A geometric attractor mechanism for self-organization of entorhinal grid modules

Louis Kang[1, 2, *] and Vijay Balasubramanian[1, †]

[1] *David Rittenhouse Laboratories, University of Pennsylvania, Philadelphia, Pennsylvania, USA*
[2] *Redwood Center for Theoretical Neuroscience, University of California, Berkeley, Berkeley, California, USA*

(Dated: June 3, 2018)

Grid cells in the medial entorhinal cortex (mEC) respond when an animal occupies a periodic lattice of "grid fields" in the environment. The grids are organized in modules with spatial periods clustered around discrete values separated by constant ratios reported in the range 1.3–1.8. We propose a mechanism for dynamical self-organization in the mEC that can produce this modular structure. In attractor network models of grid formation, the period of a single module is set by the length scale of recurrent inhibition between neurons. We show that grid cells will instead form a hierarchy of discrete modules if a continuous increase in inhibition distance along the dorso-ventral axis of the mEC is accompanied by excitatory interactions along this axis. Moreover, constant scale ratios between successive modules arise through geometric relationships between triangular grids, whose lattice constants are separated by $\sqrt{3} \approx 1.7$, $\sqrt{7}/2 \approx 1.3$, or other ratios. We discuss how the interactions required by our model might be tested experimentally and realized by circuits in the mEC.

A grid cell has a spatially modulated firing rate that peaks when an animal reaches certain locations in its environment [1]. These locations of high activity form a regular triangular grid of a particular spatial scale and orientation. Every animal has many grid cells that collectively span a wide range of scales, with smaller scales enriched dorsally and larger scales ventrally along the longitudinal axis of the mEC [2]. Instead of being smoothly distributed, grid scales cluster around particular values and thus grid cells are partitioned into modules [2]. Successive modules in individual animals have fixed scale ratios that are measured in the range 1.3–1.7 [2–4].

Encoding spatial information through grid cells with a geometric sequence of scales is thought to provide animals with an efficient way of representing their position within an environment [5–11]. Moreover, periodic representations of space permit a novel mechanism for precise error correction against neural noise [12] and are learned by machines seeking to navigate open environments [13, 14]. These findings provide motivation, but a mechanism for forming a modular grid system with a fixed scale ratio is unknown. Continuous attractor networks [15, 16], a leading model for producing grid cells, would currently require discrete changes in scales to be directly imposed as sharp changes in parameters, as would the oscillatory interference model [17, 18] or hybrid models [19]. In contrast, many sensory and behavioral systems have smooth tuning distributions, such as preferred orientation in visual cortex [20] and preferred head direction in the mEC [21]. A self-organizing map model with stripe cell inputs [22] and a firing rate adaptation model with place cell inputs [23] can generate discrete grid scales, but scale ratios between modules are not constant as they are in nature, unless explicitly tuned.

Here, we present a simple extension of the continuous attractor model that adds excitatory connections between a series of attractor networks along the dorso-ventral axis of the mEC, accompanied by an increase in the distance of inhibition. The inhibition gradient drives an increase in grid scale along the mEC axis. Meanwhile, the excitatory coupling discourages changes in grid scales and orientation unless they occur through geometric relationships with fixed scale ratios and defined orientation differences. Competition between these effects of longitudinal excitation and lateral inhibition self-organizes the complete network into a discrete hierarchy of geometrically spaced modules.

Continuous attractors are a powerful general method for self-organizing neural dynamics. To our knowledge, our results are the first demonstration of a mechanism for producing a discrete hierarchy of modules in a continuous attractor system.

RESULTS

The standard grid cell attractor model is not modular

We assemble a series of networks along the longitudinal mEC axis, numbering them $z = 1, 2, \ldots, 12$ from dorsal to ventral (**Fig. 1a**). Each network contains the standard 2D continuous attractor architecture of the Burak-Fiete model [16]. Namely, neurons are arranged in a 2D sheet with positions (x, y), receive broad excitatory drive from the hippocampus (Ref. 24 and **Fig. 1b**), and inhibit one another at a characteristic separation on the neural sheet (**Fig. 1c**). This inhibition distance l is constant within each network but increases from one network to the next along the longitudinal axis of the mEC (**Methods** and **Table 1**). With these features alone, the population activity in each network self-organizes into a triangular grid whose lattice points correspond to peaks in neural activ-

* louis.kang@berkeley.edu
† vijay@physics.upenn.edu

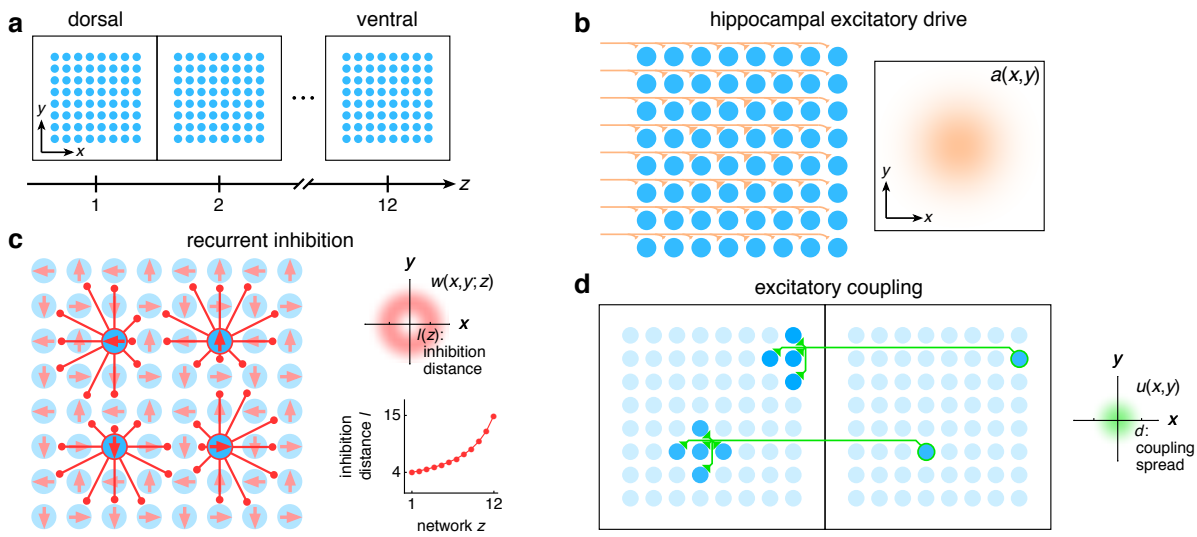

Figure 1. The entorhinal grid system as coupled 2D continuous attractor networks (**Methods**). (**a**) Each network z corresponds to a region along the dorso-ventral mEC axis and contains a 2D sheet of neurons with positions (x, y). (**b**) Neurons receive excitatory drive $a(x, y)$ from the hippocampus that is greatest at the network center and decays toward the edges. (**c**) Neurons inhibit neighbors within the same network with a weight $w(x, y; z)$ that peaks at a distance of $l(z)$ neurons, which increases as a function of z. Each neuron has its inhibitory outputs shifted slightly in one of four preferred network directions and receives slightly more hippocampal drive when the animal moves along its preferred spatial direction. (**d**) Each neuron at position (x, y) in network z excites neurons located within a spread d of (x, y) in network $z - 1$.

ity (first row of **Fig. 2a**). Importantly, the scale of each network's grid, which we call $\lambda(z)$, is proportional to that network's inhibition distance $l(z)$ ("uncoupled" simulations in **Fig. 2c**). Also, network grid orientations θ show no consistent pattern across scales and among replicate simulations with different random initial firing rates.

