANIZATION OF ECOLOGICAL NICHES THROUGH PEAK SELECTION

The general principle of topological peak selection — with a smoothly spatially varying "selecting function" and a spatially constant fixed interaction that generates multiple local minima — can be applied to generate modularity across diverse settings. To illustrate this generality, we explore a biological example that is far from structure

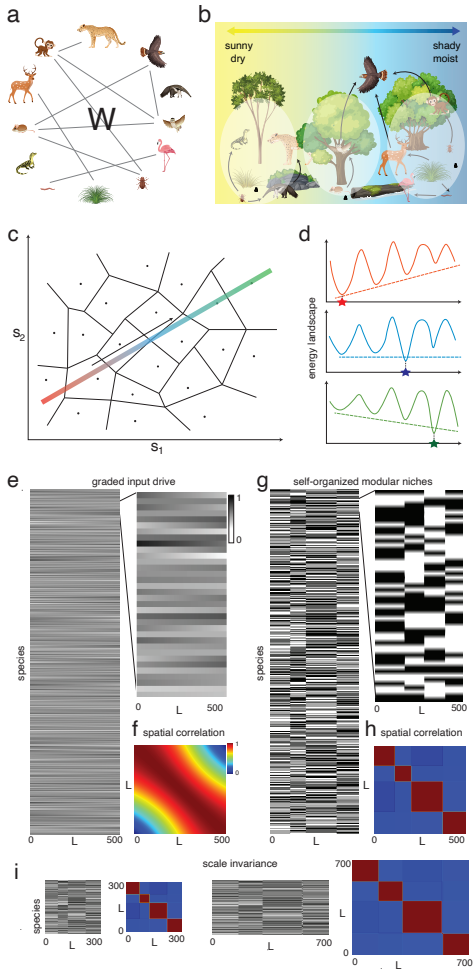


FIG. 8. **Self-organization of ecological niches through peak selection**: (a) Schematic web of competitive and cooperative interactions between species. (b) Schematic of ecological niches and smoothly varying spatial resource gradients. (c) Schematic of the state space of the interacting species system as in (a), ignoring spatial distributions of species and resources, has multiple attractor states (dots; basins boundaries depicted by black lines). Multi-colored line: One-dimensional variation in resources, passing through the basins of several attractors. (d) Energy landscape of the state space visualized at 3 distinct points along the resource gradient. The resource gradient slowly "tilts" the energy landscape, gradually varying the relative heights of the local optima and discretely changing the global optimum. (e-h) Typical dynamics of the system: no spatial structure at initialization, visualized in system state (e) and spatial correlation matrix (f), self-organizes into four ecological niches at steady state (g-h). (i) Scale invariance: changing the spatial size of the system while maintaining the end points of the resource gradient function results in the same niche structure. *See Methods for parameter and simulation details.*

emergence in the brain, at a much larger scale of organization: the emergence of spatial ecological niches in ecosystems (Fig. 8a,b).

Consider a set of species that, if co-localized, interact through cooperation and competition (Fig. 8a). Each species requires different resource distributions, and resources vary along environmental gradients. Graded environmental conditions are common features of ecosystems that have been implicated in niche formation, and include for instance sunlight, temperature, precipitation, and electrochemical gradients [63–68] as well as secondary quantities like vegetation that results from these. In aquatic ecosystems, light and nutrient availability limit the quantity of primary producers, which poses fundamental limits on the food web [69, 70]. Changes in such environmental variables and nutrient availability is known to alter population dynamics [71–74] as well as inter-species interactions [68, 75, 76], that can affect the composition of the formed niche.

We model this system with Hopfield-like dynamics [77–81], where species interact according to an interaction matrix (W). The interaction defines a set of attractor states in a rugged energy landscape, Fig. 8c, with the minima corresponding to the set of possible niches. Interactions are also modulated by spatial distances: two species interact via W only if they spatially overlap. Further, species also grow in proportion to the match between the

resource inputs and their own resource feature vectors. Formally, consider N species indexed by i and defined by M-dimensional resource feature vectors $\mathbf{b}_i$ that specify their resource needs. Each element of each resource feature vector is sampled from independent, identically distributed normal distributions (see Methods for details). A time-invariant but spatially varying resource gradient $\mathbf{r}^g(x)$ is assumed to exist for $x \in [0, L]$, that is constructed by linearly interpolating between two random vectors drawn from $\{-1, 1\}^M$. The population dynamics of species i at position x and time $t+1$ is given by:

$$s(i, x; t+1) = H\left[-\delta_i s(i, x; t) + \sum_{j,x'} W(i, j) K(x, x') s(j, x'; t) + \mathbf{b}_i \cdot \mathbf{r}^g(x)\right] \tag{8}$$

where $H[.]$ is a nonlinearity, δ_i is the death rate, $W(i, j)$ is a symmetric interaction matrix quantifying the cooperation and competition between species i and j (see Methods for details of construction), and $K(x, x') \equiv K(x - x')$ is a spatial interaction kernel that vanishes on a length scale much smaller than L. Note that while we only examine the symmetric $W(i, j)$ case here, we expect the self-organization of modular niches to hold in general for non-symmetric $W(i, j)$ as well. In such a case, the selection may be between different attractor states that need not be fixed point attractors.

This system settles into a steady state $\mathbf{s}^*$. In the energy landscape view, the graded resource inputs $\mathbf{r}^g(x)$ selectively deepen a subset of the minima. As the resource inputs continuously vary in space, the minima that are amplified vary with x, leading to a discrete selection of the global minima at each spatial location, Fig. 8d. We examine spatial structure through the normalized correlation matrix $C(x, x') = \sum_{i=1}^{N} \hat{s}^*(i, x)\hat{s}^*(i, x')$, where $\hat{s}$ is s normalized to have zero mean and unit norm. Continuously graded input drive with no initial modularity in the population state correlation matrix self-organizes into a steady state with modular niches (Fig.8e-h). The system also exhibits scale invariance if the resource gradient endpoints remain fixed as the spatial domain is scaled, even though the spatial interaction scale (set by $K(x, x')$) is not (Fig. 8i).

DISCUSSION

Summary We have combined positional and spontaneous pattern forming (or more generally, symmetry breaking) mechanisms to show that structure can robustly emerge at multiple scales from purely local interactions, without requiring modular genetic specification or modular connectivity. The resulting mechanism exhibits useful features of both processes [12]: structure is noise-tolerant and parameter invariant, self-organizing with relatively little detailed genetic instruction, and self-scaling. The mechanism involves only smooth (low information) genetically specified gradients combined with local competitive interactions whose parameters are modulated by the gradient(s), from which sharp boundaries and modules emerge.

In the context of grid cells, this is the first work to provide a framework for connecting gene expression (smoothly varying gradient(s)), structure (local neural connectivity profiles), and physiology with each other. It therefore allows future physiology, connectomics, and transcriptomics experiments to inform and constrain each other. The mechanism of module emergence generates a detailed prediction of period ratios between modules and provides a novel result on increased robustness of single continuous attractor network models with the addition of a wider local interaction.

Predictions The theoretical relationships – connecting simple properties of the local interactions and their spatial gradients to formed module properties – provide low-dimensional knobs for experimental testing and manipulation. The analytical and numerical models can be directly queried for predictions about the effects of varying any of the knobs. Here, we enumerate a few example predictions for connectomics (C), transcriptomics (T), development (D), and physiology (P): 1) Discrete modules are initially underpinned by merely smooth gradients in gene expression and in the interaction weights. (D,T,C) 2) The existence of an interaction that is fixed across the DV axis of MEC, and another that smoothly varies in scale (strength or width) across it. These two types of interactions may be in different cells and synapses, or the same synapses, for instance a single scaled-width weight profile with an axonal or dendritic cutoff radius that is invariant across the DV axis. (C) 3) The detailed and precise relationship of adjacent grid periods, as given by successive integer ratios, are distinct from other theoretically interestingpredictions [54, 82, 83], and deviations from successive integer ratios should be predictable by alterations in the coarse form of the (fixed-scale) lateral interaction profile (P). 4) Invariance of module number and periods to detailed local interaction shape and to the shape of the global gradient, and the shifting of module boundaries with gradient shape (T,C,P). 5) Self-scaling with system size (invariance of module number and periods despite MEC size variations if extremal values of the gradient fixed), dependence of these quantities on extremal values of the gradient, and dependence of module boundary positions on the functional form of the gradation in lateral interactions. (T,C,P) 6) Bidirectional computable relationship between lateral interaction shape and period ratios, through the scalar variable ϕ. (C,P) (3-6)

Could be tested across individuals and species. 7) Following MEC-wide silencing, activity patterning in all modules should re-emerge independently, rather than sequentially [61]. (P) 8) Quantifiably higher robustness of dynamics to connectivity inhomogeneity, beyond the capabilities of conventional attractor models [45]. (C) 9) Independence of dynamics between modules and high robustness of dynamics to activity perturbation within and across modules: Effects of perturbation to activity are localized to the module it is applied to, without a cascading effect across modules [61]. Entirely suppressing one module should not alter others, and suppressing half a module boundary should not shift the rest of the boundary. (P) 10) General peak selection mechanism through manipulation of genetic gradients: quantitative predictions for altering the number of formed modules, periods, and positions of module boundaries with gene expression modulation. (T,D)

Related work This work directly extends and robustifies continuous attractor models of grid cells [45], from single modules to multiple modules, from dependence on specific interaction profiles to an infinite set of kernels for grid emergence, and from high dependence on homogeneous weights to a weaker dependence on weight homogeneity. It connects to observed DV gradients in MEC [32, 38, 42, 84–87], and also potentially more generally to observed gradients that underlie discontinuous function in cortex [31, 43]. Its focus on the theory and mechanisms of emergence of structure and function from given weights is complementary to work that models the *learning* of weights in MEC, through biologically plausible Hebbian-like rules [88] or backpropagation-based learning [83, 89–91]. In fact, the learning models generally do not produce multiple modules, and in the rare cases where they may seem to, the circuit mechanisms that produce them are unknown [89]. We know of only other work that proposes a network mechanism for multi-grid module emergence [61]. It has a distinct (stacked) initial and final architecture, and its predictions on several of the quantities noted above are interestingly and distinguishably different.

The general peak-selection principle for module emergence is both an instance and a generalization of the idea of spatial bifurcation for the emergence of discrete function from smooth gradients [31, 92]. It permits a number distinct modules to form from smooth variations in the spatial dimension, but the broader theoretical framework generalizes to variations along abstract parametric dimensions. We have shown three different flavors of peak selection-based modularity emergence: peak selection in a pattern-forming process interacting with smooth gradients in interaction parameters for grid cells; peak selection via a smoothly varying regularization term in dynamics on a rough landscape in the general Lyapunov function approach; and peak selection in a symmetry-breaking process (that is more general than Turing-like pattern formation) based on initial condition or resource gradients that tilt the landscape in a generalized Hopfield model for niche formation.

These concepts provide dynamical and mechanistic principles for *how* modular structure can emerge, in contrast to normative models that focus on *why* or *when* modular structure is favored [4, 9, 82, 93–100].

A mechanistic understanding of how discrete structure could emerge from graded precursors connects with extensive literature on the existence of smooth biophysical gradients in the brain and body [28, 31–35, 37–44, 92] even when function varies discretely.

Many works explore how structure emergence can occur with precision in the presence of noise [28, 29, 101, 102]. Some solutions within the positional hypothesis involve spatial or temporal integration of noisy gradients [12, 26, 101]. Pattern forming mechanisms have some built-in robustness to noise because the patterned state is much lower-dimensional than the overall state space [103]. Our observation that peak selection confers substantial additional robustness raises the question of whether such a mechanism might assist in tandem with positional mechanisms during developmental morphogenesis [104].

Acknowledgements

We are grateful to Mehran Kardar for helpful discussions and to John Widloski, Akhilan Boopathy and Ling Dong for comments on the manuscript. This work was supported by the MathWorks Science Fellowship to MK, the Simons Foundation through the Simons Collaboration on the Global Brain, the ONR, and the Howard Hughes Medical Institute through the Faculty Scholars Program to IRF.

Conceptualization, writing: MK, SC, IRF; Coding, analysis: MK, SC.

[1] David Thouless. *Topological quantum numbers in nonrelativistic physics*. World Scientific, 1998.

[2] Marta Sales-Parda. The importance of being modular. *Science*, 6347(357):128–129, 2017.

[3] Anurag Limdi, Alfonso Pérez-Escudero, Aming Li, and Jeff Gore. Asymmetric migration decreases stability but increases resilience in a heterogeneous metapopulation. *Nat Commun*, 9(1):2969, 07 2018.

[4] Jeong-Man Park, Man Chen, Dong Wang, and Michael W Deem. Modularity enhances the rate of evolution in a rugged fitness landscape. *Phys Biol*, 12(2):025001, Mar 2015.

[5] Wilhelm von Humboldt. On the diversity of human language construction and its influence on the mental development of the human species. Cambridge University Press, 2005.

[6] J A Fodor and Z W Pylyshyn. Connectionism and cognitive architecture: a critical analysis. *Cognition*, 28(1-2):3–71, Mar 1988.

[7] Ila R Fiete, Yoram Burak, and Ted Brookings. What grid cells convey about rat location. *J Neurosci*, 28(27):6858–71, Jul 2008.

[8] Sameet Sreenivasan and Ila Fiete. Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nat Neurosci*, 14(10):1330–7, Sep 2011.

[9] Ferran Alet, Tomas Lozano-Perez, and Leslie P. Kaelbling. Modular meta-learning. In Aude Billard, Anca Dragan, Jan Peters, and Jun Morimoto, editors, *Proceedings of The 2nd Conference on Robot Learning*, volume 87 of *Proceedings of Machine Learning Research*, pages 856–868. PMLR, 29–31 Oct 2018.

[10] Man Yi Yim, Lorenzo A Sadun, Ila R Fiete, and Thibaud Taillefumier. Place-cell capacity and volatility with grid-like inputs. *Elife*, 10, May 2021.

[11] Lewis Wolpert. Positional information and the spatial pattern of cellular differentiation. *Journal of theoretical biology*, 25(1):1–47, 1969.

[12] Jeremy BA Green and James Sharpe. Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development*, 142(7):1203–1211, 2015.

[13] Lucia Durrieu, Daniel Kirrmaier, Tatjana Schneidt, Ilia Kats, Sarada Raghavan, Lars Hufnagel, Timothy E Saunders, and Michael Knop. Bicoid gradient formation mechanism and dynamics revealed by protein lifetime analysis. *Mol Syst Biol*, 14(9):e8355, 09 2018.

[14] AM Turing. The chemical basis of morphogenesis. *Philos Trans R Soc Lond BB*, 237(641), 1952.

[15] Jelena Raspopovic, Luciano Marcon, Laura Russo, and James Sharpe. Digit patterning is controlled by a bmp-sox9-wnt turing network modulated by morphogen gradients. *Science*, 345(6196):566–570, 2014.

[16] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E.I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005.

[17] Yi Gu, Sam Lewallen, Amina A Kinkhabwala, Cristina Domnisoru, Kijung Yoon, Jeffrey L Gauthier, Ila R Fiete, and David W Tank. A map-like micro-organization of grid cells in the medial entorhinal cortex. *Cell*, 175(3):736–750.e30, Oct 2018.

[18] Yoram Burak and Ila R Fiete. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput Biol*, 5(2):e1000291, Feb 2009.

[19] K.J. Yoon, M.A. Buice, R. Barry, C.and Hayman, N. Burgess, and I.R. Fiete. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nat Neurosci*, 16(8):1077–84, Aug 2013.

[20] S.G. Trettel, J.B. Trimper, E. Hwaun, I.R. Fiete, and L.L. Colgin. Grid cell co-activity patterns during sleep reflect spatial overlap of grid fields during active behaviors. *Nat Neurosci*, 22(4):609–617, 04 2019.

[21] Richard J Gardner, Li Lu, Tanja Wernle, May-Britt Moser, and Edvard I Moser. Correlation structure of grid cells is preserved during sleep. *Nat Neurosci*, 22(4):598–608, 04 2019.

[22] Richard J Gardner, Erik Hermansen, Marius Pachitariu, Yoram Burak, Nils A Baas, Benjamin A Dunn, May-Britt Moser, and Edvard I Moser. Toroidal topology of population activity in grid cells. *Nature*, pages 1–6, 2022.

[23] Mikail Khona and Ila R Fiete. Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, 23:744–766, 2022.

[24] H. Stensola, T. Stensola, T. Solstad, K. Fröland, M. Moser, and E. Moser. The entorhinal grid map is discretized. *Nature*, 492(7427):72–8, Dec 2012.

[25] Michel Kerszberg and Lewis Wolpert. Specifying positional information in the embryo: looking beyond morphogens. *Cell*, 130(2):205–209, 2007.

[26] Gašper Tkačik and Thomas Gregor. The many bits of positional information. *Development*, 148(2):dev176065, 2021.

[27] François Schweisguth and Francis Corson. Self organization in pattern formation. *Dev. Cell*, (49):659–677, Jun 2019.

[28] Thomas Gregor, David W Tank, Eric F Wieschaus, and William Bialek. Probing the limits to positional information. *Cell*, 130(1):153–164, 2007.

[29] Shawn C Little, Mikhail Tikhonov, and Thomas Gregor. Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell*, 154(4):789–800, 2013.

[30] Naama Barkai and Ben-Zion Shilo. Variability and robustness in biomolecular systems. *Molecular cell*, 28(5):755–760, 2007.

[31] Xiao-Jing Wang. Macroscopic gradients of synaptic excitation and inhibition in the neocortex. *Nature Reviews Neuroscience*, 21(3):169–178, 2020.

[32] P. Beed, A. Gundlfinger, S. Schneiderbauer, J. Song, C. Bohm, A. Burgalossi, M. Brecht, I. Vida, and D. Schmitz. Inhibitory gradient along the dorsoventral axis in the medial entorhinal cortex. *Neuron*, 79:1197–1207, 2013.

[33] Andrew Miri, Kayvon Daie, Aristides B Arrenberg, Herwig Baier, Emre Aksay, and David W Tank. Spatial gradients and multidimensional dynamics in a neural integrator circuit. *Nature neuroscience*, 14(9):1150–1159, 2011.

