From smooth cortical gradients to discrete modules: spontaneous and topologically robust emergence of modularity in grid cells

Mikail Khona [†1,2], Sarthak Chandra [†1], and Ila R. Fiete [*1]

[1]Department of Brain and Cognitive Sciences & McGovern Institute, MIT
[2]Department of Physics, MIT
[†]Denotes equal contribution

Abstract

The grid cell system is a paradigmatic example of the computational advantages of modular representations. Modular responses emerge early in juvenile animals along a strip of cortex with smooth gradients in several biophysical neural properties. The mechanism that produces grid modules from smooth gradients is unknown. We show that spatially and functionally discrete grid modules can spontaneously emerge from lateral interactions that are continuously graded in scale, when combined with a fixed-scale interaction. We derive a comprehensive analytic theory to reveal that modularization is highly generic and robust, arising from a topological process. The theory generates universal predictions for the sequence of grid period ratios that strikingly do not depend on any microscopic details and furnish the most accurate explanation of grid cell data to date. Altogether, this work reveals novel principles by which simple local interactions and smooth gradients lead to macroscopic modular organization in cortical circuits.

Modular structures are robust to localized perturbations [1, 2], faster to adapt if the world requires sparse or modular changes [3], relatively expressive because of compositionality [4, 5, 6], and probably for these reasons empirically favored by the arrow of evolution, which drives systems toward greater modularization [7]. In this sense, modularity is the crux of biological organization.

However, the mechanisms driving modularization in circuits and brains are not well understood. Most models in which modular structure emerges involve extensive supervision, selection, or feedback [8, 9, 10, 11, 12]: it is therefore a slow, top-down process based on direct function optimization. Some of these top-down models of modularization involve evolution in modularly changing environments [8, 13, 14] or learning on multiple modular tasks [10, 11, 12, 9][1]. Though selection is a key mechanism for the emergence of complex structure in biology, it is too slow to explain structure emergence within the lifetime of single individuals from simpler biophysical biases [17]. Similarly, extensive supervised training in modularly changing environments or on multiple modular tasks is often interpreted as meta-learning on evolutionary timescales rather than as a mechanism for the rapid formation of architectural features and modules in single brains.

Here we explore modularity emergence as a process of self-organization through spontaneous symmetry breaking. Just as simple local interactions lead to pattern formation for morphogenesis and memory dynamics in brains [18, 19, 20, 21, 22], we show with grid cells that similar processes can also provide critical inductive biases or predispositions that rapidly and robustly drive systems toward global modularization without feedback.

Grid cells, found in the mammalian cortex (specifically the medial entorhinal cortex or MEC, Fig. 1a, left), generate spatially periodic activity patterns as a function of animal location during navigation [23] (Fig. 1b, left). They are believed to compute updated estimates of location through integration of self-motion cues [23, 24, 21]. Multiple cellular biophysical properties vary smoothly and continuously along the MEC long axis, including cell resistance, activation threshold, time-constant, and local inhibition (Fig. 1a, right) [25, 26]. Nevertheless, the dynamics of grid cells vary in discrete, modular fashion along the same axis: Nearby cells exhibit identical spatial tuning periods and orientations with offset phases [27] that are preserved across time, environments and behavioral states [28], and there are a few discrete populations of grid cells with different periods [29], as predicted by grid cell models [30, 31]. These discrete populations, 5-6 in total per animal, are defined as modules

*Corresponding author

[1]Some models exhibit modularity emergence through motif duplication followed by specialization [15, 16], though the resulting modules typically require selection or supervision under the conditions noted earlier to remain stable.

[30, 29] and are functionally modular [32, 33], contributing separate representations of animal location through a multi-periodic population code.

This multi-periodic, non-local, modular encoding of a non-periodic and local variable (animal location) has distinct advantages relative to non-modular codes: it generates, at fixed resolution, a coding range that grows exponentially in the number of modules and thus in the number of neurons [34, 35, 36], is a strong intrinsic error-correcting code [37], and can in principle flexibly represent higher-dimensional spaces by overcoming the curse of dimensionality [38].

How does modular dynamics emerge in the grid cell system, and how does it relate to the underlying biophysical gradients in the system? Explaining how modules originate from smooth biophysical gradients is critical to understanding how relatively low-information genetic codes may be translated into complex biological structures.

Continuous attractor models of grid cells capture the empirically measured properties of single modules and generate a number of novel predictions about population dynamics [21] that have been verified [28, 33, 32, 39, 40, 41]. Models trained with simple local unsupervised spike-time dependent plasticity can explain the emergence of grid patterning, attractor dynamics, and path integration within a module [42]. Training recurrent networks with supervised learning to perform path integration and localization can also result in grid-like activity patterns with multiple scales [43, 44, 45]. These results depend on multiple hyperparameter choices; this is apparent since there are non-grid-like solutions to path integration and localization objectives [46]. In supervised RNN models, emergence requires extensive supervised training inconsistent with the rapid emergence of modular grid activity in animals, involves biologically implausible learning rules, does not relate to underlying biophysical gradients, and the grid-like patterns are highly sensitive to post-hoc modeling decisions about network size, shape, and the detailed functional form of target readout activity.

Two recent works address the question of rapid module emergence, and are moreover based on the existence of biophysical gradients: In the first, grid tuning is derived from a pattern formation process acting upon location-coding place cell inputs with a continuum of adaptation timescales [47], and in the second, grid module formation proceeds from a stacked multi-sheet architecture in which each sheet has lateral grid pattern-forming connectivity with a scale that increases vertically, and the sheets are coupled vertically through 1-1 connections between matching locations on each sheet [48]. In the former, the resulting grid-like cells lack the path integration functionality posited as central to grid cells [49, 30] and lack their demonstrated low-dimensional structure in which cell-cell relationships are preserved across environments and states [50, 28, 40, 51]. The latter possesses path integration functionality, but the model assumes a pre-existing proto-modular architecture with structured and tuned one-to-one linkages of neurons across modules [48]. Both works lack a theoretical analysis of the dynamics of module emergence.

We present a novel pathway for the spontaneous and rapid emergence of modularity from smooth cortical gradients acting in concert with a non-graded interaction. Modularity arises through a topological mechanism that confers extreme invariance to nearly all details, parameters, and perturbations, and does not require supervision or learning. The mechanism admits a complete analytical description of why modules emerge, which reveals the genericity of the process and suggests that the insights should generalize from grid cells to other systems in biology and machine learning. Finally, we generate detailed and almost entirely parameter-independent predictions about the entire sequence of grid module period ratios that, furthermore, provide the best fit to data to date.

Results

Smooth biophysical gradients alone produce graded period variations

We begin by incorporating smooth biophysical gradients along the long axis of MEC (Fig. 1a) into continuous attractor models of grid cell emergence, to examine whether these gradients combined can produce discrete modules. In continuous attractor models, strong local inhibitory interactions [30, 52, 22] drive a pattern formation process that generates stable activity bumps organized in a triangular lattice on the cortical sheet [33]. The periodicity of the cortical pattern is set by the spread of the lateral interactions, SI Fig 9 [21]. We refer to these interactions as the pattern-forming interactions.

We consider how the prominent but smooth biophysical gradients in cellular and circuit properties along the long axis of MEC (Fig. 1a) affect continuous attractor models, and how they can together result in the formation of a discrete set of grid cell modules with distinct population-level dynamics and functionality in each module (Fig. 1b). A smooth gradient in the lateral pattern forming interaction width (W^g, with scale $\sigma(n_{DV})$, where n_{DV}

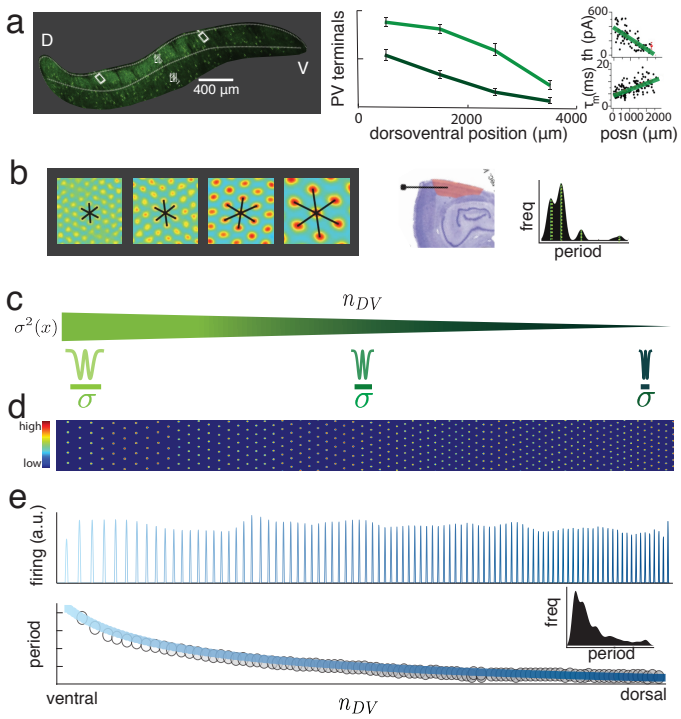


Figure 1: **Smooth gradients underlie entorhinal grid cell modules.** (a) Smooth gradient in inhibitory (PV-immunoreactive) axon terminals along the dorsoventral (DV) axis in layers II and III of the medial entorhinal cortex (MEC) [25], where grid cells are found [49]. (b) The spatial tuning curves of grid cells (left: autocorrelation of spatial responses of four cells; intensity plot as a function of two spatial dimensions) vary in spatial period along the DV axis in MEC (region and electrode penetration angle shown in middle), with period increasing ventrally. However, the variation is discrete: there are a few discrete modules with a few periods, and nearby cells have the same period (right) (adapted from [29]). (c) In a continuous attractor model of grid cells formed by pattern formation through lateral inhibition [30], we model the biophysical DV gradients in MEC by a continuous gradient in the width of lateral inhibition. (d-e) The smooth gradient of (c) results in pattern formation with smoothly varying periodicity in the population pattern, with no modularity (in (d): neural population activity pattern in a 2-dimensional neural strip; (e): top, pattern in a 1-dimensional neural strip; bottom, extracted period as a function of DV location; inset, histogram of extracted periods).

refers to position along the long axis of MEC, Fig.1c) results in the formation of hexagonal activity patterns across MEC, but the pattern period varies smoothly with the interaction gradient: there is no modularization (Fig. 1d-e).

Interplay of lateral interactions with fixed and graded scales leads to modularization

We next consider the addition of a fixed-scale interaction constraint across MEC (W^f, with scale d that is fixed but wider than $\sigma(n_{DV})$ everywhere): this scale could represent a general constraint on the extent of neural arborization, independent of position, Fig. 2a (orange). With the addition of a fixed-scale interaction, the network spontaneously and robustly decomposes into discrete modules, with spatially periodic activity patterns and discontinuous jumps in period in-between, Fig. 2b-c.

The formed modules are not in some simple way given by the larger-scale fixed interaction: each module is generally much larger than both interaction scales $d, \sigma(n_{DV})$ (cf. modules in Fig. 2b with the width of the graded and fixed interactions shown to scale in the insets, green and orange). The same combination of a smoothly graded interaction and a fixed-scale interaction produces robust and spontaneous decomposition of dynamics into discrete modules across 1D and 2D network models, Fig. 2c,g. (See SI Sec. 3.8 for additional numerical results for 2D.)

Further and remarkably, the results are strikingly similar across diverse shapes of the local interactions, both for the graded and fixed scales (cf. Fig. 2a-f), and regardless of whether the fixed-range interaction is localized at a distance d (Fig. 2a) or spatially diffuse over d (Fig. 2d), suggesting that a general, robust, and relatively parameter-independent mechanism underlies module emergence (see SI Fig. 10 for simulations with variety of additional lateral interactions shapes, including both inhibitory and excitatory fixed-scale interactions).