Following the standard attractor model [16], the inhibitory connections in each network are slightly modulated by the animal's velocity such that the population activity pattern of each network translates proportionally to animal motion at all times (**Methods**). This modulation allows each network to encode the animal's displacement through a process known as path-integration, and projects the network grid pattern onto spatial rate maps of single neurons. That is, a recording of a single neuron over the course of an animal trajectory would show high activity in spatial locations that form a triangular grid with scale Λ (third row of **Fig. 2a**). Moreover, $\Lambda(z)$ for a neuron from network z is proportional to that network's population grid scale $\lambda(z)$, and thus also proportional to its inhibition distance $l(z)$ ("uncoupled" simulations in **Fig. 2d**). To be clear, we call Λ the "spatial grid scale"; it corresponds to a single neuron's activity over the course of a simulation and has units of physical distance in space. By contrast, λ, the "network grid scale" described above, corresponds to the population activity at a single time and has units of separation on the neural sheet. Similarly, $\Theta(z)$ describes the orientation of the spatial grid of activity produced over time by a neuron in the network z; we call Θ the "spatial orientation". Like the network orientations θ discussed above, spatial orientations of grids showed no clustering ("uncoupled" simulations in **Fig. 2d**).

With an inhibition distance $l(z)$ that increases gradually from one network to the next (**Fig. 1c**), proportional changes in network and spatial scales $\lambda(z)$ and $\Lambda(z)$ lead to a smooth distribution of grid scales ("uncoupled" simulations in **Fig. 2c,d**). To reproduce the experimentally observed jumps in grid scale between modules, the inhibition length would also have to undergo discrete, sharp jumps between some adjacent networks. A further mechanism would be needed to enforce the preferred orientation differences that are observed between modules.

In summary, a grid system created by disjoint attractor networks will not self-organize into modules unless the modular structure is specifically imposed through external constraints.

Coupled attractor networks produce modules

Module self-organization can be achieved with one addition to the established features listed above: We introduce excitatory connections from each neuron to those in the preceding (more dorsal) network with approximately corresponding neural sheet positions (**Fig. 1d**). That is, a neuron in network z (more ventral) with position (x, y) will excite neurons in network $z - 1$ (more dorsal) with positions that are within a coupling distance d of position (x, y). In other words, the distance d is the "spread" of excitatory connections, and we choose a constant value across all networks comparable to the inhibition distance

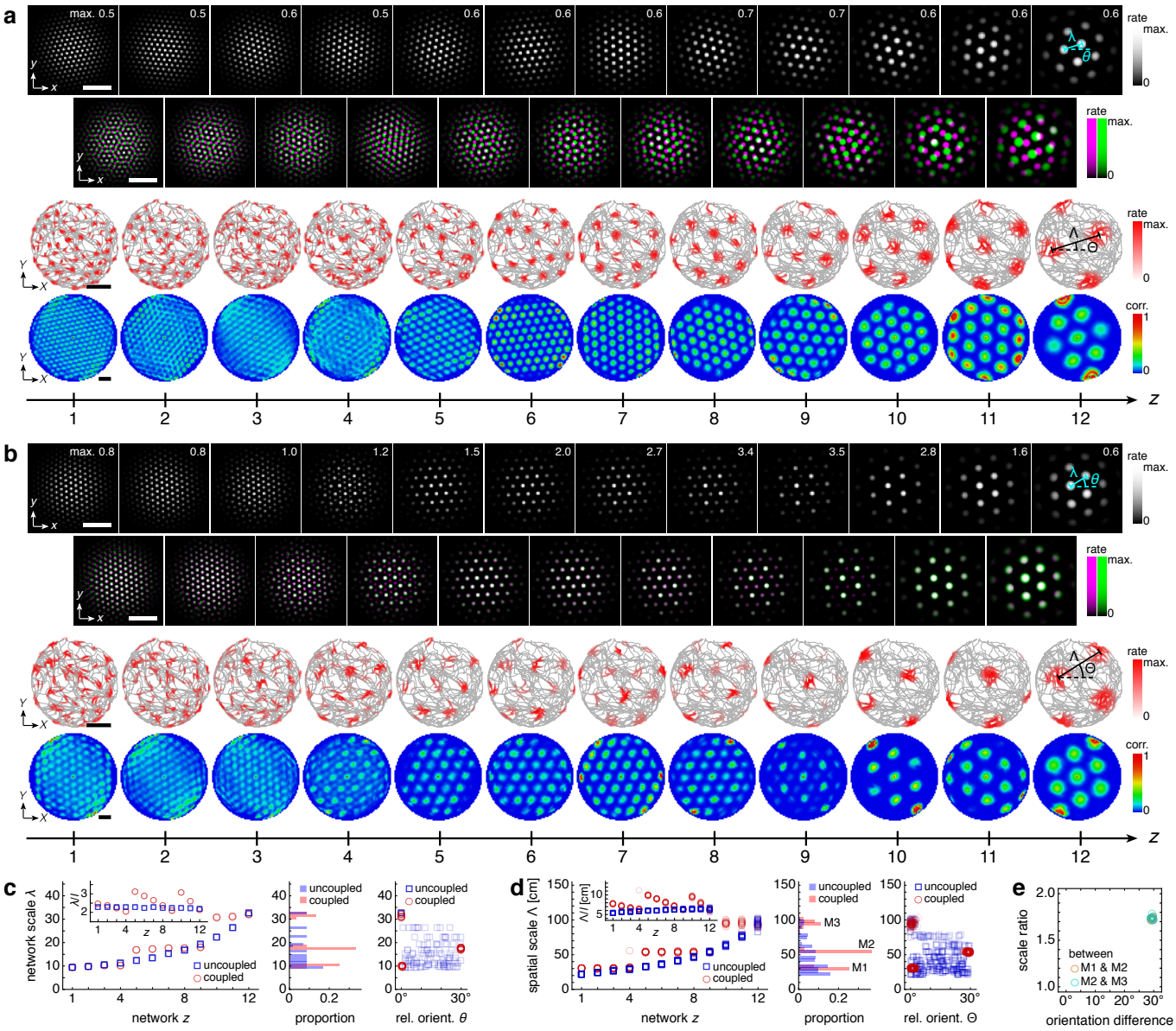

Figure 2. Coupling naturally induces modularity with fixed scale ratios and orientation differences. (**a**) A representative simulation without coupling. Top row: network activities at the end of the simulation. Numbers indicate maximum firing rates of each network in arbitrary units. Second row: activity overlays between adjacent networks depicted in the top row. In each panel, the network with smaller (larger) z is depicted in magenta (green), so white indicates regions of activity in both networks. Third row: spatial rate map of a single neuron for each z superimposed on the animal's trajectory. Bottom row: spatial autocorrelations of the rate maps depicted in the third row. (**b**) Same as **a** but for a representative simulation with coupling. (**c**–**e**) Data from 10 replicate coupled and uncoupled simulations subject to a gridness cutoff of 0.6. (**c**) Left: network grid scales $\lambda(z)$. For each network, there are 10 closely spaced red circles and 10 closely spaced blue squares corresponding to replicate simulations. Inset: $\lambda(z)$ divided by the inhibition distance $l(z)$. Middle: histogram for network grid scale λ collected across all networks. Right: network grid orientations θ relative to the network in the same simulation with largest scale. (**d**) Left: spatial grid scales $\Lambda(z)$. For each z, there are up to 30 red circles and 30 blue squares corresponding to 3 neurons recorded during each simulation. Inset: $\Lambda(z)$ divided by the inhibition distance $l(z)$. Middle: histogram for Λ collected across all networks. In the coupled model, grid cells are clustered into three modules. Right: spatial grid orientations Θ relative to the grid cell in the same simulation with largest scale. (**e**) Spatial grid scale ratios and orientation differences between adjacent modules for the coupled model. White scale bars, 50 neurons. Black scale bars, 50 cm. Standard parameter values provided in **Table 1**.