[34] Joshua B Burt, Murat Demirtaş, William J Eckner, Natasha M Navejar, Jie Lisa Ji, William J Martin, Alberto Bernacchia, Alan Anticevic, and John D Murray. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nature neuroscience*, 21(9):1251–1259, 2018.

[35] Bertha Vázquez-Rodríguez, Laura E Suárez, Ross D Markello, Golia Shafiei, Casey Paquola, Patric Hagmann, Martijn P Van Den Heuvel, Boris C Bernhardt, R Nathan Spreng, and Bratislav Misic. Gradients of structure–function tethering across neocortex. *Proceedings of the National Academy of Sciences*, 116(42):21219–21227, 2019.

[36] Hugh Pastoll, Derek L Garden, Ioannis Papastathopoulos, Gülşen Sürmeli, and Matthew F Nolan. Inter-and intra-animal variation in the integrative properties of stellate cells in the medial entorhinal cortex. *Elife*, 9:e52258, 2020.

[37] Ben D Fulcher, John D Murray, Valerio Zerbi, and Xiao-Jing Wang. Multimodal gradients across mouse cortex. *Proceedings of the National Academy of Sciences*, 116(10):4689–4695, 2019.

[38] Derek L F Garden, Paul D Dodson, Cian O'Donnell, Melanie D White, and Matthew F Nolan. Tuning of synaptic integration in the medial entorhinal cortex to the organization of grid cell firing fields. *Neuron*, 60(5):875–89, Dec 2008.

[39] Katherine W Rogers and Alexander F Schier. Morphogen gradients: from generation to interpretation. *Annu Rev Cell Dev Biol*, 27:377–407, 2011.

[40] François Schweisguth and Francis Corson. Self-organization in pattern formation. *Dev Cell*, 49(5):659–677, 06 2019.

[41] Leigh Wilson and Malcolm Maden. The mechanisms of dorsoventral patterning in the vertebrate neural tube. *Dev Biol*, 282(1):1–13, Jun 2005.

[42] Sabine Grosser, Federico J Barreda, Prateep Beed, Dietmar Schmitz, Sam A Booker, and Imre Vida. Parvalbumin interneurons are differentially connected to principal cells in inhibitory feedback microcircuits along the dorsoventral axis of the medial entorhinal cortex. *Eneuro*, 8(1), 2021.

[43] BRAIN Initiative Cell Census Network (BICCN). A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature*, 598:86–102, 2021.

[44] Meng Zhang, Stephen W Eichhorn, Brian Zingg, Zizhen Yao, Kaelan Cotter, Hongkui Zeng, Hongwei Dong, and Xiaowei Zhuang. Spatially resolved cell atlas of the mouse primary motor cortex by merfish. *Nature*, 598(7879):137–143, 2021.

[45] Yoram Burak and Ila R Fiete. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput Biol*, 5(2):e1000291, Feb 2009.

[46] D St Johnston and C Nüsslein-Volhard. The origin of pattern and polarity in the drosophila embryo. *Cell*, 68(2):201–19, Jan 1992.

[47] Stephen W Paddock, Eric J Hazen, and Peter J DeVries. Methods and applications of three-color confocal imaging. *BioTechniques*, 22(1):120–126, 1997.

[48] Jonathan J Couey, Aree Witoelar, Sheng-Jia Zhang, Kang Zheng, Jing Ye, Benjamin Dunn, Rafal Czajkowski, May-Britt Moser, Edvard I Moser, Yasser Roudi, and Menno P Witter. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nat Neurosci*, 16(3):318–24, Mar 2013.

[49] H S Seung. How the brain keeps the eyes still. *Proc Natl Acad Sci U S A*, 93(23):13339–13344, 1996.

[50] Bailu Si, Sandro Romani, and Misha Tsodyks. Continuous attractor network model for conjunctive position-by-velocity tuning of grid cells. *Plos computational biology*, 10(4):e1003558, 2014.

[51] Mark S Goldman, Joseph H Levine, Guy Major, David W Tank, and H Sebastian Seung. Robust persistent neural activity in a model integrator with multiple hysteretic dendrites per neuron. *Cerebral cortex*, 13(11):1185–1195, 2003.

[52] Sung Soo Kim, Hervé Rouault, Shaul Druckmann, and Vivek Jayaraman. Ring attractor dynamics in the drosophila central brain. *Science*, 356(6340):849–853, 05 2017.

[53] Philip W Anderson. Absence of diffusion in certain random lattices. *Physical review*, 109(5):1492, 1958.

[54] Xue-Xin Wei, Jason Prentice, and Vijay Balasubramanian. A principle of economy predicts the functional architecture of grid cells. *Elife*, 4:e08362, Sep 2015.

[55] Martin Stemmler, Alexander Mathis, and Andreas VM Herz. Connecting multiple spatial scales to decode the population activity of grid cells. *Science Advances*, 1(11):e1500816, 2015.

[56] David M Umulis and Hans G Othmer. Mechanisms of scaling in pattern formation. *Development*, 140(24):4830–4843, 2013.

[57] K Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci*, 16(6):2112–26, Mar 1996.

[58] Arseny Finkelstein, Dori Derdikman, Alon Rubin, Jakob N Foerster, Liora Las, and Nachum Ulanovsky. Three-dimensional head-direction coding in the bat brain. *Nature*, 517(7533):159–164, 2015.

[59] Ran Darshan and Alexander Rivkind. Learning to represent continuous variables in heterogeneous neural networks. *Cell Reports*, 39(1):110612, 2022.

[60] Alfonso Renart, Pengcheng Song, and Xiao-Jing Wang. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron*, 38(3):473–485, 2003.

[61] Louis Kang and Vijay Balasubramanian. A geometric attractor mechanism for self-organization of entorhinal grid modules. *Elife*, 8, 08 2019.

[62] Andrew L Krause, Václav Klika, Thomas E Woolley, and Eamonn A Gaffney. From one pattern into another: analysis of turing patterns in heterogeneous domains via wkbj. *J R Soc Interface*, 17(162):20190621, 01 2020.

[63] Jordan G Okie, David J Van Horn, David Storch, John E Barrett, Michael N Gooseff, Lenka Kopsova, and Cristina D Takacs-Vesbach. Niche and metabolic principles explain patterns of diversity and distribution: theory and a case study with soil bacterial communities. *Proceedings of the Royal Society B: Biological Sciences*, 282(1809):20142630, 2015.

[64] Diego Alarcón and Lohengrin A Cavieres. Relationships between ecological niche and expected shifts in elevation and latitude due to climate change in south american temperate forest plants. *Journal of Biogeography*, 45(10):2272–2287, 2018.

[65] David Schellenberger Costa, Friederike Gerschlauer, Ralf Kiese, Markus Fischer, Michael Kleyer, and Andreas Hemp. Plant niche breadths along environmental gradients and their relationship to plant functional traits. *Diversity and Distributions*, 24(12):1869–1882, 2018.

[66] Alex Salas-López, Cyrille Violle, François Munoz, Florian Menzel, and Jérôme Orivel. Effects of habitat and competition on niche partitioning and community structure in neotropical ants. *Insect Communities: Diversity Patterns and their Driving Forces*, 16648714, 2023.

[67] Lu Luan, Yuji Jiang, Francisco Dini-Andreote, Thomas W Crowther, Pengfa Li, Mohammad Bahram, Jie Zheng, Qinsong Xu, Xue-Xian Zhang, and Bo Sun. Integrating ph into the metabolic theory of ecology to predict bacterial diversity in soil. *Proceedings of the National Academy of Sciences*, 120(3):e2207832120, 2023.

[68] Meaghan Conway. *Niche Evolution Along a Gradient of Ecological Specialization*. The University of Maine, 2019.

[69] Simon A Levin, Stephen R Carpenter, H Charles J Godfray, Ann P Kinzig, Michel Loreau, Jonathan B Losos, Brian Walker, and David S Wilcove. *The Princeton guide to ecology*. Princeton University Press, 2009.

[70] L Citores, L Ibaibarriaga, D-J Lee, MJ Brewer, M Santos, and G Chust. Modelling species presence–absence in the ecological niche theory framework using shape-constrained generalized additive models. *Ecological Modelling*, 418:108926, 2020.

[71] James H Brown, James F Gillooly, Andrew P Allen, Van M Savage, and Geoffrey B West. Toward a metabolic theory of ecology. *Ecology*, 85(7):1771–1789, 2004.

[72] Aleksej Zelezniak, Sergej Andrejev, Olga Ponomarova, Daniel R Mende, Peer Bork, and Kiran Raosaheb Patil. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences*, 112(20):6449–6454, 2015.

[73] Brian J Enquist, Evan P Economo, Travis E Huxman, Andrew P Allen, Danielle D Ignace, and James F Gillooly. Scaling metabolism from organisms to ecosystems. *Nature*, 423(6940):639–642, 2003.

[74] Brian J Enquist, James H Brown, and Geoffrey B West. Allometric scaling of plant energetics and population density. *Nature*, 395(6698):163–165, 1998.

[75] Simon Lax, Clare I Abreu, and Jeff Gore. Higher temperatures generically favour slower-growing bacterial species in multispecies communities. *Nature ecology & evolution*, 4(4):560–567, 2020.

[76] Christoph Ratzke, Julien Barrere, and Jeff Gore. Strength of species interactions determines biodiversity and stability in microbial communities. *Nature ecology & evolution*, 4(3):376–383, 2020.

[77] Daniel A Power, Richard A Watson, Eörs Szathmáry, Rob Mills, Simon T Powers, C Patrick Doncaster, and BłaŻej Czapp. What can ecosystems learn? expanding evolutionary ecology with learning theory. *Biology direct*, 10:1–24, 2015.

[78] Giulio Biroli, Guy Bunin, and Chiara Cammarota. Marginally stable equilibria in critical ecosystems. *New Journal of Physics*, 20(8):083051, 2018.

[79] Michael Turelli. Niche overlap and invasion of competitors in random environments i. models without demographic stochasticity. *Theoretical Population Biology*, 20(1):1–56, 1981.

[80] Giulia Garcia Lorenzana and Ada Altieri. Well-mixed lotka-volterra model with random strongly competitive interactions. *Physical Review E*, 105(2):024307, 2022.

[81] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[82] Andrea *et al* Banino. Vector-based navigation using grid-like representations in artificial agents. *Nature*, 557(7705):429–433, 2018.

[83] Christopher J. Cueva and Xue-Xin Wei. Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. In *International Conference on Learning Representations*, 2018.

[84] Lisa M Giocomo, Syed A Hussaini, Fan Zheng, Eric R Kandel, May-Britt Moser, and Edvard I Moser. Grid cells use hcn1 channels for spatial scaling. *Cell*, 147(5):1159–1170, 2011.

[85] Lisa M Giocomo, Eric A Zilli, Erik Fransén, and Michael E Hasselmo. Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. *Science*, 315(5819):1719–1722, 2007.

[86] John J Tukker, Prateep Beed, Michael Brecht, Richard Kempter, Edvard I Moser, and Dietmar Schmitz. Microcircuits for spatial coding in the medial entorhinal cortex. *Physiological reviews*, 102(2):653–688, 2022.

[87] James G Heys, Lisa M Giocomo, and Michael E Hasselmo. Cholinergic modulation of the resonance properties of stellate cells in layer ii of medial entorhinal cortex. *Journal of Neurophysiology*, 104(1):258–270, 2010.

[88] John Widloski and Ila R Fiete. A model of grid cell development through spatial exploration and spike time-dependent plasticity. *Neuron*, 83(2):481–495, Jul 2014.

[89] Andrea Banino, Caswell Barry, Benigno Uria, Charles Blundell, Timothy Lillicrap, Piotr Mirowski, Alexander Pritzel, Martin J Chadwick, Thomas Degris, Joseph Modayil, Greg Wayne, Hubert Soyer, Fabio Viola, Brian Zhang, Ross Goroshin, Neil Rabinowitz, Razvan Pascanu, Charlie Beattie, Stig Petersen, Amir Sadik, Stephen Gaffney, Helen King, Koray Kavukcuoglu, Demis Hassabis, Raia Hadsell, and Dharshan Kumaran. Vector-based navigation using grid-like representations in artificial agents. *Nature*, 557(7705):429–433, 05 2018.

[90] Ben Sorscher, Gabriel Mel, Surya Ganguli, and Samuel Ocko. A unified theory for the origin of grid cells through the lens of pattern formation. In *Advances in Neural Information Processing Systems*, pages 10003–10013, 2019.

[91] Rylan Schaeffer, Mikail Khona, and Ila R Fiete. No free lunch from deep learning in neuroscience: A case study through models of the entorhinal-hippocampal circuit. *Advances in neural information processing systems*, 2022.

[92] Xiao-Jing Wang. Theory of the multiregional neocortex: large-scale neural dynamics and distributed cognition. *Annual review of neuroscience*, 45:533–560, 2022.

[93] Guang Chen, Byungwoo Kang, Jack Lindsey, Shaul Druckmann, and Nuo Li. Modularity and robustness of frontal cortical networks. *Cell*, 184(14):3717–3730, 2021.

[94] Nadav Kashtan, Avi E Mayo, Tomer Kalisky, and Uri Alon. An analytically solvable model for rapid evolution of modular structure. *PLoS computational biology*, 5(4):e1000355, 2009.

[95] Giambattista Parascandolo, Niki Kilbertus, Mateo Rojas-Carulla, and Bernhard Schölkopf. Learning independent causal mechanisms. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 4036–4044. PMLR, 2018.

[96] Guangyu Robert Yang, Madhura R Joglekar, H Francis Song, William T Newsome, and Xiao-Jing Wang. Task representations in neural networks trained to perform many cognitive tasks. *Nat. Neurosci.*, 22(2):297–306, 2019.

[97] Jun Sun and Michael W. Deem. Spontaneous emergence of modularity in a model of evolving individuals. *Physical Review Letters*, 99(228107), 2007.

[98] Jacob Andreas, Dan Klein, and Sergey Levine. Modular multitask reinforcement learning with policy sketches. In *International Conference on Machine Learning*, pages 166–175. PMLR, 2017.

[99] Nadav Kashtan and Uri Alon. Spontaneous evolution of modularity and network motifs. *Proc Natl Acad Sci U S A*, 102(39):13773–8, Sep 2005.

[100] Trevor McCourt, Ila R Fiete, and Isaac L Chuang. Noisy dynamical systems evolve error correcting codes and modularity. *arXiv preprint arXiv:2303.14448*, 2023.

[101] Brian A Camley. Collective gradient sensing and chemotaxis: modeling and recent developments. *Journal of Physics: Condensed Matter*, 30(22):223001, 2018.

[102] Carlos S Galvan-Ampudia, Guillaume Cerutti, Jonathan Legrand, Géraldine Brunoud, Raquel Martin-Arevalillo, Romain Azais, Vincent Bayle, Steven Moussu, Christian Wenzl, Yvon Jaillais, et al. Temporal integration of auxin information for the regulation of patterning. *Elife*, 9:e55832, 2020.

[103] Yoram Burak and Ila R Fiete. Fundamental limits on persistent activity in networks of noisy neurons. *Proc Natl Acad Sci U S A*, 109(43):17645–50, Oct 2012.

[104] Avishai Gavish, Arkadi Shwartz, Abraham Weizman, Eyal Schejter, Ben-Zion Shilo, and Naama Barkai. Periodic patterning of the drosophila eye is stabilized by the diffusible activator scabrous. *Nature communications*, 7(1):1–10, 2016.

[105] Mark S Goldman, A Compte, and Xiao-Jing Wang. Neural integrator models. *Encyclopedia of neuroscience*, pages 165–178, 2010.

[106] Sophie Deneve, Peter E Latham, and Alexandre Pouget. Reading population codes: a neural implementation of ideal observers. *Nature neuroscience*, 2(8):740–745, 1999.

[107] Nicolas Brunel. Is cortical connectivity optimized for storing information? *Nature neuroscience*, 19(5):749–755, 2016.

[108] KiJung Yoon, Sam Lewallen, Amina A Kinkhabwala, David W Tank, and Ila R Fiete. Grid cell responses in 1d environments assessed as slices through a 2d lattice. *Neuron*, 89(5):1086–99, Mar 2016.

[109] Kijung Yoon, Michael A Buice, Caswell Barry, Robin Hayman, Neil Burgess, and Ila R Fiete. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nat Neurosci*, 16(8):1077–84, Aug 2013.

[110] J. J. Couey, A. Witoelar, S. Zhang, K. Zheng, J. Ye, B. Dunn, R. Czajkowski, M. Moser, E. Moser, Y. Roudi, and M. Witter. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nat Neurosci*, 16(3):318–24, Mar 2013.

[111] M V Tsodyks and M V Feigelman. The enhanced storage capacity in neural networks with low activity level. *Europhys. Lett.*, 6(2):101–105, May 1988.

[112] A. Gierer and H. Meinhardt. A theory of biological pattern formation. *Kybernetik*, 12(1):30–39, dec 1972.

[113] M. C. Cross and P. C. Hohenberg. Pattern formation outside of equilibrium. *Reviews of Modern Physics*, 65(3), 1993.

[114] James Dickson Murray. *Mathematical Biology*. Number Ch. 16. Springer, Berlin, 2003.

[115] JE Widloski. Grid cell attractor networks: development and implications. UT Ph.D. Theses and Dissertations.

[116] Michael Cross and Henry Greenside. *Pattern formation and dynamics in nonequilibrium systems*. Cambridge University Press, 2009.

[117] Ben Sorscher, Gabriel C Mel, Samuel A Ocko, Lisa M Giocomo, and Surya Ganguli. A unified theory for the computational and mechanistic origins of grid cells. *Neuron*, 2022.

[118] Ernest Oliver Tuck. On positivity of fourier transforms. *Bulletin of the Australian Mathematical Society*, 74(1):133–138, 2006.

[119] BG Giraud and R Peschanski. On positive functions with positive fourier transforms. *arXiv preprint math-ph/0504015*, 2005.

[120] Bertrand G Giraud and Robi Peschanski. On the positivity of fourier transforms. *arXiv preprint arXiv:1405.3155*, 2014.