Modularity in function: Independent velocity integration per module

When all neurons in the MEC network model are driven by a velocity input, the activity pattern flows across the cortical sheet (reflecting the velocity-integration function of grid cells). However, the boundaries between

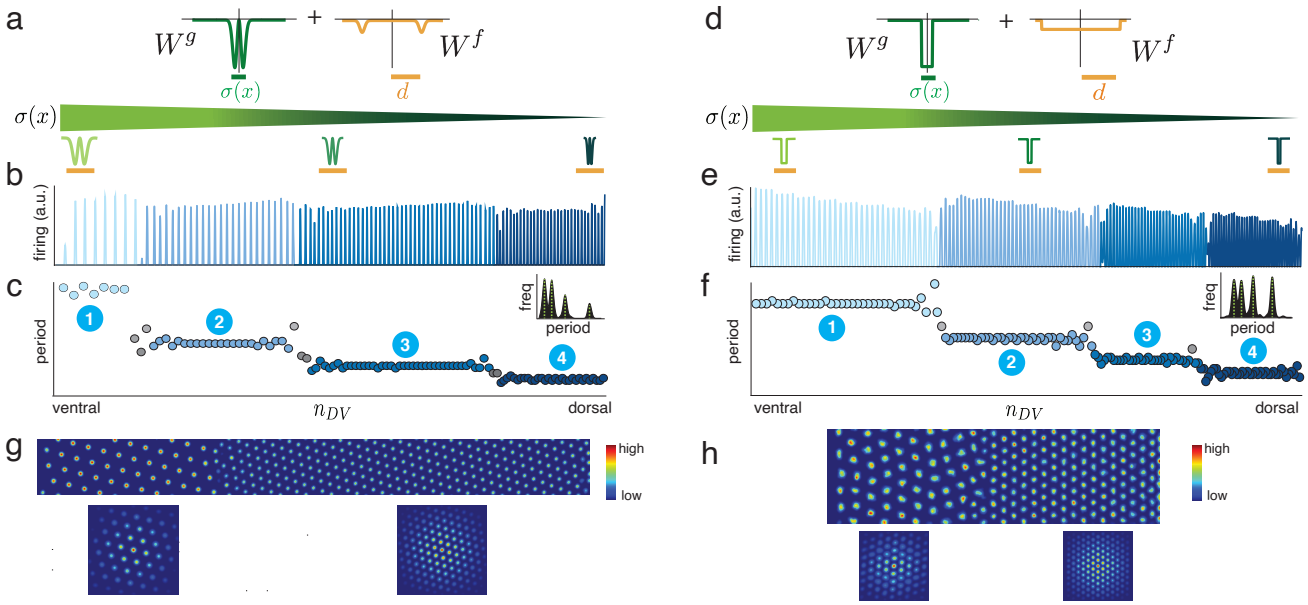


Figure 2: **A fixed connectivity range combined with smooth gradients in lateral connectivity leads to spontaneous module emergence.** (a) Two kinds of local lateral interactions: one whose width ($\sigma(n_{DV})$) varies smoothly along the DV axis (W^g, green) and one whose width (d) remains fixed along the neural strip. (b-c) The interactions from (a) lead to self-organized module formation (activity pattern in 1-dimensional neural strip with smaller periods more dorsally, (b); extracted periods, (c)). (d-f) Same as (a-c), but for a very different pair of W^g, W^f. (g-h) 2-dimensional pattern formation and module emergence: same as (b), for a 2-dimensional neural strip. (Bottom) 2d autocorrelation function of the local (single-module) patterns in the neural strip.

modules remain sharp and fixed at the same cortical locations despite pattern flow, Fig. 3a. A shifting frontier would result in non-regular spatial tuning both at the boundaries of modules and also likely in the interior since shifting boundaries would also affect patterning in the bulk. The preservation of a sharp boundary at a fixed location results in veridical velocity integration within and across all modules, and periodic spatial tuning curves per cell, Fig. 3b, but with discretely different spatial periods per module. Thus, the network is not only modular in cortical pattern formation, but also fully modular in its velocity integration functionality.

Theory of modularizaton: genericity and robustness of mechanism with topological protection

We have seen in Fig.2 that the decomposition of a network with smoothly graded and fixed-scale lateral interactions into modules is remarkably robust, suggesting a general principal at work. We seek an analytical view of module formation that sheds light on its mechanisms and robustness.

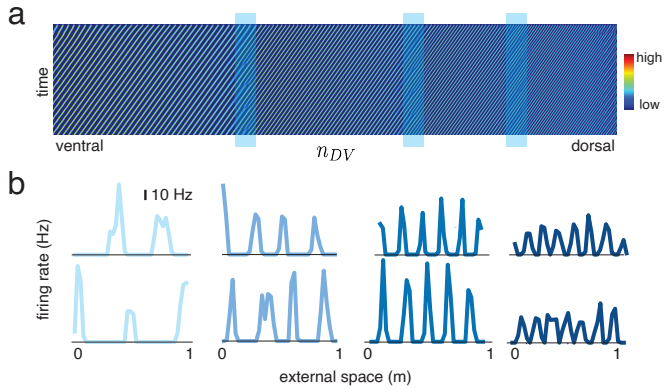


Figure 3: **The formed modules integrate velocity inputs independently and thus function as independent modules.** (a) Network response to a constant velocity signal injected globally into all cells: Within a module, bumps move coherently to maintain the pattern period in response to velocity input. Between modules, the boundary locations remain fixed (highlighted by vertical light-blue bars) while relative between-module phases change as each module's phase updates differentially relative to its neighbors. ; (b) Spatial tuning curves of 2 example cells per module (top and bottom). Tuning curves are generated by collecting spikes in response to a sinusoidally varying velocity input corresponding to multiple left-right traversals of a 1-dimensional environment and plotting the histogram of spike counts as a function of position in the environment.

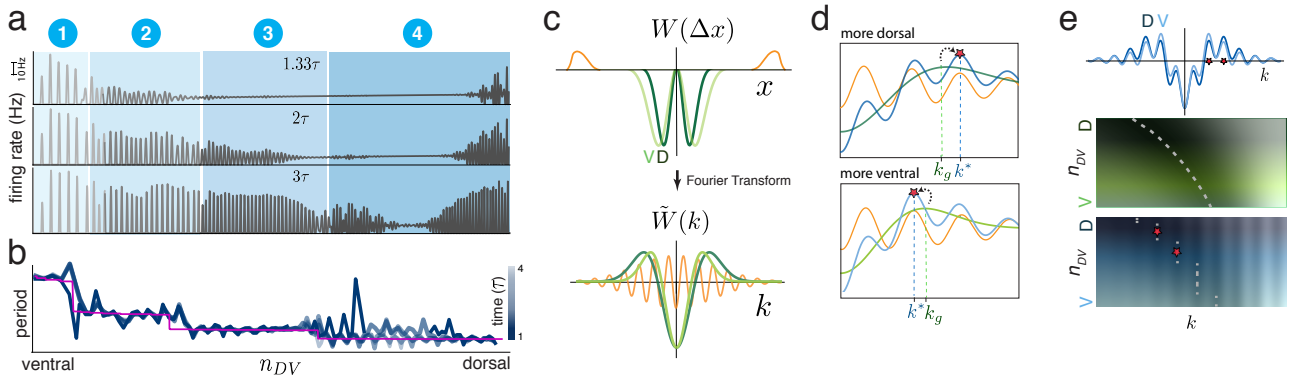


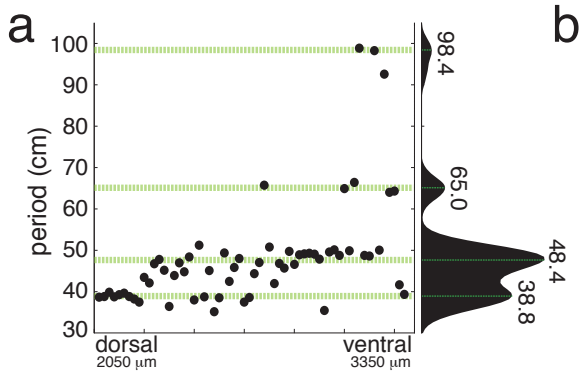
Figure 4: **Theory of module formation: multi-scale linear instability and topological peak selection.** (a) Snapshots of population activity in the neural strip within a few neural time-constants (τ) of running the dynamics. (b) Modules appear at the same time as the local pattern, and both appear before most neurons have hit their nonlinear thresholds. (c) Top: Schematic of the fixed-range (orange) and graded (green) interactions from two different DV locations on the neural strip (top), and bottom: their Fourier transforms. Fourier peaks of the fixed-range interaction remain unchanged dorsoventrally but the graded interaction peak slides smoothly inward when moving ventrally. (d) Peak selection process: the global maximum in Fourier space is based on the combination (blue) of the shallow graded-interaction peak (green) and the multiple narrow fixed-interaction peaks (orange). The smoothly shifting graded-interaction peak "selects" which of the local maxima from the fixed interaction is the global maximum. Thus, the global maximum jumps abruptly from one local maximum to the next. (e) The Fourier transform of the combined interactions, for a dorsal and a ventral location (top), and the Fourier-space location of the global maximum of just the graded interaction as a function of DV location (middle) and the combined interaction (bottom).

The dynamics of the network after initialization in a uniform activity state exhibits immediate signs of modularization, beginning at the earliest time-steps (within 1-3 τ, where $\tau \approx 30$ms is the time constant of the network), Fig. 4a-b. Modularization begins before most neurons have crossed their nonlinear thresholds and unfolds concurrently with the process of local patterning (Fig. 4a,b). This observation provided us with the critical hint that it should be possible to explain both local pattern formation and larger-scale module emergence within a unified linear instability framework.

The presence of localized eigenvectors is also consistent with the emergent modularization in the system: the eigenvector of the interaction kernel is a characteristic function that grows exponentially in a linearized approximation of the system — localization of the eigenvectors is thus another signature of the local modularization of the neural sheet (cf. SI sec 6).

Multiscale perturbation theory Accordingly, we perform an analytical derivation of the dynamics of module formation through multiscale linear instability (details in SI Sec.1). For a sufficiently slow gradation in the width of the pattern-forming interaction, the system is locally translation-invariant. Thus, we can decompose the local dynamics at each location n_{DV} in the neural strip into Fourier activity modes and examine their stability. Local pattern formation is dominated by the fastest-growing mode, which is given by where the Fourier-transformed combined interaction has its peak ($k^*(n_{DV}) = \arg\max_k \tilde{W}(k; n_{DV}) = \arg\max_k \{\tilde{W}^f(k) + \tilde{W}^g(k; n_{DV})\}$ if $\tilde{W}(k^*) > 1/\tau$) [53, 54, 55, 56, 30, 57]. The result is a locally patterned state with period $\lambda^*(n_{DV}) = 2\pi/k^*(n_{DV})$.

To understand the spatial structure of these periodic solutions along the neural strip, note that $\tilde{W}^f(k)$ prominently contains a set of maxima, narrowly spaced every $1/d$, because of the dominant spatial scale d in W^f. Because W^f is fixed along the neural strip, the set of local maxima in $\tilde{W}^f(k)$ remains the same (Fig. 4c, orange). On the other hand, the narrow graded interaction $W^g(n_{DV})$ exhibits a very broad Fourier peak (of scale $\sim 1/\sigma(n_{DV}) \gg 1/d$) that smoothly contracts more ventrally (Fig. 4c, green). Though the graded interaction drives pattern formation and is ultimately responsible for any variation in spatial period along the neural sheet, its peak is too shallow to fully determine the global maximum of the combined interaction. Rather, the narrow peaks from $\tilde{W}^f(k)$ determine the set of potential locations of the global maximum, while the smoothly moving peak from $\tilde{W}^g(k)$ performs "peak selection" (cf. SI Movie 1) on these possibilities through a "stepping mechanism": for positions along the neural strip, one of the narrowly spaced fixed-interaction peaks will be the global maximum; moving ventrally causes the peak of $\tilde{W}^g(k)$ to move smoothly leftward while the global optimum remains fixed, until the leftward shift has progressed far enough that the next peak of $\tilde{W}^f$ on the left becomes the global maximum, Fig. 4d-e. Thus, moving smoothly along the neural strip results in a constant-period intervals, with discrete and rapid steps through discrete peak selection. The theoretical derivation of all module periods accurately matches those from numerical simulation, Fig. 4b (see SI Fig. 10 for a comparison of theoretical predictions and numerical simulation for a variety of lateral interactions shapes).



b

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

Figure 5: **Detailed match between prediction and measurement of the sequence of module period ratios.** (a) Observed periods of grid cells from multiple modules [29] (b) 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).

Topological robustness of peak selection Is it possible to determine what properties of the network and interaction kernels determine the periods of the formed modules, their sizes, and how many modules will form? Our results reveal that these quantities emerge through a topological process, and thus that they are topologically protected — robust to nearly all parametric variations.

In particular, the periods of different modules are given by different peaks in $\tilde{W}^f$, the Fourier-transformed fixed-scale interaction, and if this interaction has a dominant scale $d \gg \sigma(n_{DV})$, these peaks will occur at multiples of $1/d$. Thus, the module formation process and module periods are independent of details of the shape of the W^f and W^g, and very large variations in the shape still result similar outcomes as seen in Fig. 2 and the additional profiles in SI (SI Fig.10).

The number of formed modules is independent of the shape of the lateral interactions, of the length of the neural sheet, and of the shape of the function describing how the lateral interaction scale $\sigma(n_{DV})$ varies across the neural sheet. The number of modules is given by how many integer multiples of $1/d$ fit into the interval $[1/\min(\sigma), 1/\max(\sigma)]$ (SI Sec. 3.7). If the extremal widths are fixed, the modularization process generates the same number of modules regardless of neural sheet size (cf. SI Fig.15) and the details of the boundary conditions (periodic or aperiodic [21]). Thus, the number of modules depends on the graded interaction parameters only through the extremal widths $\min(\sigma), \max(\sigma)$, and not on any details of how the interaction scale varies between the extremal values (SI Sec. 3.7 and SI Fig.14). As a corollary, because the number of modules depends only on the extremal values of the gradient but not the network size, it follows that the average module size grows linearly with network size (if the extremal values are held fixed). This explains why the formed modules in Fig. 2b,e are much larger than and unrelated to the length-scales of both the lateral interactions. Thus, module formation exhibits self-scaling with the size of the neural sheet, producing the same relative module sizes and same module periods regardless of the size of the physical substrate. Furthermore, small changes in the extremal values also have no effect on the number of modules, until a critical point at which another module would be added; thus, the number of modules is a topological invariant [58].

In sum, the module formation mechanism is at its essence a topological process with each module arising — through a combination of interactions with a smoothly varying scale and a fixed larger scale — as a topologically protected phase[58] that is independent of most details and parametric variations (SI Sec.3.7).

Energy landscape view Finally, the peak-selection process can be reformulated as minimization on an energy landscape (SI Sec. 4). In other words, whenever two functions together define an optimum that governs the local dynamics, one with a fixed set of narrowly spaced minima and another with a broader but shallower minimum that moves smoothly as a function of some spatial gradient, the dynamical system will form a step-like solution as a function of the spatial gradient. Thus, the peak-selection principle for module emergence applies to a range of dynamical systems beyond those that exhibit spatially periodic pattern-forming solutions as we have seen for grid cells, SI Sec. 5.