$l(z)$ (**Methods** and **Table 1**). Similar results are obtained with dorsal-to-ventral or bidirectional excitatory couplings (below) or with a spread $d(z)$ that increases with the inhibition distance $l(z)$ (**Supp. Fig. 1**).

The self-organization of triangular grids in the neural sheet and the faithful path-integration that projects these grids onto single neuron spatial rate maps persist after introduction of inter-network couplings (**Fig. 2b**). Network and spatial scales $\lambda(z)$ and $\Lambda(z)$ still increase from network $z = 1$ (dorsal) to network $z = 12$ (ventral). Yet, **Fig. 2c,d** shows that for the coupled model, these scales exhibit plateaus that are interrupted by large jumps, disrupting their proportionality to inhibition distance $l(z)$, which is kept identical to that of the uncoupled system (**Fig. 1c**). Collecting scales across all networks illustrates that they cluster around certain values in the coupled system while they are smoothly distributed in the uncoupled system. We identify these clusters with modules M1, M2, and M3 of increasing scale. Moreover, coupling causes grid cells that cluster around a certain scale to also cluster around a certain relative orientation (**Fig. 2c,d**), as seen in experiment [2] and as required for avoiding gaps in spatial coverage by each module [10]. The uncoupled system does not demonstrate co-modularity of orientation with scale, i.e., two networks with similar grid scales need not have similar orientations unless this is imposed by an external constraint.

In summary, excitatory coupling between grid attractor networks dynamically induces discreteness in grid scales that is co-modular with grid orientation, as seen in experiment [2–4].

Modular geometry is fixed by lattice geometry

Not only does excitatory coupling produce modules, it does so with consistent scale ratios and orientation differences. For the coupled system depicted in **Fig. 2**, scale ratios and orientation differences between pairs of adjacent modules consistently take values around 1.7 and 30°, respectively (**Fig. 2e**). If we perturb the inhibition distance profile $l(z)$ by making it less or more concave, these scale ratios and orientation differences are unchanged (**Fig. 3a,b**). Concavity only affects the number of grid cells in each module, which can be tuned to match experimental observations. The same scale ratios and orientation differences also persist after changes to the directionality and spread of excitatory connections. For example, we replace the ventral-to-dorsal connections with bidirectional coupling and decrease the coupling spread d such that a neuron in network z excites only a single neuron in both networks $z-1$ and $z+1$; scale ratios and orientation differences remain at 1.7 and 30°, respectively (**Fig. 3c**). Representative network activities and single neuron rate maps for these simulations are provided in **Supp. Fig. 2**. Data for simulations with only dorsal-to-ventral connections are provided in **Supp. Fig. 3**; they also exhibit the same scale ratios and orientation differences.

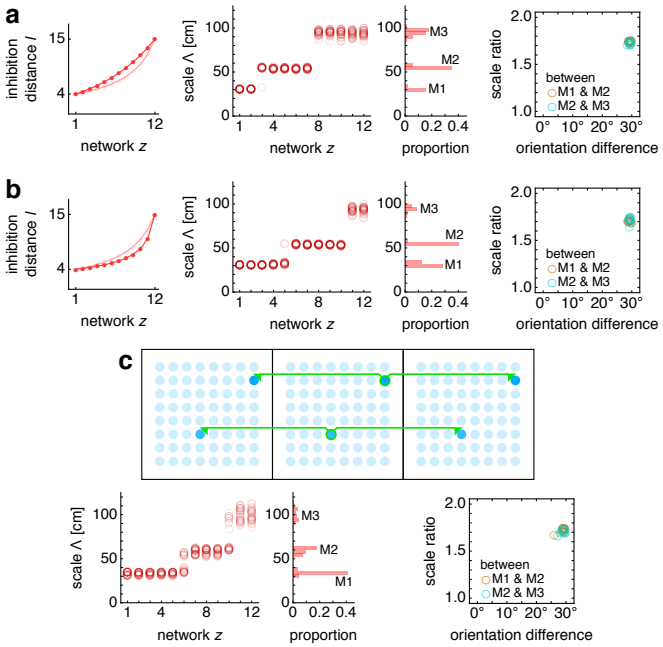

Figure 3. Modules are robust to parameter perturbations (10 replicates in each subfigure subject to a gridness cutoff of 0.6). (**a**) Left: we use a less concave inhibition distance profile $l(z)$ (dark filled circles) compared to that used in **Fig. 2** (light empty circles). Middle: spatial grid scales exhibit modules when collected in a histogram across networks. Right: modules have the same scale ratios and orientation differences as in **Fig. 2e**. (**b**) Same as **a**, but with a more concave $l(z)$. (**c**) Simulations with bidirectional point-to-point coupling instead of the unidirectional spread coupling in **Fig. 2**. Top: schematic of the neuron at position (x, y) in network z exciting only the neuron at (x, y) in networks $z-1$ and $z+1$. Bottom left/right: same as middle/right in **a**. Standard parameters are in **Table 1**. In **a**, inhibition distance exponent $l_{\text{exp}} = 0$. In **b**, inhibition distance exponent $l_{\text{exp}} = -2$. In **c**, coupling spread $d = 1$ and coupling strength $u_{\text{mag}} = 0.4$ in both directions.

We can intuitively understand this precise and robust modularity through the competition between lateral inhibition within networks and longitudinal excitation across networks. In the uncoupled system, grid scales decrease proportionally as the inhibition distance $l(z)$ decreases from $z = 12$ to $z = 1$. However, coupling causes areas of high activity in network z to preferentially excite corresponding areas in network $z-1$, which encourages adjacent networks to share the same grid pattern. Thus, coupling adds rigidity to the system and provides an opposing "force" against the changing inhibition distance that attempts to drive changes in grid scale. This rigidity produces the plateaus in network and spatial scales $\lambda(z)$ and $\Lambda(z)$ that delineate modules across multiple networks.

At interfaces between modules, coupling can no longer fully oppose the changing inhibition distance, and the

grid pattern changes. However, the rigidity fixes a geometric relationship between the grid patterns of the two networks spanning the interface. In the coupled system of **Fig. 2**, module interfaces occur between networks $z = 4$ and 5 and between $z = 9$ and 10. The network population activity overlays of **Fig. 2b** reveal overlap of many activity peaks at these interfaces. However, the more dorsal network (with smaller z) at each interface contains additional small peaks between the shared peaks. This way, adjacent networks still share many corresponding areas of high activity, as favored by coupling, but the grid scale changes, as favored by a changing inhibition distance. Pairs of grids whose lattice points demonstrate regular registry are called commensurate lattices [25] and have precise scale ratios and orientation differences, here respectively $\sqrt{3} \approx 1.7$ and 30°, which match the results in **Fig. 2e** and **Fig. 3**.