[121] Note that in our convention the module number m is ordered such that the largest grid period module is the first module. This is opposite to the numbering usually used in the literature, such as in [24].

[122] Julija Krupic, Marius Bauza, Stephen Burton, Caswell Barry, and John O'Keefe. Grid cell symmetry is shaped by environmental geometry. *Nature*, 518(7538):232–235, 2015.

[123] Tor Stensola, Hanne Stensola, May-Britt Moser, and Edvard I Moser. Shearing-induced asymmetry in entorhinal grid cells. *Nature*, 518(7538):207–212, 2015.

[124] Guifen Chen, Daniel Manson, Francesca Cacucci, and Thomas Joseph Wills. Absence of visual input results in the disruption of grid cell firing in the mouse. *Current Biology*, 26(17):2335–2342, 2016.
[125] John J Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc Natl Acad Sci U S A*, 81:3088–3092, May 1984.
[126] Rishidev Chaudhuri, Alberto Bernacchia, and Xiao-Jing Wang. A diversity of localized timescales in network activity. *Elife*, 3:e01239, 2014.

MATERIALS AND METHODS

We use a continuous attractor network (CAN) model [105–107] for grid cells [17, 18, 108, 109], with neural dynamics obeying

$$\frac{\partial s(i,t)}{\partial t} + \frac{s(i,t)}{\tau} = \phi\left[\sum_j W_0(i,j)s(j,t) + B(i,t)\right], \tag{9}$$

where $s(i,t)$ represents the synaptic activation of neuron i at time t, $W_0(i,j)$ represents the synaptic strength of the coupling from neuron j to neuron i, $B(i,t)$ represents the feed-forward bias to neuron i, and ϕ is a non-decreasing nonlinearity, for which we use the rectification function ($\phi(z) = [z]_+ = z$ for $z > 0$ and 0 otherwise). Each neuron i has a preferred direction θ_i that is used to perform velocity integration. In the one-dimensional version of our setup, each spatial location $\mathbf{x}$ on the neural sheet has two neurons, with preferred directions $\theta = 0$ and $\theta = \pi$. Correspondingly, in the two-dimensional version of our setup, each location on the neural sheet has four neurons, with preferred directions $\theta = n\pi/4$ for $n \in \{0, 1, 2, 3\}$. The synaptic weights $W_0(i,j)$ are defined via an interaction kernel $W(\Delta x)$ such that

$$W_0(i,j) = W(|\mathbf{x}_i - \mathbf{x}_j - \mathbf{\Delta s}(\theta_j)|), \tag{10}$$

where $\mathbf{x}_i$ represents the spatial location of neuron i, and $\mathbf{l}(\theta)$ is a vector with length Δs oriented parallel to the angle θ. The feed-forward bias $B(i,t;\theta)$ is given by

$$B(i,t) = b + b_{vel}|\mathbf{v}|\cos(\theta_\mathbf{i} - \psi), \tag{11}$$

where ψ is the direction of the input velocity signal and $|\mathbf{v}|$ is the speed. This results in neurons with direction preference θ driving activity in the network towards the direction of their outgoing weight shifts $\mathbf{\Delta s}(\theta)$. This mechanism is responsible for velocity integration by the network [18].

We first described the dynamics under fixed arbitrary kernels, demonstrating that they result in hexagonal pattern formation. These arbitrary kernels were constructed by interpolating between random points via the following protocol: First, we construct 'x-values' by considering $n + n_{zero}$ uniformly spaced points from $-L$ to L, which are then perturbed by the addition of a randomly sampled number from $-L/4n$ to $L/4n$ (this perturbation makes the points less regular, while disallowing consecutive points to be extremely close to each other). Second, we construct n 'y-values' sampled from a uniform distribution from -1 to 1, and define the remaining n_{zero} y-values to be 0 (the n_{zero} values at zero ensure that the interpolated function decays to zero). Then, a cubic spline interpolation (top row of Fig. 10a) or a linear interpolation (bottom row of Fig. 10a) is performed between the y-values and the x-values to generate an arbitrary function $\omega(x)$. This generated function is however not symmetric, as is required for kernel functions — thus, we construct the interaction kernel as $W(\Delta x) = \omega(\Delta x) + \omega(-\Delta x)$. Kernels whose dynamics lead to infinitely diverging firing rates are rejected and resampled. These kernels were simulated on sheets with 256×256 neurons with aperiodic boundary conditions[18]. n was randomly chosen between $2, 3$ or 4, and L was scaled as necessary to obtain a large number of activity bumps on the sheet to prevent finite-size effects from distorting the hexagonal lattice of activity.

For the case of module formation through peak selection, the interaction weight kernel W is given by the sum of two components $W = W^g_{n_{DV}} + W^f$. The first, $W^g_{n_{DV}}$ drives local pattern formation, and has a spatial scale $\sigma(n_{DV})$, which varies smoothly in a gradient along the dorso-ventral axis, and the second, W^f has a fixed spatial scale d everywhere on the neural sheet. A variety of functions $W^g_\mathbf{x}$ can drive local pattern formation. For concreteness, we use two specific examples: the Mexican-hat profile[18] (used in Figs. 2a-c,j,k, 11 and SI Fig. 12)

$$W^g_{\text{mexican-hat}}(\Delta x) = \alpha_E \exp\left[-\gamma\frac{(\Delta x)^2}{2\sigma_{mh}(n_{DV})^2}\right] - \alpha_I \exp\left[-\frac{(\Delta x)^2}{2\sigma_{mh}(n_{DV})^2}\right], \tag{12}$$

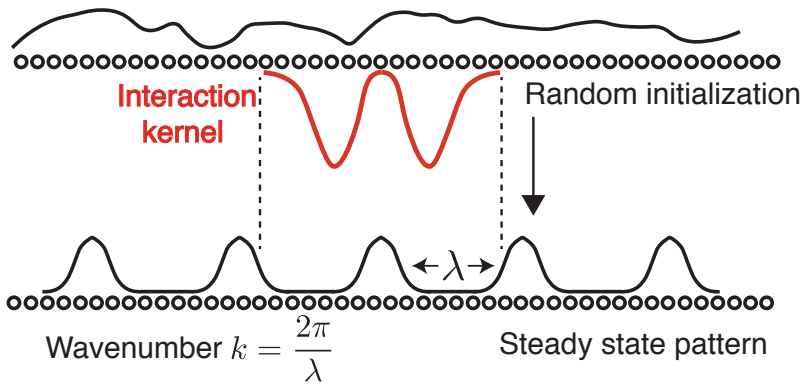


FIG. 9. Local pattern formation in continuous attractor models of grid cells: Through local amplification of random fluctuations, the lateral interaction forms periodic patterns.

and the box-function profile[110] (used in Fig. 2 and SI Fig. 12)

$$W^g_{\text{box}}(\Delta x) = \alpha_0 \times \mathbb{1}_{|\Delta x|<\sigma_b(\mathbf{x})} = \begin{cases} \alpha_0 & \text{if } |\Delta x| < \sigma_b(n_{DV}), \\ 0 & \text{if } |\Delta x| \geq \sigma_b(n_{DV}). \end{cases} \tag{13}$$

For the fixed-width interaction $W^f(\Delta x)$, we implement 3 main types — localized (used in Figs. 2,3 and SI Fig. 12), diffuse (used in Fig. 2 and SI Fig. 12) and decaying (used in SI Fig. 12).

$$\begin{aligned} W^f_{\text{localized}}(\Delta x) &= \alpha_S \exp\left[-\frac{(|\Delta x| - d_{loc})^2}{2\epsilon_S^2}\right], \\ W^f_{\text{diffuse}}(\Delta x) &= \alpha_1 \times \mathbb{1}_{|\Delta x|<d_{dif}}, \\ W^f_{\text{decaying}}(\Delta x) &= \alpha_T \times [d_{dec} - |\Delta x|]_+. \end{aligned}$$

In particular,

- In Figs. 11 we use only a smoothly varying Mexican-hat pattern forming kernel $W = W^g_{\text{mexican-hat}}$
- In Figs. 2a-c,g, j,k we use $W = W^g_{\text{mexican-hat}} + W^f_{\text{localized}}$
- In Figs. 2d-f,h we use a 'Lincoln hat' profile $W = W^g_{\text{box}} + W^f_{\text{diffuse}}$
- and, in SI Fig. 12 we present numerical simulations of other combinations of pattern forming and fixed-scale kernels.

To construct spatially heterogeneous kernels for analyzing the robustness to inhomogeneity in Fig. 6 we use the box function to construct

$$\begin{aligned} W^g_\xi[\mathbf{x}, \mathbf{x}'] &= \alpha_0 \times \mathbb{1}_{|\mathbf{x}-\mathbf{x}'|\times(1+\xi^g_1(\mathbf{x}'))<\sigma^0_b(1+\xi^g_2(\mathbf{x}'))}, \\ W^f_\xi[\mathbf{x}, \mathbf{x}'] &= \alpha_1 \times \mathbb{1}_{|\mathbf{x}-\mathbf{x}'|\times(1+\xi^f_1(\mathbf{x}'))<d_{dif}(1+\xi^f_2(\mathbf{x}'))}, \end{aligned}$$

where $\xi^{g,f}_{1,2}(\mathbf{x}')$ are independent random numbers chosen uniformly from $\epsilon^{g,f}_{1,2} \times [-1, 1]$. For the particulars of Fig. 6b, $\epsilon^g_2 = 0.1$ and other noise terms are set to zero (In the one-dimensional case ϵ_1 and ϵ_2 have the same effect); for Fig. 6c, $\epsilon^g_1 = 0.2$, $\epsilon^g_2 = 0.3$, $\epsilon^f_1 = 0.05$ and $\epsilon^f_2 = 0$ (See SI Sec. D 9 for more details).

In Table I we present a list of common parameters used across all numerical grid-cell simulations. Then, in Tables II,III we present the parameter values used for the kernels used in our numerical simulations

Parameter	Value
τ	30
dt	0.05
b	$\begin{cases} 70 & \text{in 1D} \\ 1 & \text{in 2D} \end{cases}$
b_{vel}	$\begin{cases} 105 & \text{in 1D} \\ 1 & \text{in 2D} \end{cases}$
Δs	2

TABLE I. Parameters held constant across all numerical simulations

$W^g_{\text{mexican-hat}}$ parameters	Value
α_E	1000
α_I	1000
γ	1.05
N^{1D}	3000
N_y^{2D}	100
N_x^{2D}	1000
$\sigma_{mh}(n_{DV})$	$1/\sqrt{2\beta(n_{DV})}$
$\beta(n_{DV})$	$\beta_0 + (\beta_1 - \beta_0)n_{DV}/N'$
N'	$\begin{cases} N^{1D} & \text{in 1D} \\ N_x^{2D} & \text{in 2D} \end{cases}$
β_0	$\begin{cases} 2.5 \times 10^{-2} & \text{in 1D} \\ 3/676 & \text{in 2D} \end{cases}$
β_1	$\begin{cases} 2.5 \times 10^{-1} & \text{in 1D} \\ 9/338 & \text{in 2D} \end{cases}$
W^g_{box} parameters	Value
N^{1D}	5000
α_0	-40
$\sigma_b(n_{DV})$	$15 + 30n_{DV}/N$

TABLE II. Pattern forming kernel parameters used for numerical simulations

For the modular niche formation, we consider the setup as described in Eq. 8, with $N = 1000$ species, each characterized by a random $M = 2000$ dimensional random feature vector indicating resource preference. We numerically simulate our setup on a discrete lattice $x \in \{0, L\}$ for $L = 300$ in Fig. 8i left, and $L = 500$ otherwise.We instantiate the nonlinearity H as a shifted Heaviside function, $H[x] = 1$ for $x \geq 0.5$, and $H[x] = 0$ otherwise, and choose the death rate $\delta_i = 0.1$ for all species. To construct $W(i, j)$ as an interaction matrix that quantifies the cooperation and competition between species, we follow a set up similar to a Hopfield model with $\{0, 1\}$ activity[111]. We first choose a set of random points in N-dimensional species space $\mathbf{s}_q$ for $q \in \{1, \cdots, Q\}$, denoting *potential* niches. We choose Q

$W^f_{\text{localized}}$ parameters	Value
α_S	4
d_{loc}	$\begin{cases} 84 & \text{in 1D} \\ 50 & \text{in 2D} \end{cases}$
ϵ_S	$\begin{cases} 4.77 & \text{in 1D} \\ 1.6 & \text{in 2D} \end{cases}$
W^f_{diffuse} parameters	Value
α_{dif}	-0.25
d_{dif}	135
W^f_{decaying} parameters	Value
α_T	25
d_{dec}	150

TABLE III. Fixed-scale kernel parameters used for numerical simulations

such that $1 \ll Q \ll N$. Each $\mathbf{s}_q$ vectors consists of a $+1$ at elements corresponding to species that may co-exist, and -1 otherwise. In practice, we draw each element uniformly from the set $\{0, 1\}$, constructing an $N \times Q$ matrix. The weight matrix $W(i, j)$ is then constructed as

$$W(i,j) = 0.015 \times \begin{cases} \sum_{i=1}^{Q} s_q(i)s_q(j), & \text{for } i \neq j \\ c, & \text{for } i = j \end{cases},$$

where c is a positive constant set to 10.

The spatial interaction kernel $K(x, x') = K(x - x')$ is chosen to be a Gaussian function with standard deviation 1.75 (which is much smaller than the entire spatial extent of the system, L). The end points of the resource gradient are chosen as two random $M = 2000$ dimensional vectors with elements draw independently from i.i.d. Gaussian distributions with zero mean and standard deviation $2/N$, and the preference vectors $\mathbf{b}_i$ are drawn from i.i.d. Gaussian distributions with zero mean and unit standard deviation.

The initial condition for the simulation is set to be the uniform state $s(i) = 0.5$ for all i, and the simulation is run until the dynamics reach a fixed point state. The final formed fixed point state is examined by calculating the correlation matrix

$$C(x, x') = \sum_{i=1}^{N} \cap s^*(i, x) \cap s^*(i, x')$$

where

$$\cap s(i) = [s(i) - \langle s \rangle] / \sqrt{\sum_i [s(i) - \langle s \rangle]^2}$$

where $\langle s \rangle = (1/N) \sum_j s(j)$.

SUPPLEMENTARY TEXT

The supplemental information is structured as follows: First, in SI Sec. A we present the mathematical analysis for pattern formation, and generalize the theory of CAN models of grid cells to show analytically and numerically that an infinite set of local interaction kernels can generate a grid cell network, as shown in Fig. 2 and Fig. 10. Second, we demonstrate analytically and numerically in SI Sec. B that simply introducing a gradient in the pattern forming kernel of the continuous attractor model is *not* sufficient to result in modularization, as demonstrated in Fig. 1 of the main text. Third, in Sec. B 1 we show how the addition of a Gaussian localized kernel results in self-organized modularization. Fourth, we show in Sec. D that among arbitrary kernels, those with simple shapes result in a simple equation describing the detailed period ratios of the formed grid modules as shown in Fig. 4. Fifth, this will lead to simple estimates for the number of modules and their sizes in terms of other system parameters, which we derive in SI Sec. D 7. Sixth, after having described our results primarily for the case of one-dimensional grid cells, we then demonstrate in Sec. D 8 that our arguments extend naturally to two dimensions, and we present numerical results demonstrating the same. Seventh, in SI Sec. D 10 we then demonstrate that our results and predictions of grid period ratios are consistent with available data sources to a large extent. Finally, we generalize our result to the context of dynamics on a rough energy landscape (SI Sec. E), and provide broader perspectives of our results in the contexts of general loss optimization (Sec. F) and eigenvector localization (SI Sec. G).

Appendix A: Generalization of grid cell CAN dynamics theory: infinite set of interactions produce grid cells

Existing continuous attractor grid cell models use one of two specific interaction profiles: a center excitation-surround inhibition (Mexican hat) shape which for stability has been implemented as a purely inhibitory donut-shaped surround [18], or a limited-radius uniform inhibition shape (which we term a "Lincoln hat") [48].

Here we extend these results by defining an infinite set of distinct interaction profiles that can give rise to grid cell-like patterning. Consider the standard equations for the dynamics of recurrently connected neurons (expressed for notational simplicity in the continuum or large neural number limit):

$$\frac{\partial s(\mathbf{x}, t)}{\partial t} + \frac{s(\mathbf{x}, t)}{\tau} = \phi \left[\int_{-\infty}^{+\infty} W(\mathbf{x}, \mathbf{x}') s(\mathbf{x}', t) d\mathbf{x}' + B \right], \tag{A1}$$

where $s(\mathbf{x})$ is the synaptic activation of the neuron at the vector position $\mathbf{x}$ on a 2-dimensional neural sheet, $W(\mathbf{x}, \mathbf{x}')$ is coupling strength from a neuron at $\mathbf{x}'$ to a neuron at $\mathbf{x}$, τ is the biophysical time-constant of individual neurons, ϕ is a non-negative monotonic transfer function, and B is a uniform feedforward input to all neurons. Further, we assume that $\int W(\mathbf{x}, \mathbf{x}')d\mathbf{x} < 0$, consistent with experiments suggesting that grid pattern formation is dominated by recurrent inhibitory circuitry[110]. In this continuum limit, $s(\mathbf{x})$ and $W(\mathbf{x}, \mathbf{x}')$ are continuous functions of $\mathbf{x}$ and $\mathbf{x}'$. For simplicity, the neural nonlinearity is given by the rectification function ($\phi(z) = [z]_+ = z$ for $z > 0$ and 0 otherwise. This choice is not crucial, as we will see later.