Detailed prediction of pairwise module period ratios

Returning to grid cells, the theory generates a quantitative and topologically robust prediction about how the entire sequence of grid module period ratios. We assume that the interaction kernel W^f is *simple*, i.e., it is primarily characterized by the length-scale $d \gg \sigma(n_{DV})$, without other prominent features, and any other length-scales in

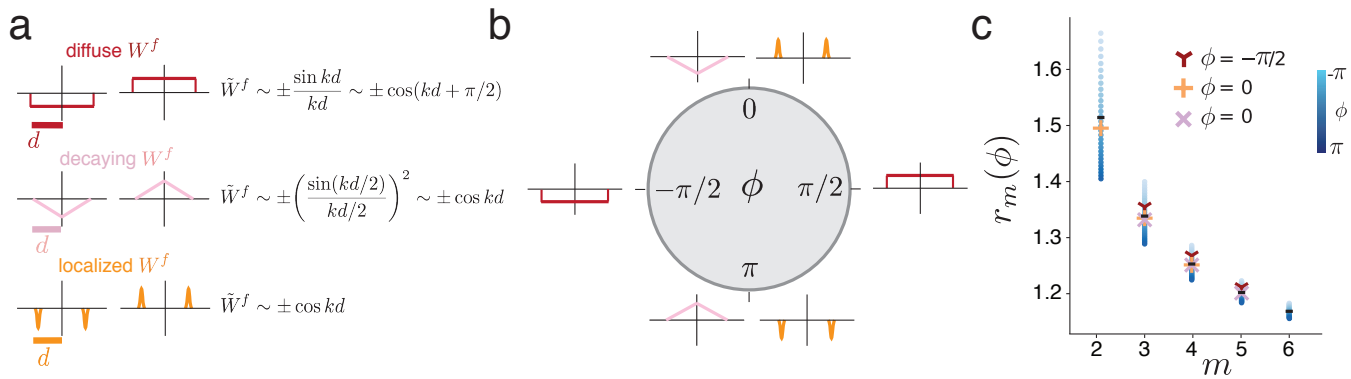


Figure 6: **Module period ratios depend weakly and predictably on the fixed-scale interaction shape.** (a) Left: examples of diverse fixed-scale interaction profiles that produce modularization: the profiles can be described as diffuse, decaying, or localized; right: the dominant terms in their Fourier transforms $\tilde{W}^f$. (b) The phases of the Fourier transforms for the diverse simple interactions in (a), all of which follow the same formula, Eq. 1, for period ratios. (c) The theoretically predicted sequence of period ratios for all values of ϕ (blue) for module numbers 2-6, together with numerical simulation of neural circuit models with three of the interaction profiles shown in (b). The variation in period ratio as a function of ϕ (and hence the sensitivity to the shape of the fixed-scale interaction) is smaller for higher module numbers (smaller module periods).

W^f are much smaller than d. Then, except for a set of measure zero in the space of functions W^f (SI Sec. 3.5.1 for details), its Fourier transform has an approximate form ($\sim \cos(kd-\phi)$) with relatively evenly spaced peaks occurring every m/d for integer m (Fig. 6a) together with a single phase ϕ; as we discuss next, this phase depends only on gross properties of the shape of the fixed-range interaction. Since module periods correspond to successive peaks of this function, it follows that the period ratio is given by (SI Sec. 3.6):

$$r_m = r_m(\phi) = (m+1+\phi/2\pi)/(m+\phi/2\pi). \tag{1}$$

The predicted average period ratio from this formula for the first three period ratios is 1.37 (averaging also over all ϕ), in good agreement with known results [29]. But further, and unlike existing experimental fits in which the period ratios $r_m = \lambda_{m+1}/\lambda_m$ for all m are described by a single parameter $\alpha \sim 1.4$, the prediction of the present model gives further detail on the sequence of successive period ratios as a function of m. We compare our more fine-grained successive period ratio prediction with experimental results, Fig.5b. The prediction with $\phi = 0$ matches the sequence of observed period ratios strikingly well for these data and others for which multiple grid periods are available from the same individual (SI Sec. 3.10).

Period ratios depend only slightly and predictably on lateral interaction shapes We derive analytically that a small number of characteristics of the fixed-width interaction W^f determine the value of ϕ in the prediction of period ratios, as illustrated in Fig. 6. 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 lead 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 described forms of fixed-scale lateral interactions (See Fig. 10 for examples of numerical simulations for several cases).

Different values of ϕ inferred from experimental period ratio observations might therefore be used to constrain an inferred shape of the fixed-scale lateral interaction kernel (under the assumption that W^f has a single dominant scale; in the case of multiple dominant scales, more-complex kernels can also be consistent with $\phi = 0$, SI Fig.12).

Robustness to inhomogeneities and local perturbations

Peak selection-based modularization is invariant to global variations in parameters, function shapes, and parameter gradients, as seen above. We now show that it is also robust to several forms of local heterogeneity and perturbation.

Robustness to spatial heterogeneity in lateral interactions Since emergent modules are determined by the set of maxima of $\tilde{W}^f(k)$ with selection by $\tilde{W}^g(k; n_{DV})$, our theoretical expectation is that spatial heterogeneities in W^g should not affect module periods. To test this, we construct noisy pattern-forming interaction weights, independently per neuron, that break radial symmetry and introduce spatial heterogeneities in scale (See SI Sec. 3.9 for more details)

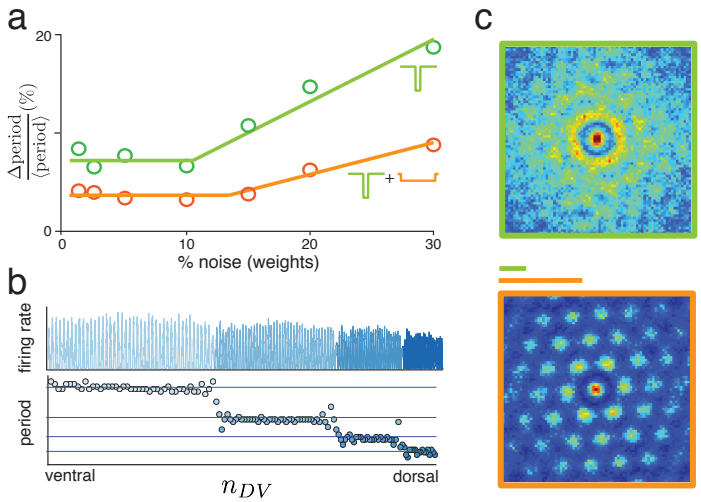


Figure 7: **The fixed-scale lateral interaction confers within-module robustness to weight noise and heterogeneity.** (a) The addition of independent random noise to all pairwise weights between neurons in a 1-dimensional grid cell model (single module, with only the pattern-forming lateral interaction W_g, as in standard continuous attractor models [21]) results in local variations in the formed pattern. Pattern variation is measured as the ratio of the standard deviation to the mean of the pattern period (green). The addition of the second (fixed-scale) lateral interaction suppresses the effects of noise in the pattern forming weights (orange). (b) Module formation in a 1-dimensional network with noise in W_g: the within-module periods remain flat and the between-module transitions remain sharp. (c) Top: Spatial patterning that is destroyed by heterogeneity in W_g (shown is the pattern autocorrelation) is rescued by the addition of a secondary interaction W_f, even with noise in both W_f and W_g (bottom). Simulation performed on a 100x100 neuron network. The pattern forming kernel W_g in a-c is from Fig. 2d-f.

In accord with the theoretical expectation, the peak selection mechanism strongly suppresses the effects of this noise, Fig. 7a, and the periods across comodular cells remain constant and continue to correspond well to analytical predictions, Fig. 7b. Robustness in the module-formation setting with the interaction of a fixed and graded-scale interaction greatly exceeds that seen in the corresponding conventional (single-module, single-scale interaction) network, Fig. 7a-b.

Additional heterogeneities in the fixed-scale interaction can result in deviations from periodicity in $\tilde{W}^f(k)$, and thus lead to variations in the emergent grid period. However, as we show in SI 3.9, the variation in the mth grid module is suppressed by the factor $\sim 1/m$, resulting in additional robustness to grid modules corresponding to larger m (smaller periods). This results in robust pattern formation even in the presence of spatial heterogeneities in all lateral interactions, Fig. 7c.

Robustness to large localized perturbations in activity Since the formed modules emerge locally and are functionally independent (cf. Fig. 3), we expect that localized perturbations to even large fractions of a module will not affect the periodicity and function of the rest of the modules. Indeed, upon silencing activity in one module (mimicking optogenetic inactivation) the other modules, module periods, and module boundaries remain stable, Fig.8a. When a subset of cells in one module are driven to maintain randomly chosen values between 0 and 20Hz, patterned activity is disrupted in adjacent regions of the module, Fig.8b; when the perturbation is lifted, the module period is restored at the site of perturbation as well. If a region that spans 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.8c. In all cases, the pre-perturbation periodicities are restored within $\approx \tau$ timesteps. In sum, our model predicts that localized perturbations of grid cells should leave nearby grid cells and other grid modules unaffected, in contrast to other models of grid cell modularization [48].

Discussion

We have found that the combination of a fixed-range lateral interaction with a spatially graded lateral interaction in continuous attractor grid cell models leads spontaneously to the formation of discrete modules. The process involves the selection of a global peak in Fourier space from a set of fixed local maxima; the spatially graded interaction defines a smoothly shifting shallow optimum that causes the global maximum to jump from one local maximum to the next. This mechanism of "peak selection" is a novel recipe for bottom-up modularization in neural circuits, which leads to functionally independent modules. The number of formed modules is independent

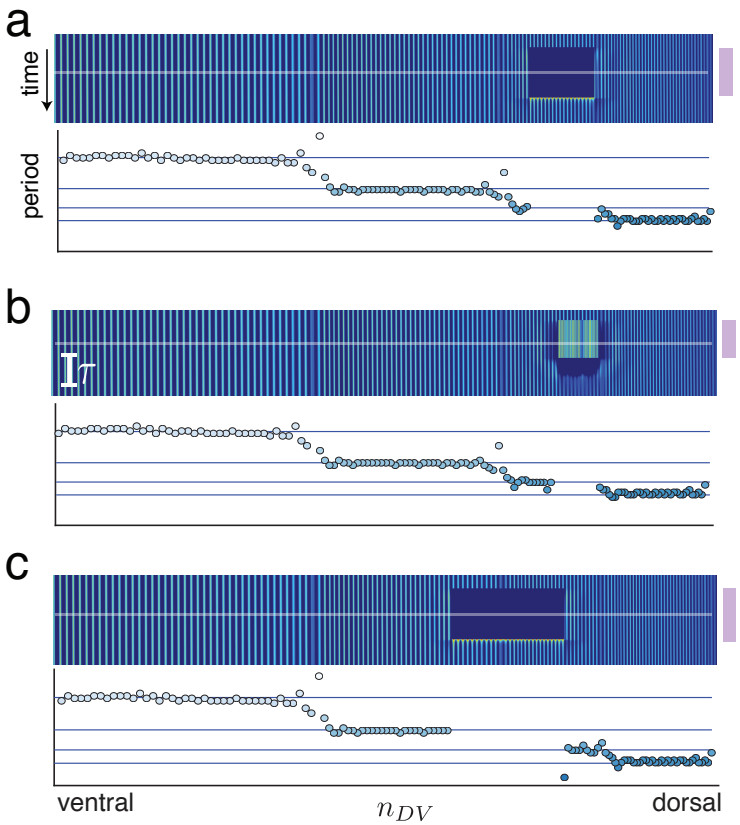


Figure 8: **Robustness of modularization and patterning to large activity perturbations.**(a) All neurons within one module are transiently silenced for 50 ms (duration of perturbation shown by purple bar at right). The existence of other modules, their locations and boundaries in the overall network, and their dynamics and periods remain unchanged despite this large perturbation. (b) Similar to a, but with large-amplitude fixed and independent random perturbations to a subset of cells in one module. This perturbation results in a surrounding local zone of suppressed activity, but other modules remain unchanged and the perturbed module recovers within the cellular biophysical time-constant of $\approx \tau$. (c) A silencing perturbation that spans two modules. Dynamics in the remaining portions of each module remain unaffected, with preserved periods and period changes across the silenced boundary. After the perturbation is lifted, the module boundary reapppears at approximately the same location.

of the shape of the spatial gradient and the size of the network, if the extremal values of the lateral interaction are kept fixed this is due to the topological nature of module emergence. Complementary mechanisms to maintain the extremal value difference could therefore lead to total invariance to system size [59, 60].

The topological roots of module emergence also endow the network with other invariances and robustness in various forms, including in the overall shape of the graded and fixed-scale interactions, heterogeneity and noise in the interactions, and random activity perturbation in submodular regions or entire modules – testable predictions for future experiments. Independent of the problem of multiple module formation, the addition of a fixed-scale lateral interaction renders individual grid modules substantially more robust to inhomogeneities in the weights, paralleling recent findings in morphogenesis [61]. Thus, it reduces the tuning requirements for continuous attractor dynamics.

The sound match of our model with experimental data suggests that MEC may contain a fixed-width lateral interaction across its dorsolateral extent, in addition to the gradients in cellular and network properties along that axis. The fact that $\phi = 0$ provides the best description of the data suggests possible forms of the fixed-width lateral interaction, which are testable with connectomic reconstructions of the network [62, 63, 64]. The two distinct types of lateral interaction in our model could potentially correspond to different cell types in the MEC [65].

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 [66, 67, 68].