Thus, excitatory coupling can compete against a changing inhibition distance to produce a rigid grid system whose "fractures" exhibit stereotyped commensurate lattice relationships. These robust geometric relationships lead to discrete modules with fixed scale ratios and orientation differences.

In our model, commensurate lattice relationships naturally lead to field-to-field firing rate variability in single neuron spatial rate maps, another experimentally observed feature of the grid system [26, 27] (for example, $z = 8$ in the third row of **Fig. 2b**). At interfaces between two commensurate lattices, only a subset of population activity peaks in the grid of smaller scale overlap with, and thus receive excitation from, those in the grid of larger scale. The network with smaller grid scale will contain activity peaks of different magnitudes; this heterogeneity is then projected onto the spatial rate maps of its neurons. Our model also suggests that mean firing rates can vary across modules. In **Fig. 2b**, for example, neurons at intermediate z have the highest firing rates, although the precise dependence of firing rate on depth along the dorso-ventral axis can be modified by changing the strength of coupling or allowing it to vary with z.

Excitation-inhibition balance sets modular geometry

Reported scale ratios between adjacent grid modules range from 1.3 to 1.7, and the majority of reported orientation differences are less than 10°, although measured values span the full range from 0° to 30° [2–4]. Results from the model with parameters chosen as in **Fig. 2** and **Table 1** agree with the upper limit of both ranges. However, we reasoned that changing the excitatory-inhibitory balance could lead to other modular geometries. To test this, we systematically tuned the competition between excitation and inhibition in a coupled model with just two networks, $z = 1$ and 2. We varied three parameters: the coupling spread d, the coupling strength u_{mag}, and the ratio of inhibition distances between the two networks $l(2)/l(1)$ (**Methods**). For each set of parameters, we measured network scale ratios and orientation differences produced by multiple replicate simulations (**Supp. Fig. 4**). We found that as the excitation-inhibition balance is varied by changing u_{mag} and $l(2)/l(1)$, a number of discretely different "phases" appear (**Fig. 4a** for coupling spread less than or equal to the inhibition distances of both networks and **Fig. 4b** for coupling spread greater than or equal to both inhibition distances). Each phase represents a distinctive modular geometry with characteristic scale ratios and orientation differences between modules.

In many regions of network parameter space, the competition between excitation and inhibition produces commensurate lattice relationships between the population activities of adjacent networks (**Fig. 4c**). Each commensurate phase has a characteristic scale ratio and orientation difference between adjacent networks. When parameters are chosen near phase boundaries, replicate simulations may belong to either phase or occasionally be trapped in other metastable lattice relationships due to variations in initial conditions (**Supp. Fig. 4**). In addition to the $[\sqrt{3}, 30°]$ relationship in **Fig. 2** and the trivial $[1, 0°]$ relationship of networks belonging to the same module, there are three other commensurate phases, two of which, $[\sqrt{7}/2, 19°]$ and $[\sqrt{7/3}, 11°]$, are compatible with experimental measurements [2–4].

In both phase diagrams with smaller (**Fig. 4a**) and larger (**Fig. 4b**) coupling spreads, a higher excitatory coupling strength provides more rigidity against gradients in inhibition distance. This is why at larger u_{mag}, there are fewer phases as $l(2)/l(1)$ varies. However, a larger coupling spread causes network $z = 2$ to excite a broader set of neurons in network $z = 1$, softening the rigidity imposed by cross-network coupling, and producing a wider variety of phases in **Fig. 4b** than **Fig. 4a**. Also in **Fig. 4b**, when excitation is weak and approaching the uncoupled limit, there is a noticeable region dominated by *incommensurate* lattices, in which the two grids lack consistent registry or relative orientation, and the grid scale is largely determined by the inhibition distance.

Figure 4b also contains a larger region of "discommensurate" lattices, in which network activities closely overlap in certain areas that are separated by a mesh of regions lacking overlap called discommensurations (**Fig. 4d**). We call this the *discommensurate* phase, although strictly speaking, in condensed matter physics, it would be termed a commensurate phase with discommensurations [25]. Discommensurate phases exhibit ranges of scale ratios 1.1–1.4 and orientation differences 0°–10° that ultimately arise from a single source, the density of discommensurations, whose properties can also be explained through excitation-inhibition competition. Stronger excitation drives more activity overlap, which favors sparser discommensurations and lowers the scale ratio and orientation difference. However, a larger inhibition distance ratio drives the two networks to differ more in grid scale, which favors denser discommensurations. To better accommodate the discommensurations,