The interaction weights $W(\mathbf{x}, \mathbf{x}')$ are assumed to possess a continuous translation symmetry, $W(\mathbf{x}, \mathbf{x}') \equiv W(|\mathbf{x} - \mathbf{x}'|; \sigma)$ where σ is the characteristic length-scale of the function (kernel) W. In words, the weight between two neurons will depend only on the separation between them (not on their absolute locations). This condition yields activity states that are themselves translationally invariant. In this network, the state with roughly uniform activity becomes unstable and evolves to a periodic activity pattern via a linear (Turing) instability, when two conditions hold on the Fourier transform of the interaction weights $W(\mathbf{x})$ [14, 18, 112–115]. Since $W(\mathbf{x}, \mathbf{x}') = W(|\mathbf{x} - \mathbf{x}'|)$, it follows that $\tilde{W}(\mathbf{k}) = \tilde{W}(|\mathbf{k}|) = \tilde{W}(k)$, and we can state these conditions in terms of the radial part of the Fourier Transform, k. First, the maximum of $\tilde{W}(k)$ should fall at $k \neq 0$, i.e.,

$$k^* \equiv \arg\max_k \tilde{W}(k) > 0. \tag{A2}$$

Second, this maximum must be sufficiently large, i.e.,

$$\tilde{W}(k^*) > 1/\tau. \tag{A3}$$

The emergent activity patterns have a periodicity of $2\pi/k^*$ [14, 18, 112–115], which scales as σ, the characteristic width of the interaction kernel W. The non-negativity of the transfer function results in patterns with *hexagonal* rather than other symmetries [113, 116, 117]. Thus, interaction kernels that satisfy the two conditions above will give rise grid-like patterning.

It is known that Mexican hat-like kernels [18] and Lincoln hat-style kernels [48] generate grid patterning. We next provide a general way to explain why, and to determine other kernel shapes that are consistent with grid emergence through continuous attractor dynamics. The key for grid-like patterning are lateral interactions W whose Fourier transforms are not everywhere negative, and are not only non-negative at $k = 0$. In fact, an infinite set of kernel functions W satisfy these conditions, with rare exceptions. First some exceptions: The Fourier transforms of Gaussian and Lorentzian functions have either a single peak at $k = 0$ (when the functions are positive), or are negative everywhere (when the functions are negative), and thus fail to satisfy the first condition. However, as we argue in Sec. C, making small perturbations to the kernels results in the conditions being satisfied.

We next performed numerical experiments to test the hypothesis that randomly generated functions will generically have Fourier Transforms that are not negative everywhere or only non-negative at 0, and therefore might generate grid-like patterning (see Methods for details of random sampling of kernel functions). We found that indeed randomly constructed kernel functions satisfied the hypothesized property for their Fourier transforms: we generated 10^6 random localized kernel functions, and all of these satisfied the conditions of being not negative everywhere or being non-negative only at $k = 0$ (SI Fig.10). We further found that these kernel functions, under the further condition that they did not produce diverging neural activity, generated hexagonal patterns. Some of these are shown in Fig. 2a. In sum, an infinite set of local interaction profiles will generate grid cell-like activation patterns. Such candidate profiles can be generated at random and with very high probability generate grid-like patterning.

Motivated by the experimental observations described in the main text, we modify the Mexican-hat function to introduce a smooth gradient in the characteristic interaction widths σ_E, σ_I.

$$W^g_{n_{DV}}(\mathbf{\Delta x}) = \alpha_E \exp\left(-\frac{\mathbf{\Delta x}^2}{2\sigma_E(n_{DV})}\right) - \alpha_I \exp\left(-\frac{\mathbf{\Delta x}^2}{2\sigma_I(n_{DV})}\right), \tag{A4}$$

where $\sigma_E(n_{DV})$ and $\sigma_I(n_{DV})$ are now functions that depend on position in the neural sheet, and encode the smoothly varying characteristic scale of the Mexican-hat interaction along the dorso-ventral axis:

$$\sigma_{E/I}(n_{DV}) = \sigma_{E/I} + \sigma'_{E/I}(0) \cdot n_{DV}. \tag{A5}$$

For such graded kernels, we will use $W(\mathbf{x}, \mathbf{x}')$ and $W_{n_{DV}}(\mathbf{x} - \mathbf{x}') = W_{n_{DV}}(\mathbf{\Delta x})$ interchangeably. In this case, Eq. A1 then becomes

$$\frac{\partial s(\mathbf{x}, t)}{\partial t} + \frac{s(\mathbf{x}, t)}{\tau} = \phi\left[\int_{-\infty}^{+\infty} W_{n_{DV}}(\mathbf{x} - \mathbf{x}')s(\mathbf{x}', t)d\mathbf{x}' + B(\mathbf{x})\right], \tag{A6}$$

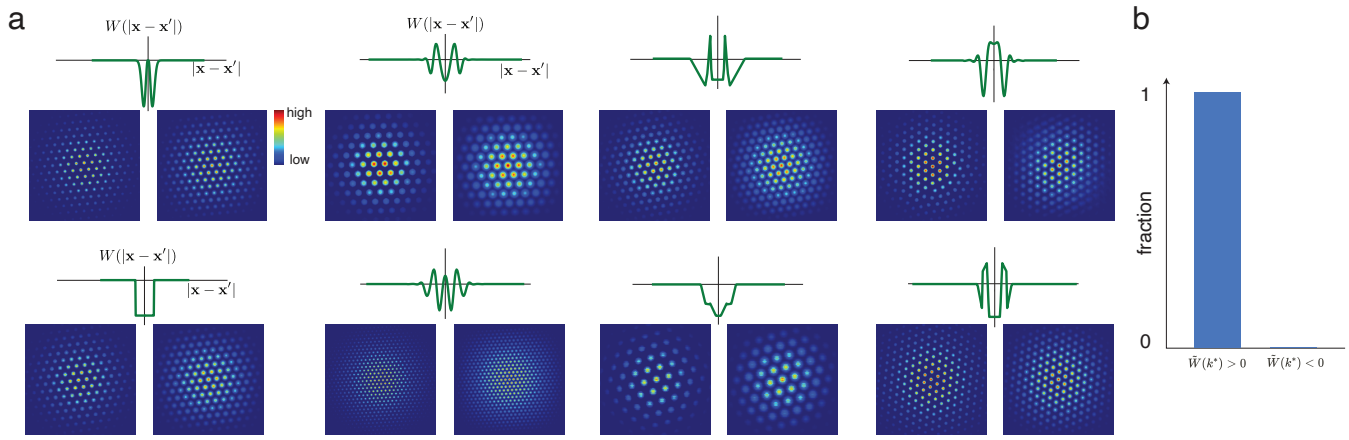


FIG. 10. All kernels satisfying the conditions laid out in the main text can result in pattern formation, with appropriate scaling.

Appendix B: Pattern formation with graded kernels

Under this approximation, we perform a linear stability analysis of the neural dynamics, to identify the the growing periodic modes locally at the position on the neural sheet n_{DV}.

We first identify an unstable steady-state solution to Eq. (A6), which we denote as $s_0(\mathbf{x})$. This solution satisfies

$$\frac{s(n_{DV})}{\tau} = \phi\left[\int_{-\infty}^{+\infty} W_{n_{DV}}(\mathbf{x}-\mathbf{x}')s_0(\mathbf{x}')d\mathbf{x}' + B(\mathbf{x})\right]. \tag{B1}$$

In the limit of very slowly varying changes in $W_{n_{DV}}(\mathbf{\Delta x})$ as a function of n_{DV}, the unstable steady state solution will be

$$s_0(n_{DV}) = \frac{\tau\bar{B}}{1-\tau\bar{W}}, \tag{B2}$$

where $\bar{B} = \int B(\mathbf{x})d\mathbf{x}$ and $\bar{W} = \int W_{n_{DV}}(\mathbf{x}-\mathbf{x}')d\mathbf{x}'$. (For $\tau\bar{W}_{\mathbf{x}} > 1$, the only locally homogeneous steady state is $s_0(n_{DV}) = 0$ due to the rectifying nonlinearity, which as we justify shortly cannot support periodic pattern formation due to being a stable fixed point).

We then consider a perturbative analysis, by examining the evolution of $s(\mathbf{x},t) = s_0(n_{DV}) + \epsilon(\mathbf{x},t)$. We apply our analysis to the early time evolution of this initial condition, such that $\epsilon(\mathbf{x},t) \ll s_0(n_{DV})$. Inserting our form of $s(\mathbf{x},t)$ in Eq. (A6), we obtain

$$\frac{\partial\epsilon(\mathbf{x},t)}{\partial t} + \frac{\epsilon(\mathbf{x},t)}{\tau} = \phi'(\bar{W}s_0(n_{DV}) + \bar{B})\int_{-\infty}^{\infty} W_{n_{DV}}(\mathbf{x}-\mathbf{x}')\epsilon(\mathbf{x}',t)d\mathbf{x}'. \tag{B3}$$

Since $W_{n_{DV}}(\mathbf{x}-\mathbf{x}')$ is a local kernel, we approximate the above integral with one evaluated over the region $\{\mathbf{x}' : |\mathbf{x}-\mathbf{x}'| < l\}$, with l much larger than the length-scale of the kernel $W_{n_{DV}}$ at all $\mathbf{x}$. Over this interval, we posit that $\epsilon(\mathbf{x}',t) = \epsilon e^{i\mathbf{k}\cdot\mathbf{x}'+\alpha(\mathbf{k})t}$, where $\alpha(\mathbf{k})$ denotes the growth rate of this ϵ perturbation. Inserting this form into Eq. (B3) yields,

$$\alpha(\mathbf{k}) + 1/\tau = \phi'(\bar{W}s_0(n_{DV}) + \bar{B})\int_{\infty}^{\infty} W_{n_{DV}}(\mathbf{x}-\mathbf{x}')e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}d\mathbf{x}', \tag{B4}$$

$$= \phi'[\bar{W}s_0(n_{DV}) + \bar{B}]\mathcal{F}[W_{n_{DV}}(\mathbf{x}-\mathbf{x}')], \tag{B5}$$

$$= \phi'[\bar{W}s_0(n_{DV}) + \bar{B}]\mathcal{F}W_{n_{DV}}(\mathbf{k}) \tag{B6}$$

where $\mathcal{F}[W_{n_{DV}}(\mathbf{x}-\mathbf{x}')] = \mathcal{F}W_{n_{DV}}(\mathbf{k})$ is the Fourier transform of the interaction kernel corresponding to position n_{DV} on the neural sheet. For the rectifying nonlinearity $\phi' = 1$, and the requirement for the periodic perturbation to be growing is $\alpha(\mathbf{k}) = \mathcal{F}W_{n_{DV}}(\mathbf{k}) - 1/\tau > 0$.

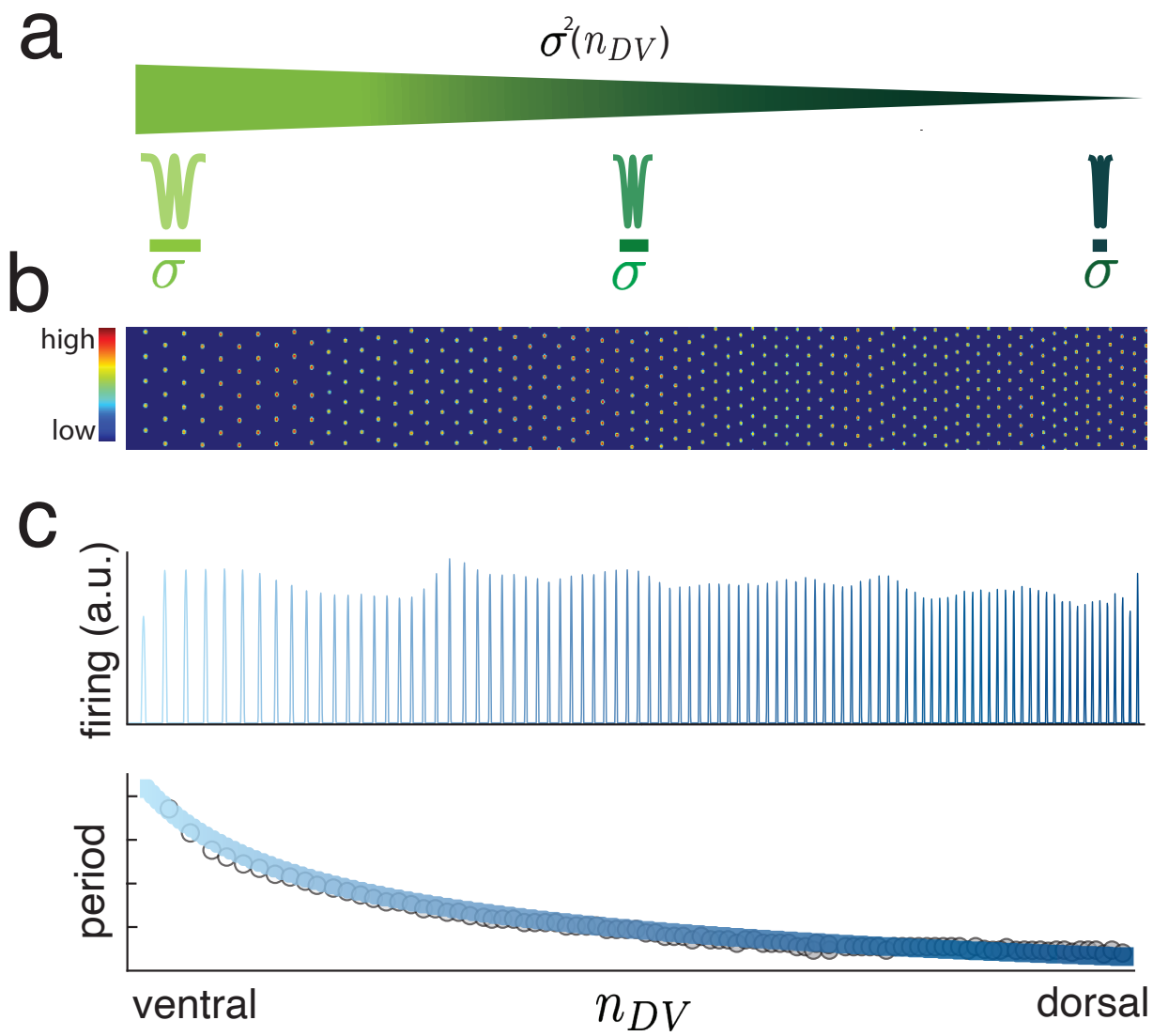


FIG. 11. (a-c) Naive merger of the two mechanisms by smoothly scaling the width of the pattern-forming lateral interaction (j) in the grid cell CAN model [45] does not generate global modularity in 2-dimensional (b) or 1-dimensional (c) grid models: the result is one smoothly varying periodic pattern.

Note that since $W_{n_{DV}}(\mathbf{\Delta x})$ is a kernel, it is a radially-symmetric real function, and hence the Fourier transform $\mathcal{F}W_{n_{DV}}(\mathbf{k})$ will also be real function that is radially-symmetric in k. Thus, for simplicity, we will only focus on the magnitude of $\mathbf{k}$, which we denote as $k = |\mathbf{k}| \geq 0$ (In this context, for the two-dimensional case, one may re-interpret the radial component of the Fourier transform of $W_{n_{DV}}(\mathbf{\Delta x})$ as the Hankel transform of $W_{n_{DV}}(|\mathbf{\Delta x}|)$).

By definition, the magnitude of the wave vector k^* that corresponds to the fastest growing mode locally around position $\mathbf{x}$ on the neural sheet will be the $\mathbf{k}$ that maximizes $\alpha(\mathbf{k})$. Under the approximation of slow changes in the length-scale of the interaction kernel $W_{n_{DV}}(\mathbf{\Delta x})$, we see from Eq. (B6) that

$$k^*(n_{DV}) = \arg\max_k \mathcal{F}W_{n_{DV}}(k), \tag{B7}$$

since $W_{n_{DV}}(\mathbf{\Delta x})$ (and hence $s_0(n_{DV})$) has been assumed to have a negligible dependence on n_{DV}.

For $W_{n_{DV}}(\mathbf{\Delta x})$ given by Eq. (A4), i.e., without any additional fixed-scale interaction, we obtain from Eq. (B7)

$$[k^*(n_{DV})]^2 = \frac{2}{\sigma_E(n_{DV})^2 - \sigma_I(n_{DV})^2} \log\left(\frac{\alpha_E \sigma_E(n_{DV})^3}{\alpha_I \sigma_I(n_{DV})^3}\right). \tag{B8}$$

If we assume that $\sigma_{E/I}(\mathbf{x}) = \eta_{E/I}\sigma(n_{DV})$, where η_E and η_I are $\mathbf{x}$-independent constants, then we obtain

$$k^*(n_{DV}) \propto 1/\sigma(n_{DV}), \tag{B9}$$

and hence

$$\lambda^*(n_{DV}) \propto \sigma(n_{DV}), \tag{B10}$$

where $\lambda^*(n_{DV})$ is the periodicity of the grid pattern formed locally around position n_{DV}. This results in a smooth change of grid period, corresponding to the observation in Fig. 1g of the main text.

Note that this result is generally true for any pattern forming kernel $W^g_{n_{DV}}(\mathbf{\Delta x})$ that has a Fourier transform with at least one local maximum, and *does not* rely on the specific form of a Mexican-hat interaction. Indeed, Eq. (B10) holds for any kernel $W^g_{n_{DV}}(\Delta x)$ that depends on a length-scale $\sigma(n_{DV})$. As an example, we present the corresponding analysis for the box-shaped kernel employed for pattern formation in Ref. [110].