Our work contrasts with supervised learning and evolutionary selection approaches to module emergence

[12, 8, 9, 10, 11, 8, 13, 14, 15, 16] in a number of ways: Selection and supervised learning act globally and require top-down instructive or selective feedback – they are not local processes while module formation through peak selection proceeds entirely from local interactions. Second, supervised learning approaches employ biologically implausible learning rules while the peak selection process is the result of activity dynamics with no learning. Third, while supervised learning proceeds slowly through extensive training and is thus interpreted as evolutionary time-scale learning in biological systems, the whole process of learning acts on the same set of synaptic weights rather than on intermediate genetic codes. Here, the smooth gradients may be viewed as arising through simple genetically encoded instructions that could be the targets of evolutionary time-scale selection, while modularization in the presence of the smooth gradient is a nearly instantaneous emergent dynamics.

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.

References

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

[2] 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.

[3] 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.

[4] 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.

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

[6] 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.

[7] Pavlicev Mihaela Wagner, Gunter and James M. Cheverud. The road to modularity. *Nature Reviews Genetics*, 9:921–931, 2007.

[8] 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.

[9] 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.

[10] 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.

[11] 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, 10–15 Jul 2018.

[12] 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, 02 2019.

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

[14] Dirk M Lorenz, Alice Jeng, and Michael W Deem. The emergence of modularity in biological systems. *Phys Life Rev*, 8(2):129–60, Jun 2011.

[15] Ricard V Solé and Sergi Valverde. Spontaneous emergence of modularity in cellular networks. *J R Soc Interface*, 5(18):129–33, Jan 2008.

[16] Fabrizio Damicelli, Claus C Hilgetag, Marc-Thorsten Hütt, and Arnaud Messé. Topological reinforcement as a principle of modularity emergence in brain networks. *Netw Neurosci*, 3(2):589–605, 2019.

[17] Anthony M Zador. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature communications*, 10(1):1–7, 2019.

[18] Kenneth D Irvine and Boris I Shraiman. Mechanical control of growth: ideas, facts and challenges. *Development*, 144(23):4238–4248, 12 2017.

[19] Alan. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, August 1952.

[20] 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.

[21] 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.

[22] Mikail Khona Ila R. Fiete. Attractor and integrator networks in the brain. preprint at https://arxiv.org/abs/2112.03978. 2021.

[23] Torkel Hafting, Marianne Fyhn, Sturla Molden, May-Britt Moser, and Edvard I Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–6, Aug 2005.

[24] Bruce L McNaughton, Francesco P Battaglia, Ole Jensen, Edvard I Moser, and May-Britt Moser. Path integration and the neural basis of the 'cognitive map'. *Nat Rev Neurosci*, 7(8):663–678, 2006.

[25] 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.

[26] 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.

[27] Marianne Fyhn, Torkel Hafting, Alessandro Treves, May-Britt Moser, and Edvard I Moser. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446(7132):190–4, Mar 2007.

[28] 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.

[29] 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.

[30] 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.

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

[32] 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.

[33] 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.

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

[35] A. Mathis, A. Herz, and M. Stemmler. Optimal population codes for space: grid cells outperform place cells. *Neural Comp.*, 24:2280–2317, 2012.

[36] 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.

[37] 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.

[38] Mirko Klukas, Marcus Lewis, and Ila Fiete. Efficient and flexible representation of higher-dimensional cognitive variables with grid cells. *PLOS Comp. Bio., in press*, page 578641, 2020.

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

[40] Sean G Trettel, John B Trimper, Ernie Hwaun, Ila R Fiete, and Laura Lee 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.

[41] Richard J Gardner, Erik Hermansen, Marius Pachitariu, Yoram Burak, Nils A Baas, Benjamin J Dunn, May-Britt Moser, and Edvard I Moser. Toroidal topology of population activity in grid cells. *bioRxiv*, 2021.

[42] 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, 2014.

[43] journal=Nature volume=557 number=7705 pages=429–433 year=2018 publisher=Nature Publishing Group Banino, Andrea *et al*. Vector-based navigation using grid-like representations in artificial agents.

[44] 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.

[45] 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.

[46] I. Kanitscheider and I. R. Fiete. Training recurrent networks to generate hypotheses about how the brain solves hard navigation problems. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[47] Eugenio Urdapilleta, Bailu Si, and Alessandro Treves. Self organization of modular activity of grid cells. *Hippocampus*, 27(11):1204–1213, 11 2017.

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

[49] 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.

[50] Marianne Fyhn, Torkel Hafting, Alessandro Treves, May-Britt Moser, and Edvard I Moser. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446(7132):190–194, 2007.

[51] 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.

[52] 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.

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

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

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

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

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

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

[59] María Almuedo-Castillo, Alexander Bläßle, David Mörsdorf, Luciano Marcon, Gary H Soh, Katherine W Rogers, Alexander F Schier, and Patrick Müller. Scale-invariant patterning by size-dependent inhibition of nodal signalling. *Nat Cell Biol*, 20(9):1032–1042, 09 2018.

[60] Steffen Werner, Tom Stückemann, Manuel Beirán Amigo, Jochen C Rink, Frank Jülicher, and Benjamin M Friedrich. Scaling and regeneration of self-organized patterns. *Phys Rev Lett*, 114(13):138101, Apr 2015.

[61] 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.

[62] Scheffer *et al.* A connectome and analysis of the adult drosophila central brain. *Elife*, 9(57443):129–33, Sept 2020.

[63] Larry F Abbott, Davi D Bock, Edward M Callaway, Winfried Denk, Catherine Dulac, Adrienne L Fairhall, Ila Fiete, Kristen M Harris, Moritz Helmstaedter, Viren Jain, et al. The mind of a mouse. *Cell*, 182(6):1372–1376, 2020.

[64] J Alexander Bae, Mahaly Baptiste, Agnes L Bodor, Derrick Brittain, JoAnn Buchanan, Daniel J Bumbarger, Manuel A Castro, Brendan Celii, Erick Cobos, Forrest Collman, et al. Functional connectomics spanning multiple areas of mouse visual cortex. *bioRxiv*, 2021.

[65] Chenglin Miao, Qichen Cao, May-Britt Moser, and Edvard I Moser. Parvalbumin and somatostatin interneurons control different space-coding networks in the medial entorhinal cortex. *Cell*, 171(3):507–521, 2017.

[66] Xiao-Jing Wang. Macroscopic gradients of synaptic excitation and inhibition in the neocortex. *Nat Rev Neurosci*, 21(3):169–178, 03 2020.

[67] 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.

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

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

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

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

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

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

[74] 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.

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

[76] 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.

[77] 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.

[78] Rishidev Chaudhuri, Alberto Bernacchia, and Xiao-Jing Wang. A diversity of localized timescales in network activity. *Elife*, 3:e01239, 2014.

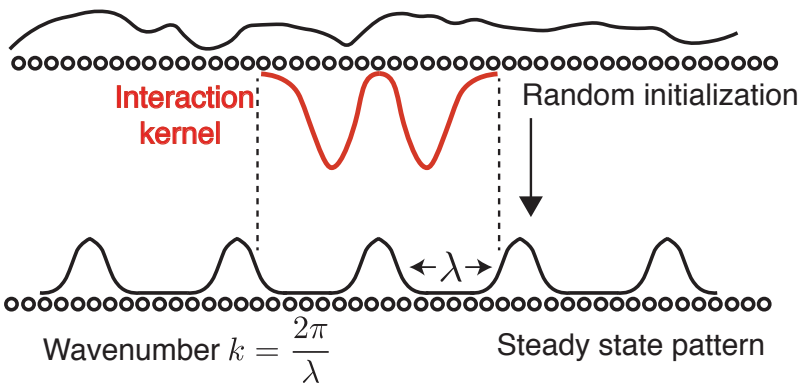


Figure 9: Local pattern formation in continuous attractor models of grid cells: Through local amplification of random fluctuations, the lateral interaction forms periodic patterns.

Supplementary materials

Materials and Methods

We use a continuous attractor network (CAN) model [69, 70, 71] for grid cells [30, 32, 33, 28], with neural dynamics obeying

$$\frac{\partial s(i,t)}{\partial t} + \frac{s(i,t)}{\tau} = \phi\left[\sum_j W_0(i,j)s(j,t) + B(i,t)\right], \tag{2}$$

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

$$W_0(i,j) = W(|\mathbf{x}_i - \mathbf{x}_j - \mathbf{l}(\theta_j)|), \tag{3}$$

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

$$B(i,t) = b + b_{vel}|\mathbf{v}|\cos(\theta_\mathbf{i} - \psi), \tag{4}$$

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 $\mathbf{l}(\theta)$. This mechanism is responsible for velocity integration by the network [30].

As noted in the main text, the interaction weight kernel W is given by the sum of two components $W = W^g_\mathbf{x} + W^f$. The first, $W^g_\mathbf{x}$ drives local pattern formation, and has a spatial scale $\sigma(\mathbf{x})$, 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[30] (used in Figs. 2a-c, 3, 1 and SI Fig. 10)

$$W^g_{\text{mexican-hat}}(\Delta x) = \alpha_E \exp\left[-\gamma\frac{(\Delta x)^2}{2\sigma_{mh}(\mathbf{x})^2}\right] - \alpha_I \exp\left[-\frac{(\Delta x)^2}{2\sigma_{mh}(\mathbf{x})^2}\right], \tag{5}$$

and the box-function profile[52] (used in Fig. 2 and SI Fig. 10)

$$W^g_{\text{box}}(\Delta x) = \alpha_0 \times \mathbb{1}_{|\Delta x| < \sigma_b(\mathbf{x})} = \begin{cases} \alpha_0 & \text{if } |\Delta x| < \sigma_b(\mathbf{x}), \\ 0 & \text{if } |\Delta x| \geq \sigma_b(\mathbf{x}). \end{cases} \tag{6}$$

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

Table 1: Parameters held constant across all numerical simulations

For the fixed-width interaction $W^f(\Delta x)$, we implement 3 main types — localized (used in Figs. 2,3 and SI Fig. 10), diffuse (used in Fig. 2 and SI Fig. 10) and decaying (used in SI Fig. 10).

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

In particular,

- In Figs. 1d-e we use only a smoothly varying Mexican-hat pattern forming kernel $W = W^g_{\text{mexican-hat}}$
- In Figs. 2a-c,g and 3 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. 10 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. 7 we use the box function to construct

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

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. 7b, $\epsilon^g_2 = 0.1$ and other noise terms are set to zero (In the one-dimensional case ϵ_1 and ϵ_2 have the same effect); for Fig. 7c, $\epsilon^g_1 = 0.2$, $\epsilon^g_2 = 0.3$, $\epsilon^f_1 = 0.05$ and $\epsilon^f_2 = 0$ (See SI Sec. 3.9 for more details).

In Table 1 we present a list of common parameters used across all numerical simulations. Then, in Tables 2,3 we present the parameter values used for the kernels used in our numerical simulations

Supplementary Text

The supplemental information is structured as follows: We will first present the mathematical analysis for pattern formation, and demonstrate in SI Sec. 1 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. Then, in Sec. 1.1 we show how the addition of a Gaussian localized kernel results in self-organized modularization. We then generalize this result in SI Sec. 2, to show that arbitrary fixed-scale kernels will almost always result in module formation as demonstrated in Fig. 3. Among arbitrary kernels, we will show in Sec. 3 that kernels with a simple features result in a simple equation describing the detailed period ratios of the formed grid modules as shown in Fig. 4. 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. 3.7. After having described our results primarily for the case of one-dimensional grid cells, we then demonstrate in Sec. 3.8 that our arguments extend naturally to two dimensions,

$W^g_{\text{mexican-hat}}$ parameters	Value
α_E	1000
α_I	1000
γ	1.05
N^{1D}	3000
N_y^{2D}	100
N_x^{2D}	1000
$\sigma_{mh}(x)$	$1/\sqrt{2\beta(x)}$
$\beta(x)$	$\beta_0 + (\beta_1 - \beta_0)x/N'$
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
$\sigma_b(x)$	$15 + 30x/N$

Table 2: Pattern forming kernel parameters used for numerical simulations

$W^f_{\text{localized}}$ 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 3: Fixed-scale kernel parameters used for numerical simulations

and we present numerical results demonstrating the same. In SI Sec. 3.10 we then demonstrate that our results and predictions of grid period ratios are consistent with available data sources to a large extent. We then end with some brief remarks relating our result to the context of energy minimization in neural networks (SI Sec. 4), and providing broader perspectives of our results in the contexts of general loss optimization (Sec. 5) and eigenvector localization (SI Sec. 6).

We will largely restrict our analysis to the continuum limit of a large number of neurons (a neural field model) for mathematical convenience. However, as shown in the main text, the results hold for and accurately predict the dynamics of networks of discrete neurons. We will use a continuous index to represent position along the neural sheet. In this limit, the dynamics of rate-based neurons is given by:

$$\frac{\partial s(\mathbf{x},t)}{\partial t} + \frac{s(\mathbf{x},t)}{\tau} = \phi\left[\int_{-\infty}^{+\infty} W(\mathbf{x},\mathbf{x}')s(\mathbf{x}',t)d\mathbf{x}' + B(\mathbf{x})\right], \tag{7}$$

where $s(\mathbf{x})$ is the synaptic activation of the neuron at the vector position $\mathbf{x}$ on the neural sheet, τ is the biophysical time-constant of individual neurons, ϕ is a monotonic non-decreasing nonlinearity, and $B(\mathbf{x})$ is the feedforward input to the neuron. For simplicity, we will use the the rectification function ($\phi(z) = [z]_+ = z$ for $z > 0$ and 0 otherwise) as the nonlinearity in all of our results.