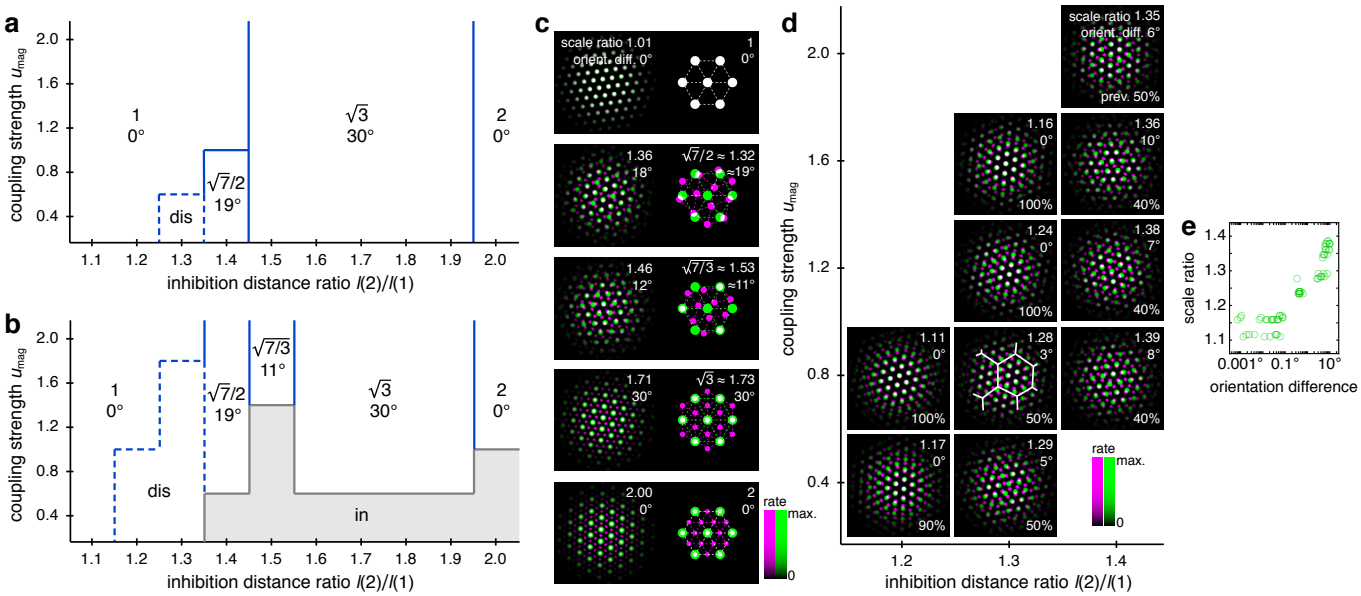

Figure 4. Diverse lattice relationships emerge over wide ranges in simulation parameters. In models with only two networks $z = 1$ and 2, we vary the coupling strength u_{mag} and the ratio of inhibition distances $l(2)/l(1)$ for two different coupling spreads d. (**a,b**) Approximate phase diagrams based on 10 replicate simulations for each set of parameters subject to a gridness cutoff of 0.6, with the mean of $l(1)$ and $l(2)$ fixed to be 9. The most frequently occurring scale ratio and orientation difference are indicated for each phase; coexistence between multiple lattice relationships may exist at drawn boundaries. (**a**) Phase diagram for small coupling spread $d = 6$. Solid lines separate four regions with different commensurate lattice relationships labeled by scale ratio and orientation difference, and dotted lines mark one region of discommensurate lattice relationships. (**b**) Phase diagram for large coupling spread $d = 12$. There are five different commensurate regions, a discommensurate region, as well as a region containing incommensurate lattices (gray). (**c**) Network activity overlays for representative observed (left) and idealized (right) commensurate lattice relationships. Numbers at the top right of each image indicate network scale ratios $\lambda(2)/\lambda(1)$ and orientation differences $\theta(2) - \theta(1)$. Networks $z = 1$ and 2 in magenta and green, respectively. (**d**) Expanded region of **b** displaying discommensurate lattice statistics. For each set of parameters, a representative overlay for the most prevalent discommensurate lattice relationship is shown. The number in the lower right indicates the proportion of replicate simulations with scale ratio within 0.01 and orientation difference within 3° of the values shown at top right. In one overlay, discommensurations are outlined by white lines. (**e**) The discommensurate relationships described in **d** demonstrate positive correlation between scale ratio and the logarithm of orientation difference (Pearson correlation coefficient $\rho = 0.91$). Parameter values provided in **Supp. Info.**

grids rotate slightly as observed previously in a crystal system [28]. **Figure 4e** confirms that scale ratios and orientation differences vary together as the discommensuration density changes.

In a full 12-network grid system, discommensurate relationships produce scale ratios and orientation differences within experimentally reported ranges, but toward their lower ends, opposite to the results of **Fig. 2** with the commensurate $[\sqrt{3}, 30°]$ relationship. The phase diagrams of **Fig. 4** provide guidance for modifying a 12-network system that exhibits a $[\sqrt{3}, 30°]$ relationship to produce discommensurate relationships instead. We make the inhibition distance profile $l(z)$ shallower (**Fig. 5a**) and increase the coupling spread d by 50%. Network activity overlays of these new simulations reveal grids obeying discommensurate relationships (**Fig. 5b**). These network activities are projected onto single neuron spatial rate maps through faithful path-integration (**Supp. Fig. 5**). Across all replicate simulations with identical parameter values but different random initial firing rates, the discommensurate system demonstrates greater variation in scale ratio and relative orientation (**Fig. 5c**) than the commensurate systems of **Figs. 2** and **3** do. However, separate analysis of each replicate simulation reveals clustering with well-defined modules (**Fig. 5d**). These modules have scale ratio 1.40 ± 0.15 and orientation difference $7.9 \pm 5.6°$ (**Fig. 5e**), which agrees well with the observations of Ref. 2. Conceptually, the excitatory coupling, especially when the spread is larger, provides enough rigidity in the discommensurate system to cluster scale ratios and orientation differences but not enough to prevent some variation in these values across replicate simulations.

Thus, networks in the discommensurate phase also lead to well-defined scale ratios and orientation differences between modules, typically at the low end of the measurements. Variations in these values between replicates are consistent with the range of experimentally observed deviations in scale ratios from their mean and in orientation differences from their mode [2–4].

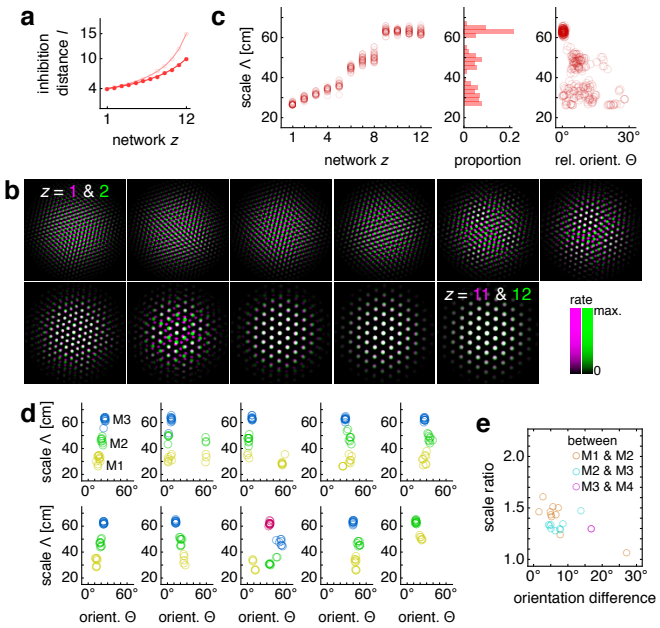

Figure 5. Discommensurate lattice relationships can also produce realistic modules. (**a**) We use a shallower inhibition distance profile $l(z)$ (dark filled circles) compared to **Fig. 2** (light empty circles). (**b**) Activity overlays between adjacent networks. In each panel, the network at smaller (larger) z is depicted in magenta (green), so white indicates regions of activity in both networks. Overlays increase in z from left to right, then from top to bottom. (**c–e**) Data from 10 replicate simulations subject to a gridness cutoff of 0.6. (**c**) Left: spatial grid scales $\Lambda(z)$. For each network, there are up to 30 red circles corresponding to 3 neurons recorded during each simulation. Middle: histogram for Λ collected across all networks, demonstrating three modules. Right: spatial grid orientations Θ relative to the grid cell in the same simulation with largest scale. (**d**) Distribution of spatial grid scales and orientations relative to an arbitrary axis for each replicate simulation. Due to 6-fold symmetry, orientation is a periodic variable modulo 60°. Different colors indicate separate modules. The first panel corresponds to the overlays in **b**. (**e**) Spatial grid scale ratios and orientation differences between adjacent modules. Maximum inhibition distance $l_{\max} = 10$, coupling spread $d = 12$. We use larger network size $n \times n = 230 \times 230$ to allow for discommensurate relationships whose periodicities span longer distances on the neural sheets. Other parameter values are in **Table 1**.

DISCUSSION

We propose that the geometric hierarchy of grid modules in the medial entorhinal cortex is self-organized by competition in an attractor network between long-range excitation along the longitudinal mEC axis and lateral inhibition. We showed that with such an architecture, an inhibition length scale that increases smoothly along the mEC axis causes the network to cluster into modules reflecting four key features of the mammalian grid system.

First, grid scales cluster around discrete values as seen in experiment [2–4]. The clusters are distinct and well-separated which implies that they will be robust to perturbations from variations in connectivity, inhomogeneities in network parameters, and spiking noise. Other models of grid module formation [22, 23] do not produce well-separated clusters, and thus neural and anatomical variability threatens to disrupt their modularity.

Second, scale ratios between adjacent modules in our model can share a universal constant value, which is consistent with experimental reports of ratios in the range 1.64 ± 0.09 from 6 rats [3], 1.42 ± 0.17 from 24 rats [2], and $\approx 1.6 \pm 0.1$ from 11 rats [4]. In some parameter regimes, commensurate lattice relationships enforce these constant scale ratios and naturally resist perturbations from noise. Many commensurate relationships produce scale ratios that are compatible with the reported data, e.g. $\sqrt{7}/2$, $\sqrt{7/3}$, and $\sqrt{3}$. Parameter choices that give rise to discommensurate lattices permit a range of possible scale ratios with many in the experimental range, but every choice induces module interfaces with approximately the same ratio. One choice that we investigated explicitly has mean scale ratio 1.4, matching the data in [2], and the small spread around the mean in each realization is consistent with experiment [2–4]. By contrast to these findings, other models require tuning of parameters to produce a constant scale ratio [22, 23].