In this case

$$W^g_{n_{DV}}(\mathbf{\Delta x}) = -W_0 \mathbb{1}_{\mathbf{\Delta x} \leq \sigma(n_{DV})}. \tag{B11}$$

As discussed above, the quantity of interest is $\mathcal{F}W^g_{n_{DV}}(\mathbf{k})$

$$\mathcal{F}W^g_{n_{DV}}(\mathbf{k}) = \int_{-\infty}^{\infty} -W_0 \mathbb{1}_{|\mathbf{x}| \leq \sigma(n_{DV})} e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x} \tag{B12}$$

$$= -W_0 \int_{|\mathbf{x}| \leq \sigma(n_{DV})} e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x}. \tag{B13}$$

The above integral can be calculated in a one-dimensional setup to obtain

$$\mathcal{F}W^g_{n_{DV}}(k) = -2W_0 \frac{\sin(k\sigma(n_{DV}))}{k} \tag{B14}$$

and can be calculated in a two-dimensional setup to obtain

$$\mathcal{F}W^g_{n_{DV}}(k) = -2\pi W_0 \sigma(n_{DV}) \frac{J_1(k\sigma(n_{DV}))}{k}. \tag{B15}$$

In both of the above cases, note that $k^* \propto 1/\sigma(n_{DV})$ since $\sigma(n_{DV})$ is the only length-scale characterizing the kernel $W^g_{n_{DV}}$. In particular, numerical maximization yields

$$k^* \approx \begin{cases} 4.493/\sigma(n_{DV}) & \text{on a one-dimensional sheet, and} \\ 5.136/\sigma(n_{DV}) & \text{on a two-dimensional sheet.} \end{cases} \tag{B16}$$

1. Fixed-scale interactions and modularization

We now claim that the addition of a fixed-scale kernel, $W^f(\mathbf{\Delta x})$ is sufficient to result in modularization of grid periods, with discrete changes in grid period as a function of spatial position along the dorso-ventral axis. This set of interactions can effectively be implemented by two populations of interneurons - one with fixed arborization and weaker synaptic connections and one with varying arborization length and stronger synaptic connections.

For simplicity, we shall present the specific Fourier transform computations for the one-dimensional problem, although we note that all of the qualitative results hold in two dimensions as well, with the Fourier transforms of the relevant functions replaced with their Hankel transforms (as shown in Sec. D 8).

We include an additional weak interaction term W^f that critically does *not* depend on the neural sheet position x. For reasons that will become apparent soon, we choose kernels $W^f(\Delta x)$ such that the Fourier transform changes sign a sufficiently large number of times. We hypothesize that this requirement is not particularly restrictive, and will demonstrate that this holds for most kernels W^f.

The entire interaction profile is then given by

$$W_{n_{DV}}(\Delta x) = W^g_{n_{DV}}(\Delta x) + W^f(\Delta x). \tag{B17}$$

We first demonstrate our result with an example of a simple kernel, to justify how Eq. (B7) leads to the emergence of discrete grid modules. Consider the localized excitatory interaction

$$W^f(\Delta x) = \alpha_S \exp\left(-\frac{(\Delta x - d)^2}{2\epsilon_S^2}\right) + \alpha_S \exp\left(-\frac{(\Delta x + d)^2}{2\epsilon_S^2}\right). \tag{B18}$$

Corresponding to our interpretation of $W^f(\Delta x)$ above being a localized kernel, we choose $\epsilon_S \ll d$.

This choice of $W_{n_{DV}}(\Delta x) = W^g_{n_{DV}}(\Delta x) + W^f(\Delta x)$ leads to the the Fourier transform,

$$\begin{aligned}
\mathcal{F}W_{n_{DV}}(k) &= \mathcal{F}W^g_{n_{DV}}(k) + \mathcal{F}W^f(k), && \text{(B19)}\\
&= \sqrt{2\pi}\left[\alpha_E\sigma_E(n_{DV})\exp\left(-\frac{\sigma_E(n_{DV})^2k^2}{2}\right)\right.\\
&\quad - \alpha_I\sigma_I(n_{DV})\exp\left(-\frac{\sigma_I(n_{DV})^2k^2}{2}\right)\\
&\quad \left.+2\alpha_S\epsilon_S\cos(kd)\exp\left(-\frac{\epsilon_S^2k^2}{2}\right)\right]. && \text{(B20)}
\end{aligned}$$

In our model, the magnitude of the $W^f(\Delta x)$, i.e., α_S, is chosen to be smaller than the magnitude of the Mexican-hat interaction. Thus we interpret $\mathcal{F}W^f(k)$ in Eq. (B20) as being a small perturbation to the Fourier transform of the standard Mexican-hat interaction, $\mathcal{F}W^g_{n_{DV}}(k)$. Further, since d is assumed to be much larger than the scale of the Mexican-hat, $\sigma_{E/I}$, then the term $\cos(kd)$ in $\mathcal{F}W^f(k)$ oscillates at a k-scale much smaller than the relevant scales of $\mathcal{F}W^g_{n_{DV}}(k)$ (see Fig. 3b-c of the main text). Additionally, since $\epsilon_S \ll d$, the gaussian envelope multiplying the rapidly oscillating term has a scale $1/\epsilon$, which is much larger than the periodicity $1/d$.

Thus, in k-space, the rapidly oscillating term, $\mathcal{F}W^f(k)$ can be thought of as predefining a set $S = \{k_1, k_2, \ldots\}$ of local maxima. Under the approximations made above, the addition of the smoother function $\mathcal{F}W^g_{n_{DV}}(k)$, will not change the position of the local maxima. This results in the *local* maxima of $\mathcal{F}W_{n_{DV}}(k)$ also being the same set S. Importantly, we note that since S was predefined purely via $\mathcal{F}W^f(k)$, *there is no n_{DV} dependence on the set S.*

Following Eq. (B7), the wave-vector corresponding to the pattern formation at point x on the neural sheet corresponds to the *global* maxima of $\mathcal{F}W_{n_{DV}}(k)$. Thus, at all points, the pattern formation corresponds to one of the discrete set of choices of wave vectors, $S = \{k_1, k_2 \ldots\}$. As can be seen from Fig. 3c, the smoothly varying gradient in the Mexican-hat term, $\mathcal{F}W^g_{n_{DV}}$ as a function of x picks different choices of k_i depending on the position n_{DV} — the $k \in S$ that is nearest to the maxima of $\mathcal{F}W^g_{n_{DV}}(k)$ will be chosen as the global maxima, and will be the wave vector corresponding to the pattern at n_{DV}. We refer to this mechanism as "peak selection".

For our particular choice of $W^f(x)$ made in Eq. (B18), we obtained

$$\mathcal{F}W^f(k) = 2\alpha_S\epsilon_S\cos(kd)\exp\left(-\frac{\epsilon_S^2k^2}{2}\right). \tag{B21}$$

We can then approximate the local maxima of $\mathcal{F}W^f(k)$ as occurring at

$$S = \left\{\frac{2m\pi}{d}\middle| m \in \mathbb{Z}^+\right\}. \tag{B22}$$

This immediately indicates that the ratios of periods of successive grid modules will be given by

$$\frac{\lambda_{m+1}}{\lambda_m} = \frac{m+1}{m}. \tag{B23}$$

Thus, the addition of a fixed-scale interaction, W^f such as Eq. (B18) results in discrete grid modules. We now show that this peak-selection mechanism, and hence modularization, occurs for arbitrary choices of the fixed-scale interaction kernel $W^f(\Delta x)$.

Appendix C: Kernels that lead to modularization

The peak-selection modularization mechanism described above arises naturally from the presence of the rapidly oscillating term in $\mathcal{F}W^f(k)$. In fact, for discrete grid modules to occur, the only constraints imposed on the fixed-scale kernel W^f are: (a) the Fourier transform $\mathcal{F}W^f(k)$ must have a sufficiently large number of maxima (at least 4 maxima, corresponding to the 4 grid modules observed in experimental observations); and, (b) these maxima must be at scales smaller than $1/\sigma$ in k-space. Here we argue that this is generally true for arbitrary kernels, modulo a single scaling parameter.

We hypothesize and give support, without formal proof, that almost every arbitrarily chosen kernel $W^f(\Delta x)$ will have a Fourier transform with multiple maxima satisfying condition (a). We will then argue that this kernel can always be scaled to satisfy condition (b).

To motivate our hypothesis, we first note that it is actually possible to construct specific kernels $W^f(\Delta x)$ whose Fourier transform does not present multiple maxima. For example, the Gaussian kernel, $W_{gauss}(\Delta x) = \exp[-(\Delta x)^2/2]$, results in a Fourier transform that is unimodal. However, we hypothesize that such functions are rare in the space of all continuous functions in L^2. Indeed, we can construct a function that is arbitrarily close to the Gaussian kernel whose Fourier transform will have an infinite number of maxima: Let $f_0(\Delta x) = \mathbb{1}_{[-1,1]}$ be the box function. Define

$$f_n = f * f_{n-1}$$

for all $n \geq 1$, where $f * g$ represents the convolution of functions f and g. By the central limit theorem, $\sqrt{n} f_n(\sqrt{n}\Delta x)$ will approach $W_{gauss}(\Delta x)$. However,

$$\mathcal{F} f_n(k) = [2\sin(k)/k]^n, \tag{C1}$$

which clearly has an infinite number of maxima. Thus, even though the Gaussian kernel has a unimodal Fourier transform, we can construct a function $g_n(\Delta x) = \sqrt{n} f_n(\sqrt{n}\Delta x)$ that is arbitrarily close to the Gaussian kernel (for sufficiently large n) but has a Fourier transform that presents an infinite number of maxima.

In this context, we claim that almost every arbitrarily chosen kernel $W^f(\Delta x)$ will have a Fourier transform with multiple maxima. This may be intuited as follows: First note that Fourier space is a dual space, and hence instead of considering arbitrary kernels in real space we may equivalently choose arbitrary kernels in Fourier space. Further assuming that $\mathcal{F}W^f(k)$ is a smooth function, we hypothesize that generically smooth functions that are in L^2 will almost always have multiple maxima and minima. Note that this heuristic also applies to the pattern forming kernel as well — we hypothesize that generic L^2 smooth functions will have some maxima and minima with a global maxima that exists at $k > 0$ with probability 1, and will not be always negative (in which case a rescaling will make the maxima larger than the constant specified by requirement 2 for pattern forming kernels in the main text). Thus we expect that kernels will generically result in hexagonal pattern formation, as demonstrated in Fig. 10.

Thus condition (a) may be satisfied for arbitrary kernels $W^f(\Delta x)$.

Next, note that scaling a function in real space results in an inverse scaling of the Fourier transform, i.e., $\mathcal{F}[W^f(a\Delta x) = \mathcal{F}W^f(k/a)$. Hence, we can always scale the function $W^f(\Delta x)$ to obtain a Fourier transform with maxima that are within any desired scale, allowing condition (b) to be satisfied.

In Fig. 12, we show examples of modularization arising from different combinations of graded pattern forming kernels (W^g) and fixed-scale kernels (W^f). In each case, we also present the expected periodicity in each module as a function of spatial position as given by the perturbative analysis Eq. (B7). The analytical result based on linear stability provides an excellent prediction of the pattern periods per module (see also Main text, Fig. 3e). It also predicts the locations of the module boundaries (see also Main text, Fig. 3e) though not as accurately: module boundary predictions tend to be slightly but systematically offset relative to the simulated dynamics, due to the effects of nonlinearity in the later stages of pattern formation.

Appendix D: Simple kernels and period ratios

What kinds of fixed-scale interactions might be present in the medial-entorhinal cortex? As described in the main text, in the context of biology, we might expect *simple* interaction kernels W^f to be relevant i.e., the fixed-scale interaction profile W^f has the following characteristics: (a) there exists a *single* length-scale d that primarily characterizes the shape of W^f; (b) any other length-scales relevant to W^f, say scales $\epsilon_1, \epsilon_2, ...$ are each much smaller than the primary length scale d. Further, we assume that the primary length-scale associated with the fixed-scale interaction is larger than the length-scales of the pattern forming kernel, i.e., $d \gg \sigma_{E/I}(n_{DV})$.

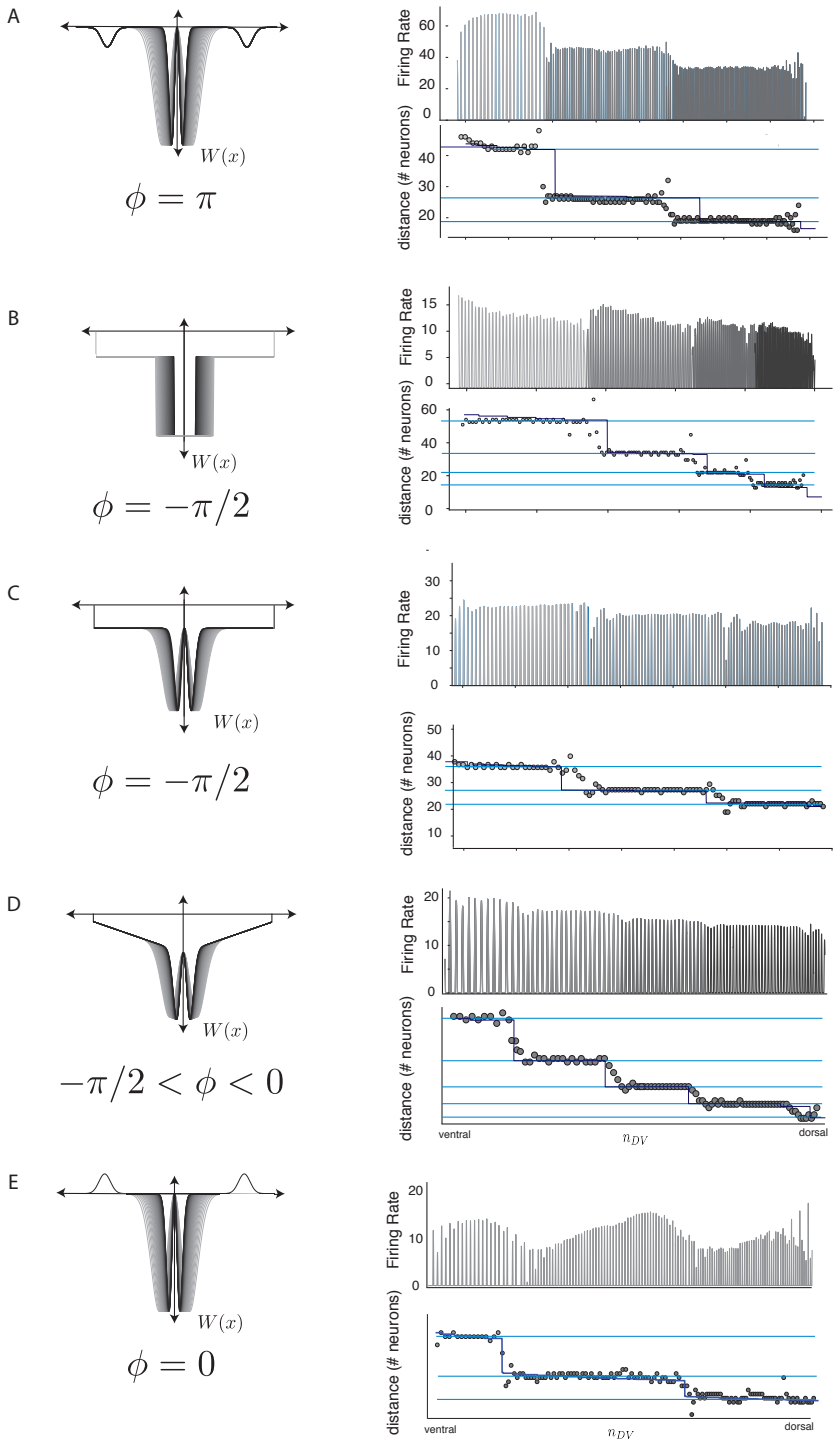


FIG. 12. Examples of modularization and population activity (right column) with various pattern forming and fixed-scale lateral interactions (left column). In each case the dark-blue curve shows the predicted value of the grid period from Eq. (B7), and is in close agreement with the numerical simulation of the population activity. Each of the fixed-scale interactions has a qualitatively different shape, spanning different values of ϕ (see Fig. 3)

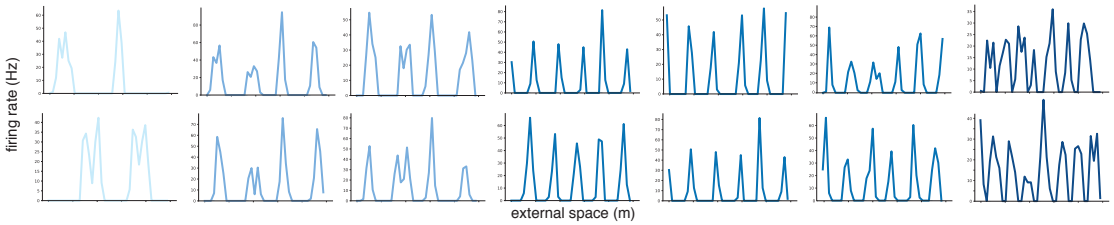


FIG. 13. Sample tuning curves from several neurons in all modules from the network of Fig 2a.

We will demonstrate that *simple* fixed-scaled interaction kernels result in analytic expressions for grid periods that are characterized by a single angular variable ϕ

$$\frac{\lambda_{m+1}}{\lambda_m} = \frac{m + 1 + \phi/(2\pi)}{m + \phi/(2\pi)}. \tag{D1}$$

Before filling in the details of our argument, we present an intuitive explanation of the general idea:
Consider the following basic classes of *simple* kernels that satisfy the above-described criteria corresponding to a length-scale d:

(a) $g(|\Delta x| - d)$, for arbitrary functions $g(\rho)$ that are nonzero only over scales $|\rho| < \epsilon_i$ (a *localized* kernel), and,

(b) A constant term, that is uniform everywhere up to $\Delta x = d$, after which it falls to zero (a *diffuse* kernel),

(c) A decaying term, that decreases from a constant value at $\Delta x = 0$ to zero at $\Delta x = d$ (a *decaying* kernel).

We also define *short-range* kernels, as any arbitrary function $h(\Delta x)$ that is nonzero only over scales $|x| < \epsilon_i$.
Any *simple* kernel $W^f(\Delta x)$ can be generally constructed as a linear combination of the above basic classes. In addition, *simple* kernels may also contain an added component of a *short-range* kernel.
To see that *simple* kernels will generally result in grid period ratios corresponding to Eq. (D1), we will examine the approximate Fourier transform structure for each component of the linear combination of *simple* kernels corresponding to a given length-scale d. We first demonstrate that each of the basic *simple* kernels will result in Fourier transforms that are sinusoidal functions with phase shifts and decaying envelopes and hence each basic *simple* kernel will satisfy Eq. (D1). We then show that short-range kernels present Fourier transforms that vary only at large scales, and can be ignored in our analyses of *simple* kernels. We then use these results to demonstrate that all *simple* kernels constructed as the above-described linear combination will have sinusoidal Fourier transforms and will satisfy Eq. (D1).