In nonlinear continuous attractor models the interaction weights $W(\mathbf{x},\mathbf{x}')$ are chosen to have a continuous symmetry, usually a translation-invariant symmetry such that $W(\mathbf{x},\mathbf{x}') = K(|\mathbf{x}-\mathbf{x}'|)$, where the weight between neurons at locations $\mathbf{x}$ and $\mathbf{x}'$ depends only on their separation, and not on their absolute locations. One example of such a weight K is the so-called Mexican-hat function described by a difference-of-Gaussians, in which neurons excite their immediate neighbors and inhibit those slightly further away; there is no interaction between faraway neurons:

$$W(\mathbf{x},\mathbf{x}') = W(\mathbf{\Delta x}) = \alpha_E \exp\left(-\frac{\mathbf{\Delta x}^2}{2\sigma_E}\right) - \alpha_I \exp\left(-\frac{\mathbf{\Delta x}^2}{2\sigma_I}\right). \tag{8}$$

Continuous attractor grid cell models of individual modules rely principally on a local inhibitory interaction (the excitatory center interaction is dispensable) [30, 52].

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.

$$W^g_{\mathbf{x}}(\mathbf{\Delta x}) = \alpha_E \exp\left(-\frac{\mathbf{\Delta x}^2}{2\sigma_E(\mathbf{x})}\right) - \alpha_I \exp\left(-\frac{\mathbf{\Delta x}^2}{2\sigma_I(\mathbf{x})}\right), \tag{9}$$

where $\sigma_E(\mathbf{x})$ and $\sigma_I(\mathbf{x})$ are now functions that depend on position in the neural sheet, and encode the smoothly varying characteristic scale of the Mexican-hat interaction, say

$$\sigma_{E/I}(\mathbf{x}) = \sigma_{E/I} + \sigma'_{E/I}(0)\cdot\mathbf{x}. \tag{10}$$

For such graded kernels, we will use $W(\mathbf{x},\mathbf{x}')$ and $W_{\mathbf{x}}(\mathbf{x}-\mathbf{x}') = W_{\mathbf{x}}(\mathbf{\Delta x})$ interchangeably.

1 Pattern formation with graded kernels

In the limit of very slow changes in the length-scale of the interaction kernel $W_{\mathbf{x}}(\mathbf{\Delta x})$ along the dorso-ventral axis of the MEC, we can analyze the pattern formation process at position $\mathbf{x}$ by assuming that $W_{\mathbf{x}}(\mathbf{\Delta x})$ is locally constant.

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 $\mathbf{x}$.

We first identify an unstable steady-state solution to Eq. (7), which we denote as $s_0(\mathbf{x})$. This solution satisfies

$$\frac{s(\mathbf{x})}{\tau} = \phi\left[\int_{-\infty}^{+\infty} W_{\mathbf{x}}(\mathbf{x}-\mathbf{x}')s_0(\mathbf{x}')d\mathbf{x}' + B(\mathbf{x})\right]. \tag{11}$$

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

$$s_0(\mathbf{x}) = \frac{\tau\bar{B}}{1-\tau\bar{W}}, \tag{12}$$

where $\bar{B} = \int B(\mathbf{x})d\mathbf{x}$ and $\bar{W} = \int W_{\mathbf{x}}(\mathbf{x}-\mathbf{x}')d\mathbf{x}'$.

We then consider a perturbative analysis, by examining the evolution of $s(\mathbf{x},t) = s_0(\mathbf{x}) + \epsilon(\mathbf{x},t)$. We apply our analysis to the early time evolution of this initial condition, such that $\epsilon(\mathbf{x},t) \ll s_0(\mathbf{x})$. Inserting our form of $s(\mathbf{x},t)$ in Eq. (7), we obtain

$$\frac{\partial \epsilon(\mathbf{x},t)}{\partial t} + \frac{\epsilon(\mathbf{x},t)}{\tau} = \phi'(\gamma \bar{W} s_0(\mathbf{x}) + B) \int_{-\infty}^{\infty} W_{\mathbf{x}}(\mathbf{x}-\mathbf{x}')\epsilon(\mathbf{x}',t)d\mathbf{x}' \tag{13}$$

Since $W_{\mathbf{x}}(\mathbf{x}-\mathbf{x}')$ 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 $W_{\mathbf{x}}$ 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. (13) yields,

$$\alpha(\mathbf{k}) + 1/\tau = \phi'(\bar{W} s_0(\mathbf{x}) + B) \int_{\infty}^{\infty} W_{\mathbf{x}}(\mathbf{x}-\mathbf{x}')e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}d\mathbf{x}', \tag{14}$$

$$= \phi'[\bar{W} s_0(\mathbf{x}) + B]\mathscr{F}[W_{\mathbf{x}}(\mathbf{x}-\mathbf{x}')], \tag{15}$$

$$= \phi'[\bar{W} s_0(\mathbf{x}) + B]\mathscr{F}W_{\mathbf{x}}(\mathbf{k}) \tag{16}$$

where $\mathscr{F}[W_{\mathbf{x}}(\mathbf{x}-\mathbf{x}')] = \mathscr{F}W_{\mathbf{x}}(\mathbf{k})$ is the Fourier transform of the interaction kernel corresponding to position $\mathbf{x}$ on the neural sheet.

Note that since $W_{\mathbf{x}}(\mathbf{\Delta x})$ is a kernel, it is a radially-symmetric real function, and hence the Fourier transform $\mathscr{F}W_{\mathbf{x}}(\mathbf{k})$ 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 $W_{\mathbf{x}}(\mathbf{\Delta x})$ as the Hankel transform of $W_{\mathbf{x}}(|\mathbf{\Delta x}|)$).

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 $W_{\mathbf{x}}(\mathbf{\Delta x})$, we see from Eq. (16) that

$$k^*(\mathbf{x}) = \arg\max_k \mathscr{F}W_{\mathbf{x}}(k), \tag{17}$$

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

For $W_{\mathbf{x}}(\mathbf{\Delta x})$ given by Eq. (9), i.e., without any additional fixed-scale interaction, we obtain from Eq. (17)

$$[k^*(\mathbf{x})]^2 = \frac{2}{\sigma_E(\mathbf{x})^2 - \sigma_I(\mathbf{x})^2} \log\left(\frac{\alpha_E \sigma_E(\mathbf{x})^3}{\alpha_I \sigma_I(\mathbf{x})^3}\right). \tag{18}$$

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

$$k^*(\mathbf{x}) \propto 1/\sigma(\mathbf{x}), \tag{19}$$

and hence

$$\lambda^*(\mathbf{x}) \propto \sigma(\mathbf{x}), \tag{20}$$

where $\lambda * (\mathbf{x})$ is the periodicity of the grid pattern formed locally around position $\mathbf{x}$. 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 $W^g_{\mathbf{x}}(\mathbf{\Delta x})$ 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. (20) holds for any kernel $W^g_x(\Delta x)$ that depends on a length-scale $\sigma(x)$. As an example, we present the corresponding analysis for the box-shaped kernel employed for pattern formation in Ref. [52]. In this case

$$W^g_{\mathbf{x}}(\mathbf{\Delta x}) = -W_0 \mathbb{1}_{\mathbf{\Delta x} \leq \sigma(\mathbf{x})}. \tag{21}$$

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

$$\mathscr{F}W^g_{\mathbf{x}}(\mathbf{k}) = \int_{-\infty}^{\infty} -W_0 \mathbb{1}_{|\mathbf{x}| \leq \sigma(\mathbf{x})} e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x} \tag{22}$$

$$= -W_0 \int_{|\mathbf{x}| \leq \sigma(\mathbf{x})} e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x}. \tag{23}$$

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

$$\mathscr{F}W_x^g(k) = -2W_0 \frac{\sin(k\sigma(x))}{k} \tag{24}$$

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

$$\mathscr{F}W_{\mathbf{x}}^g(k) = -2\pi W_0 \sigma(\mathbf{x}) \frac{J_1(k\sigma(\mathbf{x}))}{k}. \tag{25}$$

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

$$k^* \approx \begin{cases} 4.493/\sigma(x) & \text{if } x \text{ is one-dimensional, and} \\ 5.136/\sigma(\mathbf{x}) & \text{if } \mathbf{x} \text{ is two-dimensional.} \end{cases} \tag{26}$$

1.1 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. 3.8).

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

$$W_x(\Delta x) = W_x^g(\Delta x) + W^f(\Delta x). \tag{27}$$

We first demonstrate our result with an example of a simple kernel, to justify how Eq. (17) leads to the emergence of discrete grid modules. Consider the localized excitatory interaction

$$W^f(\Delta x) = \alpha_S \exp\left(-\frac{(\Delta x - d)^2}{2\sigma_S^2}\right) + \alpha_S \exp\left(-\frac{(\Delta x + d)^2}{2\epsilon_S^2}\right). \tag{28}$$

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

This choice of $W_x(\Delta x) = W_x^g(\Delta x) + W^f(\Delta x)$ leads to the the Fourier transform,

$$\mathscr{F}W_x(k) = \mathscr{F}W_x^g(k) + \mathscr{F}W^f(k), \tag{29}$$

$$= \sqrt{2\pi}\left[\alpha_E \sigma_E(x) \exp\left(-\frac{\sigma_E(x)^2 k^2}{2}\right) - \alpha_I \sigma_I(x) \exp\left(-\frac{\sigma_I(x)^2 k^2}{2}\right) + 2\alpha_S \epsilon_S \cos(kd) \exp\left(-\frac{\epsilon_S^2 k^2}{2}\right)\right]. \tag{30}$$

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 $\mathscr{F}W^f(k)$ in Eq. (30) as being a small perturbation to the Fourier transform of the standard Mexican-hat interaction, $\mathscr{F}W_x^g(k)$. 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 $\mathscr{F}W^f(k)$ oscillates at a k-scale much smaller than the relevant scales of $\mathscr{F}W_x^g(k)$ (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, $\mathscr{F}W^f(k)$ 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 $\mathscr{F}W_x^g(k)$, will not change the position of the local maxima. This results in the *local* maxima of $\mathscr{F}W_x(k)$ also being the same set S. Importantly, we note that since S was predefined purely via $\mathscr{F}W^f(k)$, *there is no x dependence on the set S.*

Following Eq. (17), the wave-vector corresponding to the pattern formation at point x on the neural sheet corresponds to the *global* maxima of $\mathscr{F}W_x(k)$. 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, $\mathscr{F}W_x^g$ as a function of x picks different choices of k_i depending on the position x — the $k \in S$ that is nearest to the maxima of $\mathscr{F}W_x^g(k)$ will be chosen as the global maxima, and will be the wave vector corresponding to the pattern at x. We refer to this mechanism as "peak selection".

For our particular choice of $W^f(x)$ made in Eq. (28), we obtained

$$\mathscr{F}W^f(k) = 2\alpha_S \epsilon_S \cos(kd) \exp\left(-\frac{\epsilon_S^2 k^2}{2}\right). \tag{31}$$

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

$$S = \left\{ \frac{2m\pi}{d} \middle| m \in \mathbb{Z}^+ \right\}. \tag{32}$$

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

$$\frac{\lambda_{m+1}}{\lambda_m} = \frac{m+1}{m}. \tag{33}$$

Thus, the addition of a fixed-scale interaction, W^f such as Eq. (28) 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 $W^f(\Delta x)$.

2 Kernels that lead to modularization

The peak-selection modularization mechanism described above arises naturally from the presence of the rapidly oscillating term in $\mathscr{F}W^f(k)$. In fact, for discrete grid modules to occur, the only constraints imposed on the fixed-scale kernel W^f are: (a) the Fourier transform $\mathscr{F}W^f(k)$ 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,

$$\mathscr{F}f_n(k) = [2\sin(k)/k]^n, \tag{34}$$

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 $\mathscr{F}W^f(k)$ is a smooth function, we hypothesize that generically smooth functions that are in L^2 will almost always have multiple maxima and minima.

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., $\mathscr{F}[W^f(a\Delta x) = \mathscr{F}W^f(k/a)$. 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. 10, 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. (17). 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.

3 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 $\epsilon_1, \epsilon_2, \ldots$ 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., $d \gg \sigma_{E/I}(x)$.

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

$$\frac{\lambda_{m+1}}{\lambda_m} = \frac{m+1+\phi/(2\pi)}{m+\phi/(2\pi)}. \tag{35}$$

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. (35), 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. (35). 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. (35).