Third, grid orientation is co-modular with scale as observed in experiment [2]. This co-modularity results from the coupling between networks and would otherwise have to be imposed by external constraints. Quantitative characterization of orientation differences between modules is limited but suggests preferences for values approximately less than 10° or close to 30° [2, 4], which can be achieved by discommensurate relationships and certain commensurate lattice relationships: $[\sqrt{7/3}, 11°]$ and $[\sqrt{3}, 30°]$. Other models can generate orientation differences less than 10° but not close to 30° [22, 23].

Note that the orientation measurements we report all come from animals that received testing in square enclosures, which may tend to anchor grid cells to certain orientations relative to the enclosure walls [4, 29]. Our model does not account for this anchoring, which may also partially explain the effects of environmental deformations on grids [2, 30, 31], and which could be implemented by border cell inputs to grid cells [32–36], or through the integration of other visual and external cues [37, 38] and internal reference frames such as boundary-vector cells [39]. Such boundary anchoring would modify our results for orientation differences between modules. Thus, further experiments on the effects of environmental geometry on grid formation would help to frame our model's predictions.

Lastly, some grid cells have firing rates that are consistently different from one field to another [26, 27]. In our model, variability arises at module interfaces from the selective excitation of certain network activity peaks in the smaller-scale grid by the overlapping activity peaks

of the larger-scale grid. Such an explanation for firing rate variability has been hinted by Ref. [26] and would be further supported by observing spatial periodicity in the variability that corresponds to the scale of the larger grids. The weaker prediction of a bimodal firing rate distribution among grid fields may be more easily tested. Other models require place cell inputs to produce field-to-field firing rate variability [27].

Other indirect signatures of coupling can be pursued experimentally. The presence of noise correlations across modules, as previously investigated but not fully characterized [40, 41], would suggest the existence and direction of connections between modules. Lesion experiments in which the destruction of grids of a certain scale cause perturbations in larger or smaller scales would do the same. Calcium imaging of the mEC has already supported the continuous attractor paradigm by observing activity correlations between grid cells that vary with anatomic distance [42]; if multiple modules could be imaged, correlations in activity across modules could indicate the presence of excitatory coupling or even commensurate lattice relationships. The effect of environmental manipulations on grid relationships has demonstrated both independence [2] and dependence [4] across modules. However, such experiments may not be ideal tests of our model due to confounding boundary effects [32–35] on grids.

A direct test for coupling would involve patch clamp experiments akin to those used to identify local inhibition and excitation between principal cells in superficial mEC layers [43–45]. Our model requires excitatory coupling between grid cells at different locations along the longitudinal mEC axis, either through direct excitation or disinhibition [44]. Note that the ventro-medial to dorsoventral path of layer I axons is consistent with the direction of excitation in our model [46]; to our knowledge, the function of these axons has not been precisely established. In general, understanding of grid modularity would benefit from further characterization of grid cell microcircuits across the longitudinal mEC axis.

Place cells have been observed to have different firing fields in different environments [47]. Experiments that observe this remapping after perturbing or inactivating the mEC [48–50] suggest a mechanism that involves grid cells from different modules changing their relative phase [51]. However, these findings do not show that remapping requires mEC input nor that it arises from the mEC in the unperturbed brain. In fact, recent work reports that place cell remapping still occurs with complete lesion of the mEC [52]. Thus, although the geometric relationships of our model do constrain relative phases between grid cells belonging to different modules, this property can be compatible with place cell remapping, which may occur independently of the mEC in the intact brain. Moreover, discommensurate relationships allow for a range of phase relationships that depend on how the discommensurations cut through the lattices.

Since spatial grid scales are both proportional to inhibition length scale l and inversely proportional to velocity gain α (Ref. 16 and **Methods**), we also simulated excitatorily-coupled networks with a depth-dependent velocity gain $\alpha(z)$ and a fixed inhibition length l (**Supp. Info.**). In contrast to simulations in one dimension [53], while we observed module self-organization, the system gave inconsistent results among replicate simulations and lacked fixed scale ratios (**Supp. Figs. 6** and **7** and **Supp. Video**).

The competition generated between excitatory and inhibitory connections bears a strong resemblance to the Frenkel-Kontorova model of condensed matter physics, in which a periodic potential of one scale acts on particles that prefer to form a lattice of a different, competing scale [54]. This model has a rich literature with many deep theoretical results, including the calculation of complicated phase diagrams involving "devil's staircases" [25, 55] which mirror those of our model (**Fig. 4**). Under certain conditions, our model produces networks with quasicrystalline approximant grids that are driven by networks with standard triangular grids at other scales (**Supp. Fig. 8**). Quasicrystalline order lacks periodicity, but contains more nuanced positional order [56]. This phenomenon wherein quasicrystalline structure is driven by crystalline order in a coupled system was recently observed for the first time in thin-film materials that contain Frenkel-Kontorova-like interactions [57–59].

Commensurate lattice relationships are a robust and versatile mechanism for self-organizing a hierarchy of grid modules with constant scale ratios and a preferred range of orientation differences. We demonstrated this mechanism in a basic extension of the continuous attractor model with excitatory connections between networks. This model is amenable to extensions that capture other features of the grid system, such as spiking dynamics, the union of our separate networks into a single network spanning the entire mEC, and the addition of border cell inputs or recurrent coupling between modules to correct path-integration errors or react to environmental deformations [34, 35, 60–62].

ACKNOWLEDGMENTS

We are grateful to Xue-Xin Wei, Tom Lubensky, Ila Fiete, and John Widloski for their thoughtful ideas and suggestions, and to the Honda Research Institute and the NSF (grant PHY-1734030) for research support. L.K. is also supported by the Miller Institute at the University of California, Berkeley. Work on this project at the Aspen Center for Physics was supported by NSF grant PHY-1607611.

Table 1. Main model parameters and their values unless otherwise noted.

Parameter	Variable	Value
Number of networks	h	12
Number of neurons per network	$n \times n$	160×160
Neurons recorded per network		3
Animal speed	$\lvert\mathbf{V}\rvert$	0–1 m/s
Diameter of enclosure		180 cm
Simulation time		500 s
Simulation timestep	Δt	1 ms
Neural relaxation time	τ	10 ms
Hippocampal input strength	a_{mag}	1
Hippocampal input falloff	a_{fall}	4
Inhibition distance minimum	l_{min}	4
Inhibition distance maximum	l_{max}	15
Inhibition distance exponent	l_{exp}	-1
Inhibition strength	w_{mag}	2.4
Subpopulation shift	ξ	1
Coupling spread	d	8
Coupling strength	u_{mag}	2.6
Velocity gain	α	0.3 s/m