1. Localized kernels

For a general localized kernel $W^f(\Delta x) = g(|\Delta x| - d)$ we obtain

$$\mathcal{F}W^f(k) = \Re[e^{-ikd}\mathcal{F}g(k)]. \tag{D2}$$

Since $g(x)$ is supported over a scale ϵ, the Fourier transform $\mathcal{F}g(k)$ will only vary at scales $k \sim 1/\epsilon \gg 1/d$. Thus for $1/d \ll k \ll 1/\epsilon$, we can approximate Eq. (D2) as

$$\mathcal{F}W^f(k) = |\mathcal{F}g(k)| \cos(kd - \psi), \tag{D3}$$

where $\psi = \arg[\mathcal{F}g(k)]$. The local maxima of $\mathcal{F}W^f(k)$ will then occur at

$$S = \left\{ \frac{2m\pi + \psi}{d} \middle| m \in \mathbb{Z}^+ \right\}, \tag{D4}$$

resulting in period ratios described by

$$\frac{\lambda_{m+1}}{\lambda_m} = \frac{m + 1 + \psi/(2\pi)}{m + \psi/(2\pi)}, \tag{D5}$$

which is identical to Eq. (D1) for $\phi = \psi$. We also note that we can now ascribe an interpretation to the phase angle ϕ — it is the phase difference between $\mathcal{F}W^f(k)$ and $\cos(kd)$.

2. Diffuse kernels

We model a diffuse interaction kernel $W^f(n_{DV})$ as

$$W^f(n_{DV}) = -W_0 \mathbb{1}_{[-d,d]} = \begin{cases} -W_0 & \text{if } |n_{DV}| \leq d \\ 0 & \text{if } |n_{DV}| > d \end{cases}. \tag{D6}$$

Corresponding to the discussion above, we look at the Fourier transform $\mathcal{F}W^f(k)$

$$\mathcal{F}W^f(k) = \int_{-\infty}^{+\infty} -W_0 \mathbb{1}_{[-d,d]} e^{ikx} dx = \int_{-d}^{+d} -W_0 e^{ikx} dx \tag{D7}$$

$$= -2W_0 \frac{\sin(kd)}{k} = -2W_0 d \operatorname{sinc}(kd). \tag{D8}$$

Note that once again, similar to Eqn. (B21), we obtain a functional form consisting of a periodic function ($\sin(kd)$) that is multiplied by a decaying envelope $1/(kd)$. Ignoring the effects of the envelope function, the maxima of this function occur at

$$S \approx \left\{ \frac{2m\pi - \pi/2}{d} \middle| m \in \mathbb{Z}^+ \right\}, \tag{D9}$$

which immediately results in period ratios of the form

$$\frac{\lambda_{m+1}}{\lambda_m} \approx \frac{m+1-1/4}{m-1/4}, \tag{D10}$$

which corresponds to the result in Eq. (D1) for $\phi = \pi/2$.

More precisely, the extrema of $\mathcal{F}W^f(k)$ occur at $k_m d = q - 1/q - 2/3q^3 + O(q^{-5})$ where $q = \left(m + \frac{1}{2}\right)\pi$. Notably, the errors decay approximately as $1/(\pi m)$, and thus for modules generated corresponding to $m \gtrsim 2$ will result in period ratios that approximate Eq. (D1) closely.

3. Decaying kernels

Decaying kernels with a scale d may be modeled as any monotonically decreasing function that decays from some constant W_0 at $\Delta x = 0$, to zero, at $\Delta x = d$. For simplicity, we consider the simplest linear approximation to such a kernel, modeled as a triangular kernel. For additional subtleties in the treatment of other decaying kernels, see D 5 a. The triangular kernel can be written as:

$$W^f(\Delta n_{DV}) = \begin{cases} W_0(\Delta n_{DV} - d)/d & \text{if } \Delta n_{DV} < d \\ 0 & \text{if } \Delta n_{DV} \geq d \end{cases} \tag{D11}$$

This function can be written as the convolution of 2 diffuse box functions:

$$W^f(\Delta n_{DV}) = (-W_0 \mathbb{1}_{[-d/2,d/2]}) * (W_0 \mathbb{1}_{[-d/2,d/2]}).$$

Thus, its Fourier transform is:

$$\mathcal{F}W^f(k) = -W_0^2 d^2 \left(\frac{\sin(kd/2)}{(kd/2)} \right)^2$$

$$= -\frac{2W_0^2}{k^2}[1 - \cos(kd)].$$

Once again, we obtain a simple trigonometric function, with maxima at

$$S \approx \left\{ \frac{2m\pi}{d} \middle| m \in \mathbb{Z}^+ \right\}, \tag{D12}$$

which immediately results in period ratios of the form

$$\frac{\lambda_{m+1}}{\lambda_m} \approx \frac{m+1}{m}, \tag{D13}$$

which corresponds to the result in Eq. (D1) for $\phi = 0$.

4. Short-range kernels

For the case of a short-range kernel $W^f(\Delta x)$ that extends upto a scale ϵ, we note from the Fourier uncertainty principle that the characteristic k-scales of $\mathcal{F}W^f(k)$ will $\sim 1/\epsilon \gg 1/d$. Thus, unlike the three other types of simple kernels discussed above, short range kernels do not have structure at the scale of $1/d$. Since all relevant scales are much larger than $1/d$, adding short range kernels to any of the other types of *simple* kernels will *not* change the structure of local maxima at scales of $1/d$.

5. Arbitrary simple kernels

We now consider a general form for *simple* kernels, by constructing linear combinations of the above described three basic classes of *simple* kernels each corresponding to the same length scale d and additional short-range kernels.

$$W^f = a_{\text{local}} W^f_{\text{local}} + a_{\text{diffuse}} W^f_{\text{diffuse}} + a_{\text{decaying}} W^f_{\text{decaying}} + a_{\text{short}} W^f_{\text{short}}. \tag{D14}$$

As demonstrated in the preceding sections, the Fourier transform $\mathcal{F}W^f(k)$ will be given as

$$\mathcal{F}W^f(k) = a_{\text{local}} |\mathcal{F}g(k)| \cos(kd - \psi) - 2W_0 a_{\text{diffuse}} \sin(kd)/k - 2W_0^2 a_{\text{decaying}} (1 - \cos(kd))/k + \mathcal{F}h(k) \tag{D15}$$

$$= H_0(k) + \sum_{i=0}^{3} H_i(k) \cos(kd + \phi_i) \tag{D16}$$

for some constants ϕ_i, and some envelope functions $H_i(k)$ for $i = 0, 1, 2, 3$ that are slowly varying for $kd \gtrsim \mathcal{O}(1)$. Under this approximation, $\mathcal{F}W^f(k)$ is simply the sum of multiple sinusoidal waves with different phases and identical frequencies. Thus,

$$\mathcal{F}W^f(k) \approx \cos(kd - \phi) \tag{D17}$$

for some ϕ and $kd \gtrsim \mathcal{O}(1)$. Hence, the maxima of $\mathcal{F}W^f(k)$ occur at

$$S \approx \left\{ \frac{2n\pi + \phi}{d} \middle| n \in \mathbb{Z}^+ \right\}, \tag{D18}$$

which immediately results in period ratios of the form Eq. (D1). Note that the approximations made above imply that there may be deviations from our results for the maxima corresponding to small k values — this may manifest as deviations in the largest period grid module away from Eq. D1.

a. *Caveats*

Clearly there exist *simple* kernels with Fourier transforms that are not given by $\mathcal{F}W^f(k) \approx \cos(kd - \phi)$. For example the Gaussian kernel, $W^f(\Delta x) = \exp[-\Delta x^2/(2d^2)]/(d\sqrt{2\pi})$ is a *simple* decaying kernel (since it has only a single scale d). Yet, its Fourier transform is simply $\mathcal{F}W^f(k) = \exp[-k^2 d^2/2]$, which has only a single maximum! However, as we have shown earlier, there exist kernels that are arbitrarily close to the Gaussian kernel, whose Fourier transforms are given by powers of trigonometric functions, and hence have multiple regularly-spaced maxima with a spacing of $\sim 1/d$. Similarly, there exist additional *simple* functions[118–120], $f(\Delta x)$, (like the Gaussian kernel) whose Fourier transforms $\mathcal{F}f(k)$ have a small number of maxima. We hypothesize that for all such functions $f(\Delta x)$ there exist *simple* kernels $g(\Delta x)$ that are arbitrarily close to $f(\Delta x)$ and possess regularly spaced maxima.

6. Period ratios

Having demonstrated analytically that *simple* kernels result in a sequence of period ratios given by Eq. (D1), we now address the question of the mean period ratio over the sequence and over different values of ϕ. In the main text we have demonstrated that setting $\phi = 0$ results in a detailed period ratio sequence that is in close agreement with the sequence of experimentally observed values. Here we consider the period ratios obtained for other values of ϕ, to demonstrate that the experimental observation of mean period ratios being approximated by 1.4 [24] emerges naturally from our setup.

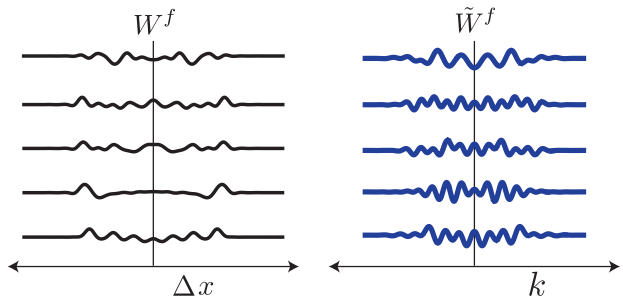


FIG. 14. Randomly constructed fixed-scale interactions (left column) and their Fourier transforms (right column), in addition to the hand-designed ones in Fig.3, that give $\phi = 0$.

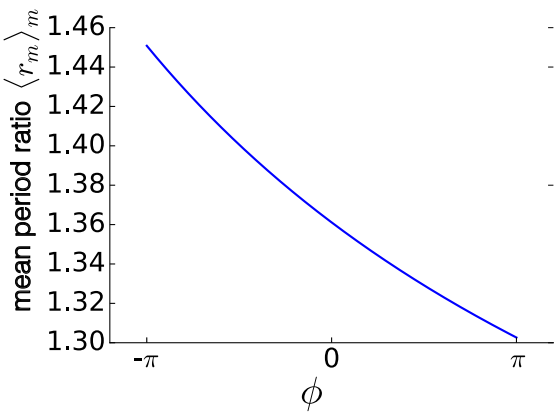


FIG. 15. **Mean grid-period ratios** Ratios of grid periods averaged over 4 modules as a function of the phase shift ϕ in Eq. (D1)

From Eq. (D1), we obtained that the period ratio, $r_m = \lambda_{m+1}/\lambda_m$ can be written as

$$r_m = 1 + 1/(m + f), \tag{D19}$$

where $f = \phi/(2\pi)$. We ignore $m = 1$, since that results in a period ratio close to 2, which does not correspond to experimental observations. Averaging the period ratio over the next 4 modules (corresponding to r_m for $m \in \{2 \ldots 4\}$) results in

$$\langle r_m \rangle_m = 1 + \frac{1}{3}\left(\frac{1}{f+2} + \frac{1}{f+3} + \frac{1}{f+4}\right) \tag{D20}$$

As can be seen in Fig. 15, this mean period ratio lies in the range [1.3,1.45], indicating that at all values of ϕ, the period ratio obtained from Eq. (D1) matches well with experimental observations. The average of these period ratios over all values of ϕ can also be calculated as

$$\langle r_m \rangle_{\phi,m} = 1 + \frac{1}{3}\left[\log\left(\frac{5}{3}\right) + \log\left(\frac{7}{5}\right) + \log\left(\frac{9}{7}\right)\right] \tag{D21}$$

which is approximately equal to 1.37.

7. Module size; number of modules as a topological quantity

As discussed in the main text, peak-selection for modularization is a highly robust mechanism that is largely indifferent to system parameters such as the the particular forms of the fixed-scale interaction and the shape of the

gradient. Here we provide an analysis of the number of modules, the scaling of module sizes, and the positions of module boundaries, which also exhibit the same robustness. Further, we also describe how this robustness may be interpreted as arising from a topological origin, similar to topological robustness in other physical systems like the quantum hall effect.

Recall that for the continuously graded kernel $W^g_{n_{DV}}(\Delta x)$ with characteristic spatial scale $\sigma(n_{DV})$ at position n_{DV}, the wave-vector of the formed pattern was proportional to $1/\sigma(n_{DV})$:

$$k^{*g}(n_{DV}) = \eta/\sigma(n_{DV}), \tag{D22}$$

where η is an n_{DV}-independent constant that depends on only the particular form of the graded kernel. Let the spatial extent of the system be $n_{DV} \in [0, L]$, with $\sigma(n_{DV})$ monotonic such that $\sigma_{\min} = \sigma(0) \leq \sigma(n_{DV}) \leq \sigma(L) = \sigma_{\max}$.

We assume for simplicity that the fixed-scale lateral interaction is a *simple* kernel, such that $\mathcal{F}W^f(k) \sim \cos(kd - \phi)$. Thus, the local maxima generated by $\mathcal{F}W^f(k)$ occur at $k_n \approx (2n\pi + \phi)/d$, where n are the natural numbers. As discussed in the main text, each of these local maxima is 'selected' in turn by the moving broad peak of the Fourier transform of the graded kernel, whose position according to Eq. D22 occurs at $k^{*g}(n_{DV}) = \eta/\sigma(n_{DV})$.

Notably, the selected maximum k_m will be robust to small perturbations in the selection function $\mathcal{F}W^g_{n_{DV}}(k)$, since k_m will remain quantized to one of the discrete values prespecified by the set $\{k_n \| n \in \mathbb{N}\}$. In this sense, the chosen maximum k_m (and hence the corresponding module) presents the hallmarks of a topologically protected state[1]. The topological number corresponding to a given module is the module number m, which is a topological invariant similar to a winding number[1][121].

The set of modules expressed through the length of the system corresponds to the set of local maxima k_n that lie within the range $[\eta/\sigma_{\max}, \eta/\sigma_{\min}]$ that is delineated by the range of peak positions of the graded interaction. It follows that the maxima selected by the graded interaction obey:

$$\frac{\eta}{\sigma_{\max}} \leq \frac{2n\pi + \phi}{d} \leq \frac{\eta}{\sigma_{\min}}. \tag{D23}$$

Thus, the set of formed modules are determined by the set of integers n that fit in the following interval:

$$\frac{-\phi + \eta d/\sigma_{\max}}{2\pi} \leq n \leq \frac{-\phi + \eta d/\sigma_{\min}}{2\pi} \tag{D24}$$

and hence the number of modules N_{mod} is:

$$\begin{aligned} \#\text{ modules} &\equiv N_{mod} \\ &= \left\lfloor \frac{-\phi + \eta d/\sigma_{\min}}{2\pi} \right\rfloor - \left\lceil \frac{-\phi + \eta d/\sigma_{\max}}{2\pi} \right\rceil \\ &= \left\lfloor \frac{-\phi + k^{*g}(0) d}{2\pi} \right\rfloor - \left\lceil \frac{-\phi + k^{*g}(L) d}{2\pi} \right\rceil \end{aligned} \tag{D25}$$

where $\lfloor\ \rfloor$, $\lceil\ \rceil$ indicate the floor and ceiling operations, respectively.

The above result leads to the following observations: First, the central quantity essential for determining the number of modules is the difference in the integer ratios of the fixed-scale interaction width to the extremal lateral interaction widths, $d/\sigma_{\min}, d/\sigma_{\max}$. Second, the number of modules depends only on the end-point values $\sigma_{\min}, \sigma_{\max}$ of the smoothly varying width $\sigma(n_{DV})$ the graded interaction; notably, it does not depend on the detailed shape of $\sigma(n_{DV})$. Moreover, if $\sigma_{\min}, \sigma_{\max}$ are varied smoothly (while d is held fixed), or if d is varied smoothly (while $\sigma_{\min}, \sigma_{\max}$ are held fixed), the number of modules will remain fixed, until the change becomes large enough to accommodate one additional or one less module. Thus, the number of modules is also a topological invariant of the system, through the module number m. Third, the number of modules does not depend on the system size L, or the number of neurons n_{DV} the system is discretized into (cf. Fig. 3f). Fourth, since the average module size will be L/N_{mod}, the module sizes are extensive in L. Thus, for sufficiently large L, the module sizes can be orders of magnitude larger than the scales of the lateral interaction d and σ.

Note that the above argument on topological robustness of the modularization of the system is not restricted to the case of *simple* fixed-scale kernels. Indeed, for any fixed-scale interaction W^f, the topological number m for any given expressed module will correspond to selecting the mth maximum of $\mathcal{F}W^f(k)$, for $k > 0$.

a. Module boundary locations

Following the peak-selection arguments made earlier, the module boundaries will occur at spatial locations that have $k^{*g}(n_{DV})$ in between k_n and k_{n+1} (the specific location will depend on the particular forms of the kernels). As

a zeroth order approximation, we can assume that the module boundaries will occur near $(k_n + k_{n+1})/2$,

$$k^{*g}(n_{\text{boundary}}) \approx \frac{(2n+1)\pi + \phi}{d} \tag{D26}$$

and thus

$$n_{\text{boundary}} \approx \sigma^{-1}\left(\frac{\eta d}{(2n+1)\pi}\right). \tag{D27}$$

where σ^{-1} is the inverse function of $\sigma(n_{DV})$, $\sigma^{-1} \circ \sigma(x) = x$. Thus, while the specific positions of the module boundaries are dependent on the shape of the gradient $\sigma(n_{DV})$, qualitative features such as the number of modules, module periods and module sizes are indifferent to the particular forms of the gradient (cf. Fig. 3f).

In (Fig. 4d), we vary the width of the $\sigma(x)$ in two different ways: linearly along and in a square root along n_{DV}. This leads to a shift in the module boundary locations that is predicted by fourier theory.