3.1 Localized kernels

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

$$\mathscr{F}W^f(k) = \Re[e^{-ikd}\mathscr{F}g(k)]. \tag{36}$$

Since $g(x)$ is supported over a scale ϵ, the Fourier transform $\mathscr{F}g(k)$ 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. (36) as

$$\mathscr{F}W^f(k) = |\mathscr{F}g(k)|\cos(kd - \psi), \tag{37}$$

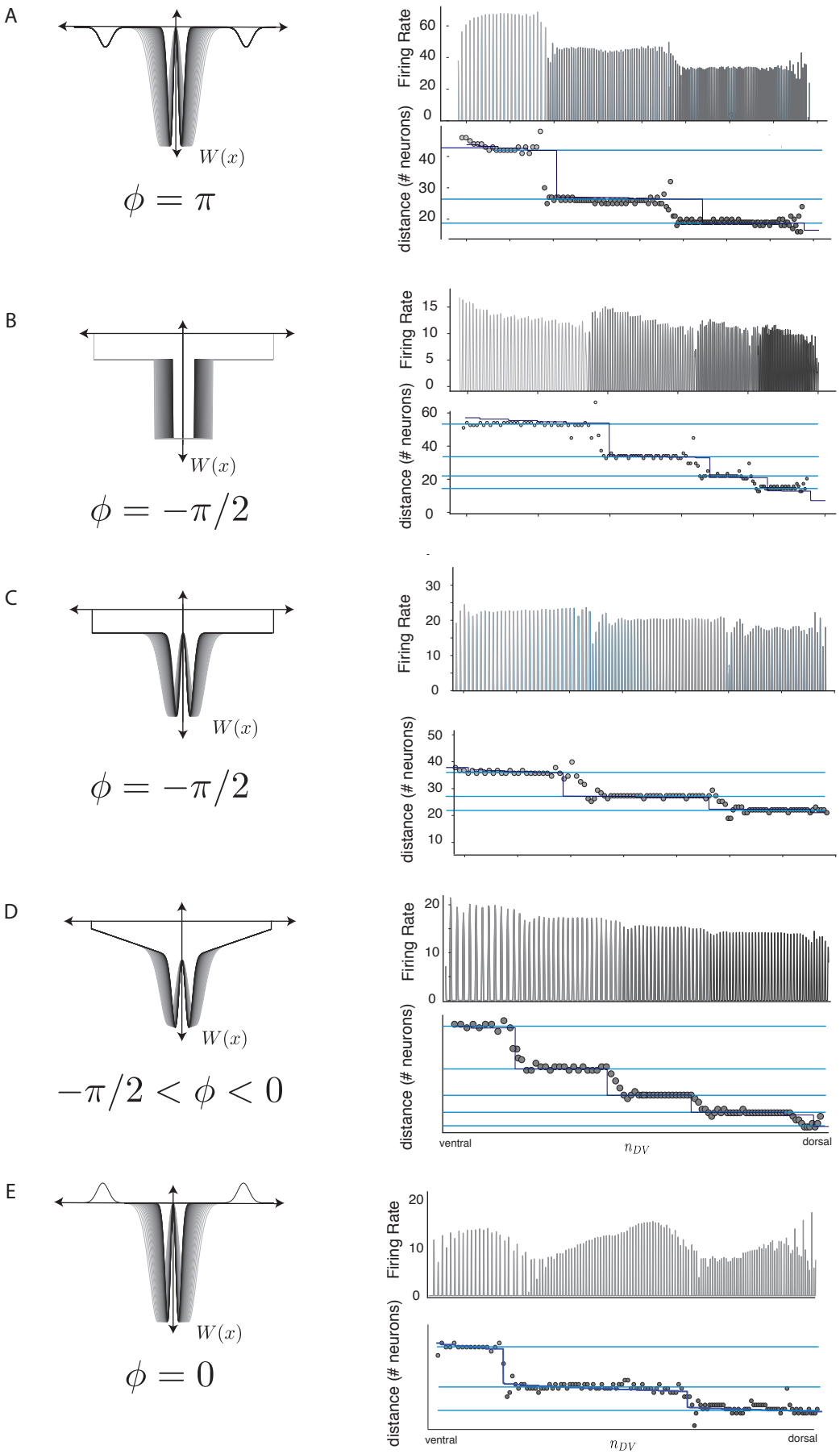


Figure 10: Examples of modularization and population activity (right column) with various pattern forming and fixed-scale lateral interactions (left column). In each case the dark-blue curve shows the predicted value of the grid period from Eq. (17), and is in close agreement with the numerical simulation of the population activity. Each of the fixed-scale interactions has a qualitatively different shape, spanning different values of ϕ (see Fig. 6)

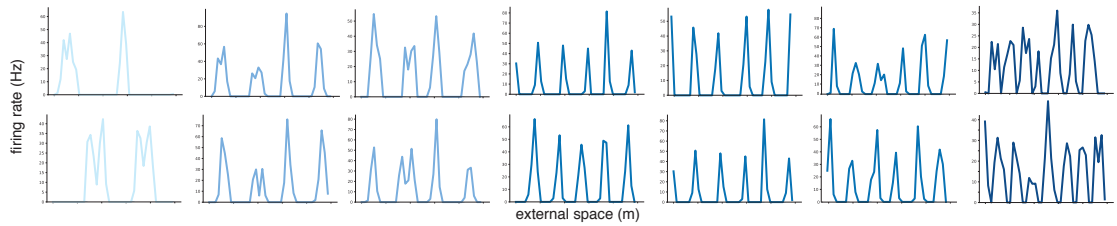


Figure 11: Sample tuning curves from several neurons in all modules from the network of Fig 2a.

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

$$S = \left\{ \frac{2m\pi + \psi}{d} \middle| m \in \mathbb{Z}^+ \right\}, \tag{38}$$

resulting in period ratios described by

$$\frac{\lambda_{m+1}}{\lambda_m} = \frac{m + 1 + \psi/(2\pi)}{m + \psi/(2\pi)}, \tag{39}$$

which is identical to Eq. (35) for $\phi = \psi$. We also note that we can now ascribe an interpretation to the phase angle ϕ — it is the phase difference between $\mathscr{F} W^f(k)$ and $\cos(kd)$.

3.2 Diffuse kernels

We model a diffuse interaction kernel $W^f(x)$ as

$$W^f(x) = -W_0 \mathbb{1}_{[-d,d]} = \begin{cases} -W_0 & \text{if } |x| \le d \\ 0 & \text{if } |x| > d \end{cases}. \tag{40}$$

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

$$\mathscr{F} W^f(k) = \int_{-\infty}^{+\infty} -W_0 \mathbb{1}_{[-d,d]} e^{ikx} dx = \int_{-d}^{+d} -W_0 e^{ikx} dx \tag{41}$$

$$= -2W_0 \frac{\sin(kd)}{k} = -2W_0 d \ \mathrm{sinc}(kd). \tag{42}$$

Note that once again, similar to Eqn. (31), 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

$$S \approx \left\{ \frac{2m\pi - \pi/2}{d} \middle| m \in \mathbb{Z}^+ \right\}, \tag{43}$$

which immediately results in period ratios of the form

$$\frac{\lambda_{m+1}}{\lambda_m} \approx \frac{m + 1 - 1/4}{m - 1/4}, \tag{44}$$

which corresponds to the result in Eq. (35) for $\phi = \pi/2$.

More precisely, the extrema of $\mathscr{F} W^f(k)$ 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. (35) closely.

3.3 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 3.5.1 The triangular kernel can be written as:

$$W^f(\Delta x) = \begin{cases} W_0(\Delta x - d)/d & \text{if } \Delta x < d \\ 0 & \text{if } \Delta x \geq d \end{cases} \tag{45}$$

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

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

Thus, its Fourier transform is:

$$\begin{aligned} \mathscr{F}W^f(k) &= -W_0^2 d^2 \left(\frac{\sin(kd/2)}{(kd/2)} \right)^2 \\ &= -\frac{2W_0^2}{k^2}[1 - \cos(kd)]. \end{aligned}$$

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

$$S \approx \left\{ \frac{2m\pi}{d} \middle| m \in \mathbb{Z}^+ \right\}, \tag{46}$$

which immediately results in period ratios of the form

$$\frac{\lambda_{m+1}}{\lambda_m} \approx \frac{m+1}{m}, \tag{47}$$

which corresponds to the result in Eq. (35) for $\phi = 0$.

3.4 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 $\mathscr{F}W^f(k)$ 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 $1/d$.

3.5 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.

$$W^f = a_{\text{local}} W^f_{\text{local}} + a_{\text{diffuse}} W^f_{\text{diffuse}} + a_{\text{decaying}} W^f_{\text{decaying}} + a_{\text{short}} W^f_{\text{short}}. \tag{48}$$

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

$$\mathscr{F}W^f(k) = a_{\text{local}} |\mathscr{F}g(k)| \cos(kd - \psi) - 2W_0 a_{\text{diffuse}} \sin(kd)/k - 2W_0^2 a_{\text{decaying}} (1 - \cos(kd))/k + \mathscr{F}h(k) \tag{49}$$

$$= H_0(k) + \sum_{i=0}^{3} H_i(k) \cos(kd + \phi_i) \tag{50}$$

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

$$\mathscr{F}W^f(k) \approx \cos(kd - \phi) \tag{51}$$

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

$$S \approx \left\{ \frac{2n\pi + \phi}{d} \middle| n \in \mathbb{Z}^+ \right\}, \tag{52}$$

which immediately results in period ratios of the form Eq. (35). 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. 35.

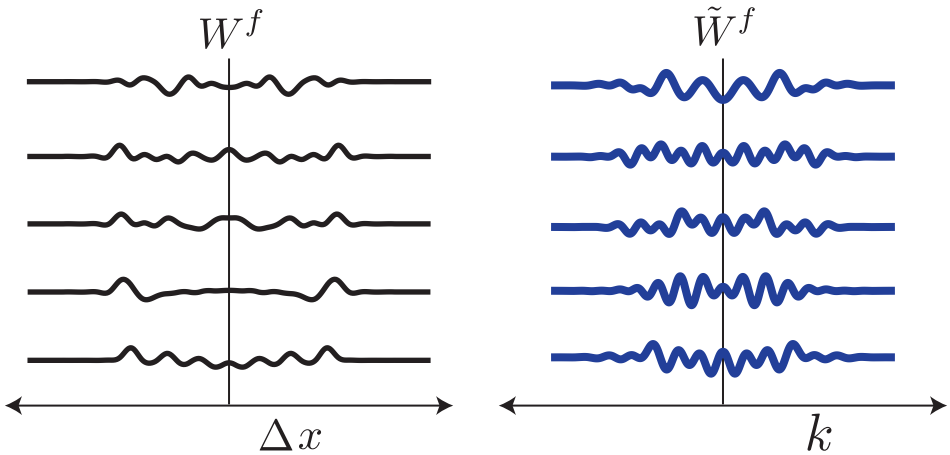


Figure 12: Randomly constructed fixed-scale interactions (left column) and their Fourier transforms (right column), in addition to the hand-designed ones in Fig.6a, that give $\phi = 0$.

3.5.1 Caveats

Clearly there exist *simple* kernels with Fourier transforms that are not given by $\mathscr{F}W^f(k) \approx \cos(kd - \phi)$. 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 $\mathscr{F}W^f(k) = \exp[-k^2d^2/2]$, 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[72, 73], $f(\Delta x)$, (like the Gaussian kernel) whose Fourier transforms $\mathscr{F}f(k)$ 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.

3.6 Period ratios

Having demonstrated analytically that *simple* kernels result in a sequence of period ratios given by Eq. (35), 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 [29] emerges naturally from our setup.

From Eq. (35), we obtained that the period ratio, $r_m = \lambda_{m+1}/\lambda_m$ can be written as

$$r_m = 1 + 1/(m + f), \tag{53}$$

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

$$\langle r_m \rangle_m = 1 + \frac{1}{3}\left(\frac{1}{f+2} + \frac{1}{f+3} + \frac{1}{f+4}\right) \tag{54}$$

As can be seen in Fig. 13, 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. (35) matches well with experimental observations. The average of these period ratios over all values of ϕ can also be calculated as

$$\langle r_m \rangle_{\phi,m} = 1 + \frac{1}{3}\left[\log\left(\frac{5}{3}\right) + \log\left(\frac{7}{5}\right) + \log\left(\frac{9}{7}\right)\right] \tag{55}$$

which is approximately equal to 1.37.

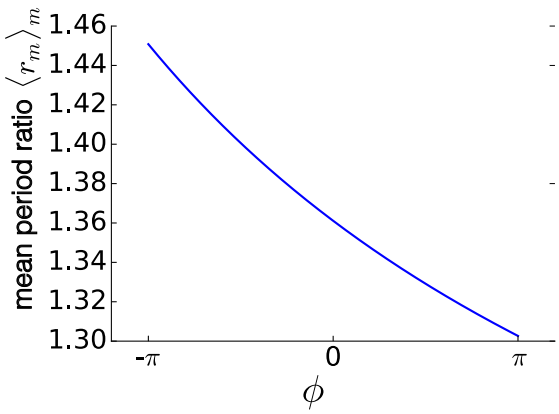


Figure 13: **Mean grid-period ratios** Ratios of grid periods averaged over 4 modules as a function of the phase shift ϕ in Eq. (35)

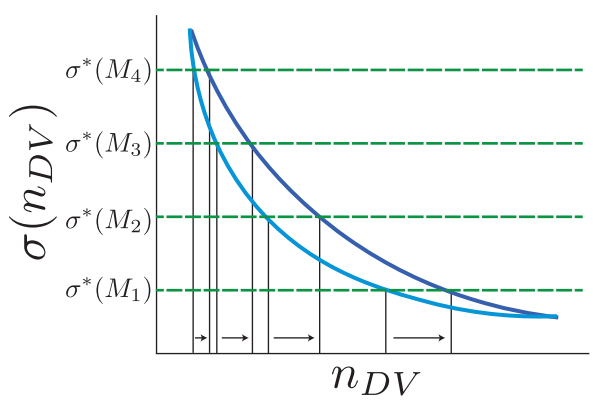


Figure 14: **The number of modules is a topological invariant and hence invariant to the gradient shape (the function $\sigma(x)$)** since it is set only by the fixed interaction width and the extremal/endpoint values of $\sigma(x)$, even though changing the gradient shape shifts the locations of module boundaries.