METHODS

Model setup and dynamics

We implemented the Burak-Fiete model as follows [16]. Networks $z = 1, \ldots, h$ each contain a 2D sheet of neurons with indices $\mathbf{r} = (x, y)$, where $x = 1, \ldots, n$ and $y = 1, \ldots, n$. Neurons receive broad excitatory input $a(\mathbf{r})$ from the hippocampus, and, to prevent edge effects, those toward the center of the networks receive more excitation than those toward the edges. Each neuron also inhibits others that lie around a length scale of $l(z)$ neurons away in the same network z Moreover, every neuron belongs to one of four subpopulations that evenly tile the neural sheet. Each subpopulation is associated with both a preferred direction $\hat{\mathbf{e}}$ along one of the network axes $\pm\hat{\mathbf{x}}$ or $\pm\hat{\mathbf{y}}$ and a corresponding preferred direction $\hat{\mathbf{E}}$ along an axis $\pm\hat{\mathbf{X}}$ or $\pm\hat{\mathbf{Y}}$ in its spatial environment. A neuron at position $\mathbf{r}$ in network z has its inhibitory outputs $w(\mathbf{r}; z)$ shifted slightly by ξ neurons in the $\hat{\mathbf{e}}(\mathbf{r})$ direction and its hippocampal excitation modulated by a small amount proportional to $\hat{\mathbf{E}}(\mathbf{r}) \cdot \mathbf{V}$, where $\mathbf{V}$ is the spatial velocity of the animal. Note that lowercase letters refer to attractor networks at each depth z in which distances have units of neurons, and uppercase letters refer to the animal's spatial environment in which distances have physical units, such as centimeters.

In addition to these established features [16], we introduce excitatory connections $u(\mathbf{r})$ from every neuron $\mathbf{r}$ in network z to neurons located within a spread d of the same $\mathbf{r}$ but in the preceding network with depth $z - 1$. $u(\mathbf{r})$ is constant for all networks except for the last one $z = h$, which has $u(\mathbf{r}) = 0$. These components lead to the following dynamical equation for the dimensionless neural firing rates $s(\mathbf{r}, z, t)$:

$$\begin{aligned}\tau \frac{s(\mathbf{r}, z, t + \Delta t) - s(\mathbf{r}, z, t)}{\Delta t} + s(\mathbf{r}, z, t) \\ = \Bigg\{ \sum_{\mathbf{r}'} w(\mathbf{r} - \mathbf{r}' + \xi \hat{\mathbf{e}}(\mathbf{r}'); z) s(\mathbf{r}', z, t) \\ + \sum_{\mathbf{r}'} u(\mathbf{r} - \mathbf{r}') s(\mathbf{r}', z + 1, t) \\ + a(\mathbf{r}) \left[1 + \alpha \hat{\mathbf{E}}(\mathbf{r}) \cdot \mathbf{V}(t) \right] \Bigg\}_+ . \end{aligned} \tag{1}$$

Inputs to each neuron are rectified by $\{c\}_+ = 0$ for $c < 0$, c for $c \geq 0$. Δt is the simulation time increment, τ is the neural relaxation time, and α is the velocity gain that describes how much the animal's velocity $\mathbf{V}$ modulates the hippocampal inputs $a(\mathbf{r})$. Note that s can be treated as a dimensionless variable because **Eq. 1** is invariant to scaling of s and a by the same factor.

We use velocities $\mathbf{V}(t)$ corresponding to a real rat trajectory [1, 16]. Details are provided in **Supp. Info.**

Inhibitory and excitatory connections

The hippocampal input is

$$a(\mathbf{r}) = \begin{cases} a_{\mathrm{mag}} \mathrm{e}^{-a_{\mathrm{fall}} r_{\mathrm{scaled}}^2} & r_{\mathrm{scaled}} < 1 \\ 0 & r_{\mathrm{scaled}} \geq 1, \end{cases} \tag{2}$$

where $r_{\mathrm{scaled}} = \sqrt{\left(x - \frac{n+1}{2}\right)^2 + \left(y - \frac{n+1}{2}\right)^2} / \frac{n}{2}$ is a scaled radial distance for the neuron at $\mathbf{r} = (x, y)$, a_{mag} is the magnitude of the input, and a_{fall} is a falloff parameter. The inhibition distance for network z is

$$l(z) = \left[l_{\mathrm{min}}^{l_{\mathrm{exp}}} + \left(l_{\mathrm{max}}^{l_{\mathrm{exp}}} - l_{\mathrm{min}}^{l_{\mathrm{exp}}} \right) \frac{z - 1}{h - 1} \right]^{1/l_{\mathrm{exp}}}, \tag{3}$$

which ranges from $l_{\mathrm{min}} = l(1)$ to $l_{\mathrm{max}} = l(h)$ with concavity tuned by l_{exp}. More negative values of l_{exp} lead to greater concavity; for $l_{\mathrm{exp}} = 0$, we use the limiting expression $l(z) = l_{\mathrm{min}}^{(h-z)/(h-1)} l_{\mathrm{max}}^{(z-1)/(h-1)}$. The recurrent inhibition profile for network z is

$$w(\mathbf{r}; z) = \begin{cases} -\frac{w_{\mathrm{mag}}}{l(z)^2} \frac{1 - \cos[\pi r / l(z)]}{2} & r < 2l(z) \\ 0 & r \geq 2l(z), \end{cases} \tag{4}$$

where w_{mag} is the magnitude of inhibition. We scale this magnitude by $l(z)^{-2}$ to make the integrated inhibition constant across z. The excitatory coupling is

$$u(\mathbf{r}) = \begin{cases} \frac{u_{\mathrm{mag}}}{d^2} \frac{1 + \cos[\pi r / d]}{2} & r < d \\ 0 & r \geq d, \end{cases} \tag{5}$$

where u_{mag} and d are the magnitude and spread of coupling, respectively. In analogy to w_{mag}, we scale u_{mag} by d^{-2}.

Overview of data analysis techniques

To determine spatial grid scales, orientations, and gridness, we consider an annular region of the spatial autocorrelation map that contains the 6 peaks closest to the origin. Grid scale is the radius with highest value, averaging over angles. Grid orientation and gridness are determined by first averaging over radial distance and analyzing the sixth component of the Fourier series with respect to angle [63]. The power of this component divided by the total Fourier power measures "gridness" and its complex phase measures the orientation. For each replicate simulation, we cluster its grid cells with respect to scale and orientation using a k-means procedure with k determined by kernel smoothed densities [2]. See **Supp. Info.** for full details.