8. 2D analysis

We have presented a majority of the above analysis for the case of one-dimensional grid cells. Here we briefly present the analogous computations for the Fourier transforms in two dimensions. We first demonstrate a classical result relating the Fourier transform of radially symmetric functions to the Hankel transform, which we shall then use to compute the relevant transforms. Consider the Fourier transform of a function $f(\mathbf{x}) = f(x, y)$

$$\mathcal{F}f(\mathbf{k}) = \int f(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}d\mathbf{x}$$
$$\mathcal{F}f(k_x, k_y) = \int f(x, y)e^{ik_x x + ik_y y}dxdy.$$

Define polar coordinates in real and Fourier space such that:

$$x = r\cos\theta$$
$$y = r\sin\theta$$
$$k_x = k\cos\phi$$
$$k_y = k\sin\phi$$

This leads to the dot product $\mathbf{k}\cdot\mathbf{x}$ to be simplified as

$$\begin{aligned} k_x x + k_y y &= rk(\cos\theta\cos\phi + \sin\theta\sin\phi) \\ &= rk\cos(\theta - \phi) \end{aligned}$$

Thus,

$$\mathcal{F}f(k_x, k_y) = \mathcal{F}f(k, \phi) = \int_0^\infty \int_0^{2\pi} rdrd\theta f(r, \theta)e^{ikr\cos(\theta-\phi)}$$

In all cases of interest, the function f is a kernel, and is hence a radially-symmetric real function $f(r, \theta) = f(r)$. Similarly, the Fourier transform $\mathcal{F}f$ will also be a real radially-symmetric function $\mathcal{F}f(k, \phi) = \mathcal{F}f(k)$. Thus

$$\mathcal{F}f(k) = \int_0^\infty \int_0^{2\pi} rdrd\theta f(r)e^{ikr\cos(\theta-\phi)}, \tag{D28}$$
$$= \int_0^\infty rdrf(r)\int_0^{2\pi} e^{ikr\cos(\theta-\phi)}d\theta, \tag{D29}$$
$$= 2\pi\int_0^\infty rf(r)J_0(kr)dr, \tag{D30}$$

where J_0 is the Bessel function of the first kind, defined by

$$J_0(x) = \frac{1}{2\pi}\int_0^{2\pi} e^{ix\cos(\theta-\phi)}d\theta.$$

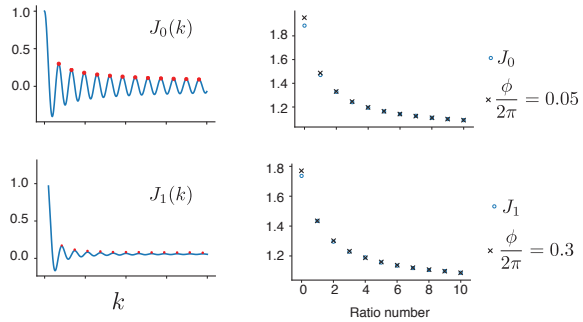


FIG. 16. Bessel functions (left column) and period ratios for Bessel function maxima (right column) with their best-fit values of ϕ for the period ratios corresponding to Eq. (D1)

Equation (D30) defines the Hankel transform (of order zero) of $f(r)$ — the radial component of the Fourier transform of the kernel $f(\mathbf{x})$ is simply the Hankel transform of $f(|\mathbf{x}|)$.

For the localized gaussian secondary interaction, we can calculate the Fourier transform analytically.

$$\begin{aligned}\mathcal{F}W_{\text{local}}(k) &= 2\pi \int_0^\infty r\left[\alpha_E e^{-r^2/2\sigma_E^2} - \alpha_I e^{-r^2/2\sigma_I^2} + \alpha_S e^{-(r-d)^2/2\sigma_S^2}\right] J_0(kr)dr \\ &= 2\pi\left[\alpha_E \sigma_E^2 e^{-k^2\sigma_E^2/2} - \alpha_I \sigma_I^2 e^{-k^2\sigma_I^2/2} + \alpha_S J_0(kd)\sigma_S^2 e^{-k^2\sigma_S^2/2}\right]\end{aligned}$$

We can also analytically calculate the Fourier transform for a box-like interaction:

$$\begin{aligned}\mathcal{F}W_{\text{diffuse}}(k) &= 2\pi W \int_0^d r J_0(kr)dr \\ &= \frac{2\pi W}{k^2}\int_0^{kd} \rho J_0(\rho)dr \\ &= \frac{2\pi W}{k^2}\left[kdJ_1(kd)\right] \\ &= \frac{2\pi W d^2 J_1(kd)}{kd}\end{aligned}$$

We can similarly also define a two-dimensional equivalent of the decaying kernel, as the convolution of the half-sized circular box kernel with itself. Thus, by applying convolution theorem to the result on diffuse kernels we obtain

$$\mathcal{F}W_{\text{decaying}}(k) = \left[\frac{\pi W d J_1(kd/2)}{k}\right]^2.$$

Note that $J_0(x)$ and $J_1(x)$ display qualitatively similar behavior to $\cos(x)$ and $\sin(x)$ respectively, apart from an amplitude modulation of the peaks — particularly, we note that the Bessel functions display approximately periodic maxima, which was the central property required for all of our results on modularization and peak selection to apply. We demonstrate this in Fig.16, where we show that the maxima of the Bessel functions are approximately periodic, and fit the form of Eq. (D1) well. In particular, note that the best-fit value of ϕ for $J_0(k)$ is approximately 0, which is similar to $\cos(k)$, and the best-fit value of ϕ for $J_1(k)$ is approximately $\pi/4$, which is similar to $\sin(k)$.

We implemented a 2d simulation that generates 3 discrete modules as shown in Figure 18. For computational feasibility, the simulation was performed in 2 parts: one with $x \in [0, 0.6N_x^{2d}]$ and the other with $x \in [0.6N_x^{2d}, N_x^{2d}]$. The weight matrices for each network were of size 100x1000 each. The weight matrix for a single large 100x2000 network would have contained $4x10^{10}$ elements, which we found prohibitively difficult and slow to run.

Fig 19(a) shows another instance of a modular 2d network, the only difference being the value of d_{loc}, which changed from 50 to 45. Fig 19(b) shows the same simulation with 2 distinct random initializations. The pair of resulting modules in each simulation have different relative orientations. Because finite size effects from our simulations also partially constrain the orientations of the modules (data not shown), we cannot make predictions about the relative orientations of the grid modules found in experiments [24].

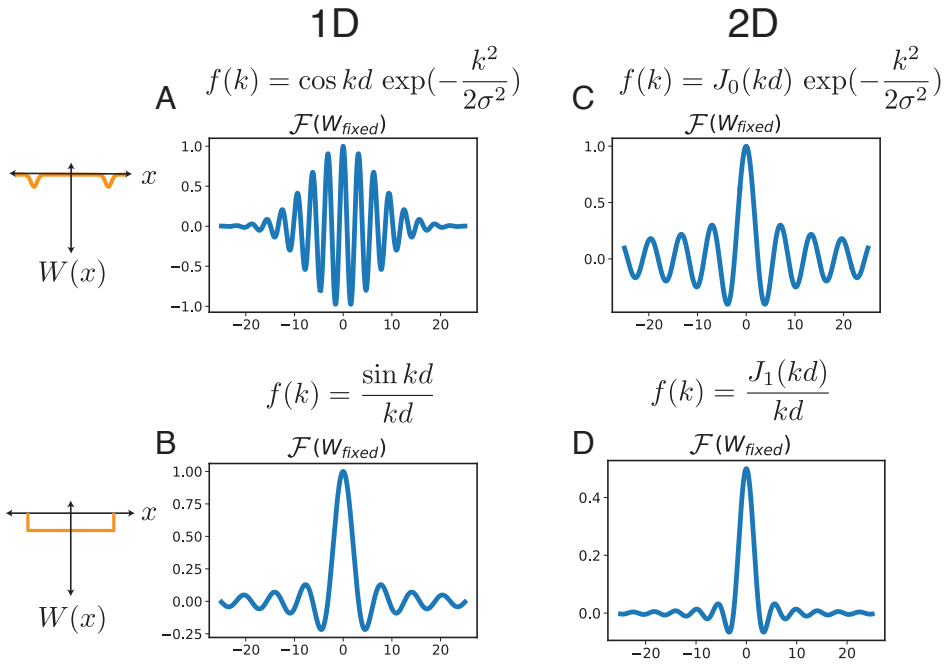


FIG. 17. Fixed interactions(left, in orange) and their oscillatory Fourier transforms in 1D (left column) and 2D (right column).

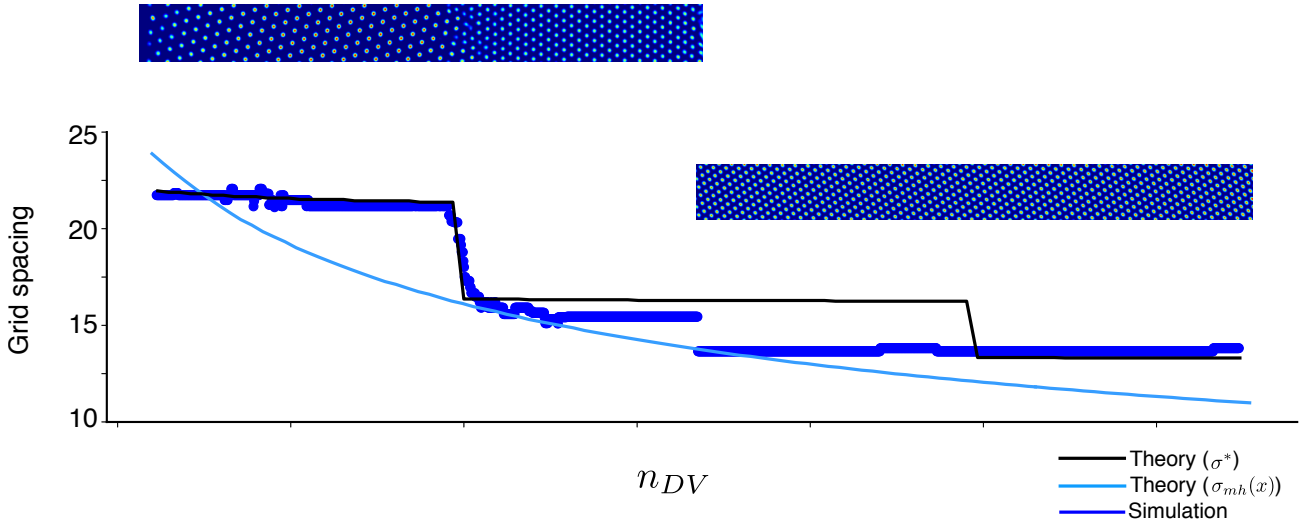


FIG. 18. **2d simulation with 3 modules:** (top) Snapshots of population activity showing 3 discrete 2d grid modules, (bottom) plot of grid spacing and comparision with Hankel transform predictions. Grid spacing determined by calculating the (neural) spatial auto-correlation of the population firing activity.

9. Robustness to spatial noise

In the main text, we discussed how the topological robustness properties of peak selection result in the formed modules being stable to several forms of noise. Particularly, here we focus on the robustness to spatial heterogeneities in the lateral interaction kernels.

We first examine the robustness to spatial heterogeneities in the pattern forming kernel W^g. To construct such an inhomogeneous pattern-forming interaction, we construct the noisy kernel at location $\mathbf{x}$, by replacing the spatially homogeneous kernel $W^g[\mathbf{x}, \mathbf{x}'] = W^g[\mathbf{x} - \mathbf{x}']$, with a spatially heterogeneous kernel $W^g_\xi[\mathbf{x}, \mathbf{x}'] = W^g[|\mathbf{x} - \mathbf{x}'| + \xi(\mathbf{x}')]$,

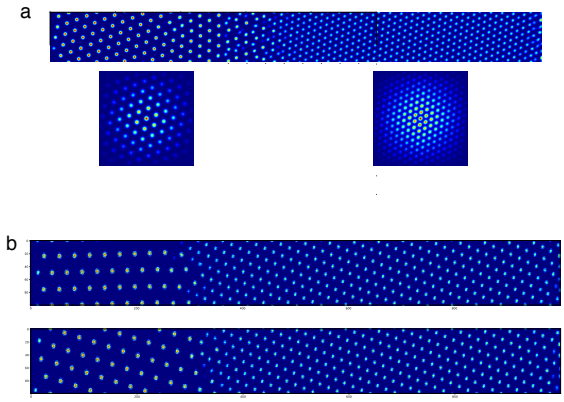


FIG. 19. (a) Another instance of a spontaneously formed two dimensional network with parameters given in Table 4. (b) Two different random initializations of the network from Fig 2h show different relative orientations between the 2 formed modules.

where $\xi(\mathbf{x}')$ is a random number sampled independently for each spatial location $\mathbf{x}'$ with mean zero and variance ϵ^2. In Fig. 20d we present examples of such kernels for the case of $W^g[\mathbf{x}, \mathbf{x}']$ described by the box function Eq. (13). Note how the independent sampling of $\xi(\mathbf{x}')$ at each location results in a heterogenous kernel W^g_ξ that varies in scale at different $\mathbf{x}$, and is no longer radially symmetric.

Recall that peak selection entails that the grid period at any location n_{DV} is dependent on the set of potential maxima defined by $\mathcal{F}W^f(k)$, with a selection between these maxima performed by the broader peak of $\mathcal{F}W^g(k)$. If noise in the form of spatial heterogeneities are only introduced in W^g (and hence introduced in $\mathcal{F}W^g$) this results in a noisy selection function. However, since the same maxima will be chosen for a range of selection functions (See Fig. 20a-b), the heterogeneity in W^g will not be manifested in the emergent grid period.

We next consider the addition of similar heterogeneities in the fixed-scale interaction as well, W^f (such as in Fig. 20c). Note that maxima induced by *simple* W^f are at $k_n \approx (2n\pi + \phi)/d$, where n are the natural numbers, and hence the grid periodicity of the nth module is given by $\lambda_n \approx d/(n + \phi/2\pi)$. If we consider $\mathcal{O}(\epsilon)$ noise added to W^f in the form of spatial heterogeneities, this would result in an $\mathcal{O}(\epsilon)$ error in the effective fixed-scale d. However, since λ_n is approximately d/n, thus the effective noise in periodicity of the nth module, λ_n, will be $\mathcal{O}(\epsilon/n)$. Thus, higher module numbers (corresponding to modules with smaller grid periods) have additional error correction beyond the robustness conferred by the topological nature of the peak selection process. This results in clean hexagonal firing fields despite inhomogeneities introduced in all lateral interactions as shown in Fig. 6.

10. Comparison of experimental observations with predicted period ratios

The general mechanism of peak-selection presented above describes how discrete modules can spontaneously arise in the presence of continuous gradients, by consideration of an additional fixed-scale lateral interaction W^f. However, this mechanism does not provide any testable predictions for the ratio of grid periods unless additional assumptions are made. If indeed we assume that W^f is a *simple* kernel, i.e., W^f is primarily defined by a single spatial scale, then we demonstrated in SI Sec. D that the period ratios will be given by the simple formula, Eq. D1. In this section, we show that experimental observations of grid periods largely appear to match our predicted period ratios for *simple* kernels with $\phi = 0$.

For verification of our main results on the predicted form of period ratios, we examine the literature for grid period measurements for multiple simultaneously measured grid modules in rats[24, 122–124]. We note that a large fraction of experimental observations of grid cells with more than one module measure only two modules. For a single pair of grid periods λ_1 and $\lambda_2 > \lambda_1$, we can always explicitly solve for ϕ and m in Eq. (D1), to obtain

$$\frac{\phi}{2\pi} = \left\{ \frac{\lambda_2}{\lambda_1 - \lambda_2} \right\}, m = \left\lfloor \frac{\lambda_2}{\lambda_1 - \lambda_2} \right\rfloor, \tag{D31}$$

where $\{x\}$ represents that fractional part of x, and $\lfloor x \rfloor = x - \{x\}$ represents the integer part of x. Thus, a single ratio, because it can always be fit by Eq. (D1), imposes no constraints on the accuracy of the expression.

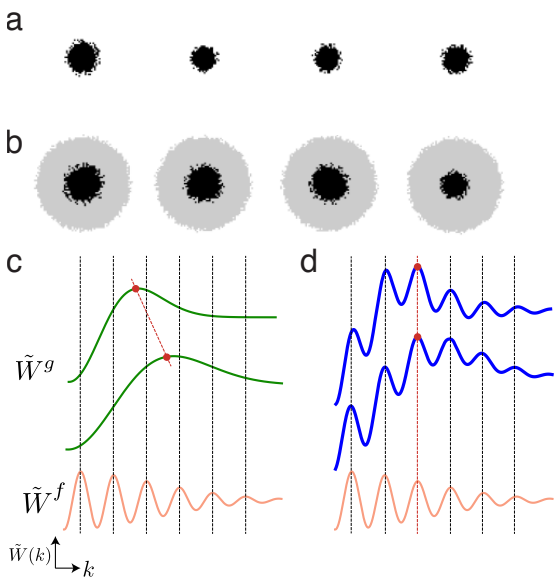


FIG. 20. **Noise robustness in peak selection process** demonstrating how the additional of the smaller oscillatory fourier transform of the fixed interaction leads to no change in maxima despite smooth movement of the primary peak. (a) Example pattern forming interaction kernels from 4 neurons without a secondary fixed scale interaction (b) Example composite kernels from 4 neurons showing both the pattern forming interaction (black) and fixed scale interaction (grey)(c-d) The movement of the pattern forming interaction leads to a shift in the location of the gloabl maxima in the absence of a secondary interaction. This secondary interaction prevents any shift in the location of the global maxima when defined by the sum of the pattern forming interaction and the fixed scale interaction.