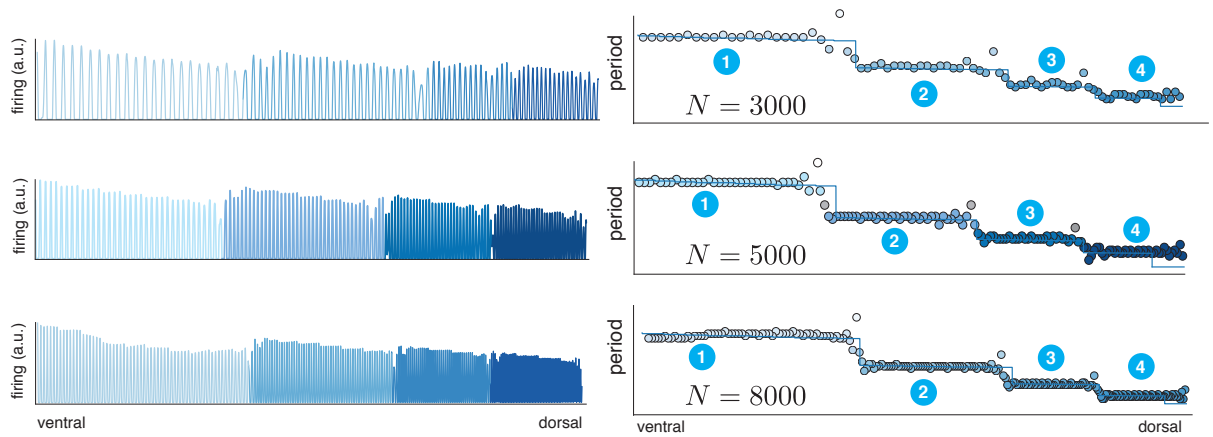


Figure 15: **The modularization process is independent of system size N and the size of formed module scales linearly with N**: 3 simulations corresponding to the same parameters as Fig.2d-f. The number of modules formed as well as the module periods remains the same as N is varied from 3000-8000.

3.7 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 $W_x^g(\Delta x)$ with characteristic spatial scale $\sigma(x)$ at position x, the wave-vector of the formed pattern was proportional to $1/\sigma(x)$:

$$k^{*g}(x) = \eta/\sigma(x), \tag{56}$$

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

We assume for simplicity that the fixed-scale lateral interaction is a *simple* kernel, such that $\mathscr{F}W^f(k) \sim \cos(kd - \phi)$. Thus, the local maxima generated by $\mathscr{F}W^f(k)$ 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. 56 occurs at $k^{*g}(x) = \eta/\sigma(x)$.

Notably, the selected maximum k_m will be robust to small perturbations in the selection function $\mathscr{F}W_x^g(k)$, 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[58]. The topological number corresponding to a given module is the module number m, which is a topological invariant similar to a winding number[58][2].

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:

$$\frac{\eta}{\sigma_{\max}} \leq \frac{2n\pi + \phi}{d} \leq \frac{\eta}{\sigma_{\min}}. \tag{57}$$

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

$$\frac{-\phi + \eta d/\sigma_{\max}}{2\pi} \leq n \leq \frac{-\phi + \eta d/\sigma_{\min}}{2\pi} \tag{58}$$

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

$$\#\text{ modules} \equiv \nu_{mod} = \left\lfloor \frac{-\phi + \eta d/\sigma_{\min}}{2\pi} \right\rfloor - \left\lceil \frac{-\phi + \eta d/\sigma_{\max}}{2\pi} \right\rceil = \left\lfloor \frac{-\phi + k^{*g}(0)d}{2\pi} \right\rfloor - \left\lceil \frac{-\phi + k^{*g}(L)d}{2\pi} \right\rceil \tag{59}$$

[2]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 [29].

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 $\sigma(x)$ the graded interaction; notably, it does not depend on the detailed shape of $\sigma(x)$. 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. SI Fig. 14). Fourth, since the average module size will be L/ν_{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 $\mathscr{F}W^f(k)$, for $k > 0$.

3.7.1 Module boundary locations

Following the peak-selection arguments made earlier, the module boundaries will occur at spatial locations that have $k^{*g}(x)$ 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 $(k_n + k_{n+1})/2$,

$$k^{*g}(x_{\text{boundary}}) \approx \frac{(2n+1)\pi + \phi}{d} \tag{60}$$

and thus

$$x_{\text{boundary}} \approx \sigma^{-1}\left(\frac{\eta d}{(2n+1)\pi}\right). \tag{61}$$

where σ^{-1} is the inverse function of $\sigma(x)$, $\sigma^{-1} \circ \sigma(x) = x$. Thus, while the specific positions of the module boundaries are dependent on the shape of the gradient $\sigma(x)$, qualitative features such as the number of modules, module periods and module sizes are indifferent to the particular forms of the gradient (cf. Fig. 14).

3.8 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 $f(\mathbf{x}) = f(x, y)$

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

Define polar coordinates in real and Fourier space such that:

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ k_x &= k\cos\phi \\ k_y &= k\sin\phi \end{aligned}$$

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,

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

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 $\mathscr{F}f$ will also be a real radially-symmetric function $\mathscr{F}f(k,\phi) = \mathscr{F}f(k)$. Thus

$$\mathscr{F}f(k) = \int_0^\infty \int_0^{2\pi} r dr d\theta f(r) e^{ikr\cos(\theta-\phi)}, \tag{62}$$

$$= \int_0^\infty r dr f(r) \int_0^{2\pi} e^{ikr\cos(\theta-\phi)} d\theta, \tag{63}$$

$$= 2\pi \int_0^\infty r f(r) J_0(kr) dr, \tag{64}$$

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

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

Equation (64) 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.

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

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

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

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

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

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. (35) 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 [29].

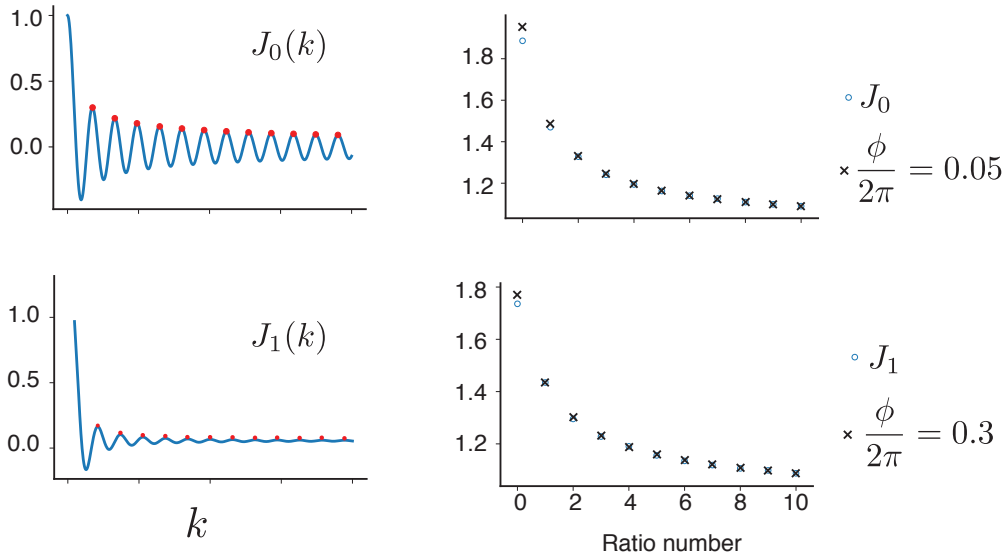


Figure 16: Bessel functions (left column) and period ratios for Bessel function maxima (right column) with their best-fit values of ϕ for the period ratios corresponding to Eq. (35)

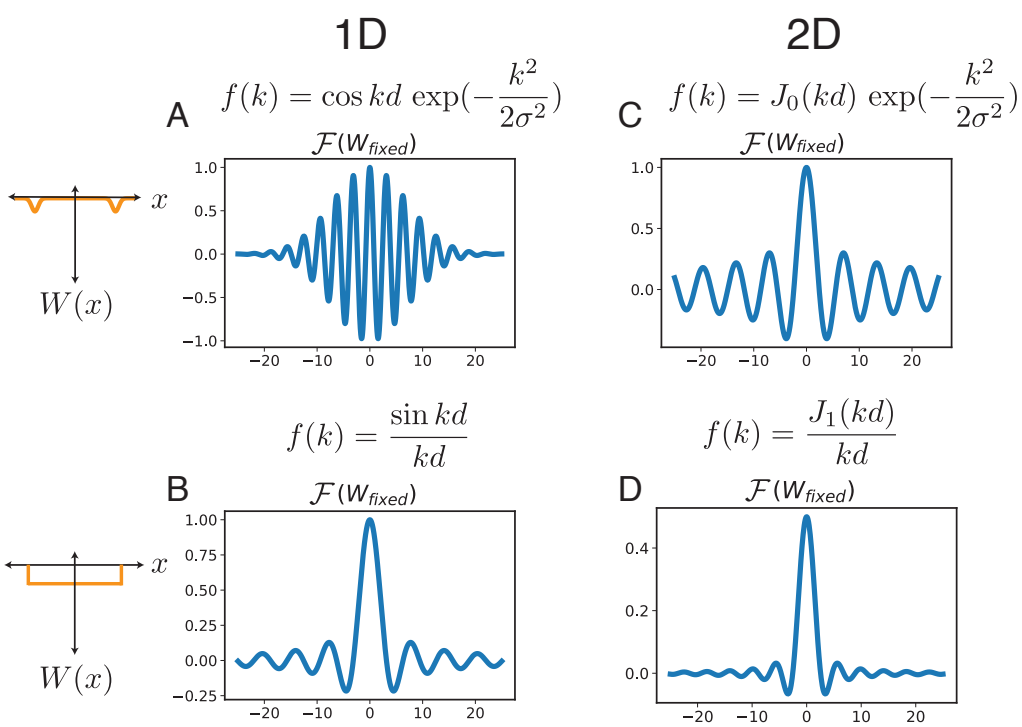


Figure 17: Fixed interactions(left, in orange) and their oscillatory Fourier transforms in 1D (left column) and 2D (right column).

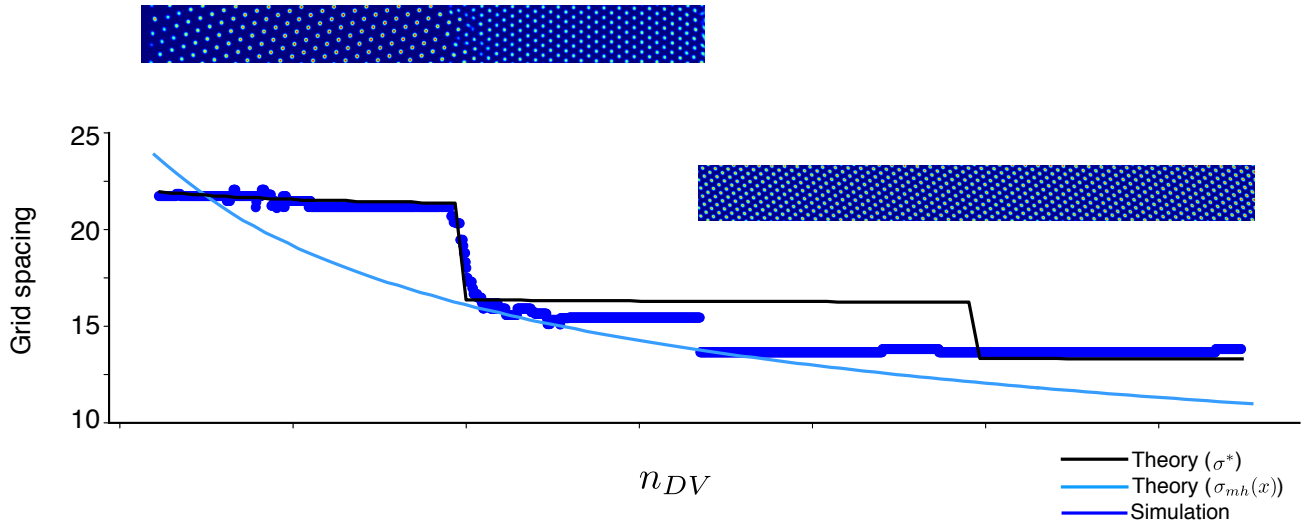


Figure 18: **2d simulation with 3 modules:** (top) Snapshots of population activity showing 3 discrete 2d grid modules, (bottom) plot of grid spacing and comparision with Hankel transform predictions. Grid spacing determined by calculating the (neural) spatial auto-correlation of the population firing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

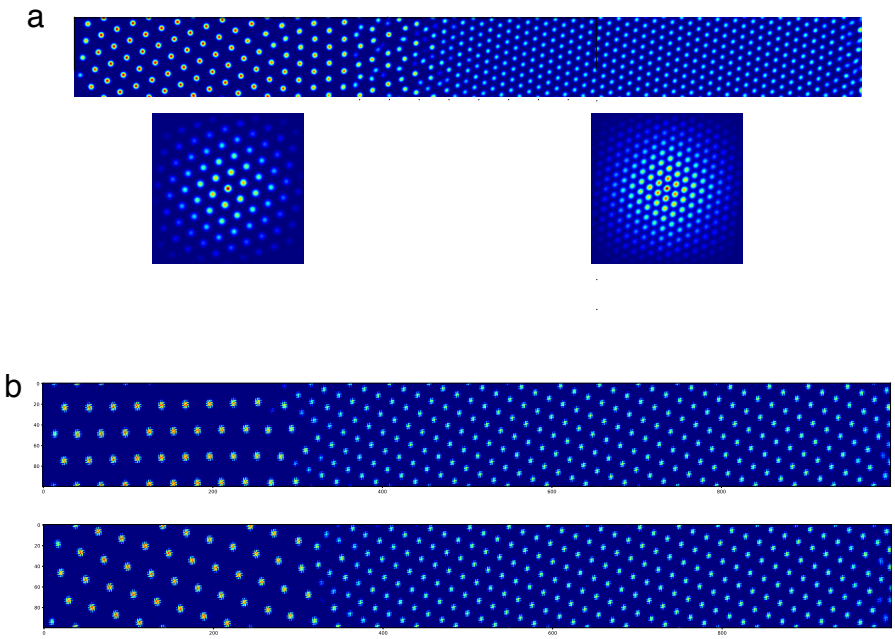


Figure 19: (a) Another instance of a spontaneously formed two dimensional network with parameters given in Table 4. (b) Two different random initializations of the network from Fig 2h show different relative orientations between the 2 formed modules.