[1] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, Nature **436**, 801 (2005).
[2] H. Stensola, T. Stensola, T. Solstad, K. Frøland, M.-B. Moser, and E. I. Moser, Nature **492**, 72 (2012).
[3] C. Barry, R. Hayman, N. Burgess, and K. J. Jeffery, Nat. Neurosci. **10**, 682 (2007).
[4] J. Krupic, M. Bauza, S. Burton, C. Barry, and J. O'Keefe, Nature **518**, 232 (2015).
[5] E. I. Moser, E. Kropff, and M.-B. Moser, Annu. Rev. Neurosci. **31**, 69 (2008).
[6] I. R. Fiete, Y. Burak, and T. Brookings, J. Neurosci. **28**, 6858 (2008).
[7] A. Mathis, A. V. M. Herz, and M. Stemmler, Neural Computation **24**, 2280 (2012).
[8] X.-X. Wei, J. Prentice, and V. Balasubramanian, eLife (2015).
[9] M. Stemmler, A. Mathis, and A. V. M. Herz, Science Advances **1**, e1500816 (2015).
[10] A. Sanzeni, V. Balasubramanian, G. Tiana, and M. Vergassola, Phys. Rev. E **94**, 599 (2016).
[11] N. Mosheiff, H. Agmon, A. Moriel, and Y. Burak, PLOS Comp. Biol. **13**, e1005597 (2017).
[12] S. Sreenivasan and I. Fiete, Nat. Neurosci. **14**, 1330 (2011).
[13] C. J. Cueva and X.-X. Wei, International Conference on Learning Representations (2018).
[14] A. Banino, C. Barry, B. Uria, C. Blundell, T. Lillicrap, P. Mirowski, A. Pritzel, M. J. Chadwick, T. Degris, J. Modayil, *et al.*, Nature , 1 (2018).
[15] M. C. Fuhs and D. S. Touretzky, J. Neurosci. **26**, 4266 (2006).
[16] Y. Burak and I. R. Fiete, PLOS Comp. Biol. **5**, e1000291 (2009).
[17] N. Burgess, C. Barry, and J. O'Keefe, Hippocampus **17**, 801 (2007).
[18] M. E. Hasselmo, L. M. Giocomo, and E. A. Zilli, Hippocampus **17**, 1252 (2007).
[19] D. Bush and N. Burgess, J. Neurosci. **34**, 5065 (2014).
[20] N. P. Issa, A. Rosenberg, and T. R. Husson, J. Neurophysiol. **99**, 2745 (2008).
[21] J. S. Taube, R. U. Muller, and J. B. Ranck, J. Neurosci. **10**, 420 (1990).
[22] S. Grossberg and P. K. Pilly, PLOS Comp. Biol. **8**, e1002648 (2012).
[23] E. Urdapilleta, B. Si, and A. Treves, Hippocampus **77**, 137 (2017).
[24] T. Bonnevie, B. Dunn, M. Fyhn, T. Hafting, D. Derdikman, J. L. Kubie, Y. Roudi, E. I. Moser, and M.-B. Moser, Nat. Neurosci. **16**, 309 (2013).
[25] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, 1995) pp. 601–620.
[26] R. Ismakov, O. Barak, K. Jeffery, and D. Derdikman, Current Biology **27**, 2337 (2017).
[27] B. Dunn, D. Wennberg, Z. Huang, and Y. Roudi, bioRxiv 10.1101/101899 (2017).
[28] J. A. Wilson, J. Phys.: Condens. Matter **2**, 1683 (1990).
[29] T. Stensola, H. Stensola, M.-B. Moser, and E. I. Moser, Nature **518**, 207 (2015).
[30] C. Barry, R. Hayman, N. Burgess, and K. J. Jeffery, Nat. Neurosci. **10**, 682 (2007).
[31] J. Krupic, M. Bauza, S. Burton, and J. O'Keefe, Science **359**, 1143 (2018).
[32] J. Krupic, M. Bauza, S. Burton, and J. O'Keefe, J. Physiol. **594**, 6489 (2016).
[33] L. M. Giocomo, J. Physiol. **594**, 6501 (2016).
[34] A. T. Keinath, R. A. Epstein, and V. Balasubramanian, bioRxiv 10.1101/174367 (2017).
[35] S. Ocko, K. Hardcastle, L. M. Giocomo, and S. Ganguli, bioRxiv 10.1101/326793 (2018).
[36] K. Hardcastle, N. Maheswaranathan, S. Ganguli, and L. M. Giocomo, Neuron **94**, 375 (2017).
[37] F. Raudies and M. E. Hasselmo, PLOS Comp. Biol. **11**, e1004596 (2015).
[38] F. Savelli, J. D. Luck, and J. J. Knierim, eLife **6** (2017).
[39] D. Bush and N. Burgess, J. Neurosci. **34**, 5065 (2014).
[40] A. Mathis, A. V. M. Herz, and M. B. Stemmler, Physical Review E **88**, 022713 (2013).
[41] G. Tocker, O. Barak, and D. Derdikman, Hippocampus **25**, 1599 (2015).
[42] J. G. Heys, K. V. Rangarajan, and D. A. Dombeck, Neuron **84**, 1079 (2014).
[43] J. J. Couey, A. Witoelar, S.-J. Zhang, K. Zheng, J. Ye, B. Dunn, R. Czajkowski, M.-B. Moser, E. I. Moser, Y. Roudi, and M. P. Witter, Nat. Neurosci. **16**, 318 (2013).
[44] E. C. Fuchs, A. Neitz, R. Pinna, S. Melzer, A. Caputi, and H. Monyer, Neuron **89**, 194 (2016).
[45] J. Winterer, N. Maier, C. Wozny, P. Beed, J. Breustedt, R. Evangelista, Y. Peng, T. D'Albis, R. Kempter, and D. Schmitz, Cell Reports **19**, 1110 (2017).
[46] S. Ray, R. Naumann, A. Burgalossi, Q. Tang, H. Schmidt, and M. Brecht, Science **343**, 891 (2014).
[47] R. U. Muller and J. L. Kubie, J. Neurosci. **7**, 1951 (1987).
[48] C. Miao, Q. Cao, H. T. Ito, H. Yamahachi, M. P. Witter, M.-B. Moser, and E. I. Moser, Neuron **88**, 590 (2015).
[49] J. W. Rueckemann, A. J. DiMauro, L. M. Rangel, X. Han, E. S. Boyden, and H. Eichenbaum, Hippocampus **26**, 246 (2016).

[50] B. R. Kanter, C. M. Lykken, D. Avesar, A. Weible, J. Dickinson, B. Dunn, N. Z. Borgesius, Y. Roudi, and C. G. Kentros, Neuron **93**, 1480 (2017).

[51] G. Buzsáki and E. I. Moser, Nat. Neurosci. **16**, 130 (2013).

[52] M. I. Schlesiger, B. L. Boublil, J. B. Hales, J. K. Leutgeb, and S. Leutgeb, Cell Reports **22**, 3152 (2018).

[53] J. Widloski and I. R. Fiete, Personal communication (2017).

[54] T. Kontorova and J. Frenkel, Zh. Eksp. Teor. Fiz. **8**, 1340 (1938).

[55] P. Bak, Rep. Prog. Phys. **45**, 587 (1982).

[56] D. Levine and P. J. Steinhardt, Phys. Rev. B **34**, 596 (1986).

[57] S. Förster, K. Meinel, R. Hammer, M. Trautmann, and W. Widdra, Nature **502**, 215 (2013).

[58] S. Förster, M. Trautmann, S. Roy, W. A. Adeagbo, E. M. Zollner, R. Hammer, F. O. Schumann, K. Meinel, S. K. Nayak, K. Mohseni, W. Hergert, H. L. Meyerheim, and W. Widdra, Phys. Rev. Lett. **117**, 1260 (2016).

[59] M. Paßens, V. Caciuc, N. Atodiresei, M. Feuerbacher, M. Moors, R. E. Dunin-Borkowski, S. Blügel, R. Waser, and S. Karthäuser, Nat. Commun. **8**, 15367 (2017).

[60] K. Hardcastle, S. Ganguli, and L. M. Giocomo, Neuron **86**, 827 (2015).

[61] E. Pollock, N. Desai, X. Wei, and V. Balasubramanian, in *Cosyne Abstracts 2017* (Salt Lake City, UT, USA, 2017).

[62] N. Mosheiff and Y. Burak, in *Cosyne Abstracts 2018* (Denver, CO, USA, 2018).

[63] S. N. Weber and H. Sprekeler, in *Cosyne Abstracts 2018* (Denver, CO, USA, 2018).