It is possible to obtain a value of ϕ from Eq. (D1) and a single pair of periods; however, the estimate obtained from a single pair is not robust: r_m depends too sensitively on ϕ. For example, in [24], Rat 13388 exhibits grid periods of ≈ 53.24 cm and ≈ 43.00 cm (as estimated from SI Fig. 12b in [24]); Eq. (D1) then yields $\phi/(2\pi) = 0.199$. Assuming a very small measurement error of $\sim 0.5cm$ in the larger period, such that if it were 53.75 cm instead of 53.24, would yield ϕ exactly equal to zero. A simple sensitivity analysis of the magnitude of error in estimating ϕ can be performed from Eq. (D31):

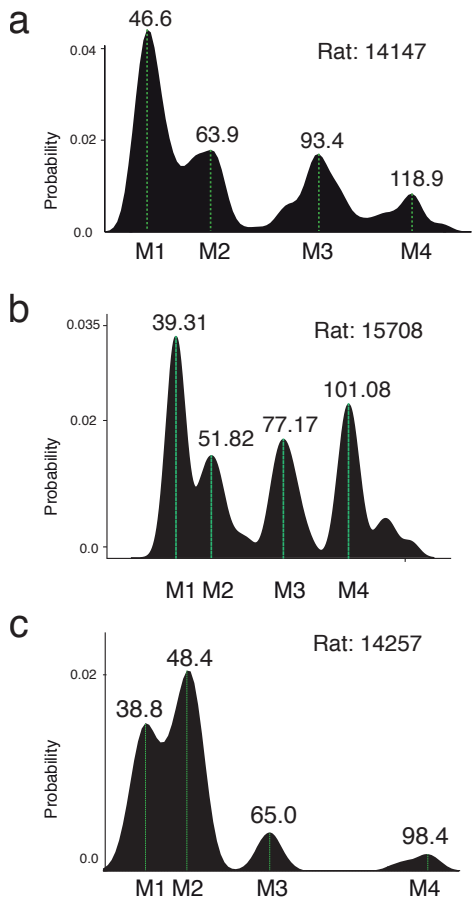
$$\delta\phi = 3\epsilon \frac{\lambda_2}{\lambda_1 - \lambda_2} \approx 3\epsilon m, \tag{D32}$$

where ϵ represents the fractional error in the estimate of grid period. Thus, particularly for smaller grid periods (corresponding to larger m), even small errors in grid period estimation can result in a large error in ϕ, making the errorbars in the estimation of ϕ from a single pair of periods large.

To obtain results with significant statistical certainty, we focus our analysis on published experimental studies that measure at least 50 grid cells per animal, spanning at least 3 distinct modules. This restriction results in grid period data sets for three rats — we present kernel density estimates of the module periods for each of them in Fig. 21 (Fig. 21c corresponds to the data presented in the main text in Fig. 5).

We have already demonstrated in Fig. 5 that Rat 14257 presents an extremely accurate match to the period ratio prediction for $\phi = 0$ (i.e., predicted period ratios of 2, 3/2, 4/3, 5/4,...); in addition, Rat 14147 (observed period ratios of 1.27 , $1.46 \approx 3/2$, $1.37 \approx 4/3$) and Rat 15708 (observed period ratios of 1.31, $1.49 \approx 3/2$, $1.32 \approx 4/3$) also match $\phi = 0$ very well (R^2 values of 0.999, 0.979, and 0.968 for Rats 14257, 15708, 14147 resp.) for all grid modules except for the module with the largest period.

Why is there an observed discrepancy for the grid module with the largest period? We propose four possible reasons for this discrepancy: Firstly, this discrepancy may be a result of the approximation made in arriving at Eq. (D17) — since the approximation is particularly accurate for $kd \gtrsim \mathcal{O}(1)$, the potentialy mismatch would primarily affect only the largest grid period module. Secondly, as demonstrated in Sec. D 9, the grid module corresponding to the largest grid period will have the least robustness to noise in the fixed-scale interaction, potentially introducing a large variance in the grid period for that module. Thirdly, as can be seen in Fig. 3h and Eq. (D32), the error in estimating the grid period for the first module ($m = 1$) is the most susceptible to errors in the value of ϕ Lastly, our predictions for grid period ratios Eq. (D1) are for the case of *simple* kernels that have a single spatial scale. A discrepancy at



observation	$r_m(1.26)$prediction	deviation
1.273	1.83	30.%
1.462	1.46	0.13%
1.371	1.31	4.65%

observation	$r_m(0.7)$prediction	deviation
1.310	1.90	31.%
1.489	1.48	0.61%
1.318	1.32	0.15%

observation	$r_m(0)$prediction	deviation
1.514	3/2 = 1.5	0.93%
1.343	4/3 = 1.33	0.75%
1.247	5/4 = 1.25	0.24%

FIG. 21. The 3 rats from Stensola *et al.* with 4 modules and their corresponding periods.

only the largest grid module may thus be suggestive of fixed-scale interactions that are primarily described by a single scale, with an additional low frequency perturbation at a larger spatial scale.

However, note that (particularly for Rats 14147 and 14257) there are relatively few grid cells observed from this largest period module, and the resulting uncertainty in period estimation may instead contribute to the error. In sum, apart from the possibility of some additional low frequency perturbations, the experimental data for rats with several simultaneously observed grid modules is largely consistent with the predicted period ratios for simple kernels with $\phi = 0$.

Skipped modules: Sometimes, neural recordings can miss a module. This can cause a large deviation from our predictions. For example, for a set of 5 modules following period ratios M4/M5 = 1.20, M3/M4 = 1.25, M2/M3 = 1.33, M1/M2 = 1.5. If recordings had missed module M4, the measured ratios would be M1/M2 = 1.5, M2/M3 = 1.33, M3/M5 = 1.5.

However, we do note that available data on multiple modules with a statistically large number of grid cells per module are quite sparse. To obtain further verification of our theoretical results, including the prediction of Eq. (D1) and even more specifically the hypothesis that ϕ is close to zero, additional data with multiple simultaneously observed grid modules will be important.

Appendix E: Lyapunov Function

The energy function of continuous time neural networks can be written as [125]:

$$E(\mathbf{s}) = -\frac{1}{2}\sum_{ij} s(i)W_{ij}s(j) + \sum_i \int_0^{s(i)} \phi^{-1}(s)\,ds - \sum_i I_i s(i), \tag{E1}$$

where $\mathbf{s}$ represents a vector of the synaptic activation at each neuron in the network, and I_i is the input bias to neuron i. For simplicity and since linear analysis does a remarkably good job in predicting the formed modules, let us restrict ourselves to the case of $\phi(x) = x$. Also, since the system is locally translationally invariant, we know that the dominant modes are going to be periodic. Hence, we may evaluate the energy function of the network dynamics (in the linearized regime) by assessing the energy of the periodic neural activity modes:

$$\mathbf{s}_k(\mathbf{x}) = A\sin(\mathbf{k}\cdot\mathbf{x} + \delta) + B, \tag{E2}$$

where $\mathbf{k} = k\hat{\mathbf{k}}$ is an arbitrary Fourier space vector, and A,B and δ are arbitrary constants. For these modes, we can write the energy function in the continuum limit as:

$$E[s_{\mathbf{k}}(\mathbf{x})] = -\frac{1}{2}\int d\mathbf{x}d\mathbf{x}' W(\mathbf{x},\mathbf{x}')s_{\mathbf{k}}(\mathbf{x})s_{\mathbf{k}}(\mathbf{x}') + \frac{1}{2}\int d\mathbf{x}s_{\mathbf{k}}(\mathbf{x})^2$$

Assuming that the system size L is large,

$$\begin{aligned}
2E[s_{\mathbf{k}}(\mathbf{x})] &= -\int W(\mathbf{x}-\mathbf{x}')[A\sin(\mathbf{k}\cdot\mathbf{x}+\delta)+B][A\sin(\mathbf{k}\cdot\mathbf{x}'+\delta)+B]d\mathbf{x}d\mathbf{x}' + \int [A\sin(\mathbf{k}\cdot\mathbf{x}+\Delta)+B]^2 d\mathbf{x} \\
&= -A^2\int d\mathbf{u}d\mathbf{v}W(\mathbf{u})\cos(\mathbf{k}\cdot\mathbf{u}) + A^2\int d\mathbf{u}d\mathbf{v}W(\mathbf{u})\cos(2\mathbf{k}\cdot\mathbf{v}+\delta) + B^2\int d\mathbf{x}d\mathbf{x}'W(\mathbf{x}-\mathbf{x}') + L(A^2/2+B^2) \\
&= -A^2L\int d\mathbf{u}e^{i\mathbf{k}\cdot\mathbf{u}}W(\mathbf{u}) + A^2\int d\mathbf{u}W(\mathbf{u})\int d\mathbf{v}\cos(2\mathbf{k}\cdot\mathbf{v}+\delta) + B^2\int d\mathbf{u}d\mathbf{v}W(\mathbf{u}) + L(A^2/2+B^2)/2 \\
&= -A^2L\tilde{W}(k) + LB^2\bar{W} + L(A^2/2+B^2), \\
&= -\text{constant}_1 \times \tilde{W}(k) + \text{constant}_2
\end{aligned}$$

where have used the simple trigonometric identity, $2\sin(C)\sin(D) = \cos(C-D) - \cos(C+D)$, and a change of variables, $\int d\mathbf{x}d\mathbf{x}' = (1/2)\int d(\mathbf{x}-\mathbf{x}')d(\mathbf{x}+\mathbf{x}') = \int d\mathbf{u}d\mathbf{v}$, with $\mathbf{u} = \mathbf{x}-\mathbf{x}'$ and $\mathbf{v} = \frac{1}{2}(\mathbf{x}+\mathbf{x}')$.

Thus, we obtain that the energy function $E[s_{\mathbf{k}}]$ is a simple linear function of the Fourier transform $\tilde{W}(k)$ of the recurrent weight matrix. The minimum energy solution corresponds to the Fourier mode that maximizes $\tilde{W}(k)$. In other words, the dynamics is dominated by the k^* that maximizes $\tilde{W}(k)$. This result, derived from an energy landscape perspective, is equivalent to the result in Eq. (B7), which we obtained earlier via perturbation analysis.

Appendix F: General formulation of module formation dynamics: Discrete peak selection via loss minimization

In Sec. E, we demonstrated how the pattern formation on the neural sheet can be derived via an energy minimization approach. Here, we use an energy landscape view to describe how loss function minimization results in modular solutions.

The key components for spatially modular solutions to arise from energy minimization are as follows: 1) A spatially-independent loss function $f(\theta)$ with multiple local maxima and minima; 2) A gradient in a spatially-dependent variable, $\theta_0(x)$; and 3) A coupling between the system parameters θ and θ_0, that results in a combined loss function

$$L(\theta, \theta_0(x)) = (1-\alpha)f(\theta) + \alpha\|\theta - \theta_0(x)\|^2 \tag{F1}$$

Under appropriate constraints on $f(\theta)$, solving the following optimization at each x

$$\theta^*(x) = \arg\max_\theta L(\theta, x) \tag{F2}$$

will produce discrete, step-like changes as a function of x. This happens because the smooth minimum given by the $||\theta - \theta_0(x)||^2$ term effectively selects one of the local minima in $f(\theta)$ as the global minimum. As the function $||\theta - \theta_0(x)||^2$ slides smoothly along with x, the peak of $f(\theta)$ selected as the global minimum remains the same for some time, then jumps abruptly. These step-like changes are modular solutions to the global optimization problem. The energy function defined in Eq. (F1) can be viewed as a regularized optimization problem, with the spatially-dependent regularizer $||\theta - \theta_0(x)||^2$ acting as a prior that selects one of the minima of $f(\theta)$ at each location (Fig. 22).

The correspondence of this general picture with the peak selection mechanism described in the main text follows directly with the following identifications: the spatially independent nonlinear loss function $f(\theta)$ with the fixed-scale interaction W^f; the spatially varying parameter prior $\theta_0(x)$ with the graded scale $\sigma(n_{DV})$ of the pattern-forming kernel; the combined loss $L(\theta, x)$ with the full kernel $W_{n_{DV}}$; and the spatially-varying, multi-step-like set of optima $\theta^*(x)$ with the grid periods $\lambda^*(x)$, respectively. Similar to peak selection for grid cells, the formed modules in this generalized setting will also inherit topological robustness and stability.

We demonstrate a numerical example of this in Fig. 7, where we construct $f(\theta)$ as a random sample from a Gaussian process with a radial basis function kernel, and simulate gradient descent dynamics on the loss function $L(\theta, \theta_0(x))$. To prevent the dynamics from getting stuck in local minima of L, we simulate the gradient descent first purely on the regularization term, with gradually increasing strength of the rugged loss function, through gradually decreasing α with increasing time.

Although we primarily focused on the peak selection process in Fourier space for multi-periodic patterning in grid cells, we also showed that it has a general formulation in terms of dynamics on an energy landscape: One (spatially invariant) interaction sets up an optimization problem with multiple local minima, while a second (spatially graded) interaction defines a locally shallow single-optimum landscape, with a smoothly shifting optimum as a function of space. Thus, the shallow optimum selects one of the narrow local optima as the global optimum, with discontinuous jumps to the next local minimum even as the parameters vary smoothly. This analytical formulation provides a simplifying mathematical perspective on how smooth gradients could lead to discrete patterning and modular specialization in the brain and body [31, 40, 62].

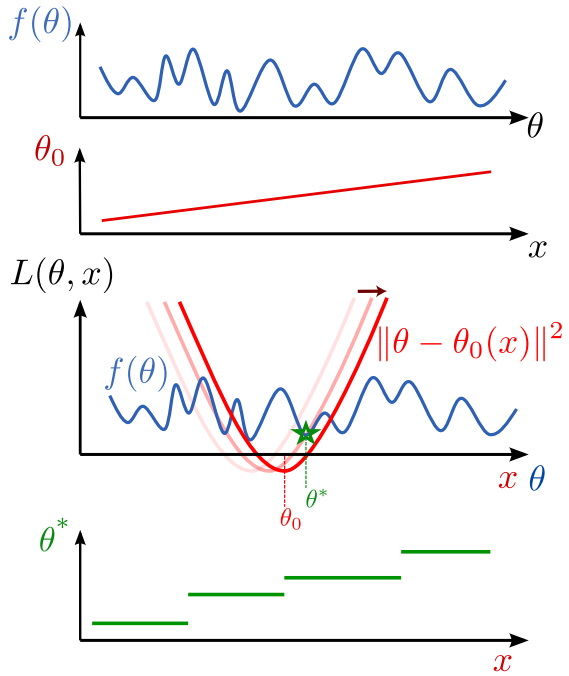


FIG. 22. **A general setting for peak-selection** Assuming a loss function $f(\theta)$ (blue) and a spatially dependent quantity θ_0 (red), a combined loss function $L(\theta, x)$ can be constructed such that the x-dependent optimizer of $L(\theta, x)$ will be modular (green), since it will be constrained to correspond to one of the minima of $f(\theta)$.

Appendix G: The emergence of modules corresponds to the formation of localized eigenvectors

As has been observed before [126], a neural network endowed with slowly varying local interactions shows diverse timescales that are spatially localized: different parts of the network respond with disparate temporal dynamics. We also find a localization of eigenvectors in our multi-module grid network, Fig. 23A. Similar to [126], our interaction matrix has a locally circulant form (due to the slowly varying gradient in lateral inhibition width). This is a signature of a phase transition, similar to the Anderson localization transition in condensed matter physics [53]. The eigenvectors for a regular pattern forming interaction in traditional continuous attractor models are delocalized fourier waves which are then transformed into localized fixed-wavelength gaussian wavepackets with the addition of the gradient and fixed scale interaction.

We find that in the resulting set of localized eigenvectors, each has a different but constant period, Fig. 23B. These periods exactly match the spatial periods of the modules formed in steady state. In sum, the locally circulant matrix gives rise to eigenvector localization, and the localized eigenvectors correspond to the modules.

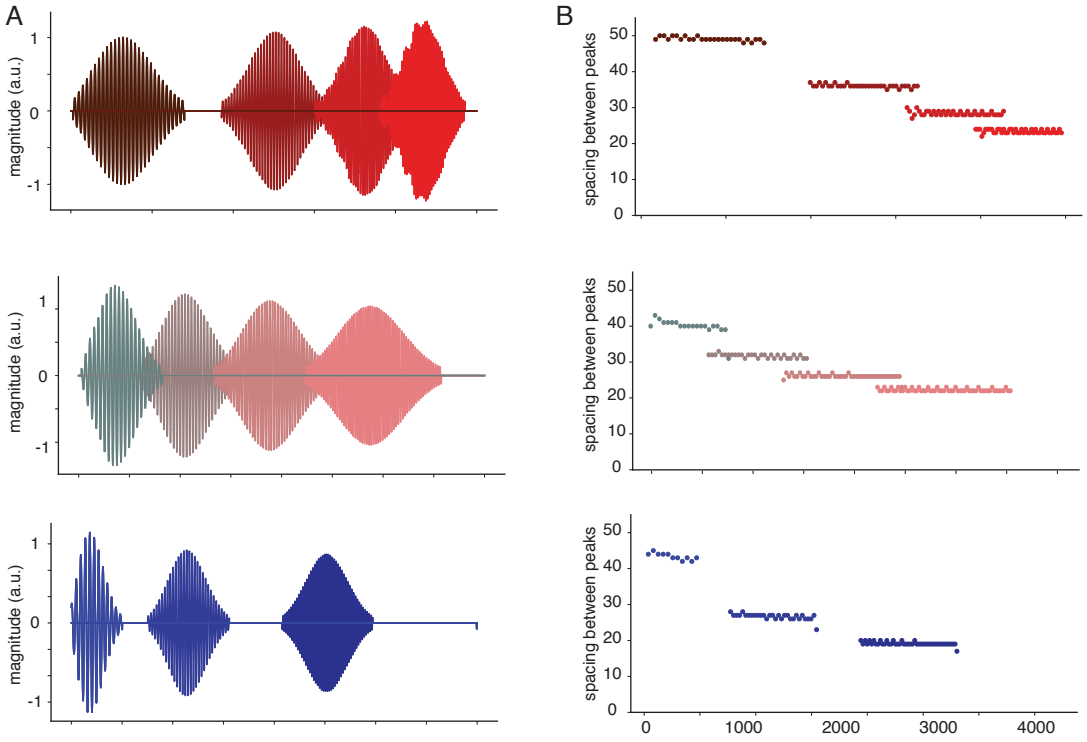


FIG. 23. **Localization of eigenvectors**: A) Eigenvectors of various one-dimensional interaction weight matrices along with the corresponding inter-peak spacings are localized, B) The periodicity within an eigenvector is constant.