3.9 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 $W^g_\xi[\mathbf{x},\mathbf{x}'] = W^g[(\mathbf{x}-\mathbf{x}')\times(1+\xi(\mathbf{x}'))]$, 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. Fig. 20d we present examples of such kernels for the case of $W^g[\mathbf{x},\mathbf{x}']$ described by the box function Eq. (6). 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 $\mathbf{x}$ is dependent on the set of potential maxima defined by $\mathscr{F}W^f(k)$, with a selection between these maxima performed by the broader peak of $\mathscr{F}W^g(k)$. If noise in the form of spatial heterogeneities are only introduced in W^g (and hence introduced in $\mathscr{F}W^g$) 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 $\mathscr{O}(\epsilon)$ noise added to W^f in the form of spatial heterogeneities, this would result in an $\mathscr{O}(\epsilon)$ 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 $\mathscr{O}(\epsilon/n)$. 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. 7.

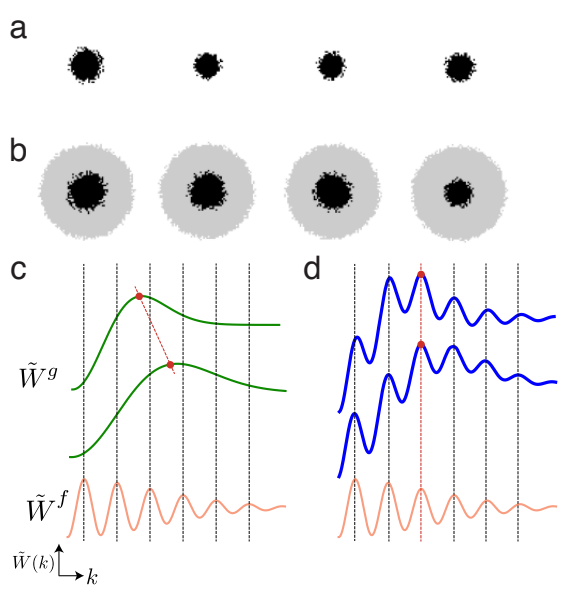


Figure 20: **Noise robustness in peak selection process** demonstrating how the additional of the smaller oscillatory fourier transform of the fixed interaction leads to no change in maxima despite smooth movement of the primary peak.

3.10 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. 3 that the period ratios will be given by the simple formula, Eq. 35. 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[29, 74, 75, 76]. 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. (35), to obtain

$$\frac{\phi}{2\pi} = \left\{\frac{\lambda_2}{\lambda_1 - \lambda_2}\right\}, m = \left\lfloor \frac{\lambda_2}{\lambda_1 - \lambda_2} \right\rfloor, \tag{65}$$

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. (35), imposes no constraints on the accuracy of the expression.

It is possible to obtain a value of ϕ from Eq. (35) 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 [29], Rat 13388 exhibits grid periods of $\approx$ 53.24 cm and $\approx$ 43.00 cm (as estimated from SI Fig. 12b in [29]); Eq. (35) 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. (65):

$$\delta\phi = 3\epsilon \frac{\lambda_2}{\lambda_1 - \lambda_2} \approx 3\epsilon m, \tag{66}$$

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).

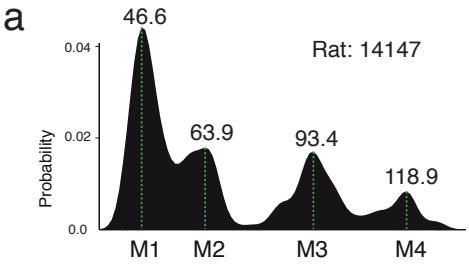
We have already demonstrated in Fig. 5 that Rat 14257 presents an extremely accurate match to the period ratio prediction for $\phi = 0$ (i.e., predicted period ratios of 2, 3/2, 4/3, 5/4,...); in addition, Rat 14147 (observed period ratios of 1.27 , 1.46 $\approx$ 3/2, 1.37 $\approx$ 4/3) and Rat 15708 (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. (51) — 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. 3.9, 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. 6c and Eq. (66), 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. (35) are for the case of *simple* kernels that have a single spatial scale. A discrepancy at 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.

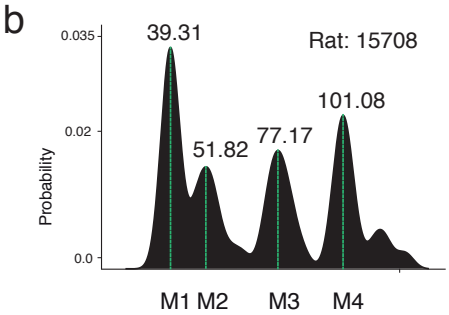
However, note that (particularly for Rats 14147 and 14257) there are relatively few grid cells observed from this largest period module, and the resulting uncertainty in period estimation may instead contribute to the error. In sum, apart from the possibility of some additional low frequency perturbations, 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.

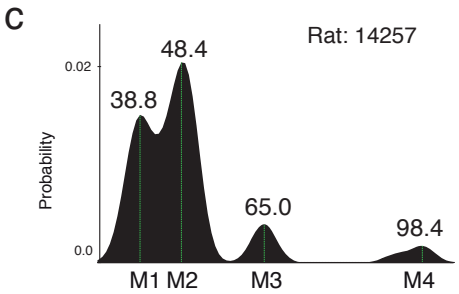
However, we do note that available data on multiple modules with a statistically large number of grid cells per module are quite sparse. To obtain further verification of our theoretical results, including the prediction of Eq. (35) and even more specifically the hypothesis that ϕ is close to zero, additional data with multiple simultaneously observed grid modules will be important.



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%

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

4 Lyapunov Function

The energy function of continuous time neural networks can be written as [77]:

$$E(\mathbf{s}) = -\frac{1}{2}\sum_{ij} s(i)W_{ij}s(j) + \sum_i \frac{1}{R_i}\int_0^{s(i)} \phi^{-1}(s)\,ds - \sum_i I_i s(i), \tag{67}$$

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:

$$\mathbf{s}_k(\mathbf{x}) = A\sin(\mathbf{k}\cdot\mathbf{x} + \delta) + B, \tag{68}$$

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. (17), which we obtained earlier via perturbation analysis.

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

In Sec. 4, 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

$$L(\theta, x) = f(\theta) + \alpha\|\theta - \theta_0(x)\|^2 \tag{69}$$

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

$$\theta^*(x) = \arg\max_\theta L(\theta, x) \tag{70}$$

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. (69) 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 $\sigma(x)$ of the pattern-forming kernel; the combined loss $L(\theta, x)$ with the full kernel W_x; 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.

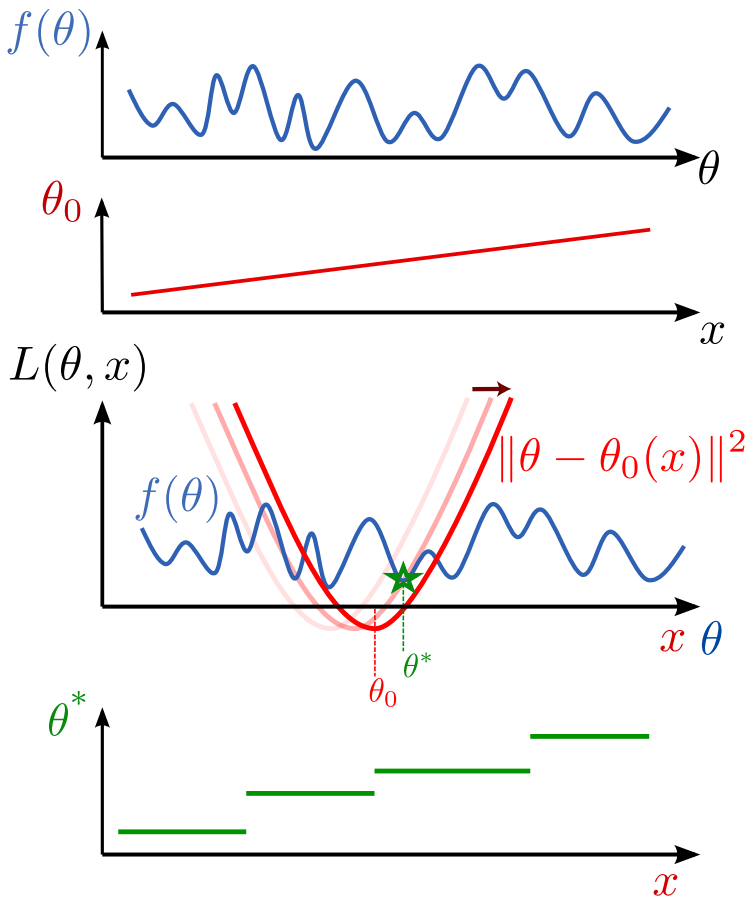


Figure 22: **A general setting for peak-selection** Assuming a loss function $f(\theta)$ (blue) and a spatially dependent quantity θ_0 (red), a combined loss function $L(\theta, x)$ can be constructed such that the x-dependent optimizer of $L(\theta, x)$ will be modular (green), since it will be constrained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

6 The emergence of modules corresponds to the formation of localized eigenvectors

As has been observed before [78], 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 [78], our interaction matrix has a locally circulant form (due to the slowly varying gradient in lateral inhibition width).

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.

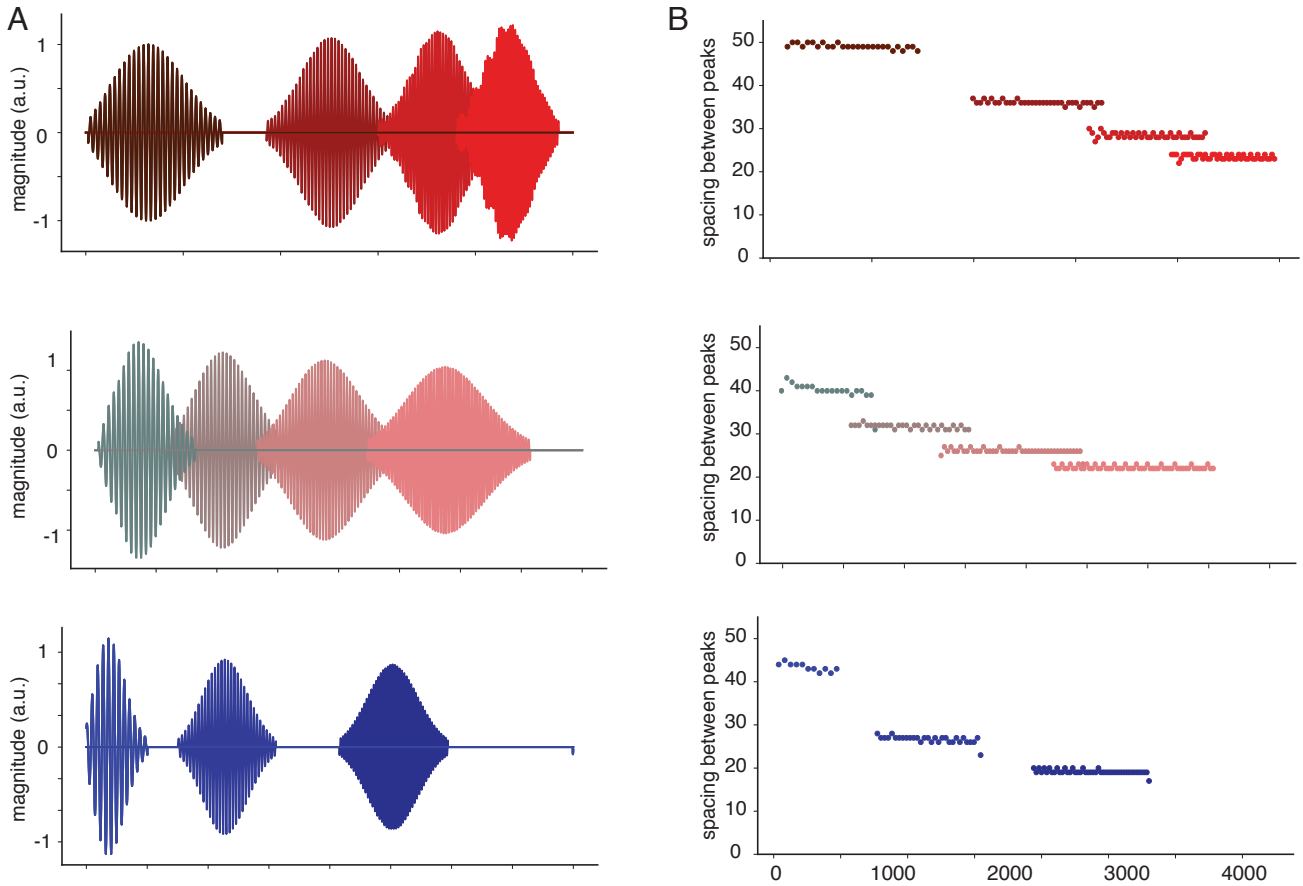


Figure 23: **Localization of eigenvectors**: A) Eigenvectors of various one-dimensional interaction weight matrices along with the corresponding inter-peak spacings are localized, B) The periodicity within an eigenvector is constant.